STIC Search Report
Biotech-Chem Library

STIC Database Tracking Number: 161233

TO: Celine Qian **Location: REM/2A64/2C70** **Art Unit: 1636** **Friday, April 22, 2005** **Case Serial Number: 09/939209**	**From: Edward Hart** **Location: Biotech-Chem Library** **REM-1A55** **Phone: 571-272-2512** **edward.hart@uspto.gov**

Search Notes

Examiner Qian,

Here are the results of the search you requested.

Please feel free to contact me if you have any questions.

Edward Hart

This Page Blank (uspto)

FOR OFFICIAL USE ONLY

ACCESS DB # 15/233

PLEASE PRINT CLEARLY

Scientific and Technical Information Center

SEARCH REQUEST FORM

Requester's Full Name: Celine Qian Examiner #: 78710 Date: 4/20/05

Art Unit: 1636 Phone Number: 2-0777 Serial Number: 09/939,209

Location (Bldg/Room#): 2A64 (Mailbox #): _____ Results Format Preferred (circle): PAPER DISK

****

To ensure an efficient and quality search, please attach a copy of the cover sheet, claims, and abstract or fill out the following:

Title of Invention: _____

Inventors (please provide full names): _____

_____

Earliest Priority Date: 8/24/2000

Search Topic:

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc., if known.

For Sequence Searches Only ***Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.***

SEE ATTACHED

****

STAFF USE ONLY

Searcher: _____

Searcher Phone #: _____

Searcher Location: _____

Date Searcher Picked Up: 4/20/05

Date Completed: 4/22/05

Searcher Prep & Review Time: _____

Online Time: _____

Type of Search

S NA Sequence (#)

_____ AA Sequence (#)

_____ Structure (#)

_____ Bibliographic

_____ Litigation

_____ Fulltext

_____ Other

Vendors and cost where applicable

_____ STN _____ Dialog

_____ Questel/Orbit _____ Lexis/Nexis

_____ Westlaw _____ WWW/Internet

Oligo In-house sequence systems

___Commercial ___Oligomer ___Score/Length

___ Interference ___ SPDI ___ Encode/Transl

_____ Other (specify)

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 20, 2005, 22:52:20 ; Search time 12187.1 Seconds
(without alignments)
19011.657 Million cell updates/sec

Title: US-09-939-209A-3_COPY_4970_11056
Perfect score: 6087
Sequence: 1 ccagtaggaaacagttcttc.....catcatcagaaaggcactat 6087

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_htc:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	344.8	5.7	578	6	CB294770	CB294770 12B22014_
c 2	322.8	5.3	510	7	CO888725	CO888725 BovGen_17
3	293.2	4.8	798	5	BX914754	BX914754 BX914754
4	293.2	4.8	825	5	BX924658	BX924658 BX924658
5	277.6	4.6	582	5	BP199581	BP199581 BP199581
6	277.6	4.6	582	5	BP335994	BP335994 BP335994
7	277.6	4.6	583	5	BP249954	BP249954 BP249954
8	276	4.5	551	5	BP213595	BP213595 BP213595
c 9	272.8	4.5	298	6	CD511718	CD511718 AGENCOURT
c 10	268.4	4.4	864	8	AQ745626	AQ745626 HS_2273_A
11	262.8	4.3	787	4	BI547160	BI547160 603190405
c 12	242.4	4.0	453	7	CO895317	BovGen_23
c 13	240.4	3.9	647	9	AG090278	AG090278 Pan trogl
14	226.4	3.7	712	8	AQ274674	AQ274674 RPCI-6-11
15	194.8	3.2	474	2	AW102832	xd38h03.x
16	189.2	3.1	734	8	AQ282818	AQ282818 RPCI11-89
c 17	181.2	3.0	852	9	CC904986	CC904986 t018121ba
c 18	179.2	2.9	437	8	AQ472918	CITBI-E1-
19	172.4	2.8	777	4	BG699761	BG699761 602681454
20	170	2.8	593	4	BG699541	BG699541 602679246
21	170	2.8	624	4	BI548918	BI548918 603189043
22	170	2.8	680	4	BI548997	BI548997 603188958
23	170	2.8	720	4	BI545493	BI545493 603187655
24	170	2.8	833	4	BI551559	BI551559 603194205
25	170	2.8	834	4	BI547188	BI547188 603190434
c 26	169.8	2.8	482	8	AQ457908	AQ457908 HS_5189_B
27	169.6	2.8	803	4	BI551703	BI551703 603197395
c 28	169.4	2.8	454	8	AQ466676	AQ466676 HS_5161_B
29	169.4	2.8	582	5	BP251245	BP251245 BP251245
30	168.4	2.8	521	7	CR556919	CR556919 DKFZp459I
31	168.4	2.8	540	7	CR539342	CR539342 DKFZp459M
32	168.4	2.8	578	7	CR538651	CR538651 DKFZp459M
33	168.4	2.8	583	5	BP348638	BP348638 BP348638
34	168.4	2.8	634	7	CR539181	CR539181 DKFZp459E
35	168.4	2.8	716	7	CR630572	CR630572 DKFZp459H
36	168.4	2.8	727	6	CB998494	CB998494 AGENCOURT
37	168.4	2.8	902	4	BG708958	BG708958 602675104
38	168.2	2.8	760	4	BI550370	BI550370 603192421
39	168	2.8	174	1	AU076999	AU076999 AU076999
40	168	2.8	461	5	BP199723	BP199723 BP199723
41	168	2.8	461	5	BP200217	BP200217 BP200217
42	168	2.8	461	5	BP200544	BP200544 BP200544
43	168	2.8	461	5	BP215732	BP215732 BP215732
44	168	2.8	461	5	BP233321	BP233321 BP233321
45	168	2.8	461	5	BP235431	BP235431 BP235431

ALIGNMENTS

RESULT 1
CB294770
LOCUS CB294770 578 bp mRNA linear EST 28-FEB-2003
DEFINITION 12B22014_rev_1_E11_r_087.ab1 Chimpanzee brain library Koos Pan troglodytes cDNA clone 12B22014_rev_1_E11_r_087.ab1 5', mRNA sequence.
ACCESSION CB294770
VERSION CB294770.1 GI:28620200
KEYWORDS EST.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE 1 (bases 1 to 578)
AUTHORS Hellmann,I., Zollner,S., Enard,W., Ebersberger,I., Nickel,B. and Paabo,S.
TITLE Selection on human genes as revealed by comparisons to chimpanzee cDNA
JOURNAL Genome Res. (2003) In press
COMMENT Contact: Paabo S
Evolutionary Genetics
Max-Planck-Institute for evolutionary Anthropology
Deutscher Platz 6, 04103 Leipzig, Germany
Tel: +49-(0)-341-3550 500
Fax: +49-(0)-341-3550 555
Email: paabo@eva.mpg.de
Seq primer: M13 reverse.
FEATURES Location/Qualifiers
source 1. .578
/organism="Pan troglodytes"
/mol_type="mRNA"
/db_xref="taxon:9598"
/clone="12B22014_rev_1_E11_r_087.ab1"
/sex="male"
/tissue_type="brain, presumably cortex"
/dev_stage="adult"
/lab_host="Epicurian Coli (TM) XL-10-Gold"
/clone_lib="Chimpanzee brain library Koos"
/note="Vector: pUChi; Site_1: SfiI-A; Site_2: SfiI-B; The library was prepared using the SMART cDNA library construction Kit (Clontech), doing only primer extension, but not PCR amplification of the cDNA. The only deviation from the published protocoll was that we cloned the cDNA into a plasmid Vector."
ORIGIN

Query Match 5.7%; Score 344.8; DB 6; Length 578;

```
  Best Local Similarity   99.4%;  Pred. No. 9.7e-72;
  Matches  346;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       4746 TTACAGGACGCAGGCATGTGAAGGAGAAACTGTCAGTGCTAGGCTGAATTCTGTTGTTAC 4805
             ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db          1 TTACAGGACGCAGGCATGTGAAGGAGAAACCGTCAGTGCTAGGCTGAATTCTGTTGTTAC 60

Qy       4806 CAAGATTTCTAGAAAAGTATTCCTCAGTCAGGTTGATTACAGATATAGCAAATCTATTTT 4865
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAGATTTCTAGAAAAGTATTCCTCAGTCAGGTTGATTACAGATATAGCAAATCTATTTT 120

Qy       4866 TCCTAGGGTAGTTTCTGTATGCTGCCGGGCTTATAACTGTCTGTCATCCAGCTATTTCTC 4925
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCCTAGGGTAGTTTCTGTATGCTGCCGGGCTTATAACTGTCTGTCATCCAGCTATTTCTC 180

Qy       4926 TCCACCTTCTTGTTTGCATAACAACCAAGGCAACTTCCGCAAATCACTGCGTGGAGACGA 4985
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCCACCTTCTTGTTTGCATAACAACCAAGGCAACTTCCGCAAATCACTGCGTGGAGACGA 240

Qy       4986 TGATCCTGCCAGCTCCCTTTTGGAAATCGTGAGGATCAGATCTTGGACCATGTATAATAT 5045
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TGATCCTGCCAGCTCCCTTTTGGAAATCGTGAGGATCAGATCTTGGACCATGTATAATAT 300

Qy       5046 GATGCTTCTAATCCAAAAGAGGAAAGGCATTGGGAGTCAGCTCCTAAG 5093
             ||||||||||||||||||||||| |||||||||||||||||||||||||
Db        301 GATGCTTCTAATCCAAAAGAGGAAGGGCATTGGGAGTCAGCTCCTAAG 348


RESULT 2
CO888725/c
LOCUS       CO888725                510 bp    mRNA    linear   EST 01-SEP-2004
DEFINITION  BovGen_17050 normal cattle brain Bos taurus cDNA clone
            RZPDp1056G0548Q 5', mRNA sequence.
ACCESSION   CO888725
VERSION     CO888725.1  GI:51819010
KEYWORDS    EST.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
            Bovinae; Bos.
REFERENCE   1  (bases 1 to 510)
  AUTHORS   Hennig,S., Janitz,M., Herwig,R. and Williams,J.
  TITLE     Generation, annotation, evolutionary analysis and database
            integration of 14969 cattle EST clusters
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Hennig S
            laboraty 123, dept.Lehrach
            Max-Planck-Institut fuer Molekulare Genetik
            Ihnestr.63-73, D-14195 Berlin, Germany
            Tel: +49 30 8413 1612
            Fax: +49 30 8413 1380
            Email: hennig@molgen.mpg.de
            The library was characterised by oligonucleotide fingerprinting
            (ONFP) to reduce sequencing redundancy. According to the ONFP
            procedure, clones that display the same hybridisation matrix with a
            battery of 200 8mer oligonucleotides are grouped into clusters. One
            clone per ONFP cluster was selected for sequencing. cDNA clones and
            filters are distributed via Deutsches Ressourcenzentrum fuer
            Genomforschung GmbH (http://www.rzpd.de).
            PCR PRimers
            FORWARD: 5' CCCCAGGCTTTACACTTTATGCTTCCGGCTCG 3' (M13RSP) 5'-seq
            BACKWARD: 5' GCTATTACGCCAGCTGGCGAAAGGGGGATGTG 3' (M13FSP) 3'-seq
            Seq primer: 5'-CCGGTCCGGAATTCCCGGGT-3' (M13RSP).
FEATURES             Location/Qualifiers
     source          1. .510
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
                     /clone="RZPDp1056G0548Q"
                     /sex="female"
                     /tissue_type="brain tissue"
                     /dev_stage="adult brain"
                     /clone_lib="normal cattle brain"
                     /note="Organ: brain; Vector: pSport1; Site_1: NotI;
                     Site_2: SalI; Random primed and directionally cloned in
                     pSport1 vector using NotI
                     (5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCC (T)15-3' and SalI 5'-
                     TCGACCCACGCGTCCG-3' adapters (Gibco BRL)"
ORIGIN

  Query Match             5.3%;  Score 322.8;  DB 7;  Length 510;
  Best Local Similarity   89.2%;  Pred. No. 1.9e-66;
  Matches  348;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy       3064 TTAAAACTCCTAAGAAAGCAAGATTTAATAGTCAAAGAAGTGAGTAAACATGAAATGCCT 3123
             |||||| || |||||||||||||||| |||||||||||||||||||| ||||| ||||
Db        509 TTAAAAGTCGTAAGAAAGCAAGATTTCATAGTCAAAGAAGTGAGTAAACTTGAAACGCCT 450

Qy       3124 GAACAGAGTAATGAGCTAAGCACAAAGTTAGAGACATGTTAGTTAATATGTCTTGAAAGC 3183
             ||| |||||||||| ||||||||||||||||||||||||||||||| ||||||||||
Db        449 GAAGAGAGTAATGAGTTAAGCACAAAGTTAGAGACATGTTAGTTAATAGGTCTTGAAAGC 390

Qy       3184 AGCAGCTCCTGCTTTCAAGGAGCAAGAACAAATTGGGCAAGTGAACACTCCTTGAATAAA 3243
             |||||||| || |||||||||||||||||||||||||||||||||||||| ||||||||
Db        389 AGCAGCTCGTGATTTCAAGGAGCAAGAACAAATTGGGCAAGTGAACACTCCGTGAATAAA 330

Qy       3244 ATGTGTAAAATTAATTTTGGGTTATGTTCTATACTGTGTATAATAGAATGATAAAAATTA 3303
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        329 ATGTGTAAAATTAATTTTGGGTTATGTTCTATACTGTGTATAATAGAATGATAAAAATTA 270

Qy       3304 TTTGACTAGCACTTTGTAGTTTAGAAATATCTCTATTTACACAGTTTACCTTATTTGATA 3363
             |||||||||||||||||    ||||| |||| |||| ||||||||||||||||||||||
Db        269 TTTGACTAGCACTTTGTAGGGAAGAAAGATCTTTATTGACACAGTTTACCTTATTTGATA 210

Qy       3364 AGACTGTTGAGTGATGGGATAGCATGGTGGACAATCCACATAACTGAGTATCGAGACACC 3423
             ||| || ||| |||||||||||||| |||||||||||||| | ||   ||    ||||| ||
Db        209 AGAATGGTGACTGATGGGATAGCAGGGTGGACAATCCATACAAGGCAGCCCTGAGACTCC 150

Qy       3424 TGTATCTGGACCCAGCTCTGTTAGTAAGAA 3453
             ||| ||| ||||||  |  |  | ||| |
Db        149 CGTACCTGTACCCAGGACCCTGGGAAAGGA 120


RESULT 3
BX914754
LOCUS       BX914754                798 bp    mRNA    linear   EST 07-MAY-2004
DEFINITION  BX914754 Sus Scrofa library (scan) Sus scrofa cDNA clone
            scan0028d.k.07 5prim, mRNA sequence.
ACCESSION   BX914754
VERSION     BX914754.1  GI:41131533
KEYWORDS    EST.
SOURCE      Sus scrofa (pig)
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 798)
  AUTHORS   Bonnet,A., Tosser-Klopp,G., Benne,F., Cabau,C., Villeger,S.,
            Soares,M., Bonaldo,F. and Hatey,F.
  TITLE     A Pig Normalised Multi-Tissue cDNA Library
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Tosser-Klopp G
            Genetique Animale
            Institut National de la Recherche Agronomique
            Chemin de Borde-Rouge - Auzeville BP27, 31326 Castanet-Tolosan
            cedex, FRANCE
            Tel: 33 (0) 5.61.28.51.14
            Fax: 33 (0) 5.61.28.53.08
            Email: tosser@toulouse.inra.fr
            Sequence cleaned of vector, adaptator and repetitions. Contact us
            at sigenasupport@jouy.inra.fr to obtain the chromatogram of this
            sequence.
            Plate: 0028  row: k  column: 7.
FEATURES             Location/Qualifiers
     source          1. .798
                     /organism="Sus scrofa"
```

```
                     /mol_type="mRNA"
                     /db_xref="taxon:9823"
                     /clone="scan0028d.k.07"
                     /tissue_type="mixed"
                     /dev_stage="from embryos to adults"
                     /clone_lib="Sus Scrofa library (scan)"
                     /note="tissues: adipose tissue, brain, kidney, liver,
                     muscle, ovary, testis,  heart, hypothalamus, pancreas,
                     skin, spleen, thymus, placenta, pituitary  gland, seminal
                     vesicle, small intestine, uterus, adrenals, bulbo  uretral
                     gland, cerebral trunk, epididymis, female gonad,
                     gall-bladder,  hippocampus, large intestine, male gonad,
                     melanocytes, stomach, udder"
ORIGIN

  Query Match             4.8%;  Score 293.2;  DB 5;  Length 798;
  Best Local Similarity   74.8%;  Pred. No. 3e-59;
  Matches  453;  Conservative    0;  Mismatches  133;  Indels   20;  Gaps    6;

Qy       5485 TATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGCCGAAGCCACAGCTCCTCCTGCCG 5544
             || |||| | |||||| ||||||||| |||||||| | ||||||||||||||||||||| |
Db          2 TAGCTTTCCGGGAGAAAAGGCAAAGTGCGCTCAAAACAGAAGCCACAGCTCCTCCTGCTG 61

Qy       5545 CATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAATAAGATGTGCAAAGGGCTTGCAGG 5604
             |||||||||||        |   |  |      |  | ||||||||||||||| || |||||
Db         62 CATTTCTTTCCTAACGTGAATCCCACA----CTTCACAAGATGTGCAAAGGACTCGCAGG 117

Qy       5605 TCTGCCGGCTTCTTGCTTGAGGAGGTAAGATTGCTTTCA-GCCATTAACCATATTAAACT 5663
             |||||||||||||||||||||||||||||||    ||||| ||  | |||| | |||||||
Db        118 TCTGCCGGCTTCTTGCTTGAGGAGGTAAGATGATTTTCACGCACTAAACCGTGTTAAACT 177

Qy       5664 TTTGGCTAGACTTTCTCAGTTATTTACATGTTGTACTTACTAACCTAGTTCTGTGCAATT 5723
             ||  | ||||.||| | |||||| | | | |  || |||||||||| ||| | |  | ||
Db        178 TTCAGATAGACTTTTTTAGTTATGTGCGTATCCTATTTACTAACCTGGTTGTTTAGAGTT 237

Qy       5724 AGAAACAGTGTGGTCAGGAGAGCACGA-CTTTCTAACTTTCCTCCAAGACTAGCTAGATA 5782
             |||||||||||||||||| |||  || | | |||  ||||||||||||||||||||||| |
Db        238 AGAAACAGTGTGGTCAGGAAAGCCTGAGCCTCCTAGCTTTCCTCCAAGACTAGCTAGAAA 297

Qy       5783 TTGTGACTTAAGACATGTGCTCCCCAAATTTCAGCCCTTATGTGTTGTTTTGTGTGACCT 5842
             |||||||||    |  ||| |||||| || ||  |  || |||||||| | ||| ||||||
Db        298 TTGTGACTTGGAAAGTGTTCTCCCCCAAATTTGGAACTCATGTGTTGCTGTGTATGACCT 357

Qy       5843 CAGTTTTGAGAACTGTTCTATTCTTTAAGCCAGGTCTAAGAAAGCTAGTTTTAATTAAGA 5902
               |||| ||| |||||||  ||| |   |||           ||||| |||  |||| |||
Db        358 TGGTTTGGAGGACTGTTCAGTTCCTCTGGCC----------AAGCTTGTTCAAATTGAGA 407

Qy       5903 AGCGAGATGAGGTTTGAG-GCTATGTACAGTGATCTGTAATATCTCCATCTGTGATTACT 5961
             || || |||||||  | ||| | |||||   |||| |||| ||| | |||||||    |
Db        408 AGTGATATAAGGTTTTTGTGCTTTATACAGCAGTCTGCAATACCTCTACCTGTGAT---T 464

Qy       5962 ACTGCTATTTGAGCATCCCTGGAGTACATAGAAGCCTGGCTCTGGGCTTTCTGATTGTAT 6021
             |||| |||||||   || |||   ||| ||||| | ||| ||||||||||||||| ||||
Db        465 ACTGTTATTTGAATGTCACTGAGCTACTTAGAATCATGGGTCTGGGCTTTCTGATGGTAT 524

Qy       6022 GCTACAACTTGTTTCAGGAAAGGTACCCCAGAATGAGGTTTGGCTCCATCATCAGAAAGG 6081
              ||| |  |  |||   ||| |  |||||||||||| ||||| || ||||| ||||||  |
Db        525 ACTAGATTTCATTTTGAGAAGGACACCCCAGAATGAAGTTTGACTACATCACCAGAAAAG 584

Qy       6082 CACTAT 6087
             ||||||
Db        585 CACTAT 590


RESULT 4
BX924658
LOCUS       BX924658                825 bp    mRNA    linear   EST 07-MAY-2004
DEFINITION  BX924658 Sus Scrofa library (scan) Sus scrofa cDNA clone
            scan0022d.p.16 5prim, mRNA sequence.
ACCESSION   BX924658
VERSION     BX924658.1  GI:41141506
KEYWORDS    EST.
SOURCE      Sus scrofa (pig)
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 825)
  AUTHORS   Bonnet,A., Tosser-Klopp,G., Benne,F., Cabau,C., Villeger,S.,
            Soares,M., Bonaldo,F. and Hatey,F.
  TITLE     A Pig Normalised Multi-Tissue cDNA Library
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Tosser-Klopp G
            Genetique Animale
            Institut National de la Recherche Agronomique
            Chemin de Borde-Rouge - Auzeville BP27, 31326 Castanet-Tolosan
            cedex, FRANCE
            Tel: 33 (0) 5.61.28.51.14
            Fax: 33 (0) 5.61.28.53.08
            Email: tosser@toulouse.inra.fr
            Sequence cleaned of vector, adaptator and repetitions. Contact us
            at sigenasupport@jouy.inra.fr to obtain the chromatogram of this
            sequence.
            Plate: 0022  row: p  column: 16.
FEATURES             Location/Qualifiers
     source          1. .825
                     /organism="Sus scrofa"
                     /mol_type="mRNA"
                     /db_xref="taxon:9823"
                     /clone="scan0022d.p.16"
                     /tissue_type="mixed"
                     /dev_stage="from embryos to adults"
                     /clone_lib="Sus Scrofa library (scan)"
                     /note="tissues: adipose tissue, brain, kidney, liver,
                     muscle, ovary, testis,  heart, hypothalamus, pancreas,
                     skin, spleen, thymus, placenta, pituitary  gland, seminal
                     vesicle, small intestine, uterus, adrenals, bulbo  uretral
                     gland, cerebral trunk, epididymis, female gonad,
                     gall-bladder,  hippocampus, large intestine, male gonad,
                     melanocytes, stomach, udder"
ORIGIN

  Query Match             4.8%;  Score 293.2;  DB 5;  Length 825;
  Best Local Similarity   74.8%;  Pred. No. 3.1e-59;
  Matches  453;  Conservative    0;  Mismatches  133;  Indels   20;  Gaps    6;

Qy       5485 TATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGCCGAAGCCACAGCTCCTCCTGCCG 5544
             || |||| | |||||| ||||||||| |||||||| | ||||||||||||||||||||| |
Db          2 TAGCTTTCCGGGAGAAAAGGCAAAGTGCGCTCAAAACAGAAGCCACAGCTCCTCCTGCTG 61

Qy       5545 CATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAATAAGATGTGCAAAGGGCTTGCAGG 5604
             |||||||||||        |   |  |      |  | ||||||||||||||| || |||||
Db         62 CATTTCTTTCCTAACGTGAATCCCACA----CTTCACAAGATGTGCAAAGGACTCGCAGG 117

Qy       5605 TCTGCCGGCTTCTTGCTTGAGGAGGTAAGATTGCTTTCA-GCCATTAACCATATTAAACT 5663
             |||||||||||||||||||||||||||||||    ||||| ||  | |||| | |||||||
Db        118 TCTGCCGGCTTCTTGCTTGAGGAGGTAAGATGATTTTCACGCACTAAACCGTGTTAAACT 177

Qy       5664 TTTGGCTAGACTTTCTCAGTTATTTACATGTTGTACTTACTAACCTAGTTCTGTGCAATT 5723
             ||  | ||||||||  | ||||||  | | |  || |||||||||| ||| | |  | ||
Db        178 TTCAGATAGACTTTTTTAGTTATGTGCGTATCCTATTTACTAACCTGGTTGTTTAGAGTT 237

Qy       5724 AGAAACAGTGTGGTCAGGAGAGCACGA-CTTTCTAACTTTCCTCCAAGACTAGCTAGATA 5782
             |||||||||||||||||| |||  || | | |||  ||||||||||||||||||||||| |
Db        238 AGAAACAGTGTGGTCAGGAAAGCCTGAGCCTCCTAGCTTTCCTCCAAGACTAGCTAGAAA 297

Qy       5783 TTGTGACTTAAGACATGTGCTCCCCAAATTTCAGCCCTTATGTGTTGTTTTGTGTGACCT 5842
             |||||||||    |  ||| |||||| || ||  |  || |||||||| | ||| ||||||
Db        298 TTGTGACTTGGAAAGTGTTCTCCCCCAAATTTGGAACTCATGTGTTGCTGTGTATGACCT 357

Qy       5843 CAGTTTTGAGAACTGTTCTATTCTTTAAGCCAGGTCTAAGAAAGCTAGTTTTAATTAAGA 5902
               |||| ||| |||||||  ||| |   |||           ||||| |||  |||| |||
Db        358 TGGTTTGGAGGACTGTTCAGTTCCTCTGGCC----------AAGCTTGTTCAAATTGAGA 407

Qy       5903 AGCGAGATGAGGTTTGAG-GCTATGTACAGTGATCTGTAATATCTCCATCTGTGATTACT 5961
             || || |||||||  | ||| | |||||   |||| |||| ||| | |||||||    |
Db        408 AGTGATATAAGGTTTTTGTGCTTTATACAGCAGTCTGCAATACCTCTACCTGTGAT---T 464
```

```
Qy       5962 ACTGCTATTTGAGCATCCCTGGAGTACATAGAAGCCTGGCTCTGGGCTTTCTGATTGTAT 6021
             |||| |||||||   || |||   ||| ||||| | ||| ||||||||||||||| ||||
Db        465 ACTGTTATTTGAATGTCACTGAGCTACTTAGAATCATGGGTCTGGGCTTTCTGATGGTAT 524

Qy       6022 GCTACAACTTGTTTCAGGAAAGGTACCCCAGAATGAGGTTTGGCTCCATCATCAGAAAGG 6081
              ||| |  |  |||   ||| |  |||||||||||| ||||| || ||||| |||||| |
Db        525 ACTAGATTTCATTTTGAGAAGGACACCCCAGAATGAAGTTTGACTACATCACCAGAAAAG 584

Qy       6082 CACTAT 6087
              ||||||
Db        585 CACTAT 590


RESULT 5
BP199581
LOCUS       BP199581                582 bp    mRNA    linear   EST 14-SEP-2004
DEFINITION  BP199581 Sugano cDNA library, brain Homo sapiens cDNA clone
            ADR08588, mRNA sequence.
ACCESSION   BP199581
VERSION     BP199581.1  GI:52047676
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 582)
  AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
  TITLE     Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
  JOURNAL   Genome Res. 14 (9), 1711-1718 (2004)
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .582
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="ADR08588"
                     /tissue_type="brain"
                     /clone_lib="Sugano cDNA library, brain"
ORIGIN

  Query Match             4.6%;  Score 277.6;  DB 5;  Length 582;
  Best Local Similarity   98.6%;  Pred. No. 1.7e-55;
  Matches  280;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       5345 CTTTCAGGGGCTGGAGAGGCAGAGGGAGACAGAGGAGCTGGTACTGCAGAGCGGTCGTCT 5404
              ||   | ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         31 CTCCTAAGGGCTGGAGAGGCAGAGGGAGACAGAGGAGCTGGTACTGCAGAGCGGTCGTCT 90

Qy       5405 GATTGGCTGGACGGTCGTAGCTGGGCTATAAAAGAGACCCCTACAGGCTTAGCAGGAAGA 5464
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         91 GATTGGCTGGACGGTCGTAGCTGGGCTATAAAAGAGACCCCTACAGGCTTAGCAGGAAGA 150

Qy       5465 CGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGCCGA 5524
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        151 CGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGCCGA 210

Qy       5525 AGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAATAAG 5584
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        211 AGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAATAAG 270

Qy       5585 ATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
              |||||||||||||||||||||||||||||||||||||||||||||
Db        271 ATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 314


RESULT 6
```

```
BP335994
LOCUS       BP335994                582 bp    mRNA    linear   EST 17-SEP-2004
DEFINITION  BP335994 Sugano cDNA library, coronary artery smooth muscle cell
            Homo sapiens cDNA clone SMR04438, mRNA sequence.
ACCESSION   BP335994
VERSION     BP335994.1  GI:52265553
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 582)
  AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
  TITLE     Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
  JOURNAL   Genome Res. 14 (9), 1711-1718 (2004)
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .582
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="SMR04438"
                     /tissue_type="coronary artery"
                     /cell_type="smooth muscle cell"
                     /clone_lib="Sugano cDNA library, coronary artery smooth
                     muscle cell"
ORIGIN

  Query Match             4.6%;  Score 277.6;  DB 5;  Length 582;
  Best Local Similarity   98.6%;  Pred. No. 1.7e-55;
  Matches  280;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       5345 CTTTCAGGGGCTGGAGAGGCAGAGGGAGACAGAGGAGCTGGTACTGCAGAGCGGTCGTCT 5404
              ||   | ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         86 CTCCTAAGGGCTGGAGAGGCAGAGGGAGACAGAGGAGCTGGTACTGCAGAGCGGTCGTCT 145

Qy       5405 GATTGGCTGGACGGTCGTAGCTGGGCTATAAAAGAGACCCCTACAGGCTTAGCAGGAAGA 5464
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        146 GATTGGCTGGACGGTCGTAGCTGGGCTATAAAAGAGACCCCTACAGGCTTAGCAGGAAGA 205

Qy       5465 CGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGCCGA 5524
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        206 CGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGCCGA 265

Qy       5525 AGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAATAAG 5584
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        266 AGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAATAAG 325

Qy       5585 ATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
              |||||||||||||||||||||||||||||||||||||||||||||
Db        326 ATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 369


RESULT 7
BP249954
LOCUS       BP249954                583 bp    mRNA    linear   EST 15-SEP-2004
DEFINITION  BP249954 Sugano cDNA library, hippocampus Homo sapiens cDNA clone
            HPR01013, mRNA sequence.
ACCESSION   BP249954
VERSION     BP249954.1  GI:52132233
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 583)
  AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
```

```
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
  TITLE     Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
  JOURNAL   Genome Res. 14 (9), 1711-1718 (2004)
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .583
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="HPR01013"
                     /tissue_type="hippocampus"
                     /clone_lib="Sugano cDNA library, hippocampus"
ORIGIN

  Query Match             4.6%;  Score 277.6;  DB 5;  Length 583;
  Best Local Similarity   98.6%;  Pred. No. 1.7e-55;
  Matches  280;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       5345 CTTTCAGGGGCTGGAGAGGCAGAGGGAGACAGAGGAGCTGGTACTGCAGAGCGGTCGTCT 5404
              ||   | ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         48 CTCCTAAGGGCTGGAGAGGCAGAGGGAGACAGAGGAGCTGGTACTGCAGAGCGGTCGTCT 107

Qy       5405 GATTGGCTGGACGGTCGTAGCTGGGCTATAAAAGAGACCCCTACAGGCTTAGCAGGAAGA 5464
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        108 GATTGGCTGGACGGTCGTAGCTGGGCTATAAAAGAGACCCCTACAGGCTTAGCAGGAAGA 167

Qy       5465 CGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGCCGA 5524
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        168 CGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGCCGA 227

Qy       5525 AGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAATAAG 5584
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        228 AGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAATAAG 287

Qy       5585 ATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
              |||||||||||||||||||||||||||||||||||||||||||||
Db        288 ATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 331


RESULT 8
BP213595
LOCUS       BP213595                 551 bp    mRNA    linear   EST 15-SEP-2004
DEFINITION  BP213595 Sugano cDNA library, cerebrum Homo sapiens cDNA clone
            CBR07747, mRNA sequence.
ACCESSION   BP213595
VERSION     BP213595.1  GI:52086496
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 551)
  AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
  TITLE     Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
  JOURNAL   Genome Res. 14 (9), 1711-1718 (2004)
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .551
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CBR07747"
                     /tissue_type="cerebrum"
                     /clone_lib="Sugano cDNA library, cerebrum"
ORIGIN

  Query Match             4.5%;  Score 276;  DB 5;  Length 551;
  Best Local Similarity   98.2%;  Pred. No. 4e-55;
  Matches  279;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       5345 CTTTCAGGGGCTGGAGAGGCAGAGGGAGACAGAGGAGCTGGTACTGCAGAGCGGTCGTCT 5404
              ||   | ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        116 CTCCTAAGGGCTGGAGAGGCAGAGGGAGACAGAGGAGCTGGTACTGCAGAGCGGTCGTGT 175

Qy       5405 GATTGGCTGGACGGTCGTAGCTGGGCTATAAAAGAGACCCCTACAGGCTTAGCAGGAAGA 5464
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        176 GATTGGCTGGACGGTCGTAGCTGGGCTATAAAAGAGACCCCTACAGGCTTAGCAGGAAGA 235

Qy       5465 CGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGCCGA 5524
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        236 CGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGCCGA 295

Qy       5525 AGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAATAAG 5584
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        296 AGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAATAAG 355

Qy       5585 ATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
              |||||||||||||||||||||||||||||||||||||||||||||
Db        356 ATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 399


RESULT 9
CD511718/c
LOCUS       CD511718                 298 bp    mRNA    linear   EST 06-JUN-2003
DEFINITION  AGENCOURT_14360795 NIH_MGC_187 Homo sapiens cDNA clone
            IMAGE:30405362 5', mRNA sequence.
ACCESSION   CD511718
VERSION     CD511718.1  GI:31443436
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 298)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
            National Cancer Institute / NIH
            Bldg. 31 Rm10A07 Bethesda, MD 20892
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Michael Brownstein and Dr. Miklos Palkovits
             cDNA Library Preparation: CLONTECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: NDCM192  row: g  column: 03
            High quality sequence stop: 298.
FEATURES             Location/Qualifiers
     source          1. .298
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:30405362"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_187"
                     /note="Organ: Blood vessels - aorta, basilar and artery;
                     Vector: pDNR-LIB; Site_1: SfiI (ggccattatggcc); Site_2:
                     SfiI (ggccgcctcggcc); 5' and 3' adaptors were used in
                     cloning as follows: 5' adaptor sequence:
                     5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
```

C, or G and N = A, C, G, or T). Average insert size 1.4 kb (range 0.5-4.0 kb). 14/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC Library."

ORIGIN

Query Match 4.5%; Score 272.8; DB 6; Length 298;
Best Local Similarity 97.5%; Pred. No. 2.1e-54;
Matches 277; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```
Qy       1931 TTTTTTTTTAATTTTTAAAATTTGTATATATTTATGGGTTATAAGTGAAATCTTTTTAGA 1990
             |||||||||  |||||    || ||||||||||||||||||||||||||||||||||||||
Db        286 TTTTTTTTTTTTTTTTTTTTTTGGTATATATTTATGGGTTATAAGTGAAATCTTTTTAGA 227

Qy       1991 TGCATAGGTTGTATAGTGATAAAATCAGGGCTTTTAGGGTATTCATCACCTGAATGATGT 2050
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        226 TGCATAGGTTGTATAGTGATAAAATCAGGGCTTTTAGGGTATTCATCACCTGAATGATGT 167

Qy       2051 ACATTGTACCCCTTAAGTAATTTCTCACCATCCGCTGACTTCTTGCCCCCTGGGTATTCA 2110
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        166 ACATTGTACCCCTTAAGTAATTTCTCACCATCCGCTGACTTCTTGCCCCCTGGGTATTCA 107

Qy       2111 TCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCTGACTTCTTG 2170
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        106 TCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCTGACTTCTTG 47

Qy       2171 CCCCCTGGGTATTCATCACCTGAATGATGTGCATTGTACCCCTT 2214
             |||||||||||||||||||||||||||||||||||||||||||||
Db          46 CCCCCTGGGTATTCATCACCTGAATGATGTGCATTGTACCCCTT 3
```

RESULT 10
AQ745626/c

```
LOCUS       AQ745626                 864 bp    DNA     linear   GSS 16-JUL-1999
DEFINITION  HS_2273_A2_B06_T7C CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=2273 Col=12 Row=C, genomic survey
            sequence.
ACCESSION   AQ745626
VERSION     AQ745626.1  GI:5523148
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 864)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
   PUBMED   10449764
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones may be purchased from Research Genetics (info@resgen.com).
            BAC end Web Server: http://www.htsc.washington.edu
            Plate: 2273  row: C  column: 12
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 864.
FEATURES             Location/Qualifiers
     source          1. .864
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="Plate=2273 Col=12 Row=C"
                     /sex="male"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
```

ORIGIN

Query Match 4.4%; Score 268.4; DB 8; Length 864;
Best Local Similarity 70.7%; Pred. No. 3e-53;
Matches 454; Conservative 0; Mismatches 156; Indels 32; Gaps 6;

```
Qy        279 TTCATTTATCTCTCCTTCTCTCTCAAAGCTGCAGTAGGGGTGAAAACGTGTGATACATTT 338
              || |  |||   | ||    || ||         ||| |||   ||| ||||||||  |||
Db        703 TTTAAATATTGTTTCTGAATTCCCATTTAGTTTGTAAGGGCTGAAAGGTGTGATATCTTT 644

Qy        339 TCTCTTCCATCATAAGGGTCGCAACCAAAACTCCTATAGTAAAAGACAGGTTAATAAGAG 398
              |||  || | ||||   || || |||| |||||||||  ||||  |||| ||| ||||
Db        643 CCTCACCCGTTATAATCATCACAGCCAACACTCCTATAACAAAAACCAGGATAACAAGAA 584

Qy        399 CAAAACCTAACAAATTTATTTAATCAAAGTTTTACATGACATGGGAGTCTTCAGAAATGA 458
              ||      |||| |||| ||||||||| ||||||||||| |||| ||||||||||||
Db        583 AAA-----AACATATTTCTTTAATCAAACTTTTACATGACATAGGAGGCTTCAGAAATGA 529

Qy        459 AGACCCAAAGACCCAGGGGAAACTGTCTGTTTTTTTTGCTGAGGTTCGATGAAGAATGGA 518
              | ||| ||| | |||||| |||   | || ||||| | || |||||| ||||||||| ||
Db        528 AAACCTAAAAATCCAGGGAAAATAATTTG-TTTTTATACTTAGGTTCAATGAAGAATAGA 470

Qy        519 TAGCATGTAGCCATGTAGATTAGACAAAAGGATATGATCTAGTGGTAAAGGACTCAGGG- 577
              |||        |   ||||||||||||| |||||||| ||||||  |||| ||||
Db        469 CAGCTGTGTAGAAATGTGATTAGACAAAAGGGTATGATCTAATGGTAACAGACTGAGGGA 410

Qy        578 GGAAACACAGCAAGGCCTGTCTATTCAGATTCTTCTTGATCTCTCTCTCTCTATGTATAG 637
              ||| |  |||||| |||||  | ||||||||||||||  |||||| |
Db        409 GGAGATCCAGCAAAGCCTGCTTGTTCAGATTCTTCTTGGCCTCTCTGT---------GCA 359

Qy        638 CATTCTTTCCTCCTGAGTATGGGGCAGGACTCTTCTTCAATGAGGGTCTTCAAGGGAGAA 697
              |  |||||||    | |||| |||||  ||| |  |||||||| ||||||| |||||
Db        358 ACATTCTTCCTCCCAGATTTGGGCCAGGATCCTTATGGAATGAGGGCCTTCAAGAGAGAA 299

Qy        698 GGGAGAA-------AGTGGCCTTTTTAGATTTTATGGCTTGCTTCGGGGAAGAGGAGTTC 750
              ||| ||       ||| |||||| ||| || |||||||||| |||| ||| |  |||
Db        298 TGGAAAAGGGAAAGAGTAGCCTTTCTAGGATTGATGGCTTGCTTTGGGGGAGAAGGATTC 239

Qy        751 TAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCTGGTTTCTGTGACTTGCTTTCATGA 810
              |||||||||||||||  || |||||  |||||||||| ||| |||||  |||     |
Db        238 TAGTTTCTATGACCCACATTAGGGAAATGGAATTCTGGTCTCTATGACTCGCTTCAGGGG 179

Qy        811 AGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGGTCAGAAAGAGACT---------TG 861
              |||| |||| | | ||||||||||||||| ||| || || | |||| ||          |
Db        178 AGAAGGAGGGGCAGGAGGCAGGAGGGCAGAAGAAGGGCAAAGAGAGGCTATGCTTCTGGG 119

Qy        862 GCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAGTACT 903
              |||| |||||||| |||||  |||||||||| || |||||||
Db        118 GCTGTTTCTGAGGCCTTCCAGTCTCCTTTAGCTCAAAGTACT 77
```

RESULT 11
BI547160

```
LOCUS       BI547160                 787 bp    mRNA    linear   EST 05-SEP-2001
DEFINITION  603190405F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:5261664 5',
            mRNA sequence.
ACCESSION   BI547160
VERSION     BI547160.1  GI:15434472
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 787)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
```

Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate: LLAM11659 row: j column: 01
High quality sequence stop: 721.

FEATURES Location/Qualifiers
source 1. .787
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5261664"
/tissue_type="hippocampus"
/lab_host="DH10B"
/clone_lib="NIH_MGC_95"
/note="Organ: brain; Vector: pBluescriptR (modified pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTVN-3', size-selected for average insert size 2.5 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIMH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."

ORIGIN

Query Match 4.3%; Score 262.8; DB 4; Length 787;
Best Local Similarity 99.2%; Pred. No. 6.7e-52;
Matches 264; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy       5363 GCAGAGGGAGACAGAGGAGCTGGTACTGCAGAGCGGTCGTCTGATTGGCTGGACGGTCGT 5422
             | || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          4 GGAGGGGGAGACAGAGGAGCTGGTACTGCAGAGCGGTCGTCTGATTGGCTGGACGGTCGT 63

Qy       5423 AGCTGGGCTATAAAAGAGACCCCTACAGGCTTAGCAGGAAGACGCTCAGAGGATTCTGAC 5482
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         64 AGCTGGGCTATAAAAGAGACCCCTACAGGCTTAGCAGGAAGACGCTCAGAGGATTCTGAC 123

Qy       5483 AATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGCCGAAGCCACAGCTCCTCCTGC 5542
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        124 AATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGCCGAAGCCACAGCTCCTCCTGC 183

Qy       5543 CGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAATAAGATGTGCAAAGGGCTTGCA 5602
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        184 CGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAATAAGATGTGCAAAGGGCTTGCA 243

Qy       5603 GGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             |||||||||||||||||||||||||||
Db        244 GGTCTGCCGGCTTCTTGCTTGAGGAG 269
```

RESULT 12
CO895317/c
LOCUS CO895317 453 bp mRNA linear EST 01-SEP-2004
DEFINITION BovGen_23642 normal cattle brain Bos taurus cDNA clone RZPDp1056K0348Q 5', mRNA sequence.
ACCESSION CO895317
VERSION CO895317.1 GI:51825631
KEYWORDS EST.
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 453)
AUTHORS Hennig,S., Janitz,M., Herwig,R. and Williams,J.
TITLE Generation, annotation, evolutionary analysis and database integration of 14969 cattle EST clusters
JOURNAL Unpublished (2004)
COMMENT Contact: Hennig S
laboraty 123, dept.Lehrach
Max-Planck-Institut fuer Molekulare Genetik
Ihnestr.63-73, D-14195 Berlin, Germany
Tel: +49 30 8413 1612
Fax: +49 30 8413 1380
Email: hennig@molgen.mpg.de
The library was characterised by oligonucleotide fingerprinting (ONFP) to reduce sequencing redundancy. According to the ONFP procedure, clones that display the same hybridisation matrix with a battery of 200 8mer oligonucleotides are grouped into clusters. One clone per ONFP cluster was selected for sequencing. cDNA clones and filters are distributed via Deutsches Ressourcenzentrum fuer Genomforschung GmbH (http://www.rzpd.de).
PCR PRimers
FORWARD: 5' CCCCAGGCTTTACACTTTATGCTTCCGGCTCG 3' (M13RSP) 5'-seq
BACKWARD: 5' GCTATTACGCCAGCTGGCGAAAGGGGGATGTG 3' (M13FSP) 3'-seq
Seq primer: 5'-CCGGTCCGGAATTCCCGGGT-3' (M13RSP).

FEATURES Location/Qualifiers
source 1. .453
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/clone="RZPDp1056K0348Q"
/sex="female"
/tissue_type="brain tissue"
/dev_stage="adult brain"
/clone_lib="normal cattle brain"
/note="Organ: brain; Vector: pSport1; Site_1: NotI; Site_2: SalI; Random primed and directionally cloned in pSport1 vector using NotI (5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCC (T)15-3' and SalI 5'-TCGACCCACGCGTCCG-3' adapters (Gibco BRL)"

ORIGIN

Query Match 4.0%; Score 242.4; DB 7; Length 453;
Best Local Similarity 93.8%; Pred. No. 5e-47;
Matches 274; Conservative 0; Mismatches 16; Indels 2; Gaps 2;

```
Qy       3044 CAAACACTTTTTTAAAAGTCTTAAAACTCCTAAGAAAGCAAGATTTAATAGTCAAAGAAG 3103
             ||||||   ||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        453 CAAACAGGGTTTTAAAAGTCTTAAAACTCCTAAGAAAGCAAGATTTTATAGTCAAAGAAG 394

Qy       3104 TGAGTAAACATGAAATGCCTGAACAGAGTAATGAGCTAAGCACAAAGTTAGAGACATGTT 3163
             ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        393 TGAGTAAACATGAAATGCCTGAACAGAGTAATGAGCTAAGCACAAAGTGAGAGACATGTT 334

Qy       3164 AGTTAATATGTCTTGAAAGCAGCAGCTCCTGCTTTCAAGGAGCAAGAACAAATTGGGCAA 3223
             |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db        333 AGTTAATATGTCTTGAAAGCAGCAGCTCTTGCTTTCAAGGAGCAAGAACAAATTGGGCAA 274

Qy       3224 GTGAACACTCCTTGAATAAAATGTGTAAAATTAATTTTGGG-TTATGTTCTATACTGTGT 3282
             |||||||||||||||||||||||||||||||||||||||| |||| || | || |||||
Db        273 GTGAACACTCCTTGAATAAAATGTGTAAAATTAATTTTGGGTTTATTTTTTTTATTGTGT 214

Qy       3283 ATAATAGAATGATAAAAATTATTTGA-CTAGCACTTTGTAGTTTAGAAATAT 3333
             |||||||||||||||||||  ||||  ||||| |||||||||||||| | ||
Db        213 ATAATAGAATGATAAAAATTTATTGATTTAGCATTTTGTAGTTTAGATAAAT 162
```

RESULT 13
AG090278/c
LOCUS AG090278 647 bp DNA linear GSS 03-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-089M15.R, genomic survey sequence.
ACCESSION AG090278
VERSION AG090278.1 GI:16642080
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE 1

```
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 647)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                 Sequencing: M13Rev
             LIBRARY
                 Vector    : pKS145
                 R.Site 1  : SacI
                 R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1. .647
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="PTB-089M15.R"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
ORIGIN

  Query Match             3.9%;  Score 240.4;  DB 9;  Length 647;
  Best Local Similarity   69.3%;  Pred. No. 1.6e-46;
  Matches  424;  Conservative    0;  Mismatches  161;  Indels   27;  Gaps    6;

Qy        334 CATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTCCTATAGTAAAAGACAGGTTAAT 393
              | ||| |||  |||||||||||||| | ||  | | |  ||||  ||||||||||||||
Db        647 CCTTTCCTCACCCATCATAAGGGTCACGACTGACATTTTTATAACAAAAGACAGGTTAAC 588

Qy        394 AAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTACATGACATGGGAGTCTTCAGA 453
               |||| ||| | |||||| ||||||||||  ||||| |||||||||   ||| ||| ||
Db        587 CAGAGAAAAGCATAACAATTTTATTTAATTGAAGTTGTACATGACACATGAGCCTTTAGG 528

Qy        454 AATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTTTTTTGCTGAGGTTCGATGAAGA 513
              |    |   || | |   || ||| |||||      ||||| |||| ||||||  ||||||
Db        527 ATGAAGACCCAAGA---CCTGGGAAAACT-----ATTTTATGCTTAGGTTCAGTGAAGA 476

Qy        514 ATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGATATGATCTAGTGGTAAAGGACTC 573
              |   | ||        |    |  | || |  |   |      |  |  || |
Db        475 ACAAACGGCCATGTAGAAATGTGGCCAAAAAACTGTGTTCTAATGGTAAGAGACCGAGTG 416

Qy        574 AGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTTCTTGATCTCTCTCTCTCTATGT 633
                | |     | || ||| |||| | ||||||||||| || | | ||        |||
Db        415 GAGAAGGCCAGCCGCGAGGACTGTGTGTTCAGATTCTTCCTGGTTTTTC--------TGT 364

Qy        634 ATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTTCTTCAATGAGGGTCTTCAAGGG 693
              ||||||||||||||||||| || | |||||||||| | |||  ||||||| ||||||| ||
Db        363 GTAGCATTCTTTCCTCCTGGGTGTCGGGCAGGACCCCTCTGGAATGAGGTTCTTCAAAGG 304

Qy        694 AGAAGGGAGAAAGTGGCCTTTTTAGATTTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAG 753
              |||||||||| ||||  |||| ||| || |||||||||||| ||||| || || ||||||
Db        303 AGAAGGGAGAGAGTGATCTTTGTAGGTTCTATGGCTTGCTTTGGGGAGGA-GAATTCTAG 245

Qy        754 TTTCTATGACCCATCTTGGGGAAGAGGAATTCTGGTTTCTGTGACTTGCTTTCATGAAGA 813
              |||||||||||  | |||||||||||||||||||||||  |  |||| ||||  || |||
Db        244 TTTCTATGACCTGTGTTGGGGAAGAGGAATTCTGGTTTGTAGGACTCGCTTTGGTGGAGA 185

Qy        814 AAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGGTCAGAAAGAGACTTGGCTGCTTCTGAG 873
              |||||| | ||||| |||||||||||| | | || ||||       |  ||||||||||
Db        184 AAGAGGGGCAAGAGACAGGAGGGCAGGAAAAGATCGGAAA-------GTTTGCTTCTGAG 132

Qy        874 GGCTTCCGCTCTCCTTTAGTTCCAAGTACTTCTTAGCATACCAAAGCACTATACTTTGGC 933
              |||||  |||||||||||  |||||    |  |  || ||||||| | ||||| |
Db        131 ATCTTCCAGTCTCCTTTAGTTTGAAGTA---TTCGGTGTATCAAAGCACCAGACTTTCGA 75

Qy        934 ATATGGTTTTCT 945
              ||| || ||||
Db         74 GTATTGTCTTCT 63


RESULT 14
AQ274674
LOCUS       AQ274674                712 bp    DNA     linear   GSS 26-JAN-2001
DEFINITION  RPCI-6-112A10Sp6 RPCI-6 Homo sapiens genomic clone
            RPCI-6-112A10Sp6, genomic survey sequence.
ACCESSION   AQ274674
VERSION     AQ274674.1  GI:3851781
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 712)
  AUTHORS   Beck,A., Cairo,A., Catanese,J., Osoegawa,K., van Geel,M., Wu,C. and
            de Jong,P.J.
  TITLE     BAC and PAC end sequence database for sequence-ready map building
  JOURNAL   Unpublished (1998)
COMMENT     Contact: de Jong, P.J.
            Children's Hospital Oakland Research Institute
            747 Fifty second Street, Oakland, CA 94609-1809, USA
            Tel: 510 450 7911
            Fax: 510 450 7924
            Email: pdejong@mail.cho.org
            PAC end sequences. For clone availability please contact Pieter de
            Jong (pdejong@.mail.cho.org). BACPAC Resources WWW site:
            www.chori.org/bacpac
            Seq primer: Sp6
            Class: PAC end.
FEATURES             Location/Qualifiers
     source          1. .712
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="RPCI-6-112A10Sp6"
                     /sex="female"
                     /cell_type="lymphocytes"
                     /lab_host="E. coli DH10B"
                     /clone_lib="RPCI-6"
                     /note="Vector: pPAC4; PAC clones in E. coli DH10B. For PAC
                     library availability, please contact Pieter de Jong
                     (BACPACorders@mail.cho.org). Clones may be purchased from
                     BACPAC Resources (www.chori.org/bacpac) or from Research
                     Genetics (info@resgen.com)"
ORIGIN

  Query Match             3.7%;  Score 226.4;  DB 8;  Length 712;
  Best Local Similarity   71.5%;  Pred. No. 4e-43;
  Matches  418;  Conservative    0;  Mismatches  137;  Indels   30;  Gaps    8;

Qy        316 GGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTCCTAT 375
              |||    || ||||||||| ||| ||| || | ||||||| |  |  |||| ||| ||||
Db         56 GGGATGGAAAGTGTGATACCTTTCCTCATCTACCATAAGGATTACGGCCAACACTGCTAT 115

Qy        376 AGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTACAT 435
              |  |||||  ||||||| ||||| ||| | |||      ||  ||       ||||| ||||
Db        116 AACAAAAG--AGGTTAACAAGAGAAAAGCATAAATGCATTTATTTCATC-AGTTTCACAT 172

Qy        436 GACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTTTTTT 495
              ||||  | || |||||||||||||| |||||||||||| ||||||     | ||| |
Db        173 GACACAGAAGCCTTCAGAAATGAAGACTCAAAGACCCAGGGAAAACTG-----TATTTGT 227

Qy        496 GCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGATATGA 555
              ||| |||||| ||||||||||  ||| |      |    |||| ||||||||| | |||
Db        228 GCTTAGGTTCCATGAAGAATGGGAAGCCTTGTAGAAATGTGATTGGACAAAAGGGTGTGA 287
```

```
Qy        556 TCTAGTGGTAAAGGACTCAGGG--GGAAACACAGCAAGGCCTGTCTATTCAGATTCTTCT 613
              |||| ||||||  |||| ||||  |||||| |||||||||||||||| ||||||||||||
Db        288 TCTACTGGTAACAGACTGAGGGGAGGAAACCCAGCAAGGCCTGTCTGTTCAGATTCTTCT 347

Qy        614 TGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTTCT 673
              |  | ||||         ||  |||||||| | |||||          || |||||   |||
Db        348 TAGTGTCTC--------TGAGTAGCATTCCTACCTCC-------CAGGTAGGACCTCTCT 392

Qy        674 TCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATGGCTTGCT 733
              |||||||| ||||||||||||||||  |||| ||||| ||| |||| |||||||||
Db        393 GGAATGAGGGTTTTCAAGGGAGAAGGGAGCGAGTGACCTTTCTAGGTTTTGTGGCTTGCT 452

Qy        734 TCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCTGGTTTCT 793
              | |  | |||||  |||||||||||||||||||  |||| ||| | |||||||| ||||||
Db        453 TTGCAGGAGAGGCATTCTAGTTTCTATGACCCACTTTGGCGAAAAAGAATTCTTGTTTCT 512

Qy        794 GTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGGTCAGAAA 853
              ||||||  | |    | ||||| |||  |  ||| | ||| |     |||| ||   |  |
Db        513 ATGACTTACGTCGGGGGAGAAAAAGGGATGGGAGACGGGGGCAGGAGAGAAGG---GTGA 569

Qy        854 GAGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAA 898
              || |||||| ||     |  |||||  |||||||||||| ||
Db        570 AAGCCTTGGCTTCT--GAGGNCCTTCCATTCTCCTTTAGTTCAAA 612
```

RESULT 15
AW102832

LOCUS	AW102832 474 bp mRNA linear EST 19-OCT-1999
DEFINITION	xd38h03.x1 NCI_CGAP_Ov23 Homo sapiens cDNA clone IMAGE:2596085 3' similar to contains MER40.b2 MER40 repetitive element ;, mRNA sequence.
ACCESSION	AW102832
VERSION	AW102832.1 GI:6073445
KEYWORDS	EST.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 474)
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL	Unpublished (1997)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-r@mail.nih.gov Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html Possible reversed clone: polyT not found Seq primer: -40UP from Gibco High quality sequence stop: 409.

FEATURES	Location/Qualifiers
source	1. .474 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="IMAGE:2596085" /tissue_type="tumor, 5 pooled (see description)" /lab_host="DH10B" /clone_lib="NCI_CGAP_Ov23" /note="Organ: ovary; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.35 kb. Tumor types include: mixed Mullerian tumor, papillary serous, clear cell, spindle cell. All are primary tumors, metastasis positive. Life Technologies catalog #: 11534-013"

ORIGIN

Query Match 3.2%; Score 194.8; DB 2; Length 474;
Best Local Similarity 73.5%; Pred. No. 1.6e-35;
Matches 322; Conservative 0; Mismatches 97; Indels 19; Gaps 5;

```
Qy        411 AATTTATTTAATCAAAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGAC 470
              ||   ||||| | |||||||||||| |||| |||| | ||||||||||||||  ||
Db         19 AACAAATTTATTTAAAGTTTTACATTACATAGGAG-CCTCAGAAATGAAGACC--AAAGA 75

Qy        471 CCAGGGGAAACTGTCTGTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCC 530
              ||  || ||| ||      ||| | |||| | ||| ||||||||||||| |||
Db         76 CCCAGGAAAAATGATCTATTTGTATGCTTATGTTTGATGAAGAATGGACAGCCATATAGA 135

Qy        531 ATGTAGATTAGACAAAAGGATATGATCTAGTGGTAAAGGACT-CAGGGGGAAACACAGCA 589
              |    | |       |||||| || |||  | || ||   |||| |||||| |||| |||||
Db        136 AATGTGTTGGACAAAAAGGGTACGATTGAATGATAGTAGACTGCAGGGGAAAACCCAGCA 195

Qy        590 AGGCCTGTCTATTCAGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTC 649
              |||||||||| |||||| ||||||||  |||||         ||| ||||||||||||||
Db        196 AGGCCTGTCTGTTCAGAGTCTTCTTGGCCTCTC--------TGTGTAGCATTCTTTCCTC 247

Qy        650 CTGAGTATGGGGCAGGACTCTTCTTCAATGAGGGTCTTCAAGGGAG-------AAGGGAG 702
              ||| || |||| |||||| | |||  ||||||||||||||||||  ||       |||||||
Db        248 CTGGGTGTGGGACAGGACCCCTCTGGAATGAGGGTCTTCAAGGAAGAAGGGCAAAGGGAG 307

Qy        703 AAAGTGGCCTTTTTAGATTTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGA 762
              | | || ||||| ||| |||| |||||||||| |||| | ||| ||||||||||||||||
Db        308 AGAATGACCTTTCTAGGTTTTTTGGCTTGCTTTGGGGGATAGGGGTTCTAGTTTCTATGG 367

Qy        763 CCCATCTTGGGGAAGAGGAATTCTGGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGT 822
              ||| |||||||||||| |||| |||||||| || ||  |||    | ||||||||  ||
Db        368 CCCGCCTTGGGGAAGAGAAATTTTGGTTTCTATGCCTCACTTCAGGGGAGAAAGAGAGGT 427

Qy        823 AAGAGGCAGGAGGGCAGG 840
              |||| |||||||||| ||
Db        428 GAGAGACAGGAGGGCGGG 445
```

RESULT 16
AQ282818

LOCUS	AQ282818 734 bp DNA linear GSS 27-APR-1999
DEFINITION	RPCI11-89G8.TV RPCI-11 Homo sapiens genomic clone RPCI-11-89G8, genomic survey sequence.
ACCESSION	AQ282818
VERSION	AQ282818.1 GI:3909290
KEYWORDS	GSS.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 734)
AUTHORS	Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
TITLE	Use of human BAC End Sequences for Sequence-Ready Map Building
JOURNAL	Unpublished (1998)
COMMENT	Other_GSSs: RPCI11-89G8.TJ Contact: Mark Adams Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: mdadams@tigr.org Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genetics (info@resgen.com). BAC end search page: http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html Seq primer: T7 Class: BAC ends.

FEATURES	Location/Qualifiers

```
     source          1. .734
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="GDB:7533943"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-89G8"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /clone_lib="RPCI-11"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
ORIGIN

  Query Match             3.1%;  Score 189.2;  DB 8;  Length 734;
  Best Local Similarity   66.0%;  Pred. No. 3.9e-34;
  Matches  289;  Conservative    0;  Mismatches  148;  Indels    1;  Gaps    1;

Qy        190 CTACCTCTTCTCCTCCATTAACTACAGAGTAAATTCTGGTCTTCAGAGTAACAAGAAACA 249
              |||    ||| | |   |||| |   |   ||  ||  | |  |  ||  ||     |
Db        280 CTAGACTTTCACTTATTTTAATTGATGCACAATGGCTTATTTAAAATGTTTCAGCCTTAA 339

Qy        250 CCCTTTAGTTCTCAGCATATTCGTGCACCTTCATTTATCTCTCCTTCTCTCTCAAAGCTG 309
              |||     || | || | ||     |  |||   | || |     |     ||    | |
Db        340 TCCTCCCTCTCCCTGCTTTTTGTGAGATTTTCTGGTCTCACCAACTACAGTTCTTTCCAG 399

Qy        310 CAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAAC 369
              | ||| || |||            |||   | |||||||||||||| || |||| ||
Db        400 GTTTGGGGATGGAAAAAGCGTGGTACCTTTTCATCCCATCATAAGGGTCACAGCCAACAC 459

Qy        370 TCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTT 429
              |||||||  |||| ||||||||| | ||| ||| | ||||| |||| ||||||||||||
Db        460 TCCTATAACAAAAAACAGGTTAACATGAGAAAAGCATAACACATTTCTTTAATCAAAGTT 519

Qy        430 TTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTT 489
              ||| |||||||||||| |||||||||||||||| |||||||||||  |||||||||  |
Db        520 TTATATGACATGGGAGGCTTCAGAAATGAAGACCTAAAGACCCAGGGAAAACTGTCT-AT 578

Qy        490 TTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAGG 549
              ||||||||| ||| || |||||||||||| |||         |    |||| ||| |||||
Db        579 TTTTTTGCTTAGGATCAATGAAGAATGGACAGCCATGTAGAAATGTGATTGGACCAAAGG 638

Qy        550 ATATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTC 609
              | ||||||| |||| |  |  |    ||||||| |||||||||||| || |||||||||
Db        639 TTGTGATCTAATGGTGATAGGTTGCATGGGAAACCCAGCAAGGCCTGACTGTTCAGATTC 698

Qy        610 TTCTTGATCTCTCTCTCT 627
              ||||||  |||| | | |
Db        699 TTCTTGGCCTCTTTGTGT 716


RESULT 17
CC904986/c
LOCUS       CC904986               852 bp    DNA     linear   GSS 08-AUG-2003
DEFINITION  t018121ba.f1 TAMBT Bos taurus genomic clone t018121ba, genomic
            survey sequence.
ACCESSION   CC904986
VERSION     CC904986.1  GI:33523919
KEYWORDS    GSS.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
            Bovinae; Bos.
REFERENCE   1  (bases 1 to 852)
  AUTHORS   Lin,S., Najar,F.Z., Adelson,D., Gill,C.A. and Roe,B.A.
  TITLE     Bovine BAC End Sequences from Library TAMBT
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Bruce A. Roe
            Advanced Center for Genome Technology
            University of Oklahoma Department of Chemistry and Biochemistry
            620 Parrington Oval, Room 208, Norman, OK 73019, USA
            Tel: 405 325 4912
            Fax: 405 325 7762
            Email: broe@ou.edu
            Class: BAC ends
            High quality sequence start: 109
            High quality sequence stop: 408.
FEATURES             Location/Qualifiers
     source          1. .852
                     /organism="Bos taurus"
                     /mol_type="genomic DNA"
                     /strain="Angus bull T A M U Shoshone Y6 11519666"
                     /db_xref="taxon:9913"
                     /clone="t018121ba"
                     /sex="Male"
                     /cell_type="Blood"
                     /clone_lib="TAMBT"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII; TAMBT Bovine BAC library (Male) produced by Texas
                     A&M University, Department of Animal Science."
ORIGIN

  Query Match             3.0%;  Score 181.2;  DB 9;  Length 852;
  Best Local Similarity   65.3%;  Pred. No. 3.5e-32;
  Matches  372;  Conservative    0;  Mismatches  168;  Indels   30;  Gaps    6;

Qy       2478 TTGCTCTTACTTCTATTTGGAATACTAATTCCTAGGCTTCTTGCATTGCTTTCTCCTTCT 2537
              | |||||||| ||||||||| ||| || ||| || |||| |  ||||   |  || | |
Db        622 TGGCTCTTACCTCTATTTGGTATA-TATTTCGTAAGCTTGTCTCATTTGGTATTCGTGGT 564

Qy       2538 CACCCATC-AAATCTCATTTTAGATACCACCTCTTCAAAGAGGGCTTTCCTGACCACCTT 2596
              || ||||| |||||||||||| ||||| ||| |  |||| |||| ||||||
Db        563 CAGCCATCAAAATCTCATTTCAGATATCACTTGGTCAAGGAGGACTTTCC---------- 514

Qy       2597 GGCTGAATTAGCCCTTCACCATCTGATTACTCTCTAGCACATCACCTGCCCATTTTATTC 2656
              | ||   | || ||      || | ||  | | || ||       |   || ||| ||
Db        513 --CAGATACACCCTTTGCTATTCAGTTTCTTTTGTATCA------GTTGTCAGTTTTTTT 462

Qy       2657 ATGGTACAGGTCAAAATCTGGAATCACCTGATTTGTTTATTTTCTGACTCCTTCTACTGA 2716
                | | |   | ||| |||| |  |   | ||||  |  ||   |  ||||||||
Db        461 TCATGATATGGGTGACTCTAGAATTATTTTTATAGTTT-CTGACTCTTTTTTTCTACTGG 403

Qy       2717 GATGAAAACTCTACTAGAGCGGAGATTTTATCTGCTTGTATCAGGTACTG---------C 2767
              | |||||||| ||||| | || ||   | | ||||||       ||||          |
Db        402 GGTGAAAACTCCACTAGGGTGGGGACCCTGTTTGCTTGCTCATCATACTAACAGCAGTTC 343

Qy       2768 TTCAAACAGCACCTGATACAGAGTAGGTGGTCAAAAGATATTTCTTAAACAAATGAACAA 2827
              ||||||| | |||||    |||||||| | |||||| |||| | |||||  || |||| ||
Db        342 TTCAAACGGTACCTGTCCCAGAGTAGGAGCTCAAAATATATGTATTAAATAAGTGAATAA 283

Qy       2828 ATAAAAAGTAGATCTTTTGAGAGTAAAGCTCTTCCACACTACCAGAGTCATTCAGGAATG 2887
              ||||||| || ||||||| ||| |||||||||| ||||  ||||| |||| ||||||||
Db        282 ATAAAAATTAAATCTTTTCAGAATAAAGCTCTTTCACATCACCAGAATCATCCAGGAATG 223

Qy       2888 ACAAATCATAGAATAACAGAATTTGATGCTTTGTGCATATCAGAGAAAGAAGGTGGAAGG 2947
              |||  || |||||||| ||| | | ||||||| ||||| |||||||||| |||||||
Db        222 ACAGGTCTTAGAATAATAGAGTGTAATGCTTTCTGCATGTCAGAGAAAGAGGGTGGAAAT 163

Qy       2948 TTGTCAAGGTATCATGATGTACCAGTCCTCGCCTCCTCAAACACAATCTGCAAGTCCCAC 3007
              |||||||||||||||||||||||| |  | ||||||  |    ||  | ||      |
Db        162 TTGTCAAGGTATCATGATGTACCAATATTTGCCTCCCATAGTGAAAAATACATTATGAAA 103

Qy       3008 AGTGAAAAAGTAAGTTAACTCATGTGAAGC 3037
              ||     | |   |     | |||||| ||
Db        102 AGGTTGCATGCCTGCAGGTTGATGTGACGC 73


RESULT 18
AQ472918/c
LOCUS       AQ472918               437 bp    DNA     linear   GSS 23-APR-1999
DEFINITION  CITBI-E1-2589A4.TR CITBI-E1 Homo sapiens genomic clone 2589A4,
            genomic survey sequence.
ACCESSION   AQ472918
VERSION     AQ472918.1  GI:4656484
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
```

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 437)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., Shizuya,H., Simon,M. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences from CalTech Libraries for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: CITBI-E1-2589A4.TF
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13 Reverse
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .437
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="2589A4"
                     /sex="male"
                     /cell_type="sperm"
                     /clone_lib="CITBI-E1"
                     /note="Vector: pBeloBAC11; Site_1: EcoRI; Site_2: EcoRI;
                     CalTech Human BAC Library D"
ORIGIN

  Query Match          2.9%;  Score 179.2;  DB 8;  Length 437;
  Best Local Similarity   72.9%;  Pred. No. 9.1e-32;
  Matches  334;  Conservative    0;  Mismatches   98;  Indels   26;  Gaps    7;

Qy        326 GTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTCCTATAGTAAAAGAC 385
              ||||||||| |||  ||  ||||||| | ||      | || |||||||||  ||||
Db        436 GTGTGATACCTTTCATCAGCCATCATTAAGG-----GCTAACACTCCTATAACAAAAACA 382

Qy        386 AGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTACATGACATGGGAG 445
              || |     ||| ||||| ||||| |||||||||||||| |||||||||||  ||  ||||
Db        381 AGTT---ATCGAGAAAAACATAACACATTTATTTAATCAGAGTTTTACATGGCACAGGAG 325

Qy        446 TCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTTTTTTGCTGAGGTTC 505
              ||||||||||||||||||||||||||||||    ||||      ||||| |||| || ||
Db        324 GCTTCAGAAATGAAGACCCAAAGACCCAGGATTAACT-----ATTTTTATGCTCAGATTT 270

Qy        506 GATGAAGAATGGATAGC-ATGTAGCCATGTAGATTAGACAAAAGGATATGATCTAGTGGT 564
              ||| ||||| |  ||| ||| ||  ||||   |     ||| || ||||||||| ||||
Db        269 AATGCAGAATAGTCAGCTATGAAGGAATGTGATTGGACAAAATGGGTATGATCTAATGGT 210

Qy        565 AAAGGACTCAGGGGGAAAC---ACAGCAAGGCCTGTCTATTCAGATTCTTCTTGATCTCT 621
              |  ||||  ||||  | |    ||||||| ||||||| ||||||||||||||  | ||
Db        209 TATAGACTGGGGGGTGAGCGGACCAGCAAGCCCTGTCTGTTCAGATTCTTCTTGGACGCT 150

Qy        622 CTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTTCTTCAATGAG 681
              |       || |||| |||  |||||| || |||||||| |||||| |||  |  |||
Db        149 C--------TGCATAGTATTTATTCCTCTTGGGTATGGGGCCGGACTCCTCTAAATGGAG 98

Qy        682 GGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTATGGCTTGCTTCGGGGAA 741
              || |||||||||||||| ||| |||| ||||| ||| |||||| | |||||| | || |
Db         97 GGCCTTCAAGGGAGAAGGAAGAGAGTGACCTTTCTAGGTTTTAT-GTTTGCTTTGAGGGA 39

Qy        742 GAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAG 779
              ||||||| |||||||||| |||     ||| |||||||
Db         38 GAGGAGTCCTAGTTTCTAAGACTGGCTTTGAGGAAGAG 1


RESULT 19
BG699761
LOCUS       BG699761                 777 bp    mRNA    linear   EST 07-MAY-2001
DEFINITION  602681454F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:4814340 5',
            mRNA sequence.
ACCESSION   BG699761
VERSION     BG699761.1  GI:13968395
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 777)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
             cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10709  row: k  column: 13
            High quality sequence stop: 772.
FEATURES             Location/Qualifiers
     source          1. .777
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4814340"
                     /tissue_type="hippocampus"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_95"
                     /note="Organ: brain; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI
                     (gtcgag); Oligo-dT primed using primer
                     5'-TTTTTTTTTTTTTTTVN-3', size-selected for average
                     insert size 2.5 kb and normalized to ROT 5. This is a
                     primary library enriched for full-length clones and
                     constructed using the Cap-trapper method (Carninci, in
                     preparation). Library constructed by M. Brownstein
                     (NIMH/NHGRI, National Institutes of Health). Note: this
                     is a NIH_MGC Library."
ORIGIN

  Query Match          2.8%;  Score 172.4;  DB 4;  Length 777;
  Best Local Similarity   99.4%;  Pred. No. 4.5e-30;
  Matches  173;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       5455 AGCAGGAAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGC 5514
              ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGCGGGAAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGC 60

Qy       5515 TCAAAGCCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCT 5574
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TCAAAGCCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCT 120

Qy       5575 GTTAAATAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GTTAAATAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 174


RESULT 20
BG699541
LOCUS       BG699541                 593 bp    mRNA    linear   EST 07-MAY-2001
DEFINITION  602679246F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:4812059 5',
            mRNA sequence.
ACCESSION   BG699541
VERSION     BG699541.1  GI:13967945
KEYWORDS    EST.
```

```
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 593)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
             cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10703  row: 1  column: 12
            High quality sequence stop: 591.
FEATURES             Location/Qualifiers
     source          1. .593
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4812059"
                     /tissue_type="hippocampus"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_95"
                     /note="Organ: brain; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI
                     (gtcgag); Oligo-dT primed using primer
                     5'-TTTTTTTTTTTTTTTVN-3', size-selected for  average
                     insert size 2.5 kb and normalized to ROT 5. This is a
                     primary library enriched for  full-length clones and
                     constructed using the Cap-trapper method (Carninci, in
                     preparation).  Library constructed by M. Brownstein
                     (NIMH/NHGRI, National Institutes of Health).  Note: this
                     is a NIH_MGC Library."
ORIGIN

  Query Match             2.8%;  Score 170;  DB 4;  Length 593;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-29;
  Matches  170;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5459 GGAAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAA 5518
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          4 GGAAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAA 63

Qy       5519 AGCCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTA 5578
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         64 AGCCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTA 123

Qy       5579 AATAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        124 AATAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 173


RESULT 21
BI548918
LOCUS       BI548918                624 bp    mRNA    linear   EST 05-SEP-2001
DEFINITION  603189043F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:5260403 5',
            mRNA sequence.
ACCESSION   BI548918
VERSION     BI548918.1  GI:15436217
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 624)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
             cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11656  row: e  column: 12
            High quality sequence stop: 624.
FEATURES             Location/Qualifiers
     source          1. .624
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5260403"
                     /tissue_type="hippocampus"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_95"
                     /note="Organ: brain; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI
                     (gtcgag); Oligo-dT primed using primer
                     5'-TTTTTTTTTTTTTTTVN-3', size-selected for  average
                     insert size 2.5 kb and normalized to ROT 5. This is a
                     primary library enriched for  full-length clones and
                     constructed using the Cap-trapper method (Carninci, in
                     preparation).  Library constructed by M. Brownstein
                     (NIMH/NHGRI, National Institutes of Health).  Note: this
                     is a NIH_MGC Library."
ORIGIN

  Query Match             2.8%;  Score 170;  DB 4;  Length 624;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-29;
  Matches  170;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5459 GGAAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAA 5518
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          4 GGAAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAA 63

Qy       5519 AGCCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTA 5578
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         64 AGCCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTA 123

Qy       5579 AATAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        124 AATAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 173


RESULT 22
BI548997
LOCUS       BI548997                680 bp    mRNA    linear   EST 05-SEP-2001
DEFINITION  603188958F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:5260358 5',
            mRNA sequence.
ACCESSION   BI548997
VERSION     BI548997.1  GI:15436309
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 680)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
             cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
```

DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11656 row: c column: 15
High quality sequence stop: 680.

FEATURES Location/Qualifiers
source 1. .680
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5260358"
/tissue_type="hippocampus"
/lab_host="DH10B"
/clone_lib="NIH_MGC_95"
/note="Organ: brain; Vector: pBluescriptR (modified pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTVN-3', size-selected for average insert size 2.5 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIMH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."

ORIGIN

Query Match 2.8%; Score 170; DB 4; Length 680;
Best Local Similarity 100.0%; Pred. No. 1.7e-29;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5459 GGAAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAA 5518
Db 4 GGAAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAA 63

Qy 5519 AGCCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTA 5578
Db 64 AGCCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTA 123

Qy 5579 AATAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
Db 124 AATAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 173

RESULT 23
BI545493
LOCUS BI545493 720 bp mRNA linear EST 05-SEP-2001
DEFINITION 603187655F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:5259084 5', mRNA sequence.
ACCESSION BI545493
VERSION BI545493.1 GI:15432805
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 720)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11652 row: n column: 13
High quality sequence stop: 714.

FEATURES Location/Qualifiers
source 1. .720
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5259084"
/tissue_type="hippocampus"
/lab_host="DH10B"
/clone_lib="NIH_MGC_95"
/note="Organ: brain; Vector: pBluescriptR (modified pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTVN-3', size-selected for average insert size 2.5 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIMH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."

ORIGIN

Query Match 2.8%; Score 170; DB 4; Length 720;
Best Local Similarity 100.0%; Pred. No. 1.7e-29;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5459 GGAAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAA 5518
Db 4 GGAAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAA 63

Qy 5519 AGCCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTA 5578
Db 64 AGCCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTA 123

Qy 5579 AATAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
Db 124 AATAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 173

RESULT 24
BI551559
LOCUS BI551559 833 bp mRNA linear EST 05-SEP-2001
DEFINITION 603194205F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:5273929 5', mRNA sequence.
ACCESSION BI551559
VERSION BI551559.1 GI:15438871
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 833)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11691 row: i column: 02
High quality sequence stop: 765.

FEATURES Location/Qualifiers
source 1. .833
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5273929"
/tissue_type="hippocampus"
/lab_host="DH10B"

/clone_lib="NIH_MGC_95"
/note="Organ: brain; Vector: pBluescriptR (modified pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTTVN-3', size-selected for average insert size 2.5 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIMH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."

ORIGIN

Query Match 2.8%; Score 170; DB 4; Length 833;
Best Local Similarity 100.0%; Pred. No. 1.8e-29;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       5459 GGAAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAA 5518
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          4 GGAAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAA 63

Qy       5519 AGCCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTA 5578
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         64 AGCCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTA 123

Qy       5579 AATAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        124 AATAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 173
```

RESULT 25
BI547188
LOCUS BI547188 834 bp mRNA linear EST 05-SEP-2001
DEFINITION 603190434F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:5261528 5', mRNA sequence.
ACCESSION BI547188
VERSION BI547188.1 GI:15434500
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 834)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11659 row: d column: 09
High quality sequence stop: 802.
FEATURES Location/Qualifiers
source 1. .834
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5261528"
/tissue_type="hippocampus"
/lab_host="DH10B"
/clone_lib="NIH_MGC_95"
/note="Organ: brain; Vector: pBluescriptR (modified pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTTVN-3', size-selected for average insert size 2.5 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIMH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."

ORIGIN

Query Match 2.8%; Score 170; DB 4; Length 834;
Best Local Similarity 100.0%; Pred. No. 1.8e-29;
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       5459 GGAAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAA 5518
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          4 GGAAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAA 63

Qy       5519 AGCCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTA 5578
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         64 AGCCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTA 123

Qy       5579 AATAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        124 AATAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 173
```

RESULT 26
AQ457908/c
LOCUS AQ457908 482 bp DNA linear GSS 23-APR-1999
DEFINITION HS_5189_B2_D04_SP6E RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=765 Col=8 Row=H, genomic survey sequence.
ACCESSION AQ457908
VERSION AQ457908.1 GI:4636548
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 482)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
PUBMED 10449764
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Resear h Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu
Plate: 765 row: H column: 8
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 482.
FEATURES Location/Qualifiers
source 1. .482
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="Plate=765 Col=8 Row=H"
/sex="male"
/clone_lib="RPCI-11 Human Male BAC Library"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the

```
                    pBACe3.6 vector at EcoRI sites"
ORIGIN

  Query Match             2.8%;  Score 169.8;  DB 8;  Length 482;
  Best Local Similarity   71.7%;  Pred. No. 1.7e-29;
  Matches  281;  Conservative    0;  Mismatches   98;  Indels   13;  Gaps    4;

Qy        389 TTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTACATGACATGGGAGTCT 448
              || | |||||  || | |||  ||||||||  ||||  |||||| ||||||||||||| ||
Db        382 TTGACAAGAGATAAGCATAATCAATTTATTGGATCATGGTTTTATATGACATGGGAGCCT 323

Qy        449 TCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTTTTTTTGCTGAGGTTCGAT 508
              |||||||| |||||||||||| || |  || |||||||| ||||| ||||| |||||  |
Db        322 TCAGAAATAAAGACCCAAAGAACCCGTAGACACTGTCTG-TTTTTATGCTGCGGTTCAGT 264

Qy        509 GAAGAATGGATAGC-ATGTAGCCATGTAGATTAGACAAAAGGATATGATCTAGTGGTAAA 567
              || ||||||  ||| | ||||  |||||  |       ||| | ||| |||||   ||||
Db        263 GATGAATGGTCAGCTACGTAGAAATGTAATTGGACATAAAAGGTATAATCTATCAGTAAT 204

Qy        568 GGACTCAG-GGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTTCTTGATCTCTCTCTC 626
               || | ||   || ||| |    | |||||| | |||| ||||||||||| |||
Db        203 AGATTGAGAATGGGAACCCCAGCAAGCCTGTTTGTTCATATTCTTCTTGACCTC------ 150

Qy        627 TCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTTCTTCAATGAGGGTCT 686
                  ||| |   | |    ||| | ||||||||||||| | | |  |||||||| |
Db        149 ----TTTATGGAGCTTTCCTATCCAGGGTATGGGGCAGGACCCCTATGGAATGAGGGTTT 94

Qy        687 TCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATGGCTTGCTTCGGGGAAGAGGA 746
              |||||||||||| ||||| || ||||| ||| || | ||||| |||  | || |||||
Db         93 TCAAGGGAGAAGGAAGAAAATGACCTTTCTAGGTTCTGTGGCTCGCTGTGAGGGAGAGGT 34

Qy        747 GTTCTAGTTTCTATGACCCATCTTGGGGAAGA 778
                |||||||| |||||  ||||||| ||||||
Db         33 TGTCTAGTTTNTATGATGCATCTTGAGGAAGA 2


RESULT 27
BI551703
LOCUS       BI551703                803 bp    mRNA    linear   EST 05-SEP-2001
DEFINITION  603197395F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:5277118 5',
            mRNA sequence.
ACCESSION   BI551703
VERSION     BI551703.1  GI:15439028
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 803)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
             cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11699  row: m  column: 23
            High quality sequence stop: 789.
FEATURES             Location/Qualifiers
     source          1. .803
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5277118"
                     /tissue_type="hippocampus"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_95"
                     /note="Organ: brain; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI
                     (gtcgag); Oligo-dT primed using primer
                     5'-TTTTTTTTTTTTTTTVN-3', size-selected for  average
                     insert size 2.5 kb and normalized to ROT 5. This is a
                     primary library enriched for  full-length clones and
                     constructed using the Cap-trapper method (Carninci, in
                     preparation).  Library constructed by M. Brownstein
                     (NIMH/NHGRI, National Institutes of Health).  Note: this
                     is a NIH_MGC Library."
ORIGIN

  Query Match             2.8%;  Score 169.6;  DB 4;  Length 803;
  Best Local Similarity   90.5%;  Pred. No. 2.2e-29;
  Matches  181;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy       5459 GGAAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAA 5518
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          4 GGAAGATGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAA 63

Qy       5519 AGCCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTA 5578
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         64 AGCCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTA 123

Qy       5579 AATAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAGGTAAGATTGC 5638
              |||||||||||||||||||||||||||||||||||||||||||||||||||    |   |
Db        124 AATAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAGTGCAACAAGA 183

Qy       5639 TTTCAGCCATTAACCATATT 5658
              | | |  |||   | |  ||
Db        184 TATGAAACATCGGCTAGGTT 203


RESULT 28
AQ466676/c
LOCUS       AQ466676                454 bp    DNA     linear   GSS 23-APR-1999
DEFINITION  HS_5161_B2_H02_SP6E RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=737 Col=4 Row=P, genomic survey sequence.
ACCESSION   AQ466676
VERSION     AQ466676.1  GI:4643771
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 454)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
   PUBMED   10449764
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 737  row: P  column: 4
            Seq primer: SP6
            Class: BAC ends
            High quality sequence stop: 454.
FEATURES             Location/Qualifiers
     source          1. .454
```

```
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="Plate=737 Col=4 Row=P"
                     /sex="male"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Male blood DNA was isolated from one randomly chosen donor
                     and partially digested with a combination of EcoRI and
                     EcoRI Methylase. Size selected DNA was cloned into the
                     pBACe3.6 vector at EcoRI sites"
ORIGIN

  Query Match           2.8%;  Score 169.4;  DB 8;  Length 454;
  Best Local Similarity   70.4%;  Pred. No. 2.1e-29;
  Matches  288;  Conservative    0;  Mismatches  106;  Indels   15;  Gaps    4;

Qy        337 TTTCTCTTCCATCATAAGGGTCGCAACCAAAACTCCTATAGTAAAAGACAGGTTAATAAG 396
              ||||    |||||||||||||| ||||  | |||| ||||  ||| | |||||||| |||
Db        427 TTTCCTCCCCATCATAAGGGTCTCAACTGACACTCATATAACAAATGTCAGGTTAACAAG 368

Qy        397 AGCAAAACCTAACAAATTTATTTAATCAAAGTTTTACATGACATGGGAGTCTTCAGAAAT 456
              || ||| | |||||||| ||||||||| |||||| |  |||||  ||   ||||||||||
Db        367 AGAAAAGCATAACAAATGTATTTAATC-AAGTTTCATGTGACACAGGGACCTTCAGAAAT 309

Qy        457 GAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTTTTTTGCTGAGGTTCGATGAAGAATG 516
              |||||||||||||||||| ||| | ||||||||      |  |  |      ||    |
Db        308 GAAGACCCAAAGACCCAGGGAAAATTATCTGTTTTGACACTTACGTATGATGAAAATGCG 249

Qy        517 GATAGCATGTAGCCATGTAGATTAGACAAAAGGATATGATCTAGTGGTAAAGGACTCAGG 576
              |    ||| ||   ||||   |     |||||| | | ||||| ||| ||  || | |||
Db        248 GCAGCCATATAAAAATGTGATTGGACAAAAAGGGTGTAATCTAATGGCAAGAGATTAAGG 189

Qy        577 GGGAAACACAGCAAGGCCTGTCTATTCAGATTCTTCTTGATCTCTCTCTCTCTATGTATA 636
              |||||||| ||| | |||||| | ||||| ||||||||   |||||         ||  |
Db        188 GGGAAACATAGC-ATGCCTGTGTGTTCAGCTTCTTCTTAGCCTCTC--------TGACTT 138

Qy        637 GCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTTCTTCAATGAGGGTCTTCAAGGGAGA 696
              | ||| ||||| | | | || | ||||| |||||||   |||||| ||||||||||||||
Db        137 GGATTTTTTCCACTTAAATAGGAGGCAGAACTCTTCTGGAATGAGAGTCTTCAAGGGAGA 78

Qy        697 -----AGGGAGAAAGTGGCCTTTTTAGATTTTATGGCTTGCTTCGGGGA 740
                  ||||||||||| ||| |    ||| |||||||||| |  ||||
Db         77 TGGAGAGGGAGAAAGTGACCTATCGGAATTGTATGGCTTGCATATGGGA 29


RESULT 29
BP251245
LOCUS       BP251245                582 bp    mRNA    linear   EST 15-SEP-2004
DEFINITION  BP251245 Sugano cDNA library, hippocampus Homo sapiens cDNA clone
            HPR06860, mRNA sequence.
ACCESSION   BP251245
VERSION     BP251245.1  GI:52133526
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 582)
  AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
  TITLE     Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
  JOURNAL   Genome Res. 14 (9), 1711-1718 (2004)
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .582
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="HPR06860"
                     /tissue_type="hippocampus"
                     /clone_lib="Sugano cDNA library, hippocampus"
ORIGIN

  Query Match           2.8%;  Score 169.4;  DB 5;  Length 582;
  Best Local Similarity   99.4%;  Pred. No. 2.3e-29;
  Matches  170;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       5458 AGGAAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCA 5517
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGAAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCA 60

Qy       5518 AAGCCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTT 5577
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AAGCCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTT 120

Qy       5578 AAATAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AAATAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 171


RESULT 30
CR556919
LOCUS       CR556919                521 bp    mRNA    linear   EST 13-JUL-2004
DEFINITION  DKFZp459I0132_r1 459 (synonym: pcor1) Pongo pygmaeus cDNA clone
            DKFZp459I0132 5', mRNA sequence.
ACCESSION   CR556919
VERSION     CR556919.1  GI:50278198
KEYWORDS    EST.
SOURCE      Pongo pygmaeus (orangutan)
  ORGANISM  Pongo pygmaeus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pongo.
REFERENCE   1  (bases 1 to 521)
  AUTHORS   Koehrer,K., Beyer,A., Mewes,H.W., Weil,B., Amid,C., Osanger,A.,
            Fobo,G., Han,M. and Wiemann,S.
  TITLE     Pongo pygmaeus mRNA (Koehrer,K., Beyer,A., Mewes,H.W., et al.)
  JOURNAL   Unpublished (2004)
COMMENT     Contact: MIPS
            MIPS
            Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
            This is the 5' sequence of the clone insert Clone from S. Wiemann,
            Molecular Genome Analysis, German Cancer Research Center (DKFZ);
            Email s.wiemann@dkfz-heidelberg.de; mforschung GmbH in Berlin,
            Germany. Please contact RZPD for ordering:
            http://www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=DKFZp459I0132
            Further information about the clone and the sequencing project is
            available at http://mips.gsf.de/projects/cdna/.
FEATURES             Location/Qualifiers
     source          1. .521
                     /organism="Pongo pygmaeus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9600"
                     /clone="DKFZp459I0132"
                     /tissue_type="cortex"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="459 (synonym: pcor1)"
                     /note="Vector: pSport1_Sfi; Site_1: SfilA; Site_2: SfilB"
ORIGIN

  Query Match           2.8%;  Score 168.4;  DB 7;  Length 521;
  Best Local Similarity   99.4%;  Pred. No. 3.8e-29;
  Matches  169;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       5459 GGAAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAA 5518
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         17 GGAAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAA 76

Qy       5519 AGCCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTA 5578
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
```

```
Db         77 AGCCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCAAATTCCAAGCTGTTA 136

Qy       5579 AATAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        137 AATAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 186


RESULT 31
CR539342
LOCUS       CR539342                540 bp    mRNA    linear   EST 06-JUL-2004
DEFINITION  DKFZp459M241_r1 459 (synonym: pcor1) Pongo pygmaeus cDNA clone
            DKFZp459M241 5', mRNA sequence.
ACCESSION   CR539342
VERSION     CR539342.1  GI:49686458
KEYWORDS    EST.
SOURCE      Pongo pygmaeus (orangutan)
  ORGANISM  Pongo pygmaeus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pongo.
REFERENCE   1  (bases 1 to 540)
  AUTHORS   Poustka,A., Albert,R., Moosmayer,P., Schupp,I., Wellenreuther,R.,
            Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and
            Wiemann,S.
  TITLE     Pongo pygmaeus mRNA (Poustka,A., Albert,R., Moosmayer,P., et al.)
  JOURNAL   Unpublished (2004)
COMMENT     Contact: MIPS
            MIPS
            Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
            This is the 5' sequence of the clone insert Clone from S. Wiemann,
            Molecular Genome Analysis, German Cancer Research Center (DKFZ);
            Email s.wiemann@dkfz-heidelberg.de; sequenced by DKFZ (German
            Cancer Research Center, Heidelberg/Germany) within the cDNA
            sequencing consortium of the German Genome Project. This clone
            (DKFZp459M241) is available at the RZPD in Berlin. Please contact
            the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
            information about the clone and the sequencing project is available
            at http://mips.gsf.de/projects/cdna/.
FEATURES              Location/Qualifiers
     source           1. .540
                      /organism="Pongo pygmaeus"
                      /mol_type="mRNA"
                      /db_xref="taxon:9600"
                      /clone="DKFZp459M241"
                      /tissue_type="cortex"
                      /dev_stage="adult"
                      /lab_host="DH10B"
                      /clone_lib="459 (synonym: pcor1)"
                      /note="Vector: pSport1_Sfi; Site_1: SfilA; Site_2: SfilB"
ORIGIN

  Query Match             2.8%;  Score 168.4;  DB 7;  Length 540;
  Best Local Similarity   99.4%;  Pred. No. 3.9e-29;
  Matches  169;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       5459 GGAAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAA 5518
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GGAAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAA 60

Qy       5519 AGCCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTA 5578
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db         61 AGCCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCAAATTCCAAGCTGTTA 120

Qy       5579 AATAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AATAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 170


RESULT 32
CR538651
LOCUS       CR538651                578 bp    mRNA    linear   EST 06-JUL-2004
DEFINITION  DKFZp459M187_r1 459 (synonym: pcor1) Pongo pygmaeus cDNA clone
            DKFZp459M187 5', mRNA sequence.
ACCESSION   CR538651
VERSION     CR538651.1  GI:49684789
KEYWORDS    EST.
SOURCE      Pongo pygmaeus (orangutan)
  ORGANISM  Pongo pygmaeus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pongo.
REFERENCE   1  (bases 1 to 578)
  AUTHORS   Ottenwaelder,B., Obermaier,B., Deutschenbaur,S., Schaipp,A.,
            Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and
            Wiemann,S.
  TITLE     Pongo pygmaeus mRNA (Ottenwaelder,B., Obermaier,B.,
            Deutschenbaur,S., et al.)
  JOURNAL   Unpublished (2004)
COMMENT     Contact: MIPS
            MIPS
            Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
            This is the 5' sequence of the clone insert Clone from S. Wiemann,
            Molecular Genome Analysis, German Cancer Research Center (DKFZ);
            Email s.wiemann@dkfz-heidelberg.de; sequenced by Medigenomix
            (Martinsried/Germany) within the cDNA sequencing consortium of the
            German Genome Project. This clone (DKFZp459M187) is available at
            the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum,
            Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email:
            clone@rzpd.de Further information about the clone and the
            sequencing project is available at
            http://mips.gsf.de/projects/cdna/.
FEATURES              Location/Qualifiers
     source           1. .578
                      /organism="Pongo pygmaeus"
                      /mol_type="mRNA"
                      /db_xref="taxon:9600"
                      /clone="DKFZp459M187"
                      /tissue_type="cortex"
                      /dev_stage="adult"
                      /lab_host="DH10B"
                      /clone_lib="459 (synonym: pcor1)"
                      /note="Vector: pSport1_Sfi; Site_1: SfilA; Site_2: SfilB"
ORIGIN

  Query Match             2.8%;  Score 168.4;  DB 7;  Length 578;
  Best Local Similarity   99.4%;  Pred. No. 3.9e-29;
  Matches  169;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       5459 GGAAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAA 5518
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GGAAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAA 60

Qy       5519 AGCCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTA 5578
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db         61 AGCCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCAAATTCCAAGCTGTTA 120

Qy       5579 AATAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AATAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 170


RESULT 33
BP348638
LOCUS       BP348638                583 bp    mRNA    linear   EST 17-SEP-2004
DEFINITION  BP348638 Sugano cDNA library, brain Homo sapiens cDNA clone
            SZR01236, mRNA sequence.
ACCESSION   BP348638
VERSION     BP348638.1  GI:52278623
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 583)
  AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
  TITLE     Sequence comparison of human and mouse genes reveals a homologous
```

```
            block structure in the promoter regions
  JOURNAL   Genome Res. 14 (9), 1711-1718 (2004)
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .583
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="SZR01236"
                     /tissue_type="brain"
                     /clone_lib="Sugano cDNA library, brain"
ORIGIN

  Query Match             2.8%;  Score 168.4;  DB 5;  Length 583;
  Best Local Similarity   99.4%;  Pred. No. 3.9e-29;
  Matches  169;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       5459 GGAAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAA 5518
             | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          5 GAAAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAA 64

Qy       5519 AGCCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTA 5578
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         65 AGCCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTA 124

Qy       5579 AATAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        125 AATAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 174


RESULT 34
CR539181
LOCUS       CR539181                634 bp    mRNA    linear   EST 06-JUL-2004
DEFINITION  DKFZp459E041_r1 459 (synonym: pcor1) Pongo pygmaeus cDNA clone
            DKFZp459E041 5', mRNA sequence.
ACCESSION   CR539181
VERSION     CR539181.1  GI:49686297
KEYWORDS    EST.
SOURCE      Pongo pygmaeus (orangutan)
  ORGANISM  Pongo pygmaeus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pongo.
REFERENCE   1  (bases 1 to 634)
  AUTHORS   Poustka,A., Albert,R., Moosmayer,P., Schupp,I., Wellenreuther,R.,
            Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and
            Wiemann,S.
  TITLE     Pongo pygmaeus mRNA (Poustka,A., Albert,R., Moosmayer,P., et al.)
  JOURNAL   Unpublished (2004)
COMMENT     Contact: MIPS
            MIPS
            Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
            This is the 5' sequence of the clone insert Clone from S. Wiemann,
            Molecular Genome Analysis, German Cancer Research Center (DKFZ);
            Email s.wiemann@dkfz-heidelberg.de; sequenced by DKFZ (German
            Cancer Research Center, Heidelberg/Germany) within the cDNA
            sequencing consortium of the German Genome Project. This clone
            (DKFZp459E041) is available at the RZPD in Berlin. Please contact
            the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
            information about the clone and the sequencing project is available
            at http://mips.gsf.de/projects/cdna/.
FEATURES             Location/Qualifiers
     source          1. .634
                     /organism="Pongo pygmaeus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9600"
                     /clone="DKFZp459E041"
                     /tissue_type="cortex"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="459 (synonym: pcor1)"
                     /note="Vector: pSport1_Sfi; Site_1: SfilA; Site_2: SfilB"
ORIGIN

  Query Match             2.8%;  Score 168.4;  DB 7;  Length 634;
  Best Local Similarity   99.4%;  Pred. No. 4e-29;
  Matches  169;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       5459 GGAAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAA 5518
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GGAAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAA 60

Qy       5519 AGCCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCAAATTCCAAGCTGTTA 5578
              |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db         61 AGCCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCAAATTCCAAGCTGTTA 120

Qy       5579 AATAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AATAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 170


RESULT 35
CR630572
LOCUS       CR630572                716 bp    mRNA    linear   EST 11-AUG-2004
DEFINITION  DKFZp459H2423_r1 459 (synonym: pcor1) Pongo pygmaeus cDNA clone
            DKFZp459H2423 5', mRNA sequence.
ACCESSION   CR630572
VERSION     CR630572.1  GI:51126652
KEYWORDS    EST.
SOURCE      Pongo pygmaeus (orangutan)
  ORGANISM  Pongo pygmaeus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pongo.
REFERENCE   1  (bases 1 to 716)
  AUTHORS   Poustka,A., Albert,R., Moosmayer,P., Schupp,I., Wellenreuther,R.,
            Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and
            Wiemann,S.
  TITLE     Pongo pygmaeus mRNA (Poustka,A., Albert,R., Moosmayer,P., et al.)
  JOURNAL   Unpublished (2004)
COMMENT     Contact: MIPS
            MIPS
            Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
            This is the 5' sequence of the clone insert Clone from S. Wiemann,
            Molecular Genome Analysis, German Cancer Research Center (DKFZ);
            Email s.wiemann@dkfz-heidelberg.de; any. Please contact RZPD for
            ordering:
            http://www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=DKFZp459H2423
            Further information about the clone and the sequencing project is
            available at http://mips.gsf.de/projects/cdna/.
FEATURES             Location/Qualifiers
     source          1. .716
                     /organism="Pongo pygmaeus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9600"
                     /clone="DKFZp459H2423"
                     /tissue_type="cortex"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="459 (synonym: pcor1)"
                     /note="Vector: pSport1_Sfi; Site_1: SfilA; Site_2: SfilB"
ORIGIN

  Query Match             2.8%;  Score 168.4;  DB 7;  Length 716;
  Best Local Similarity   99.4%;  Pred. No. 4.1e-29;
  Matches  169;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       5459 GGAAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAA 5518
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GGAAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAA 60

Qy       5519 AGCCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTA 5578
              |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
```

```
Db         61 AGCCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCAAATTCCAAGCTGTTA 120

Qy       5579 AATAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AATAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 170


RESULT 36
CB998494
LOCUS       CB998494                727 bp    mRNA    linear   EST 01-MAY-2003
DEFINITION  AGENCOURT_13650932 NIH_MGC_187 Homo sapiens cDNA clone
            IMAGE:30320704 5', mRNA sequence.
ACCESSION   CB998494
VERSION     CB998494.1  GI:30293014
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 727)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Michael Brownstein and Dr. Miklos Palkovits
             cDNA Library Preparation: CLONTECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: NDCM118  row: o  column: 17
            High quality sequence stop: 433.
FEATURES             Location/Qualifiers
     source          1. .727
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:30320704"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_187"
                     /note="Organ: blood vessels - aorta, basilar and artery;
                     Vector: pDNR-LIB; Site_1: SfiI (ggccattatggcc); Site_2:
                     SfiI (ggccgcctcggcc); 5' and 3' adaptors were used in
                     cloning as follows: 5' adaptor sequence:
                     5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C, G, or T). Average insert size 1.4 kb
                     (range 0.5-4.0 kb). 14/15 colonies contained inserts by
                     PCR. This library was enriched for full-length clones and
                     was constructed by Clontech Laboratories (Palo Alto, CA).
                     Note: this is a NIH_MGC Library."
ORIGIN

  Query Match             2.8%;  Score 168.4;  DB 6;  Length 727;
  Best Local Similarity   99.4%;  Pred. No. 4.1e-29;
  Matches  169;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       5459 GGAAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAA 5518
              || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GGGAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAA 60

Qy       5519 AGCCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTA 5578
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AGCCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTA 120

Qy       5579 AATAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AATAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 170


RESULT 37
```

```
BG708958
LOCUS       BG708958                902 bp    mRNA    linear   EST 07-MAY-2001
DEFINITION  602675104F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:4797721 5',
            mRNA sequence.
ACCESSION   BG708958
VERSION     BG708958.1  GI:13986817
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 902)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
             cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10684  row: g  column: 02
            High quality sequence stop: 727.
FEATURES             Location/Qualifiers
     source          1. .902
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4797721"
                     /tissue_type="hypothalamus"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_96"
                     /note="Organ: brain; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI
                     (gtcgag); Oligo-dT primed using primer
                     5'-TTTTTTTTTTTTTTTVN-3', size-selected for average
                     insert size 2.3 kb and normalized to ROT 5. This is a
                     primary library enriched for full-length clones and
                     constructed using the Cap-trapper method (Carninci, in
                     preparation). Library constructed by M. Brownstein
                     (NIMH/NHGRI, National Institutes of Health). Note: this is
                     a NIH_MGC Library."
ORIGIN

  Query Match             2.8%;  Score 168.4;  DB 4;  Length 902;
  Best Local Similarity   99.4%;  Pred. No. 4.4e-29;
  Matches  169;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       5459 GGAAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAA 5518
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          4 GGAAGTCGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAA 63

Qy       5519 AGCCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTA 5578
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         64 AGCCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTA 123

Qy       5579 AATAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        124 AATAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 173


RESULT 38
BI550370
LOCUS       BI550370                760 bp    mRNA    linear   EST 05-SEP-2001
DEFINITION  603192421F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:5263588 5',
            mRNA sequence.
ACCESSION   BI550370
VERSION     BI550370.1  GI:15437682
KEYWORDS    EST.
```

```
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 760)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
             cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11664  row: j  column: 05
            High quality sequence stop: 757.
FEATURES             Location/Qualifiers
     source          1. .760
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5263588"
                     /tissue_type="hippocampus"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_95"
                     /note="Organ: brain; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI
                     (gtcgag); Oligo-dT primed using primer
                     5'-TTTTTTTTTTTTTTTVN-3', size-selected for average
                     insert size 2.5 kb and normalized to ROT 5. This is a
                     primary library enriched for full-length clones and
                     constructed using the Cap-trapper method (Carninci, in
                     preparation). Library constructed by M. Brownstein
                     (NIMH/NHGRI, National Institutes of Health). Note: this
                     is a NIH_MGC Library."
ORIGIN

  Query Match             2.8%;  Score 168.2;  DB 4;  Length 760;
  Best Local Similarity   98.3%;  Pred. No. 4.7e-29;
  Matches  170;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       5456 GCAGGAAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCT 5515
              | | | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          4 GGAAGGAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCT 63

Qy       5516 CAAAGCCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTG 5575
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         64 CAAAGCCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTG 123

Qy       5576 TTAAATAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        124 TTAAATAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 176


RESULT 39
AU076999
LOCUS       AU076999                 174 bp    mRNA    linear   EST 04-MAY-2000
DEFINITION  AU076999 Sugano cDNA library Homo sapiens cDNA clone nblb0114
            similar to 5'-end region of Human RGP4 mRNA, mRNA sequence.
ACCESSION   AU076999
VERSION     AU076999.1  GI:7439562
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 174)
  AUTHORS   Suzuki,Y., Ishihara,D., Sasaki,M., Nakagawa,H., Hata,H.,
            Tsunoda,T., Watanabe,M., Komatsu,T., Ota,T., Isogai,T., Suyama,A.
            and Sugano,S.
  TITLE     Statistical analysis of the 5' untranslated region of human mRNA
            using 'Oligo-Capped' cDNA libraries
  JOURNAL   Genomics 64 (3), 286-297 (2000)
  MEDLINE   20221373
   PUBMED   10756096
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp
            Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and
            Sugano,S. Construction and characterization of a full
            length-enriched and a 5'-end-enriched cDNA library. Gene 200 (1-2),
            149-156 (1997)
             This clone was obtained from a 'full length-enriched' cDNA library
            constructed by 'Oligo-Capping' method. The coding region starts
            from the 50 bp upstream to the 3'-end.
FEATURES             Location/Qualifiers
     source          1. .174
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="nblb0114"
                     /clone_lib="Sugano cDNA library"
ORIGIN

  Query Match             2.8%;  Score 168;  DB 1;  Length 174;
  Best Local Similarity   100.0%;  Pred. No. 3.7e-29;
  Matches  168;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5461 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 5520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 60

Qy       5521 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 5580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 120

Qy       5581 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 168


RESULT 40
BP199723
LOCUS       BP199723                 461 bp    mRNA    linear   EST 14-SEP-2004
DEFINITION  BP199723 Sugano cDNA library, brain Homo sapiens cDNA clone
            ADR09171, mRNA sequence.
ACCESSION   BP199723
VERSION     BP199723.1  GI:52047962
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 461)
  AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
  TITLE     Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
  JOURNAL   Genome Res. 14 (9), 1711-1718 (2004)
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .461
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="ADR09171"
```

```
                     /tissue_type="brain"
                     /clone_lib="Sugano cDNA library, brain"
ORIGIN

  Query Match             2.8%;  Score 168;  DB 5;  Length 461;
  Best Local Similarity   100.0%;  Pred. No. 4.7e-29;
  Matches  168;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5461 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 5520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 60

Qy       5521 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 5580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 120

Qy       5581 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 168


RESULT 41
BP200217
LOCUS       BP200217                461 bp    mRNA    linear   EST 14-SEP-2004
DEFINITION  BP200217 Sugano cDNA library, amygdala Homo sapiens cDNA clone
            AMR01359, mRNA sequence.
ACCESSION   BP200217
VERSION     BP200217.1  GI:52048963
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 461)
  AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
  TITLE     Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
  JOURNAL   Genome Res. 14 (9), 1711-1718 (2004)
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .461
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="AMR01359"
                     /tissue_type="amygdala"
                     /clone_lib="Sugano cDNA library, amygdala"
ORIGIN

  Query Match             2.8%;  Score 168;  DB 5;  Length 461;
  Best Local Similarity   100.0%;  Pred. No. 4.7e-29;
  Matches  168;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5461 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 5520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 60

Qy       5521 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 5580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 120

Qy       5581 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 168


RESULT 42
BP200544
LOCUS       BP200544                461 bp    mRNA    linear   EST 14-SEP-2004
DEFINITION  BP200544 Sugano cDNA library, amygdala Homo sapiens cDNA clone
            AMR02760, mRNA sequence.
ACCESSION   BP200544
VERSION     BP200544.1  GI:52049623
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 461)
  AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
  TITLE     Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
  JOURNAL   Genome Res. 14 (9), 1711-1718 (2004)
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .461
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="AMR02760"
                     /tissue_type="amygdala"
                     /clone_lib="Sugano cDNA library, amygdala"
ORIGIN

  Query Match             2.8%;  Score 168;  DB 5;  Length 461;
  Best Local Similarity   100.0%;  Pred. No. 4.7e-29;
  Matches  168;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5461 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 5520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 60

Qy       5521 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 5580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 120

Qy       5581 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 168


RESULT 43
BP215732
LOCUS       BP215732                461 bp    mRNA    linear   EST 15-SEP-2004
DEFINITION  BP215732 Sugano cDNA library, corpus callosum Homo sapiens cDNA
            clone CCR04680, mRNA sequence.
ACCESSION   BP215732
VERSION     BP215732.1  GI:52088633
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 461)
  AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
  TITLE     Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
  JOURNAL   Genome Res. 14 (9), 1711-1718 (2004)
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .461
```

```
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CCR04680"
                     /tissue_type="corpus callosum"
                     /clone_lib="Sugano cDNA library, corpus callosum"
ORIGIN

  Query Match             2.8%;  Score 168;  DB 5;  Length 461;
  Best Local Similarity   100.0%;  Pred. No. 4.7e-29;
  Matches  168;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5461 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 5520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 60

Qy       5521 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 5580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 120

Qy       5581 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 168


RESULT 44
BP233321
LOCUS       BP233321                461 bp    mRNA    linear   EST 15-SEP-2004
DEFINITION  BP233321 Sugano cDNA library, coronary artery endothelial cell Homo
            sapiens cDNA clone HCR00121, mRNA sequence.
ACCESSION   BP233321
VERSION     BP233321.1  GI:52106231
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 461)
  AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
  TITLE     Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
  JOURNAL   Genome Res. 14 (9), 1711-1718 (2004)
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .461
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="HCR00121"
                     /tissue_type="coronary artery"
                     /cell_type="endothelial cell"
                     /clone_lib="Sugano cDNA library, coronary artery
                     endothelial cell"
ORIGIN

  Query Match             2.8%;  Score 168;  DB 5;  Length 461;
  Best Local Similarity   100.0%;  Pred. No. 4.7e-29;
  Matches  168;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5461 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 5520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 60

Qy       5521 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 5580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 120

Qy       5581 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 168


RESULT 45
BP235431
LOCUS       BP235431                461 bp    mRNA    linear   EST 15-SEP-2004
DEFINITION  BP235431 Sugano cDNA library, coronary artery endothelial cell Homo
            sapiens cDNA clone HCR05162, mRNA sequence.
ACCESSION   BP235431
VERSION     BP235431.1  GI:52108341
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 461)
  AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
  TITLE     Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
  JOURNAL   Genome Res. 14 (9), 1711-1718 (2004)
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .461
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="HCR05162"
                     /tissue_type="coronary artery"
                     /cell_type="endothelial cell"
                     /clone_lib="Sugano cDNA library, coronary artery
                     endothelial cell"
ORIGIN

  Query Match             2.8%;  Score 168;  DB 5;  Length 461;
  Best Local Similarity   100.0%;  Pred. No. 4.7e-29;
  Matches  168;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5461 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 5520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 60

Qy       5521 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 5580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 120

Qy       5581 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 168


RESULT 46
BP235747
LOCUS       BP235747                461 bp    mRNA    linear   EST 15-SEP-2004
DEFINITION  BP235747 Sugano cDNA library, coronary artery endothelial cell Homo
            sapiens cDNA clone HCR05817, mRNA sequence.
ACCESSION   BP235747
VERSION     BP235747.1  GI:52108657
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 461)
  AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
  TITLE     Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
```

```
  JOURNAL   Genome Res. 14 (9), 1711-1718 (2004)
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .461
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="HCR05817"
                     /tissue_type="coronary artery"
                     /cell_type="endothelial cell"
                     /clone_lib="Sugano cDNA library, coronary artery
                     endothelial cell"
ORIGIN

  Query Match             2.8%;  Score 168;  DB 5;  Length 461;
  Best Local Similarity   100.0%;  Pred. No. 4.7e-29;
  Matches  168;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5461 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 5520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 60

Qy       5521 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 5580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 120

Qy       5581 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 168


RESULT 47
BP235937
LOCUS       BP235937                461 bp    mRNA    linear   EST 15-SEP-2004
DEFINITION  BP235937 Sugano cDNA library, coronary artery endothelial cell Homo
            sapiens cDNA clone HCR06437, mRNA sequence.
ACCESSION   BP235937
VERSION     BP235937.1  GI:52108847
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 461)
  AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
  TITLE     Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
  JOURNAL   Genome Res. 14 (9), 1711-1718 (2004)
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .461
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="HCR06437"
                     /tissue_type="coronary artery"
                     /cell_type="endothelial cell"
                     /clone_lib="Sugano cDNA library, coronary artery
                     endothelial cell"
ORIGIN

  Query Match             2.8%;  Score 168;  DB 5;  Length 461;
  Best Local Similarity   100.0%;  Pred. No. 4.7e-29;
  Matches  168;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5461 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 5520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 60

Qy       5521 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 5580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 120

Qy       5581 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 168


RESULT 48
BP278768
LOCUS       BP278768                461 bp    mRNA    linear   EST 16-SEP-2004
DEFINITION  BP278768 Sugano cDNA library, KG-1-C Homo sapiens cDNA clone
            KGR02804, mRNA sequence.
ACCESSION   BP278768
VERSION     BP278768.1  GI:52192500
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 461)
  AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
  TITLE     Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
  JOURNAL   Genome Res. 14 (9), 1711-1718 (2004)
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .461
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="KGR02804"
                     /cell_line="KG-1-C"
                     /clone_lib="Sugano cDNA library, KG-1-C"
                     /note="glioma"
ORIGIN

  Query Match             2.8%;  Score 168;  DB 5;  Length 461;
  Best Local Similarity   100.0%;  Pred. No. 4.7e-29;
  Matches  168;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5461 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 5520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 60

Qy       5521 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 5580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 120

Qy       5581 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 168


RESULT 49
BP279422
LOCUS       BP279422                461 bp    mRNA    linear   EST 16-SEP-2004
DEFINITION  BP279422 Sugano cDNA library, KG-1-C Homo sapiens cDNA clone
            KGR05042, mRNA sequence.
ACCESSION   BP279422
VERSION     BP279422.1  GI:52193154
KEYWORDS    EST.
```

```
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 461)
  AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
  TITLE     Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
  JOURNAL   Genome Res. 14 (9), 1711-1718 (2004)
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .461
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="KGR05042"
                     /cell_line="KG-1-C"
                     /clone_lib="Sugano cDNA library, KG-1-C"
                     /note="glioma"
ORIGIN

  Query Match             2.8%;  Score 168;  DB 5;  Length 461;
  Best Local Similarity   100.0%;  Pred. No. 4.7e-29;
  Matches  168;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5461 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 5520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 60

Qy       5521 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 5580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 120

Qy       5581 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 168


RESULT 50
BP335562
LOCUS       BP335562                461 bp    mRNA    linear   EST 17-SEP-2004
DEFINITION  BP335562 Sugano cDNA library, coronary artery smooth muscle cell
            Homo sapiens cDNA clone SMR01675, mRNA sequence.
ACCESSION   BP335562
VERSION     BP335562.1  GI:52265114
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 461)
  AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
  TITLE     Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
  JOURNAL   Genome Res. 14 (9), 1711-1718 (2004)
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .461
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="SMR01675"
                     /tissue_type="coronary artery"
                     /cell_type="smooth muscle cell"
                     /clone_lib="Sugano cDNA library, coronary artery smooth
                     muscle cell"
ORIGIN

  Query Match             2.8%;  Score 168;  DB 5;  Length 461;
  Best Local Similarity   100.0%;  Pred. No. 4.7e-29;
  Matches  168;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5461 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 5520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 60

Qy       5521 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 5580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 120

Qy       5581 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 168


RESULT 51
BP335982
LOCUS       BP335982                461 bp    mRNA    linear   EST 17-SEP-2004
DEFINITION  BP335982 Sugano cDNA library, coronary artery smooth muscle cell
            Homo sapiens cDNA clone SMR04405, mRNA sequence.
ACCESSION   BP335982
VERSION     BP335982.1  GI:52265541
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 461)
  AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
  TITLE     Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
  JOURNAL   Genome Res. 14 (9), 1711-1718 (2004)
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .461
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="SMR04405"
                     /tissue_type="coronary artery"
                     /cell_type="smooth muscle cell"
                     /clone_lib="Sugano cDNA library, coronary artery smooth
                     muscle cell"
ORIGIN

  Query Match             2.8%;  Score 168;  DB 5;  Length 461;
  Best Local Similarity   100.0%;  Pred. No. 4.7e-29;
  Matches  168;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5461 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 5520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 60

Qy       5521 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 5580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 120

Qy       5581 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 168
```

```
RESULT 52
BP336529
LOCUS       BP336529                461 bp    mRNA    linear   EST 17-SEP-2004
DEFINITION  BP336529 Sugano cDNA library, coronary artery smooth muscle cell
            Homo sapiens cDNA clone SMR05907, mRNA sequence.
ACCESSION   BP336529
VERSION     BP336529.1  GI:52266104
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 461)
  AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
  TITLE     Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
  JOURNAL   Genome Res. 14 (9), 1711-1718 (2004)
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .461
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="SMR05907"
                     /tissue_type="coronary artery"
                     /cell_type="smooth muscle cell"
                     /clone_lib="Sugano cDNA library, coronary artery smooth
                     muscle cell"
ORIGIN

  Query Match             2.8%;  Score 168;  DB 5;  Length 461;
  Best Local Similarity   100.0%;  Pred. No. 4.7e-29;
  Matches  168;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5461 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 5520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 60

Qy       5521 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 5580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 120

Qy       5581 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 168


RESULT 53
BP336576
LOCUS       BP336576                461 bp    mRNA    linear   EST 17-SEP-2004
DEFINITION  BP336576 Sugano cDNA library, coronary artery smooth muscle cell
            Homo sapiens cDNA clone SMR06025, mRNA sequence.
ACCESSION   BP336576
VERSION     BP336576.1  GI:52266154
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 461)
  AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
  TITLE     Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
  JOURNAL   Genome Res. 14 (9), 1711-1718 (2004)
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .461
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="SMR06025"
                     /tissue_type="coronary artery"
                     /cell_type="smooth muscle cell"
                     /clone_lib="Sugano cDNA library, coronary artery smooth
                     muscle cell"
ORIGIN

  Query Match             2.8%;  Score 168;  DB 5;  Length 461;
  Best Local Similarity   100.0%;  Pred. No. 4.7e-29;
  Matches  168;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5461 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 5520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 60

Qy       5521 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 5580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 120

Qy       5581 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 168


RESULT 54
BP199046
LOCUS       BP199046                462 bp    mRNA    linear   EST 14-SEP-2004
DEFINITION  BP199046 Sugano cDNA library, brain Homo sapiens cDNA clone
            ADR06421, mRNA sequence.
ACCESSION   BP199046
VERSION     BP199046.1  GI:52046605
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 462)
  AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
  TITLE     Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
  JOURNAL   Genome Res. 14 (9), 1711-1718 (2004)
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .462
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="ADR06421"
                     /tissue_type="brain"
                     /clone_lib="Sugano cDNA library, brain"
ORIGIN

  Query Match             2.8%;  Score 168;  DB 5;  Length 462;
  Best Local Similarity   100.0%;  Pred. No. 4.7e-29;
  Matches  168;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5461 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 5520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 61
```

```
Qy       5521 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 5580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 121

Qy       5581 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 169


RESULT 55
BP235857
LOCUS       BP235857                 544 bp    mRNA    linear   EST 15-SEP-2004
DEFINITION  BP235857 Sugano cDNA library, coronary artery endothelial cell Homo
            sapiens cDNA clone HCR06094, mRNA sequence.
ACCESSION   BP235857
VERSION     BP235857.1  GI:52108767
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 544)
  AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
  TITLE     Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
  JOURNAL   Genome Res. 14 (9), 1711-1718 (2004)
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .544
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="HCR06094"
                     /tissue_type="coronary artery"
                     /cell_type="endothelial cell"
                     /clone_lib="Sugano cDNA library, coronary artery
                     endothelial cell"
ORIGIN

  Query Match             2.8%;  Score 168;  DB 5;  Length 544;
  Best Local Similarity   100.0%;  Pred. No. 4.8e-29;
  Matches  168;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5461 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 5520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 60

Qy       5521 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 5580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 120

Qy       5581 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 168


RESULT 56
AU280510
LOCUS       AU280510                 568 bp    mRNA    linear   EST 31-JUL-2003
DEFINITION  AU280510 N1ESE2 Homo sapiens cDNA clone N1ESE2002347 5', mRNA
            sequence.
ACCESSION   AU280510
VERSION     AU280510.1  GI:28299737
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 568)
  AUTHORS   Imabayashi,H., Mori,T., Gojo,S., Kiyono,T., Sugiyama,T., Irie,R.,
            Isogai,T., Hata,J., Tomoya,Y. and Umezawa,A.
  TITLE     Redifferentiation of dedifferentiated chondrocytes and
            chondrogenesis of human bone marrow stromal cells via chondrosphere
            formation with expression profiling by large-scale cDNA analysis
  JOURNAL   Exp. Cell Res. 288 (1), 35-50 (2003)
  MEDLINE   22760698
   PUBMED   12878157
COMMENT     Contact: Takao Isogai
            Genomics Laboratory
            Helix Research Institute
            1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            Tel: 81-438-52-3975
            Fax: 81-438-52-3986
            Email: genomics@hri.co.jp
            HRI human cDNA Project, Sugiyama,T.; Wakamatsu,A.; Irie,R.;
            Umezawa,A.; Fukuma,M.; Kusakari,S.; Hata,J.; Ishii,S.; Yamamoto,J.;
            Isono,Y.; Saito,K.; Nakamura,Y.; Masuho,Y.; Nagai,K.; Isogai,T.
             HRI human cDNA project; cDNA library construction & 5'-end one
            pass sequencing: Helix Research Institute.
FEATURES             Location/Qualifiers
     source          1. .568
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="N1ESE2002347"
                     /cell_type="mesenchymal stem cells"
                     /clone_lib="N1ESE2"
                     /note="Vector: pME18SFL3"
ORIGIN

  Query Match             2.8%;  Score 168;  DB 1;  Length 568;
  Best Local Similarity   100.0%;  Pred. No. 4.9e-29;
  Matches  168;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5461 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 5520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 60

Qy       5521 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 5580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 120

Qy       5581 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 168


RESULT 57
CB155278
LOCUS       CB155278                 574 bp    mRNA    linear   EST 29-JAN-2003
DEFINITION  K-EST0213503 B2N807043 Homo sapiens cDNA clone B2N807043-14-F07 5',
            mRNA sequence.
ACCESSION   CB155278
VERSION     CB155278.1  GI:28140388
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 574)
  AUTHORS   Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
            Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
            Kim,Y.S.
  TITLE     21C Frontier Korean EST Project 2001
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Kim YS
            Genome Research Center
            Korea Research Institute of Bioscience & Biotechnology
            52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
            Tel: +82-42-860-4470
            Fax: +82-42-860-4409
```

```
            Email: yongsung@mail.kribb.re.kr
            Plate: 14  row: F  column: 07
            High quality sequence stop: 574.
FEATURES              Location/Qualifiers
     source           1. .574
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="B2N807043-14-F07"
                      /sex="M"
                      /lab_host="Top10F'"
                      /clone_lib="B2N807043"
                      /note="Organ: Brain; Vector: pCNS-D2; Site_1: EcoRI;
                      Site_2: NotI; The poly (A)+ RNA was dephosphorylated with
                      bacterial alkaline phosphatase (BAP) and then decapped
                      with tabacco acid pyrophosphatase (TAP). The decapped
                      intact mRNA was ligated with DNA-RNA linker including
                      EcoRI site by treatment of T4 RNA ligase and the first
                      strand cDNA was synthesized from oligo dT-selected mRNA by
                      priming with dT-tailed vector. The dT-tailed vector was
                      adjusted to have about 60nt. The cDNA vector was
                      circularized with E. coli DNA ligase after digestion of
                      EcoRI which site is also included in vector. An RNA strand
                      converted to a DNA strand by Okayama-Berg method. The
                      obtained cDNA vectors were used for transformation of
                      competent cells E. coli Top10F' by electroporation method.
                      The cDNA libraries constructed by this method are
                      full-length enriched cDNA library."
ORIGIN

  Query Match              2.8%;  Score 168;  DB 6;  Length 574;
  Best Local Similarity   100.0%;  Pred. No. 4.9e-29;
  Matches  168;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5461 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 5520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 60

Qy       5521 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 5580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 120

Qy       5581 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 168


RESULT 58
BP211606
LOCUS       BP211606                581 bp    mRNA    linear   EST 15-SEP-2004
DEFINITION  BP211606 Sugano cDNA library, cerebrum Homo sapiens cDNA clone
            CBR01343, mRNA sequence.
ACCESSION   BP211606
VERSION     BP211606.1  GI:52084497
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 581)
  AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
  TITLE     Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
  JOURNAL   Genome Res. 14 (9), 1711-1718 (2004)
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp.
FEATURES              Location/Qualifiers
     source           1. .581
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="CBR01343"
                      /tissue_type="cerebrum"
                      /clone_lib="Sugano cDNA library, cerebrum"
ORIGIN

  Query Match              2.8%;  Score 168;  DB 5;  Length 581;
  Best Local Similarity   100.0%;  Pred. No. 4.9e-29;
  Matches  168;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5461 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 5520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 60

Qy       5521 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 5580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 120

Qy       5581 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 168


RESULT 59
BP342248
LOCUS       BP342248                581 bp    mRNA    linear   EST 17-SEP-2004
DEFINITION  BP342248 Sugano cDNA library, stomach mucosa Homo sapiens cDNA
            clone STM04507, mRNA sequence.
ACCESSION   BP342248
VERSION     BP342248.1  GI:52272033
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 581)
  AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
  TITLE     Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
  JOURNAL   Genome Res. 14 (9), 1711-1718 (2004)
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp.
FEATURES              Location/Qualifiers
     source           1. .581
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="STM04507"
                      /tissue_type="stomach mucosa"
                      /clone_lib="Sugano cDNA library, stomach mucosa"
ORIGIN

  Query Match              2.8%;  Score 168;  DB 5;  Length 581;
  Best Local Similarity   100.0%;  Pred. No. 4.9e-29;
  Matches  168;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5461 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 5520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 60

Qy       5521 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 5580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 120

Qy       5581 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 168
```

```
RESULT 60
BP202778
LOCUS       BP202778                582 bp    mRNA    linear   EST 14-SEP-2004
DEFINITION  BP202778 Sugano cDNA library, amygdala Homo sapiens cDNA clone
            AMR09461, mRNA sequence.
ACCESSION   BP202778
VERSION     BP202778.1  GI:52054113
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 582)
  AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
  TITLE     Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
  JOURNAL   Genome Res. 14 (9), 1711-1718 (2004)
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .582
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="AMR09461"
                     /tissue_type="amygdala"
                     /clone_lib="Sugano cDNA library, amygdala"
ORIGIN

  Query Match             2.8%;  Score 168;  DB 5;  Length 582;
  Best Local Similarity   100.0%;  Pred. No. 4.9e-29;
  Matches  168;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5461 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 5520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 60

Qy       5521 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 5580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 120

Qy       5581 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             |||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 168


RESULT 61
BP233479
LOCUS       BP233479                582 bp    mRNA    linear   EST 15-SEP-2004
DEFINITION  BP233479 Sugano cDNA library, coronary artery endothelial cell Homo
            sapiens cDNA clone HCR00513, mRNA sequence.
ACCESSION   BP233479
VERSION     BP233479.1  GI:52106389
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 582)
  AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
  TITLE     Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
  JOURNAL   Genome Res. 14 (9), 1711-1718 (2004)
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .582
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="HCR00513"
                     /tissue_type="coronary artery"
                     /cell_type="endothelial cell"
                     /clone_lib="Sugano cDNA library, coronary artery
                     endothelial cell"
ORIGIN

  Query Match             2.8%;  Score 168;  DB 5;  Length 582;
  Best Local Similarity   100.0%;  Pred. No. 4.9e-29;
  Matches  168;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5461 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 5520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 60

Qy       5521 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 5580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 120

Qy       5581 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             |||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 168


RESULT 62
BP236642
LOCUS       BP236642                582 bp    mRNA    linear   EST 15-SEP-2004
DEFINITION  BP236642 Sugano cDNA library, coronary artery endothelial cell Homo
            sapiens cDNA clone HCR08129, mRNA sequence.
ACCESSION   BP236642
VERSION     BP236642.1  GI:52109552
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 582)
  AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
  TITLE     Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
  JOURNAL   Genome Res. 14 (9), 1711-1718 (2004)
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .582
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="HCR08129"
                     /tissue_type="coronary artery"
                     /cell_type="endothelial cell"
                     /clone_lib="Sugano cDNA library, coronary artery
                     endothelial cell"
ORIGIN

  Query Match             2.8%;  Score 168;  DB 5;  Length 582;
  Best Local Similarity   100.0%;  Pred. No. 4.9e-29;
  Matches  168;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5461 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 5520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 60
```

```
Qy       5521 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 5580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 120

Qy       5581 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 168


RESULT 63
BP290624
LOCUS       BP290624                582 bp    mRNA    linear   EST 16-SEP-2004
DEFINITION  BP290624 Sugano cDNA library, lung fibroblast Homo sapiens cDNA
            clone LNF05765, mRNA sequence.
ACCESSION   BP290624
VERSION     BP290624.1  GI:52204356
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 582)
  AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
  TITLE     Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
  JOURNAL   Genome Res. 14 (9), 1711-1718 (2004)
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .582
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="LNF05765"
                     /tissue_type="lung"
                     /cell_type="fibroblast"
                     /clone_lib="Sugano cDNA library, lung fibroblast"
                     /note="normal"
ORIGIN

  Query Match             2.8%;  Score 168;  DB 5;  Length 582;
  Best Local Similarity   100.0%;  Pred. No. 4.9e-29;
  Matches  168;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5461 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 5520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 60

Qy       5521 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 5580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 120

Qy       5581 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 168


RESULT 64
BP349089
LOCUS       BP349089                582 bp    mRNA    linear   EST 17-SEP-2004
DEFINITION  BP349089 Sugano cDNA library, brain Homo sapiens cDNA clone
            SZR03078, mRNA sequence.
ACCESSION   BP349089
VERSION     BP349089.1  GI:52279074
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 582)
  AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
  TITLE     Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
  JOURNAL   Genome Res. 14 (9), 1711-1718 (2004)
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .582
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="SZR03078"
                     /tissue_type="brain"
                     /clone_lib="Sugano cDNA library, brain"
ORIGIN

  Query Match             2.8%;  Score 168;  DB 5;  Length 582;
  Best Local Similarity   100.0%;  Pred. No. 4.9e-29;
  Matches  168;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5461 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 5520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 60

Qy       5521 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 5580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 120

Qy       5581 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 168


RESULT 65
BP235750
LOCUS       BP235750                583 bp    mRNA    linear   EST 15-SEP-2004
DEFINITION  BP235750 Sugano cDNA library, coronary artery endothelial cell Homo
            sapiens cDNA clone HCR05823, mRNA sequence.
ACCESSION   BP235750
VERSION     BP235750.1  GI:52108660
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 583)
  AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
  TITLE     Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
  JOURNAL   Genome Res. 14 (9), 1711-1718 (2004)
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .583
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="HCR05823"
                     /tissue_type="coronary artery"
                     /cell_type="endothelial cell"
                     /clone_lib="Sugano cDNA library, coronary artery
                     endothelial cell"
ORIGIN
```

```
  Query Match             2.8%;  Score 168;  DB 5;  Length 583;
  Best Local Similarity   100.0%;  Pred. No. 4.9e-29;
  Matches  168;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5461 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 5520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 60

Qy       5521 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 5580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 120

Qy       5581 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 168


RESULT 66
BP236264
LOCUS       BP236264                583 bp    mRNA    linear   EST 15-SEP-2004
DEFINITION  BP236264 Sugano cDNA library, coronary artery endothelial cell Homo
            sapiens cDNA clone HCR07196, mRNA sequence.
ACCESSION   BP236264
VERSION     BP236264.1  GI:52109174
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 583)
  AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
  TITLE     Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
  JOURNAL   Genome Res. 14 (9), 1711-1718 (2004)
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .583
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="HCR07196"
                     /tissue_type="coronary artery"
                     /cell_type="endothelial cell"
                     /clone_lib="Sugano cDNA library, coronary artery
                     endothelial cell"
ORIGIN

  Query Match             2.8%;  Score 168;  DB 5;  Length 583;
  Best Local Similarity   100.0%;  Pred. No. 4.9e-29;
  Matches  168;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5461 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 5520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 60

Qy       5521 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 5580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 120

Qy       5581 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 168


RESULT 67
BP278756
LOCUS       BP278756                583 bp    mRNA    linear   EST 16-SEP-2004
DEFINITION  BP278756 Sugano cDNA library, KG-1-C Homo sapiens cDNA clone
            KGR02772, mRNA sequence.
ACCESSION   BP278756
VERSION     BP278756.1  GI:52192488
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 583)
  AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
  TITLE     Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
  JOURNAL   Genome Res. 14 (9), 1711-1718 (2004)
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .583
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="KGR02772"
                     /cell_line="KG-1-C"
                     /clone_lib="Sugano cDNA library, KG-1-C"
                     /note="glioma"
ORIGIN

  Query Match             2.8%;  Score 168;  DB 5;  Length 583;
  Best Local Similarity   100.0%;  Pred. No. 4.9e-29;
  Matches  168;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5461 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 5520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 60

Qy       5521 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 5580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 120

Qy       5581 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 168


RESULT 68
BP279792
LOCUS       BP279792                583 bp    mRNA    linear   EST 16-SEP-2004
DEFINITION  BP279792 Sugano cDNA library, KG-1-C Homo sapiens cDNA clone
            KGR06056, mRNA sequence.
ACCESSION   BP279792
VERSION     BP279792.1  GI:52193524
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 583)
  AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
  TITLE     Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
  JOURNAL   Genome Res. 14 (9), 1711-1718 (2004)
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .583
                     /organism="Homo sapiens"
```

```
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="KGR06056"
                     /cell_line="KG-1-C"
                     /clone_lib="Sugano cDNA library, KG-1-C"
                     /note="glioma"
ORIGIN

  Query Match             2.8%;  Score 168;  DB 5;  Length 583;
  Best Local Similarity   100.0%;  Pred. No. 4.9e-29;
  Matches  168;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5461 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 5520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 60

Qy       5521 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 5580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 120

Qy       5581 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 168


RESULT 69
BP335473
LOCUS       BP335473                583 bp    mRNA    linear   EST 17-SEP-2004
DEFINITION  BP335473 Sugano cDNA library, coronary artery smooth muscle cell
            Homo sapiens cDNA clone SMR01424, mRNA sequence.
ACCESSION   BP335473
VERSION     BP335473.1  GI:52265024
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 583)
  AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
  TITLE     Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
  JOURNAL   Genome Res. 14 (9), 1711-1718 (2004)
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .583
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="SMR01424"
                     /tissue_type="coronary artery"
                     /cell_type="smooth muscle cell"
                     /clone_lib="Sugano cDNA library, coronary artery smooth
                     muscle cell"
ORIGIN

  Query Match             2.8%;  Score 168;  DB 5;  Length 583;
  Best Local Similarity   100.0%;  Pred. No. 4.9e-29;
  Matches  168;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5461 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 5520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 60

Qy       5521 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 5580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 120

Qy       5581 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 168


RESULT 70
BP233940
LOCUS       BP233940                587 bp    mRNA    linear   EST 15-SEP-2004
DEFINITION  BP233940 Sugano cDNA library, coronary artery endothelial cell Homo
            sapiens cDNA clone HCR01671, mRNA sequence.
ACCESSION   BP233940
VERSION     BP233940.1  GI:52106850
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 587)
  AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
  TITLE     Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
  JOURNAL   Genome Res. 14 (9), 1711-1718 (2004)
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .587
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="HCR01671"
                     /tissue_type="coronary artery"
                     /cell_type="endothelial cell"
                     /clone_lib="Sugano cDNA library, coronary artery
                     endothelial cell"
ORIGIN

  Query Match             2.8%;  Score 168;  DB 5;  Length 587;
  Best Local Similarity   100.0%;  Pred. No. 4.9e-29;
  Matches  168;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5461 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 5520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 60

Qy       5521 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 5580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 120

Qy       5581 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 168


RESULT 71
BP250409
LOCUS       BP250409                592 bp    mRNA    linear   EST 15-SEP-2004
DEFINITION  BP250409 Sugano cDNA library, hippocampus Homo sapiens cDNA clone
            HPR02985, mRNA sequence.
ACCESSION   BP250409
VERSION     BP250409.1  GI:52132688
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 592)
  AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
  TITLE     Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
```

```
  JOURNAL   Genome Res. 14 (9), 1711-1718 (2004)
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .592
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="HPR02985"
                     /tissue_type="hippocampus"
                     /clone_lib="Sugano cDNA library, hippocampus"
ORIGIN

  Query Match             2.8%;  Score 168;  DB 5;  Length 592;
  Best Local Similarity   100.0%;  Pred. No. 4.9e-29;
  Matches  168;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5461 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 5520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 60

Qy       5521 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 5580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 120

Qy       5581 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 168


RESULT 72
BP251461
LOCUS       BP251461                608 bp    mRNA    linear   EST 15-SEP-2004
DEFINITION  BP251461 Sugano cDNA library, hippocampus Homo sapiens cDNA clone
            HPR07808, mRNA sequence.
ACCESSION   BP251461
VERSION     BP251461.1  GI:52133742
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 608)
  AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
  TITLE     Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
  JOURNAL   Genome Res. 14 (9), 1711-1718 (2004)
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .608
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="HPR07808"
                     /tissue_type="hippocampus"
                     /clone_lib="Sugano cDNA library, hippocampus"
ORIGIN

  Query Match             2.8%;  Score 168;  DB 5;  Length 608;
  Best Local Similarity   100.0%;  Pred. No. 5e-29;
  Matches  168;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5461 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 5520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 60

Qy       5521 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 5580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 120

Qy       5581 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 168


RESULT 73
BX951606
LOCUS       BX951606                581 bp    mRNA    linear   EST 01-MAR-2004
DEFINITION  DKFZp781G12152_r1 781 (synonym: hlcc4) Homo sapiens cDNA clone
            DKFZp781G12152 5', mRNA sequence.
ACCESSION   BX951606
VERSION     BX951606.1  GI:43426898
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 581)
  AUTHORS   Wambutt,R., Heubner,D., Mewes,H.W., Weil,B., Amid,C., Osanger,A.,
            Fobo,G., Han,M. and Wiemann,S.
  TITLE     EST (Wambutt,R., Heubner,D., Mewes,H.W., Weil,B., Amid,C., et al.)
  JOURNAL   Unpublished (2003)
COMMENT     Contact: MIPS
            MIPS
            Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
            This is the 5' sequence of the clone insert
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz- heidelberg.de;
            sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing
            consortium of the German Genome Project.
            No s1 sequence available.
            This clone (DKFZp781G12152) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES             Location/Qualifiers
     source          1. .581
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="DKFZp781G12152"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="781 (synonym: hlcc4)"
                     /note="Vector: pSport1_Sfi; Site_1: SfiIA; Site_2: SfiIB;
                     cDNA-collection"
ORIGIN

  Query Match             2.8%;  Score 167.4;  DB 5;  Length 581;
  Best Local Similarity   99.4%;  Pred. No. 6.9e-29;
  Matches  168;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       5460 GAAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAA 5519
             | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GGAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAA 60

Qy       5520 GCCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAA 5579
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAA 120

Qy       5580 ATAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ATAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 169


RESULT 74
BI552207
LOCUS       BI552207                958 bp    mRNA    linear   EST 05-SEP-2001
DEFINITION  603195256F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:5274876 5',
```

```
            mRNA sequence.
ACCESSION   BI552207
VERSION     BI552207.1  GI:15439519
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 958)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
             cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11693  row: p  column: 13
            High quality sequence stop: 754.
FEATURES             Location/Qualifiers
     source          1. .958
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5274876"
                     /tissue_type="hippocampus"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_95"
                     /note="Organ: brain; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI
                     (gtcgag); Oligo-dT primed using primer
                     5'-TTTTTTTTTTTTTTTVN-3', size-selected for average
                     insert size 2.5 kb and normalized to ROT 5. This is a
                     primary library enriched for full-length clones and
                     constructed using the Cap-trapper method (Carninci, in
                     preparation). Library constructed by M. Brownstein
                     (NIMH/NHGRI, National Institutes of Health). Note: this
                     is a NIH_MGC Library."
ORIGIN

  Query Match             2.8%;  Score 167.4;  DB 4;  Length 958;
  Best Local Similarity   99.4%;  Pred. No. 7.7e-29;
  Matches  168;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       5460 GAAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAA 5519
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          4 GGAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAA 63

Qy       5520 GCCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAA 5579
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         64 GCCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAA 123

Qy       5580 ATAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        124 ATAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 172


RESULT 75
BP201862
LOCUS       BP201862                460 bp    mRNA    linear   EST 14-SEP-2004
DEFINITION  BP201862 Sugano cDNA library, amygdala Homo sapiens cDNA clone
            AMR06659, mRNA sequence.
ACCESSION   BP201862
VERSION     BP201862.1  GI:52052259
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 460)
  AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
  TITLE     Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
  JOURNAL   Genome Res. 14 (9), 1711-1718 (2004)
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .460
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="AMR06659"
                     /tissue_type="amygdala"
                     /clone_lib="Sugano cDNA library, amygdala"
ORIGIN

  Query Match             2.7%;  Score 167;  DB 5;  Length 460;
  Best Local Similarity   100.0%;  Pred. No. 8.1e-29;
  Matches  167;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5462 AGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGC 5521
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGC 60

Qy       5522 CGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAAT 5581
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAAT 120

Qy       5582 AAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
              |||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 167


RESULT 76
BP236908
LOCUS       BP236908                460 bp    mRNA    linear   EST 15-SEP-2004
DEFINITION  BP236908 Sugano cDNA library, coronary artery endothelial cell Homo
            sapiens cDNA clone HCR08766, mRNA sequence.
ACCESSION   BP236908
VERSION     BP236908.1  GI:52109818
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 460)
  AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
  TITLE     Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
  JOURNAL   Genome Res. 14 (9), 1711-1718 (2004)
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .460
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="HCR08766"
                     /tissue_type="coronary artery"
                     /cell_type="endothelial cell"
                     /clone_lib="Sugano cDNA library, coronary artery
                     endothelial cell"
ORIGIN
```

```
  Query Match             2.7%;  Score 167;  DB 5;  Length 460;
  Best Local Similarity   100.0%;  Pred. No. 8.1e-29;
  Matches  167;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5462 AGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGC 5521
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGC 60

Qy       5522 CGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAAT 5581
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAAT 120

Qy       5582 AAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             |||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 167


RESULT 77
BP336320
LOCUS       BP336320                460 bp    mRNA    linear   EST 17-SEP-2004
DEFINITION  BP336320 Sugano cDNA library, coronary artery smooth muscle cell
            Homo sapiens cDNA clone SMR05360, mRNA sequence.
ACCESSION   BP336320
VERSION     BP336320.1  GI:52265883
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 460)
  AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
  TITLE     Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
  JOURNAL   Genome Res. 14 (9), 1711-1718 (2004)
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .460
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="SMR05360"
                     /tissue_type="coronary artery"
                     /cell_type="smooth muscle cell"
                     /clone_lib="Sugano cDNA library, coronary artery smooth
                     muscle cell"
ORIGIN

  Query Match             2.7%;  Score 167;  DB 5;  Length 460;
  Best Local Similarity   100.0%;  Pred. No. 8.1e-29;
  Matches  167;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5462 AGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGC 5521
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGC 60

Qy       5522 CGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAAT 5581
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAAT 120

Qy       5582 AAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             |||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 167


RESULT 78
BP337220
LOCUS       BP337220                460 bp    mRNA    linear   EST 17-SEP-2004
DEFINITION  BP337220 Sugano cDNA library, coronary artery smooth muscle cell
            Homo sapiens cDNA clone SMR07801, mRNA sequence.
ACCESSION   BP337220
VERSION     BP337220.1  GI:52266826
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 460)
  AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
  TITLE     Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
  JOURNAL   Genome Res. 14 (9), 1711-1718 (2004)
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .460
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="SMR07801"
                     /tissue_type="coronary artery"
                     /cell_type="smooth muscle cell"
                     /clone_lib="Sugano cDNA library, coronary artery smooth
                     muscle cell"
ORIGIN

  Query Match             2.7%;  Score 167;  DB 5;  Length 460;
  Best Local Similarity   100.0%;  Pred. No. 8.1e-29;
  Matches  167;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5462 AGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGC 5521
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGC 60

Qy       5522 CGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAAT 5581
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAAT 120

Qy       5582 AAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             |||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 167


RESULT 79
BP337084
LOCUS       BP337084                574 bp    mRNA    linear   EST 17-SEP-2004
DEFINITION  BP337084 Sugano cDNA library, coronary artery smooth muscle cell
            Homo sapiens cDNA clone SMR07441, mRNA sequence.
ACCESSION   BP337084
VERSION     BP337084.1  GI:52266683
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 574)
  AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
  TITLE     Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
  JOURNAL   Genome Res. 14 (9), 1711-1718 (2004)
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp.
FEATURES             Location/Qualifiers
```

```
     source          1. .574
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="SMR07441"
                     /tissue_type="coronary artery"
                     /cell_type="smooth muscle cell"
                     /clone_lib="Sugano cDNA library, coronary artery smooth
                     muscle cell"
ORIGIN

  Query Match             2.7%;  Score 167;  DB 5;  Length 574;
  Best Local Similarity   100.0%;  Pred. No. 8.6e-29;
  Matches  167;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5462 AGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGC 5521
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGC 60

Qy       5522 CGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAAT 5581
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAAT 120

Qy       5582 AAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             |||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 167


RESULT 80
BP310499
LOCUS       BP310499                582 bp    mRNA    linear   EST 17-SEP-2004
DEFINITION  BP310499 Sugano cDNA library, brain Homo sapiens cDNA clone
            NRR02679, mRNA sequence.
ACCESSION   BP310499
VERSION     BP310499.1  GI:52239474
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 582)
  AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
  TITLE     Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
  JOURNAL   Genome Res. 14 (9), 1711-1718 (2004)
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .582
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="NRR02679"
                     /tissue_type="brain"
                     /clone_lib="Sugano cDNA library, brain"
ORIGIN

  Query Match             2.7%;  Score 167;  DB 5;  Length 582;
  Best Local Similarity   100.0%;  Pred. No. 8.6e-29;
  Matches  167;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5462 AGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGC 5521
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 AGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGC 61

Qy       5522 CGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAAT 5581
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 CGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAAT 121

Qy       5582 AAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             |||||||||||||||||||||||||||||||||||||||||||||||
Db        122 AAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 168


RESULT 81
BP208499
LOCUS       BP208499                583 bp    mRNA    linear   EST 14-SEP-2004
DEFINITION  BP208499 Sugano cDNA library, cerebellum Homo sapiens cDNA clone
            CBL01281, mRNA sequence.
ACCESSION   BP208499
VERSION     BP208499.1  GI:52065652
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 583)
  AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
  TITLE     Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
  JOURNAL   Genome Res. 14 (9), 1711-1718 (2004)
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .583
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CBL01281"
                     /tissue_type="cerebellum"
                     /clone_lib="Sugano cDNA library, cerebellum"
ORIGIN

  Query Match             2.7%;  Score 167;  DB 5;  Length 583;
  Best Local Similarity   100.0%;  Pred. No. 8.6e-29;
  Matches  167;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5462 AGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGC 5521
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          3 AGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGC 62

Qy       5522 CGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAAT 5581
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         63 CGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAAT 122

Qy       5582 AAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             |||||||||||||||||||||||||||||||||||||||||||||||
Db        123 AAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 169


RESULT 82
BP234676
LOCUS       BP234676                584 bp    mRNA    linear   EST 15-SEP-2004
DEFINITION  BP234676 Sugano cDNA library, coronary artery endothelial cell Homo
            sapiens cDNA clone HCR03457, mRNA sequence.
ACCESSION   BP234676
VERSION     BP234676.1  GI:52107586
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 584)
  AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
  TITLE     Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
  JOURNAL   Genome Res. 14 (9), 1711-1718 (2004)
```

```
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .584
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="HCR03457"
                     /tissue_type="coronary artery"
                     /cell_type="endothelial cell"
                     /clone_lib="Sugano cDNA library, coronary artery
                     endothelial cell"
ORIGIN

  Query Match             2.7%;  Score 167;  DB 5;  Length 584;
  Best Local Similarity   100.0%;  Pred. No. 8.6e-29;
  Matches  167;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5462 AGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGC 5521
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGC 60

Qy       5522 CGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAAT 5581
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAAT 120

Qy       5582 AAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             ||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 167


RESULT 83
BP212771
LOCUS       BP212771                585 bp    mRNA    linear   EST 15-SEP-2004
DEFINITION  BP212771 Sugano cDNA library, cerebrum Homo sapiens cDNA clone
            CBR04780, mRNA sequence.
ACCESSION   BP212771
VERSION     BP212771.1  GI:52085662
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 585)
  AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
  TITLE     Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
  JOURNAL   Genome Res. 14 (9), 1711-1718 (2004)
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .585
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CBR04780"
                     /tissue_type="cerebrum"
                     /clone_lib="Sugano cDNA library, cerebrum"
ORIGIN

  Query Match             2.7%;  Score 167;  DB 5;  Length 585;
  Best Local Similarity   100.0%;  Pred. No. 8.6e-29;
  Matches  167;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5462 AGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGC 5521
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGC 60

Qy       5522 CGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAAT 5581
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAAT 120

Qy       5582 AAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             ||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 167


RESULT 84
BP336664
LOCUS       BP336664                598 bp    mRNA    linear   EST 17-SEP-2004
DEFINITION  BP336664 Sugano cDNA library, coronary artery smooth muscle cell
            Homo sapiens cDNA clone SMR06273, mRNA sequence.
ACCESSION   BP336664
VERSION     BP336664.1  GI:52266247
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 598)
  AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
  TITLE     Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
  JOURNAL   Genome Res. 14 (9), 1711-1718 (2004)
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .598
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="SMR06273"
                     /tissue_type="coronary artery"
                     /cell_type="smooth muscle cell"
                     /clone_lib="Sugano cDNA library, coronary artery smooth
                     muscle cell"
ORIGIN

  Query Match             2.7%;  Score 167;  DB 5;  Length 598;
  Best Local Similarity   100.0%;  Pred. No. 8.6e-29;
  Matches  167;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5462 AGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGC 5521
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGC 60

Qy       5522 CGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAAT 5581
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAAT 120

Qy       5582 AAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             ||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 167


RESULT 85
CR539890
LOCUS       CR539890                586 bp    mRNA    linear   EST 06-JUL-2004
DEFINITION  DKFZp459I175_r1 459 (synonym: pcor1) Pongo pygmaeus cDNA clone
            DKFZp459I175 5', mRNA sequence.
ACCESSION   CR539890
VERSION     CR539890.1  GI:49687006
KEYWORDS    EST.
SOURCE      Pongo pygmaeus (orangutan)
  ORGANISM  Pongo pygmaeus
```

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pongo.
REFERENCE   1  (bases 1 to 586)
  AUTHORS   Ansorge,W., Krieger,S., Regiert,T., Rittmueller,C., Schwager,B.,
            Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and
            Wiemann,S.
  TITLE     Pongo pygmaeus mRNA (Ansorge,W., Krieger,S., Regiert,T., et al.)
  JOURNAL   Unpublished (2004)
COMMENT     Contact: MIPS
            MIPS
            Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
            This is the 5' sequence of the clone insert Clone from S. Wiemann,
            Molecular Genome Analysis, German Cancer Research Center (DKFZ);
            Email s.wiemann@dkfz-heidelberg.de; sequenced by EMBL (European
            Molecular Biology Laboratories, Heidelberg/Germany) within the cDNA
            sequencing consortium of the German Genome Project. This clone
            (DKFZp459I175) is available at the RZPD in Berlin. Please contact
            the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
            information about the clone and the sequencing project is available
            at http://mips.gsf.de/projects/cdna/.
FEATURES             Location/Qualifiers
     source          1. .586
                     /organism="Pongo pygmaeus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9600"
                     /clone="DKFZp459I175"
                     /tissue_type="cortex"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="459 (synonym: pcor1)"
                     /note="Vector: pSport1_Sfi; Site_1: SfilA; Site_2: SfilB"
ORIGIN

  Query Match             2.7%;  Score 166.8;  DB 7;  Length 586;
  Best Local Similarity   98.8%;  Pred. No. 9.6e-29;
  Matches  168;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       5459 GGAAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAA 5518
             |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db          2 GGAAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAATAGGCAAAGTACGCTCAA 61

Qy       5519 AGCCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTA 5578
             |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db         62 AGCCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCAAATTCCAAGCTGTTA 121

Qy       5579 AATAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 AATAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 171


RESULT 86
CR544736
LOCUS       CR544736                600 bp    mRNA    linear   EST 07-JUL-2004
DEFINITION  DKFZp459L0324_r1 459 (synonym: pcor1) Pongo pygmaeus cDNA clone
            DKFZp459L0324 5', mRNA sequence.
ACCESSION   CR544736
VERSION     CR544736.1  GI:49896428
KEYWORDS    EST.
SOURCE      Pongo pygmaeus (orangutan)
  ORGANISM  Pongo pygmaeus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pongo.
REFERENCE   1  (bases 1 to 600)
  AUTHORS   Koehrer,K., Beyer,A., Mewes,H.W., Weil,B., Amid,C., Osanger,A.,
            Fobo,G., Han,M. and Wiemann,S.
  TITLE     Pongo pygmaeus mRNA (Koehrer,K., Beyer,A., Mewes,H.W., et al.)
  JOURNAL   Unpublished (2004)
COMMENT     Contact: MIPS
            MIPS
            Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
            This is the 5' sequence of the clone insert Clone from S. Wiemann,
            Molecular Genome Analysis, German Cancer Research Center (DKFZ);
            Email s.wiemann@dkfz-heidelberg.de; sequenced by BMFZ (Biomedical
            Research Center at the Heinrich-Heine-University,
            Duesseldorf/Germany) within the cDNA sequencing consortium of the
            German Genome Project. This clone (DKFZp459L0324) is available at
            the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum,
            Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email:
            clone@rzpd.de Further information about the clone and the
            sequencing project is available at
            http://mips.gsf.de/projects/cdna/.
FEATURES             Location/Qualifiers
     source          1. .600
                     /organism="Pongo pygmaeus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9600"
                     /clone="DKFZp459L0324"
                     /tissue_type="cortex"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="459 (synonym: pcor1)"
                     /note="Vector: pSport1_Sfi; Site_1: SfilA; Site_2: SfilB"
ORIGIN

  Query Match             2.7%;  Score 166.8;  DB 7;  Length 600;
  Best Local Similarity   98.8%;  Pred. No. 9.7e-29;
  Matches  168;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       5459 GGAAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAA 5518
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         17 GGAAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAA 76

Qy       5519 AGCCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTA 5578
             |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db         77 AGCCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCAAATTCCAAGCTGTTA 136

Qy       5579 AATAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             ||||||||||||||||||||||||| |||||||||||||||||||||||||
Db        137 AATAAGATGTGCAAAGGGCTTGCAAGTCTGCCGGCTTCTTGCTTGAGGAG 186


RESULT 87
BG708946
LOCUS       BG708946                618 bp    mRNA    linear   EST 07-MAY-2001
DEFINITION  602674887F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:4797525 5',
            mRNA sequence.
ACCESSION   BG708946
VERSION     BG708946.1  GI:13986793
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 618)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
             cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10683  row: n  column: 22
            High quality sequence stop: 615.
FEATURES             Location/Qualifiers
     source          1. .618
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4797525"
```

/tissue_type="hypothalamus"
/lab_host="DH10B"
/clone_lib="NIH_MGC_96"
/note="Organ: brain; Vector: pBluescriptR (modified pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTTVN-3', size-selected for average insert size 2.3 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIMH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."

ORIGIN

Query Match 2.7%; Score 166.8; DB 4; Length 618;
Best Local Similarity 98.8%; Pred. No. 9.7e-29;
Matches 168; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy       5459 GGAAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAA 5518
             ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          4 GGAAGGCGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAA 63

Qy       5519 AGCCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTA 5578
             |||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db         64 AGCCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTTCTTGCGAATTCCAAGCTGTTA 123

Qy       5579 AATAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        124 AATAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 173
```

RESULT 88
CR549977
LOCUS CR549977 747 bp mRNA linear EST 12-JUL-2004
DEFINITION DKFZp459B1749_r1 459 (synonym: pcor1) Pongo pygmaeus cDNA clone DKFZp459B1749 5', mRNA sequence.
ACCESSION CR549977
VERSION CR549977.1 GI:50243601
KEYWORDS EST.
SOURCE Pongo pygmaeus (orangutan)
ORGANISM Pongo pygmaeus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pongo.
REFERENCE 1 (bases 1 to 747)
AUTHORS Bahr,A., Lauber,J., Mewes,H.W., Weil,B., Amid,C., Osanger,A., Pobo,G., Han,M. and Wiemann,S.
TITLE Pongo pygmaeus mRNA (Bahr,A., Lauber,J., Mewes,H.W., et al.)
JOURNAL Unpublished (2004)
COMMENT Contact: MIPS
MIPS
Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
This is the 5' sequence of the clone insert Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by Qiagen (Hilden/Germany) within the cDNA sequencing consortium of the German Genome Project. This clone (DKFZp459B1749) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further information about the clone and the sequencing project is available at http://mips.gsf.de/projects/cdna/.
FEATURES Location/Qualifiers
source 1. .747
/organism="Pongo pygmaeus"
/mol_type="mRNA"
/db_xref="taxon:9600"
/clone="DKFZp459B1749"
/tissue_type="cortex"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="459 (synonym: pcor1)"
/note="Vector: pSport1_Sfi; Site_1: SfilA; Site_2: SfilB"

ORIGIN

Query Match 2.7%; Score 166.8; DB 7; Length 747;
Best Local Similarity 98.8%; Pred. No. 1e-28;
Matches 168; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy       5459 GGAAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAA 5518
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GGAAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAA 60

Qy       5519 AGCCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTA 5578
             |||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db         61 AGCCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCAAATTCCAAGCTGCTA 120

Qy       5579 AATAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AATAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 170
```

RESULT 89
BI666696
LOCUS BI666696 778 bp mRNA linear EST 12-SEP-2001
DEFINITION 603291914F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:5311250 5', mRNA sequence.
ACCESSION BI666696
VERSION BI666696.1 GI:15580929
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 778)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate: LLAM11788 row: 1 column: 03
High quality sequence stop: 761.
FEATURES Location/Qualifiers
source 1. .778
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5311250"
/tissue_type="hypothalamus"
/lab_host="DH10B"
/clone_lib="NIH_MGC_96"
/note="Organ: brain; Vector: pBluescriptR (modified pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTTVN-3', size-selected for average insert size 2.3 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIMH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."

ORIGIN

Query Match 2.7%; Score 166.8; DB 4; Length 778;
Best Local Similarity 98.8%; Pred. No. 1e-28;
Matches 168; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy       5459 GGAAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAA 5518
```

```
                ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          4 GGAAGGCGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAA 63

Qy       5519 AGCCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTA 5578
                ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db         64 AGCCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTTCTTGCGAATTCCAAGCTGTTA 123

Qy       5579 AATAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
                ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        124 AATAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 173


RESULT 90
BI668181
LOCUS       BI668181                845 bp    mRNA    linear   EST 12-SEP-2001
DEFINITION  603295727F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:5314927 5',
            mRNA sequence.
ACCESSION   BI668181
VERSION     BI668181.1  GI:15582414
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 845)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
             cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11798  row: e  column: 08
            High quality sequence stop: 790.
FEATURES             Location/Qualifiers
     source          1. .845
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5314927"
                     /tissue_type="hypothalamus"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_96"
                     /note="Organ: brain; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI
                     (gtcgag); Oligo-dT primed using primer
                     5'-TTTTTTTTTTTTTTTVN-3', size-selected for average
                     insert size 2.3 kb and normalized to ROT 5. This is a
                     primary library enriched for full-length clones and
                     constructed using the Cap-trapper method (Carninci, in
                     preparation). Library constructed by M. Brownstein
                     (NIMH/NHGRI, National Institutes of Health). Note: this is
                     a NIH_MGC Library."
ORIGIN

  Query Match             2.7%;  Score 166.8;  DB 4;  Length 845;
  Best Local Similarity   98.8%;  Pred. No. 1e-28;
  Matches  168;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       5459 GGAAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAA 5518
                ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          4 GGAAGGCGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAA 63

Qy       5519 AGCCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTA 5578
                ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db         64 AGCCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTTCTTGCGAATTCCAAGCTGTTA 123

Qy       5579 AATAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
                ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        124 AATAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 173


RESULT 91
CR544591
LOCUS       CR544591                567 bp    mRNA    linear   EST 07-JUL-2004
DEFINITION  DKFZp459E0424_r1 459 (synonym: pcor1) Pongo pygmaeus cDNA clone
            DKFZp459E0424 5', mRNA sequence..
ACCESSION   CR544591
VERSION     CR544591.1  GI:49896242
KEYWORDS    EST.
SOURCE      Pongo pygmaeus (orangutan)
  ORGANISM  Pongo pygmaeus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pongo.
REFERENCE   1  (bases 1 to 567)
  AUTHORS   Koehrer,K., Beyer,A., Mewes,H.W., Weil,B., Amid,C., Osanger,A.,
            Fobo,G., Han,M. and Wiemann,S.
  TITLE     Pongo pygmaeus mRNA (Koehrer,K., Beyer,A., Mewes,H.W., et al.)
  JOURNAL   Unpublished (2004)
COMMENT     Contact: MIPS
            MIPS
            Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
            This is the 5' sequence of the clone insert Clone from S. Wiemann,
            Molecular Genome Analysis, German Cancer Research Center (DKFZ);
            Email s.wiemann@dkfz-heidelberg.de; sequenced by BMFZ (Biomedical
            Research Center at the Heinrich-Heine-University,
            Duesseldorf/Germany) within the cDNA sequencing consortium of the
            German Genome Project. This clone (DKFZp459E0424) is available at
            the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum,
            Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email:
            clone@rzpd.de Further information about the clone and the
            sequencing project is available at
            http://mips.gsf.de/projects/cdna/.
FEATURES             Location/Qualifiers
     source          1. .567
                     /organism="Pongo pygmaeus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9600"
                     /clone="DKFZp459E0424"
                     /tissue_type="cortex"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="459 (synonym: pcor1)"
                     /note="Vector: pSport1_Sfi; Site_1: SfilA; Site_2: SfilB"
ORIGIN

  Query Match             2.7%;  Score 166.6;  DB 7;  Length 567;
  Best Local Similarity   97.7%;  Pred. No. 1.1e-28;
  Matches  169;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       5456 GCAGGAAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCT 5515
                || ||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         13 GCCGGGGGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCT 72

Qy       5516 CAAAGCCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTG 5575
                |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db         73 CAAAGCCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCAAATTCCAAGCTG 132

Qy       5576 TTAAATAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
                ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        133 TTAAATAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 185


RESULT 92
CR540761
LOCUS       CR540761                637 bp    mRNA    linear   EST 06-JUL-2004
DEFINITION  DKFZp459H1818_r1 459 (synonym: pcor1) Pongo pygmaeus cDNA clone
            DKFZp459H1818 5', mRNA sequence.
ACCESSION   CR540761
VERSION     CR540761.1  GI:49687877
```

```
KEYWORDS    EST.
SOURCE      Pongo pygmaeus (orangutan)
  ORGANISM  Pongo pygmaeus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pongo.
REFERENCE   1  (bases 1 to 637)
  AUTHORS   Koehrer,K., Beyer,A., Mewes,H.W., Weil,B., Amid,C., Osanger,A.,
            Fobo,G., Han,M. and Wiemann,S.
  TITLE     Pongo pygmaeus mRNA (Koehrer,K., Beyer,A., Mewes,H.W., et al.)
  JOURNAL   Unpublished (2004)
COMMENT     Contact: MIPS
            MIPS
            Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
            This is the 5' sequence of the clone insert Clone from S. Wiemann,
            Molecular Genome Analysis, German Cancer Research Center (DKFZ);
            Email s.wiemann@dkfz-heidelberg.de; sequenced by BMFZ (Biomedical
            Research Center at the Heinrich-Heine-University,
            Duesseldorf/Germany) within the cDNA sequencing consortium of the
            German Genome Project. This clone (DKFZp459H1818) is available at
            the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum,
            Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email:
            clone@rzpd.de Further information about the clone and the
            sequencing project is available at
            http://mips.gsf.de/projects/cdna/.
FEATURES             Location/Qualifiers
     source          1. .637
                     /organism="Pongo pygmaeus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9600"
                     /clone="DKFZp459H1818"
                     /tissue_type="cortex"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="459 (synonym: pcor1)"
                     /note="Vector: pSport1_SfiI; Site_1: SfilA; Site_2: SfilB"
ORIGIN

  Query Match             2.7%;  Score 166.6;  DB 7;  Length 637;
  Best Local Similarity   97.7%;  Pred. No. 1.1e-28;
  Matches  169;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       5456 GCAGGAAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCT 5515
              || ||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          6 GCCGGGGGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCT 65

Qy       5516 CAAAGCCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTG 5575
              ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db         66 CAAAGCCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCAAATTCCAAGCTG 125

Qy       5576 TTAAATAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        126 TTAAATAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 178


RESULT 93
BP234674
LOCUS       BP234674                461 bp    mRNA    linear   EST 15-SEP-2004
DEFINITION  BP234674 Sugano cDNA library, coronary artery endothelial cell Homo
            sapiens cDNA clone HCR03451, mRNA sequence.
ACCESSION   BP234674
VERSION     BP234674.1  GI:52107584
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 461)
  AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
  TITLE     Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
  JOURNAL   Genome Res. 14 (9), 1711-1718 (2004)
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .461
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="HCR03451"
                     /tissue_type="coronary artery"
                     /cell_type="endothelial cell"
                     /clone_lib="Sugano cDNA library, coronary artery
                     endothelial cell"
ORIGIN

  Query Match             2.7%;  Score 166.4;  DB 5;  Length 461;
  Best Local Similarity   99.4%;  Pred. No. 1.1e-28;
  Matches  167;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       5461 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 5520
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGACGCTCAGCGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 60

Qy       5521 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 5580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 120

Qy       5581 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 168


RESULT 94
BP236116
LOCUS       BP236116                461 bp    mRNA    linear   EST 15-SEP-2004
DEFINITION  BP236116 Sugano cDNA library, coronary artery endothelial cell Homo
            sapiens cDNA clone HCR06863, mRNA sequence.
ACCESSION   BP236116
VERSION     BP236116.1  GI:52109026
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 461)
  AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
  TITLE     Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
  JOURNAL   Genome Res. 14 (9), 1711-1718 (2004)
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .461
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="HCR06863"
                     /tissue_type="coronary artery"
                     /cell_type="endothelial cell"
                     /clone_lib="Sugano cDNA library, coronary artery
                     endothelial cell"
ORIGIN

  Query Match             2.7%;  Score 166.4;  DB 5;  Length 461;
  Best Local Similarity   99.4%;  Pred. No. 1.1e-28;
  Matches  167;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       5461 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 5520
```

```
                ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db          1 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGTAGAAGAGGCAAAGTACGCTCAAAG 60

Qy       5521 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 5580
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 120

Qy       5581 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
                ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 168


RESULT 95
BP236227
LOCUS       BP236227                461 bp    mRNA    linear   EST 15-SEP-2004
DEFINITION  BP236227 Sugano cDNA library, coronary artery endothelial cell Homo
            sapiens cDNA clone HCR07119, mRNA sequence.
ACCESSION   BP236227
VERSION     BP236227.1  GI:52109137
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 461)
  AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
  TITLE     Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
  JOURNAL   Genome Res. 14 (9), 1711-1718 (2004)
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .461
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="HCR07119"
                     /tissue_type="coronary artery"
                     /cell_type="endothelial cell"
                     /clone_lib="Sugano cDNA library, coronary artery
                     endothelial cell"
ORIGIN

  Query Match             2.7%;  Score 166.4;  DB 5;  Length 461;
  Best Local Similarity   99.4%;  Pred. No. 1.1e-28;
  Matches  167;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       5461 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 5520
                ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db          1 AAGACGCTCAGAGGATTCTGGCAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 60

Qy       5521 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 5580
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 120

Qy       5581 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
                ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 168


RESULT 96
CR774475
LOCUS       CR774475                511 bp    mRNA    linear   EST 23-SEP-2004
DEFINITION  DKFZp459G2052_r1 459 (synonym: pcor1) Pongo pygmaeus cDNA clone
            DKFZp459G2052 5', mRNA sequence.
ACCESSION   CR774475
VERSION     CR774475.1  GI:52617724
KEYWORDS    EST.
SOURCE      Pongo pygmaeus (orangutan)
  ORGANISM  Pongo pygmaeus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pongo.
REFERENCE   1  (bases 1 to 511)
  AUTHORS   Koehrer,K., Beyer,A., Mewes,H.W., Weil,B., Amid,C., Osanger,A.,
            Fobo,G., Han,M. and Wiemann,S.
  TITLE     Pongo pygmaeus mRNA (Koehrer,K., Beyer,A., Mewes,H.W., et al.)
  JOURNAL   Unpublished (2004)
COMMENT     Contact: MIPS
            MIPS
            Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
            This is the 5' sequence of the clone insert. Clone from S. Wiemann,
            Molecular Genome Analysis, German Cancer Research Center (DKFZ);
            Email s.wiemann@dkfz-heidelberg.de; mforschung GmbH in Berlin,
            Germany. Please contact RZPD for ordering:
            http://www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=DKFZp459G2052
            Further information about the clone and the sequencing project is
            available at http://mips.gsf.de/projects/cdna/.
FEATURES             Location/Qualifiers.
     source          1. .511
                     /organism="Pongo pygmaeus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9600"
                     /clone="DKFZp459G2052"
                     /tissue_type="cortex"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="459 (synonym: pcor1)"
                     /note="Vector: pSport1_Sfi; Site_1: SfilA; Site_2: SfilB"
ORIGIN

  Query Match             2.7%;  Score 166.4;  DB 7;  Length 511;
  Best Local Similarity   99.4%;  Pred. No. 1.2e-28;
  Matches  167;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       5461 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 5520
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 60

Qy       5521 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 5580
                ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db         61 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCAAATTCCAAGCTGTTAAA 120

Qy       5581 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
                ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 168


RESULT 97
BP204106
LOCUS       BP204106                582 bp    mRNA    linear   EST 14-SEP-2004
DEFINITION  BP204106 Sugano cDNA library, coronary artery smooth muscle cell
            Homo sapiens cDNA clone CAS01569, mRNA sequence.
ACCESSION   BP204106
VERSION     BP204106.1  GI:52056798
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 582)
  AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
  TITLE     Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
  JOURNAL   Genome Res. 14 (9), 1711-1718 (2004)
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .582
```

```
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CAS01569"
                     /tissue_type="coronary artery"
                     /cell_type="smooth muscle cell"
                     /clone_lib="Sugano cDNA library, coronary artery smooth
                     muscle cell"
ORIGIN

  Query Match             2.7%;  Score 166.4;  DB 5;  Length 582;
  Best Local Similarity   99.4%;  Pred. No. 1.2e-28;
  Matches  167;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       5461 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 5520
             |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGACGCTCAGAGGATTGTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 60

Qy       5521 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 5580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 120

Qy       5581 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 168


RESULT 98
BP310221
LOCUS       BP310221                582 bp    mRNA    linear   EST 17-SEP-2004
DEFINITION  BP310221 Sugano cDNA library, brain Homo sapiens cDNA clone
            NRR01886, mRNA sequence.
ACCESSION   BP310221
VERSION     BP310221.1  GI:52239196
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 582)
  AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
  TITLE     Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
  JOURNAL   Genome Res. 14 (9), 1711-1718 (2004)
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .582
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="NRR01886"
                     /tissue_type="brain"
                     /clone_lib="Sugano cDNA library, brain"
ORIGIN

  Query Match             2.7%;  Score 166.4;  DB 5;  Length 582;
  Best Local Similarity   99.4%;  Pred. No. 1.2e-28;
  Matches  167;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       5461 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 5520
             |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db          1 AAGACGCTCAGAGGATTCTGACAATATCTCTACCGGAGAAGAGGCAAAGTACGCTCAAAG 60

Qy       5521 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 5580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 120

Qy       5581 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 168


RESULT 99
BP335960
LOCUS       BP335960                634 bp    mRNA    linear   EST 17-SEP-2004
DEFINITION  BP335960 Sugano cDNA library, coronary artery smooth muscle cell
            Homo sapiens cDNA clone SMR04328, mRNA sequence.
ACCESSION   BP335960
VERSION     BP335960.1  GI:52265518
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 634)
  AUTHORS   Suzuki,Y., Yamashita,R., Shirota,M., Sakakibara,Y., Chiba,J.,
            Mizushima-Sugano,J., Nakai,K. and Sugano,S.
  TITLE     Sequence comparison of human and mouse genes reveals a homologous
            block structure in the promoter regions
  JOURNAL   Genome Res. 14 (9), 1711-1718 (2004)
COMMENT     Contact: Yutaka Suzuki
            Department of Virology
            Institute of Medical Science, University of Tokyo
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Email: ysuzuki@ims.u-tokyo.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .634
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="SMR04328"
                     /tissue_type="coronary artery"
                     /cell_type="smooth muscle cell"
                     /clone_lib="Sugano cDNA library, coronary artery smooth
                     muscle cell"
ORIGIN

  Query Match             2.7%;  Score 166.4;  DB 5;  Length 634;
  Best Local Similarity   99.4%;  Pred. No. 1.2e-28;
  Matches  167;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       5461 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 5520
             |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db          1 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGGAGAGGCAAAGTACGCTCAAAG 60

Qy       5521 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 5580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 120

Qy       5581 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 168


RESULT 100
AQ316106/c
LOCUS       AQ316106                521 bp    DNA     linear   GSS 04-MAY-1999
DEFINITION  RPCI11-105A21.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-105A21,
            genomic survey sequence.
ACCESSION   AQ316106
VERSION     AQ316106.1  GI:4047569
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 521)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
            Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
  TITLE     Use of human BAC End Sequences for Sequence-Ready Map Building
  JOURNAL   Unpublished (1998)
```

```
COMMENT     Other_GSSs: RPCI11-105A21.TV
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: SP6
            Class: BAC ends.
FEATURES              Location/Qualifiers
     source           1. .521
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                      /db_xref="GDB:7539956"
                      /db_xref="taxon:9606"
                      /clone="RPCI-11-105A21"
                      /sex="Male"
                      /cell_type="Lymphocytes"
                      /clone_lib="RPCI-11"
                      /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                      RPCI11 Human Male BAC Library"
ORIGIN

  Query Match             2.7%;  Score 166.2;  DB 8;  Length 521;
  Best Local Similarity   69.5%;  Pred. No. 1.3e-28;
  Matches  331;  Conservative    0;  Mismatches  124;  Indels   21;  Gaps    7;

Qy        312 GTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTC 371
              ||  ||||| ||| ||||||| | ||| |||   ||||||||||||   |  |||| | ||
Db        457 GTGNGGGTGGAAAGGTGTGATGCCTTTCCTC-AGCATCATAAGGGGAACTGCCAACAATC 399

Qy        372 CTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTT 431
              |||||  ||||  ||||| || |||  ||||   ||| | |||||| ||||||||  ||||
Db        398 CTATAACAAAAAGCAGGTAAACAAGGACAAATA-TAATAGATTTATATAATCAAAATTTT 340

Qy        432 ACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTT 491
                  ||| ||||||||||||||||||||||  |||||  || ||| |   |||
Db        339 TTTGTGACACAGAGCCTTCAGAAATGAAGACCCAAAGGTCCAGGAAAACCTGGC-AATTT 281

Qy        492 TTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGAT 551
              || || | ||||| ||| ||||||| | |||        |    |||  |||||   |||
Db        280 TTATGTTTAGGTTTGATAAAGAATGAACAGCCATGTCGAAATGTGATCGGACAA---GAT 224

Qy        552 ATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTT 611
              ||||||||||     ||       |||| ||   ||||||||||||||| || ||||||||
Db        223 ATGATCTAGT-----AGACTGAGGGGGAAACTCCAGCAAGGCCTGTCTGTTAAGATTCTT 169

Qy        612 CTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTT 671
              ||||        ||  | ||| | |||||| |||||||    ||||||| || || | |
Db        168 CTTG--------GTCATTTTGTGTGGCATTCCTTCCTCCCACATATGGGGTAGCACCCCT 117

Qy        672 CTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATGGCTTG 731
              ||  |||||||||||||||||| |||||   | |    ||||| ||| ||||||||||||
Db        116 CTAGAATGAGGGTCTTCAAGGGTGAAGGAGAAGAACCACCTTTCTAGGTTTTATGGCTTG 57

Qy        732 CTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCTG 787
              ||| | || | | |||||||  |||||||||||||||| |  | |   ||||||
Db         56 CTTTGAGGGACAAGAGTTCT--TTTCTATGACCCATCTCGCAGCAATAGAATTCCG 3


Search completed: April 22, 2005, 04:39:22
Job time : 12205.1 secs
```

This Page Blank (uspto)

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 20, 2005, 22:49:10 ; Search time 17029.6 Seconds
(without alignments)
17319.707 Million cell updates/sec

Title: US-09-939-209A-3_COPY_4970_11056
Perfect score: 6087
Sequence: 1 ccagtaggaaacagttcttc.....catcatcagaaaggcactat 6087

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	6087	100.0	20300	6	AX451337	AX451337 Sequence
2	6087	100.0	165329	9	AL583850	AL583850 Human DNA
c 3	6068.6	99.7	191699	2	AC031977	AC031977 Homo sapi
c 4	405.4	6.7	275631	2	AC125563	AC125563 Rattus no
5	402.6	6.6	225600	10	AC115766	AC115766 Mus muscu
6	326.2	5.4	172931	9	AC006581	AC006581 Homo sapi
c 7	317	5.2	94158	9	AC119619	AC119619 Homo sapi
c 8	317	5.2	175440	2	AC137626	AC137626 Homo sapi
c 9	311.2	5.1	154125	2	AC073227	AC073227 Homo sapi
10	311.2	5.1	196832	9	AC018694	AC018694 Homo sapi
11	307.2	5.0	176773	2	AC009867	AC009867 Homo sapi
12	305.6	5.0	89000	9	AP003059	AP003059 Homo sapi
c 13	299.4	4.9	169972	9	AL513284	AL513284 Human DNA
14	297.6	4.9	121720	9	AL591491	AL591491 Human DNA
c 15	293.8	4.8	134937	9	AC010244	AC010244 Homo sapi
16	293.8	4.8	170219	9	AC091991	AC091991 Homo sapi
c 17	293.8	4.8	182049	2	AC021239	AC021239 Homo sapi
c 18	293.8	4.8	188207	2	AC021800	AC021800 Homo sapi
c 19	292	4.8	193126	9	AL139275	AL139275 Human DNA
20	290.4	4.8	158983	2	AC073070	AC073070 Homo sapi
c 21	289.6	4.8	150145	9	AC069216	AC069216 Homo sapi
22	289.6	4.8	178168	2	AC024412	AC024412 Homo sapi
c 23	289	4.7	126394	9	AC109998	AC109998 Homo sapi
24	289	4.7	194860	9	AC091903	AC091903 Homo sapi
25	288.2	4.7	65608	9	AL450243	AL450243 Human DNA
26	288.2	4.7	181618	2	AL442129	AL442129 Homo sapi
27	288.2	4.7	190092	2	AC120502	AC120502 Lemur cat
c 28	284.4	4.7	147055	9	HS372K1	AL023580 Human DNA
c 29	282.2	4.6	178905	9	AC024153	AC024153 Homo sapi
30	280.2	4.6	98232	9	AC012271	AC012271 Homo sapi
31	280.2	4.6	152005	9	AC100834	AC100834 Homo sapi
c 32	280	4.6	27724	9	BX284678	BX284678 Human DNA
33	277.6	4.6	1238	6	CQ434203	CQ434203 Sequence
34	276.2	4.5	104011	9	AL590392	AL590392 Human DNA
35	275.8	4.5	137996	9	AP001982	AP001982 Homo sapi
36	275.8	4.5	160800	2	AP002011	AP002011 Homo sapi
c 37	275.8	4.5	162584	2	AP002763	AP002763 Homo sapi
c 38	275.8	4.5	183674	2	AC016996	AC016996 Homo sapi
c 39	275.4	4.5	142533	9	BS000038	BS000038 Pan trogl
c 40	274	4.5	228353	2	AL445283	AL445283 Homo sapi
c 41	273.8	4.5	172461	2	AC083796	AC083796 Homo sapi
c 42	273.8	4.5	173010	9	CNS01DSU	AL122013 Human chr
c 43	271.6	4.5	154336	9	AC117409	AC117409 Homo sapi
44	271.6	4.5	154341	9	HS1164K7	AL592209 Homo sapi
c 45	268.8	4.4	204662	9	CNS01RHQ	AL162191 Human chr

ALIGNMENTS

RESULT 1
AX451337
LOCUS AX451337 20300 bp DNA linear PAT 03-JUL-2002
DEFINITION Sequence 3 from Patent WO0216653.
ACCESSION AX451337
VERSION AX451337.1 GI:21698388
KEYWORDS .
SOURCE synthetic construct
ORGANISM synthetic construct
other sequences; artificial sequences.
REFERENCE 1
AUTHORS Levitt,P.R., Mirnics,K., Kodavali,V.C. and Nimgaonkar,V.L.
TITLE Methods and systems for facilitating the diagnosis and treatment of schizophrenia
JOURNAL Patent: WO 0216653-A 3 28-FEB-2002;
University of Pittsburgh (US)
FEATURES Location/Qualifiers
source 1. .20300
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="A genomic sequence containing RGS4 nucleic acid sequence and sequ ences upstream and downstream to the RGS4 nucleic acid sequence"
ORIGIN

Query Match 100.0%; Score 6087; DB 6; Length 20300;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 6087; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 CCAGTAGGAAACAGTTCTTCTAAAGCCTATTGTCAACAATACCTTCCAGATGTTAGCATT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4970 CCAGTAGGAAACAGTTCTTCTAAAGCCTATTGTCAACAATACCTTCCAGATGTTAGCATT 5029

Qy         61 TTACAATTTAAGGAACTTAAAATAGCCTTCAAACTTTTTGCCAGTTTCTCTGATATCCAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5030 TTACAATTTAAGGAACTTAAAATAGCCTTCAAACTTTTTGCCAGTTTCTCTGATATCCAA 5089

Qy        121 TCTATTCTTTTACTCTGCCTCCCAAGCTTTCTTTCTAGAATGCTAACCTGATCGGCTTAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5090 TCTATTCTTTTACTCTGCCTCCCAAGCTTTCTTTCTAGAATGCTAACCTGATCGGCTTAA 5149
```

```
Qy       181 GTACTTGAACTACCTCTTCTCCTCCATTAACTACAGAGTAAATTCTGGTCTTCAGAGTAA 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5150 GTACTTGAACTACCTCTTCTCCTCCATTAACTACAGAGTAAATTCTGGTCTTCAGAGTAA 5209

Qy       241 CAAGAAACACCCTTTAGTTCTCAGCATATTCGTGCACCTTCATTTATCTCTCCTTCTCTC 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5210 CAAGAAACACCCTTTAGTTCTCAGCATATTCGTGCACCTTCATTTATCTCTCCTTCTCTC 5269

Qy       301 TCAAAGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGC 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5270 TCAAAGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGC 5329

Qy       361 AACCAAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTA 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5330 AACCAAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTA 5389

Qy       421 ATCAAAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAA 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5390 ATCAAAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAA 5449

Qy       481 CTGTCTGTTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTA 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5450 CTGTCTGTTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTA 5509

Qy       541 GACAAAAGGATATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTA 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5510 GACAAAAGGATATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTA 5569

Qy       601 TTCAGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGG 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5570 TTCAGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGG 5629

Qy       661 GCAGGACTCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGAT 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5630 GCAGGACTCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGAT 5689

Qy       721 TTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGG 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5690 TTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGG 5749

Qy       781 AATTCTGGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGG 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5750 AATTCTGGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGG 5809

Qy       841 AGATGGTCAGAAAGAGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAGT 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5810 AGATGGTCAGAAAGAGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAGT 5869

Qy       901 ACTTCTTAGCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAACACTGC 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5870 ACTTCTTAGCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAACACTGC 5929

Qy       961 AATCATGCTAAACTCCTCTATGACCTTCAAACATTCCACTTGCTTTTATTCTTTATGGTT 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5930 AATCATGCTAAACTCCTCTATGACCTTCAAACATTCCACTTGCTTTTATTCTTTATGGTT 5989

Qy      1021 GTGATGGCATAGAGGTCAATAGCAAAGACCCTGGAGTCCCACTGTCTGAGCTGGCATAAC 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5990 GTGATGGCATAGAGGTCAATAGCAAAGACCCTGGAGTCCCACTGTCTGAGCTGGCATAAC 6049

Qy      1081 ATTACTACCACTTAATCAATGTGTAAGCTCAGGTAAGTACTTAAGTCCTCTATGCTTCAT 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6050 ATTACTACCACTTAATCAATGTGTAAGCTCAGGTAAGTACTTAAGTCCTCTATGCTTCAT 6109

Qy      1141 CTGTAAAATGAGAATCATTGAAGAACATTCTCTCAGGATGGATCATGAGGAATAAGTGAA 1200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6110 CTGTAAAATGAGAATCATTGAAGAACATTCTCTCAGGATGGATCATGAGGAATAAGTGAA 6169

Qy      1201 TTAACTGGCATATAGTGCTTAAACCAGTGCCTTGCTCAGTTAGTGACAGATAAAATCATC 1260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6170 TTAACTGGCATATAGTGCTTAAACCAGTGCCTTGCTCAGTTAGTGACAGATAAAATCATC 6229

Qy      1261 TGTTATTACTGTGCCCACTATTGTGATGCTCTTCTCTTCTTTGTACAACGACTACATCTC 1320
```

```
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6230 TGTTATTACTGTGCCCACTATTGTGATGCTCTTCTCTTCTTTGTACAACGACTACATCTC 6289

Qy      1321 TATTTATCATTTTAGGGTCTCCTTGTGAAAAACCACTCCAGATTCAAAAGATTGAGTTTA 1380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6290 TATTTATCATTTTAGGGTCTCCTTGTGAAAAACCACTCCAGATTCAAAAGATTGAGTTTA 6349

Qy      1381 ATCTCTATCCTCTGTGCTTTCCTGGAGTTTTGTAAAGTAAATCTTCACTTGACATCATGG 1440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6350 ATCTCTATCCTCTGTGCTTTCCTGGAGTTTTGTAAAGTAAATCTTCACTTGACATCATGG 6409

Qy      1441 ATAGGTTCTTGGAAACTACAACTTCAAGTGAAAGGACATAACTAAACCAATTTTTTTCTC 1500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6410 ATAGGTTCTTGGAAACTACAACTTCAAGTGAAAGGACATAACTAAACCAATTTTTTTCTC 6469

Qy      1501 ATCAACGTTATAATGAAATGGCATTGATGAAATGATGGCATTCAAGGACCTGCTGTACCT 1560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6470 ATCAACGTTATAATGAAATGGCATTGATGAAATGATGGCATTCAAGGACCTGCTGTACCT 6529

Qy      1561 TGTTTCACTTAAAGTCACTGTTTCCAATAATCTATTGATGACATTGAGGACTTACTATAT 1620
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6530 TGTTTCACTTAAAGTCACTGTTTCCAATAATCTATTGATGACATTGAGGACTTACTATAT 6589

Qy      1621 AATAATAAATATATATATAATCGACGAAACAGGAATCAAACTGCTAACTCTGCTAACTGG 1680
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6590 AATAATAAATATATATATAATCGACGAAACAGGAATCAAACTGCTAACTCTGCTAACTGG 6649

Qy      1681 TCTCCCTGCTTCCACACTCTGCCCACTCATCTCAGTCTTTCTTTCACAAGAGTCAGAATG 1740
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6650 TCTCCCTGCTTCCACACTCTGCCCACTCATCTCAGTCTTTCTTTCACAAGAGTCAGAATG 6709

Qy      1741 ATCAGATGAGACCCCTCCTCTGCTTCTGTTTCTTCCATGGATTTCCACTGCACTCTGATA 1800
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6710 ATCAGATGAGACCCCTCCTCTGCTTCTGTTTCTTCCATGGATTTCCACTGCACTCTGATA 6769

Qy      1801 AAGTCCAGCCTCTTGACCACAGCCTACAAATCCTTGCACGATCTATCGTTTACTTTTCCA 1860
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6770 AAGTCCAGCCTCTTGACCACAGCCTACAAATCCTTGCACGATCTATCGTTTACTTTTCCA 6829

Qy      1861 TCTCCTTTTATGCTACTTTCATCTTGTTCTCAATTCTCTAGCTATGCTGGCCCCTTCTTG 1920
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6830 TCTCCTTTTATGCTACTTTCATCTTGTTCTCAATTCTCTAGCTATGCTGGCCCCTTCTTG 6889

Qy      1921 TTCTTTCCCATTTTTTTTTAATTTTTAAAATTTGTATATATTTATGGGTTATAAGTGAAA 1980
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6890 TTCTTTCCCATTTTTTTTTAATTTTTAAAATTTGTATATATTTATGGGTTATAAGTGAAA 6949

Qy      1981 TCTTTTTAGATGCATAGGTTGTATAGTGATAAAATCAGGGCTTTTAGGGTATTCATCACC 2040
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6950 TCTTTTTAGATGCATAGGTTGTATAGTGATAAAATCAGGGCTTTTAGGGTATTCATCACC 7009

Qy      2041 TGAATGATGTACATTGTACCCCTTAAGTAATTTCTCACCATCCGCTGACTTCTTGCCCCC 2100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7010 TGAATGATGTACATTGTACCCCTTAAGTAATTTCTCACCATCCGCTGACTTCTTGCCCCC 7069

Qy      2101 TGGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGC 2160
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7070 TGGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGC 7129

Qy      2161 TGACTTCTTGCCCCCTGGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAA 2220
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7130 TGACTTCTTGCCCCCTGGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAA 7189

Qy      2221 TTTCTCACCATCCGCTGACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCAT 2280
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7190 TTTCTCACCATCCGCTGACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCAT 7249

Qy      2281 TCCACACTCTACATCTATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGC 2340
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7250 TCCACACTCTACATCTATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGC 7309

Qy      2341 AATATTTGTCTTTCTGTGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCT 2400
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      7310 AATATTTGTCTTTCTGTGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCT 7369

Qy      2401 AGGCTGCTGCAAAAGGCATGATTTCATTCTTTTTTATGGCTATGTTCTTTCCCAATTTAG 2460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7370 AGGCTGCTGCAAAAGGCATGATTTCATTCTTTTTTATGGCTATGTTCTTTCCCAATTTAG 7429

Qy      2461 ATAAAGAACACTCGCACTTGCTCTTACTTCTATTTGGAATACTAATTCCTAGGCTTCTTG 2520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7430 ATAAAGAACACTCGCACTTGCTCTTACTTCTATTTGGAATACTAATTCCTAGGCTTCTTG 7489

Qy      2521 CATTGCTTTCTCCTTCTCACCCATCAAATCTCATTTTAGATACCACCTCTTCAAAGAGGG 2580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7490 CATTGCTTTCTCCTTCTCACCCATCAAATCTCATTTTAGATACCACCTCTTCAAAGAGGG 7549

Qy      2581 CTTTCCTGACCACCTTGGCTGAATTAGCCCTTCACCATCTGATTACTCTCTAGCACATCA 2640
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7550 CTTTCCTGACCACCTTGGCTGAATTAGCCCTTCACCATCTGATTACTCTCTAGCACATCA 7609

Qy      2641 CCTGCCCATTTTATTCATGGTACAGGTCAAAATCTGGAATCACCTGATTTGTTTATTTTC 2700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7610 CCTGCCCATTTTATTCATGGTACAGGTCAAAATCTGGAATCACCTGATTTGTTTATTTTC 7669

Qy      2701 TGACTCCTTCTACTGAGATGAAAACTCTACTAGAGCGGAGATTTTATCTGCTTGTATCAG 2760
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7670 TGACTCCTTCTACTGAGATGAAAACTCTACTAGAGCGGAGATTTTATCTGCTTGTATCAG 7729

Qy      2761 GTACTGCTTCAAACAGCACCTGATACAGAGTAGGTGGTCAAAAGATATTTCTTAAACAAA 2820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7730 GTACTGCTTCAAACAGCACCTGATACAGAGTAGGTGGTCAAAAGATATTTCTTAAACAAA 7789

Qy      2821 TGAACAAATAAAAAGTAGATCTTTTGAGAGTAAAGCTCTTCCACACTACCAGAGTCATTC 2880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7790 TGAACAAATAAAAAGTAGATCTTTTGAGAGTAAAGCTCTTCCACACTACCAGAGTCATTC 7849

Qy      2881 AGGAATGACAAATCATAGAATAACAGAATTTGATGCTTTGTGCATATCAGAGAAAGAAGG 2940
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7850 AGGAATGACAAATCATAGAATAACAGAATTTGATGCTTTGTGCATATCAGAGAAAGAAGG 7909

Qy      2941 TGGAAGGTTGTCAAGGTATCATGATGTACCAGTCCTCGCCTCCTCAAACACAATCTGCAA 3000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7910 TGGAAGGTTGTCAAGGTATCATGATGTACCAGTCCTCGCCTCCTCAAACACAATCTGCAA 7969

Qy      3001 GTCCCACAGTGAAAAAGTAAGTTAACTCATGTGAAGCGTTTTACAAACACTTTTTTAAAA 3060
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      7970 GTCCCACAGTGAAAAAGTAAGTTAACTCATGTGAAGCGTTTTACAAACACTTTTTTAAAA 8029

Qy      3061 GTCTTAAAACTCCTAAGAAAGCAAGATTTAATAGTCAAAGAAGTGAGTAAACATGAAATG 3120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8030 GTCTTAAAACTCCTAAGAAAGCAAGATTTAATAGTCAAAGAAGTGAGTAAACATGAAATG 8089

Qy      3121 CCTGAACAGAGTAATGAGCTAAGCACAAAGTTAGAGACATGTTAGTTAATATGTCTTGAA 3180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8090 CCTGAACAGAGTAATGAGCTAAGCACAAAGTTAGAGACATGTTAGTTAATATGTCTTGAA 8149

Qy      3181 AGCAGCAGCTCCTGCTTTCAAGGAGCAAGAACAAATTGGGCAAGTGAACACTCCTTGAAT 3240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8150 AGCAGCAGCTCCTGCTTTCAAGGAGCAAGAACAAATTGGGCAAGTGAACACTCCTTGAAT 8209

Qy      3241 AAAATGTGTAAAATTAATTTTGGGTTATGTTCTATACTGTGTATAATAGAATGATAAAAA 3300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8210 AAAATGTGTAAAATTAATTTTGGGTTATGTTCTATACTGTGTATAATAGAATGATAAAAA 8269

Qy      3301 TTATTTGACTAGCACTTTGTAGTTTAGAAATATCTCTATTTACACAGTTTACCTTATTTG 3360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8270 TTATTTGACTAGCACTTTGTAGTTTAGAAATATCTCTATTTACACAGTTTACCTTATTTG 8329

Qy      3361 ATAAGACTGTTGAGTGATGGGATAGCATGGTGGACAATCCACATAACTGAGTATCGAGAC 3420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8330 ATAAGACTGTTGAGTGATGGGATAGCATGGTGGACAATCCACATAACTGAGTATCGAGAC 8389

Qy      3421 ACCTGTATCTGGACCCAGCTCTGTTAGTAAGAAGCTGTAACCTCAGCAAGTCACTTTCTC 3480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8390 ACCTGTATCTGGACCCAGCTCTGTTAGTAAGAAGCTGTAACCTCAGCAAGTCACTTTCTC 8449

Qy      3481 TTTCTGGGTCTCTATTTCCTTTTTGGTGAAATGAGAGTGTTAGGCTAGATTGCCTTTGAA 3540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8450 TTTCTGGGTCTCTATTTCCTTTTTGGTGAAATGAGAGTGTTAGGCTAGATTGCCTTTGAA 8509

Qy      3541 GTCCCATTTTGTCTTTAAAGTCCCATCTATTGCAGTGATTTATATTTAACTCATGACAAA 3600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8510 GTCCCATTTTGTCTTTAAAGTCCCATCTATTGCAGTGATTTATATTTAACTCATGACAAA 8569

Qy      3601 TCAGGCTTCTCTTATTCTAAGTGCAAGACATAAAACTTTTATTGTGGAATTTCAGGCATC 3660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8570 TCAGGCTTCTCTTATTCTAAGTGCAAGACATAAAACTTTTATTGTGGAATTTCAGGCATC 8629

Qy      3661 AGTAAATCTTTTTGGGTACTCACTTATGTTCCTGAAATCAATCTATTTGAGTGATCACTC 3720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8630 AGTAAATCTTTTTGGGTACTCACTTATGTTCCTGAAATCAATCTATTTGAGTGATCACTC 8689

Qy      3721 TTTTAGGTGCCCAGGTAAACAAAGAAGGCCATGGTCTTTCTTTGAGTGACCTTCTTTCCC 3780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8690 TTTTAGGTGCCCAGGTAAACAAAGAAGGCCATGGTCTTTCTTTGAGTGACCTTCTTTCCC 8749

Qy      3781 TTTTAATTAGTCTGACCTCTTTAATGTCAGTTCTGACTGATTCATTTCCCTGGTCCATCT 3840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8750 TTTTAATTAGTCTGACCTCTTTAATGTCAGTTCTGACTGATTCATTTCCCTGGTCCATCT 8809

Qy      3841 TCCTTGGTCTGAGGGCCTTCCTAGTTTCATATTGCACTTCAGTTCCTTCCACACCACCAT 3900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8810 TCCTTGGTCTGAGGGCCTTCCTAGTTTCATATTGCACTTCAGTTCCTTCCACACCACCAT 8869

Qy      3901 CAAGGATGGCTGTCAACATTCATTTGTTCTATGTTATAATTCAAGGAAAAGTTGCCCAGT 3960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8870 CAAGGATGGCTGTCAACATTCATTTGTTCTATGTTATAATTCAAGGAAAAGTTGCCCAGT 8929

Qy      3961 AGCTAATCCAATAAATGCCCTCTTATGGGCGGCTAGAGACTTTTTCCTATAATTTAAATG 4020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8930 AGCTAATCCAATAAATGCCCTCTTATGGGCGGCTAGAGACTTTTTCCTATAATTTAAATG 8989

Qy      4021 CATCTTCTGTAGATTATGGTCCCTCCACCACTTTACATTTGTCTGCTGTCTCCTTGCTCT 4080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      8990 CATCTTCTGTAGATTATGGTCCCTCCACCACTTTACATTTGTCTGCTGTCTCCTTGCTCT 9049

Qy      4081 GCTAGTCATGGAACGTGTTGGTAGTGGGGGCAGTGTGGGATGTTCAAGGGCACGTATTGG 4140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      9050 GCTAGTCATGGAACGTGTTGGTAGTGGGGGCAGTGTGGGATGTTCAAGGGCACGTATTGG 9109

Qy      4141 GTAGGGCCACATATGGGCATTGCTTTGTGCCATTCTTTCTATATTTTTGGTATTTTGCAT 4200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      9110 GTAGGGCCACATATGGGCATTGCTTTGTGCCATTCTTTCTATATTTTTGGTATTTTGCAT 9169

Qy      4201 CTCACTGGAACCCAACTATTTTTCATCTCTTCCACCTAAACTATTTGATGCCTCTGTTTC 4260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      9170 CTCACTGGAACCCAACTATTTTTCATCTCTTCCACCTAAACTATTTGATGCCTCTGTTTC 9229

Qy      4261 TTATATATAAAGTATAGCTCACTGTAGCCTATGATCAGGAACCTATCTGCTTTCTAAATG 4320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      9230 TTATATATAAAGTATAGCTCACTGTAGCCTATGATCAGGAACCTATCTGCTTTCTAAATG 9289

Qy      4321 AAAGCTGTTTTGGTCAGATCTAGCAATTAATTCCCTTCTTCCACTTATAGCTTTCCTCTG 4380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      9290 AAAGCTGTTTTGGTCAGATCTAGCAATTAATTCCCTTCTTCCACTTATAGCTTTCCTCTG 9349

Qy      4381 TAACTCTGGTGTAGGTATTTGGTTTATGGCTATAAGATGTGAAACACCTGAATGATTCTG 4440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      9350 TAACTCTGGTGTAGGTATTTGGTTTATGGCTATAAGATGTGAAACACCTGAATGATTCTG 9409

Qy      4441 TCCATGCAGGCATTTCAGTTCATGATATTGTATGTAAAAGATACTGATTGTCTAGGTGTT 4500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      9410 TCCATGCAGGCATTTCAGTTCATGATATTGTATGTAAAAGATACTGATTGTCTAGGTGTT 9469

Qy      4501 CAGAAACACCTATAGGGCTTAATATTCTTACAATCAGTTTGAAGGCTGGTGATACGCAAA 4560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      9470 CAGAAACACCTATAGGGCTTAATATTCTTACAATCAGTTTGAAGGCTGGTGATACGCAAA 9529
```

```
Qy       4561 GCAAACTACATATTTTTCTGCCTGCTCTCTCTCTTTCTCTCTACATCTCTCTTTCTTTAT 4620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9530 GCAAACTACATATTTTTCTGCCTGCTCTCTCTCTTTCTCTCTACATCTCTCTTTCTTTAT 9589

Qy       4621 CTTTTGAAATATCAGTTTGGAGACTTAGAATTACATAAGACATAAACCCATTTGATATAA 4680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9590 CTTTTGAAATATCAGTTTGGAGACTTAGAATTACATAAGACATAAACCCATTTGATATAA 9649

Qy       4681 GAATTGCTGTGTATATTTGCTCATCTACTCCCTCCTTTGGTCCTCGAGCTGCCGGTTTAG 4740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9650 GAATTGCTGTGTATATTTGCTCATCTACTCCCTCCTTTGGTCCTCGAGCTGCCGGTTTAG 9709

Qy       4741 ACTTTTTACAGGACGCAGGCATGTGAAGGAGAAACTGTCAGTGCTAGGCTGAATTCTGTT 4800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9710 ACTTTTTACAGGACGCAGGCATGTGAAGGAGAAACTGTCAGTGCTAGGCTGAATTCTGTT 9769

Qy       4801 GTTACCAAGATTTCTAGAAAAGTATTCCTCAGTCAGGTTGATTACAGATATAGCAAATCT 4860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9770 GTTACCAAGATTTCTAGAAAAGTATTCCTCAGTCAGGTTGATTACAGATATAGCAAATCT 9829

Qy       4861 ATTTTTCCTAGGGTAGTTTCTGTATGCTGCCGGGCTTATAACTGTCTGTCATCCAGCTAT 4920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9830 ATTTTTCCTAGGGTAGTTTCTGTATGCTGCCGGGCTTATAACTGTCTGTCATCCAGCTAT 9889

Qy       4921 TTCTCTCCACCTTCTTGTTTGCATAACAACCAAGGCAACTTCCGCAAATCACTGCGTGGA 4980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9890 TTCTCTCCACCTTCTTGTTTGCATAACAACCAAGGCAACTTCCGCAAATCACTGCGTGGA 9949

Qy       4981 GACGATGATCCTGCCAGCTCCCTTTTGGAAATCGTGAGGATCAGATCTTGGACCATGTAT 5040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9950 GACGATGATCCTGCCAGCTCCCTTTTGGAAATCGTGAGGATCAGATCTTGGACCATGTAT 10009

Qy       5041 AATATGATGCTTCTAATCCAAAAGAGGAAAGGCATTGGGAGTCAGCTCCTAAGTAAGCTC 5100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10010 AATATGATGCTTCTAATCCAAAAGAGGAAAGGCATTGGGAGTCAGCTCCTAAGTAAGCTC 10069

Qy       5101 CAGAATTCCTGCTGGTACTTTTCCTTCCAGGAAGCAACTTCCTTGATATTTTTTTTTTAC 5160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10070 CAGAATTCCTGCTGGTACTTTTCCTTCCAGGAAGCAACTTCCTTGATATTTTTTTTTTAC 10129

Qy       5161 AGGCATATGAATAAAAACTATATTTTGCAGCATTGTACACTTTTTTTCCTTTTCTAGAAA 5220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10130 AGGCATATGAATAAAAACTATATTTTGCAGCATTGTACACTTTTTTTCCTTTTCTAGAAA 10189

Qy       5221 TTCTAAACCTCTGACATTGGTGGAGACATTGAGTACATTTTTTCCCATATCCCTACTTTT 5280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10190 TTCTAAACCTCTGACATTGGTGGAGACATTGAGTACATTTTTTCCCATATCCCTACTTTT 10249

Qy       5281 CAGAAGGATTTTCTCTGCTCGTTCACTTAACATTGCTGATGCGTCAGTCTTTTCTTCCTC 5340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10250 CAGAAGGATTTTCTCTGCTCGTTCACTTAACATTGCTGATGCGTCAGTCTTTTCTTCCTC 10309

Qy       5341 ATCTCTTTCAGGGGCTGGAGAGGCAGAGGGAGACAGAGGAGCTGGTACTGCAGAGCGGTC 5400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10310 ATCTCTTTCAGGGGCTGGAGAGGCAGAGGGAGACAGAGGAGCTGGTACTGCAGAGCGGTC 10369

Qy       5401 GTCTGATTGGCTGGACGGTCGTAGCTGGGCTATAAAAGAGACCCCTACAGGCTTAGCAGG 5460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10370 GTCTGATTGGCTGGACGGTCGTAGCTGGGCTATAAAAGAGACCCCTACAGGCTTAGCAGG 10429

Qy       5461 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 5520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10430 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 10489

Qy       5521 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 5580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10490 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 10549

Qy       5581 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAGGTAAGATTGCTT 5640
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10550 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAGGTAAGATTGCTT 10609

Qy       5641 TCAGCCATTAACCATATTAAACTTTTGGCTAGACTTTCTCAGTTATTTACATGTTGTACT 5700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10610 TCAGCCATTAACCATATTAAACTTTTGGCTAGACTTTCTCAGTTATTTACATGTTGTACT 10669

Qy       5701 TACTAACCTAGTTCTGTGCAATTAGAAACAGTGTGGTCAGGAGAGCACGACTTTCTAACT 5760
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10670 TACTAACCTAGTTCTGTGCAATTAGAAACAGTGTGGTCAGGAGAGCACGACTTTCTAACT 10729

Qy       5761 TTCCTCCAAGACTAGCTAGATATTGTGACTTAAGACATGTGCTCCCCAAATTTCAGCCCT 5820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10730 TTCCTCCAAGACTAGCTAGATATTGTGACTTAAGACATGTGCTCCCCAAATTTCAGCCCT 10789

Qy       5821 TATGTGTTGTTTTGTGTGACCTCAGTTTTGAGAACTGTTCTATTCTTTAAGCCAGGTCTA 5880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10790 TATGTGTTGTTTTGTGTGACCTCAGTTTTGAGAACTGTTCTATTCTTTAAGCCAGGTCTA 10849

Qy       5881 AGAAAGCTAGTTTTAATTAAGAAGCGAGATGAGGTTTGAGGCTATGTACAGTGATCTGTA 5940
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10850 AGAAAGCTAGTTTTAATTAAGAAGCGAGATGAGGTTTGAGGCTATGTACAGTGATCTGTA 10909

Qy       5941 ATATCTCCATCTGTGATTACTACTGCTATTTGAGCATCCCTGGAGTACATAGAAGCCTGG 6000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10910 ATATCTCCATCTGTGATTACTACTGCTATTTGAGCATCCCTGGAGTACATAGAAGCCTGG 10969

Qy       6001 CTCTGGGCTTTCTGATTGTATGCTACAACTTGTTTCAGGAAAGGTACCCCAGAATGAGGT 6060
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10970 CTCTGGGCTTTCTGATTGTATGCTACAACTTGTTTCAGGAAAGGTACCCCAGAATGAGGT 11029

Qy       6061 TTGGCTCCATCATCAGAAAGGCACTAT 6087
             |||||||||||||||||||||||||||||
Db      11030 TTGGCTCCATCATCAGAAAGGCACTAT 11056
```

RESULT 2
AL583850

```
LOCUS       AL583850              165329 bp    DNA     linear   PRI 15-NOV-2001
DEFINITION  Human DNA sequence from clone RP11-430G6 on chromosome 1, complete
            sequence.
ACCESSION   AL583850
VERSION     AL583850.5  GI:16973044
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 165329)
  AUTHORS   Tracey,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Nov 16, 2001 this sequence version replaced gi:15020514.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
```

RP11-430G6 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
IMPORTANT: This sequence is not the entire insert of clone RP11-430G6 It may be shorter because we sequence overlapping sections only once, except for a short overlap.
The true right end of clone RP11-430G6 is at 165329 in this sequence. The true right end of clone RP11-331H2 is at 2000 in this sequence.

FEATURES Location/Qualifiers
source 1. .165329
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-430G6"
/clone_lib="RPCI-11.2"

ORIGIN

Query Match 100.0%; Score 6087; DB 9; Length 165329;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 6087; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 CCAGTAGGAAACAGTTCTTCTAAAGCCTATTGTCAACAATACCTTCCAGATGTTAGCATT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76651 CCAGTAGGAAACAGTTCTTCTAAAGCCTATTGTCAACAATACCTTCCAGATGTTAGCATT 76710

Qy         61 TTACAATTTAAGGAACTTAAAATAGCCTTCAAACTTTTTGCCAGTTTCTCTGATATCCAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76711 TTACAATTTAAGGAACTTAAAATAGCCTTCAAACTTTTTGCCAGTTTCTCTGATATCCAA 76770

Qy        121 TCTATTCTTTTACTCTGCCTCCCAAGCTTTCTTTCTAGAATGCTAACCTGATCGGCTTAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76771 TCTATTCTTTTACTCTGCCTCCCAAGCTTTCTTTCTAGAATGCTAACCTGATCGGCTTAA 76830

Qy        181 GTACTTGAACTACCTCTTCTCCTCCATTAACTACAGAGTAAATTCTGGTCTTCAGAGTAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76831 GTACTTGAACTACCTCTTCTCCTCCATTAACTACAGAGTAAATTCTGGTCTTCAGAGTAA 76890

Qy        241 CAAGAAACACCCTTTAGTTCTCAGCATATTCGTGCACCTTCATTTATCTCTCCTTCTCTC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76891 CAAGAAACACCCTTTAGTTCTCAGCATATTCGTGCACCTTCATTTATCTCTCCTTCTCTC 76950

Qy        301 TCAAAGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76951 TCAAAGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGC 77010

Qy        361 AACCAAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      77011 AACCAAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTA 77070

Qy        421 ATCAAAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      77071 ATCAAAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAA 77130

Qy        481 CTGTCTGTTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      77131 CTGTCTGTTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTA 77190

Qy        541 GACAAAAGGATATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      77191 GACAAAAGGATATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTA 77250

Qy        601 TTCAGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      77251 TTCAGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGG 77310

Qy        661 GCAGGACTCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      77311 GCAGGACTCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGAT 77370

Qy        721 TTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      77371 TTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGG 77430

Qy        781 AATTCTGGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      77431 AATTCTGGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGG 77490

Qy        841 AGATGGTCAGAAAGAGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAGT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      77491 AGATGGTCAGAAAGAGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAGT 77550

Qy        901 ACTTCTTAGCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAACACTGC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      77551 ACTTCTTAGCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAACACTGC 77610

Qy        961 AATCATGCTAAACTCCTCTATGACCTTCAAACATTCCACTTGCTTTTATTCTTTATGGTT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      77611 AATCATGCTAAACTCCTCTATGACCTTCAAACATTCCACTTGCTTTTATTCTTTATGGTT 77670

Qy       1021 GTGATGGCATAGAGGTCAATAGCAAAGACCCTGGAGTCCCACTGTCTGAGCTGGCATAAC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      77671 GTGATGGCATAGAGGTCAATAGCAAAGACCCTGGAGTCCCACTGTCTGAGCTGGCATAAC 77730

Qy       1081 ATTACTACCACTTAATCAATGTGTAAGCTCAGGTAAGTACTTAAGTCCTCTATGCTTCAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      77731 ATTACTACCACTTAATCAATGTGTAAGCTCAGGTAAGTACTTAAGTCCTCTATGCTTCAT 77790

Qy       1141 CTGTAAAATGAGAATCATTGAAGAACATTCTCTCAGGATGGATCATGAGGAATAAGTGAA 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      77791 CTGTAAAATGAGAATCATTGAAGAACATTCTCTCAGGATGGATCATGAGGAATAAGTGAA 77850

Qy       1201 TTAACTGGCATATAGTGCTTAAACCAGTGCCTTGCTCAGTTAGTGACAGATAAAATCATC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      77851 TTAACTGGCATATAGTGCTTAAACCAGTGCCTTGCTCAGTTAGTGACAGATAAAATCATC 77910

Qy       1261 TGTTATTACTGTGCCCACTATTGTGATGCTCTTCTCTTCTTTGTACAACGACTACATCTC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      77911 TGTTATTACTGTGCCCACTATTGTGATGCTCTTCTCTTCTTTGTACAACGACTACATCTC 77970

Qy       1321 TATTTATCATTTTAGGGTCTCCTTGTGAAAAACCACTCCAGATTCAAAAGATTGAGTTTA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      77971 TATTTATCATTTTAGGGTCTCCTTGTGAAAAACCACTCCAGATTCAAAAGATTGAGTTTA 78030

Qy       1381 ATCTCTATCCTCTGTGCTTTCCTGGAGTTTTGTAAAGTAAATCTTCACTTGACATCATGG 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78031 ATCTCTATCCTCTGTGCTTTCCTGGAGTTTTGTAAAGTAAATCTTCACTTGACATCATGG 78090

Qy       1441 ATAGGTTCTTGGAAACTACAACTTCAAGTGAAAGGACATAACTAAACCAATTTTTTTCTC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78091 ATAGGTTCTTGGAAACTACAACTTCAAGTGAAAGGACATAACTAAACCAATTTTTTTCTC 78150

Qy       1501 ATCAACGTTATAATGAAATGGCATTGATGAAATGATGGCATTCAAGGACCTGCTGTACCT 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78151 ATCAACGTTATAATGAAATGGCATTGATGAAATGATGGCATTCAAGGACCTGCTGTACCT 78210

Qy       1561 TGTTTCACTTAAAGTCACTGTTTCCAATAATCTATTGATGACATTGAGGACTTACTATAT 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78211 TGTTTCACTTAAAGTCACTGTTTCCAATAATCTATTGATGACATTGAGGACTTACTATAT 78270

Qy       1621 AATAATAAATATATATATAATCGACGAAACAGGAATCAAACTGCTAACTCTGCTAACTGG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78271 AATAATAAATATATATATAATCGACGAAACAGGAATCAAACTGCTAACTCTGCTAACTGG 78330

Qy       1681 TCTCCCTGCTTCCACACTCTGCCCACTCATCTCAGTCTTTCTTTCACAAGAGTCAGAATG 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78331 TCTCCCTGCTTCCACACTCTGCCCACTCATCTCAGTCTTTCTTTCACAAGAGTCAGAATG 78390

Qy       1741 ATCAGATGAGACCCCTCCTCTGCTTCTGTTTCTTCCATGGATTTCCACTGCACTCTGATA 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      78391 ATCAGATGAGACCCCTCCTCTGCTTCTGTTTCTTCCATGGATTTCCACTGCACTCTGATA 78450

Qy       1801 AAGTCCAGCCTCTTGACCACAGCCTACAAATCCTTGCACGATCTATCGTTTACTTTTCCA 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      78451 AAGTCCAGCCTCTTGACCACAGCCTACAAATCCTTGCACGATCTATCGTTTACTTTTCCA 78510
Qy       1861 TCTCCTTTTATGCTACTTTCATCTTGTTCTCAATTCTCTAGCTATGCTGGCCCCTTCTTG 1920
Db      78511 TCTCCTTTTATGCTACTTTCATCTTGTTCTCAATTCTCTAGCTATGCTGGCCCCTTCTTG 78570
Qy       1921 TTCTTTCCCATTTTTTTTTAATTTTTAAAATTTGTATATATTTATGGGTTATAAGTGAAA 1980
Db      78571 TTCTTTCCCATTTTTTTTTAATTTTTAAAATTTGTATATATTTATGGGTTATAAGTGAAA 78630
Qy       1981 TCTTTTTAGATGCATAGGTTGTATAGTGATAAAATCAGGGCTTTTAGGGTATTCATCACC 2040
Db      78631 TCTTTTTAGATGCATAGGTTGTATAGTGATAAAATCAGGGCTTTTAGGGTATTCATCACC 78690
Qy       2041 TGAATGATGTACATTGTACCCCTTAAGTAATTTCTCACCATCCGCTGACTTCTTGCCCCC 2100
Db      78691 TGAATGATGTACATTGTACCCCTTAAGTAATTTCTCACCATCCGCTGACTTCTTGCCCCC 78750
Qy       2101 TGGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGC 2160
Db      78751 TGGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGC 78810
Qy       2161 TGACTTCTTGCCCCCTGGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAA 2220
Db      78811 TGACTTCTTGCCCCCTGGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAA 78870
Qy       2221 TTTCTCACCATCCGCTGACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCAT 2280
Db      78871 TTTCTCACCATCCGCTGACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCAT 78930
Qy       2281 TCCACACTCTACATCTATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGC 2340
Db      78931 TCCACACTCTACATCTATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGC 78990
Qy       2341 AATATTTGTCTTTCTGTGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCT 2400
Db      78991 AATATTTGTCTTTCTGTGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCT 79050
Qy       2401 AGGCTGCTGCAAAAGGCATGATTTCATTCTTTTTTATGGCTATGTTCTTTCCCAATTTAG 2460
Db      79051 AGGCTGCTGCAAAAGGCATGATTTCATTCTTTTTTATGGCTATGTTCTTTCCCAATTTAG 79110
Qy       2461 ATAAAGAACACTCGCACTTGCTCTTACTTCTATTTGGAATACTAATTCCTAGGCTTCTTG 2520
Db      79111 ATAAAGAACACTCGCACTTGCTCTTACTTCTATTTGGAATACTAATTCCTAGGCTTCTTG 79170
Qy       2521 CATTGCTTTCTCCTTCTCACCCATCAAATCTCATTTTAGATACCACCTCTTCAAAGAGGG 2580
Db      79171 CATTGCTTTCTCCTTCTCACCCATCAAATCTCATTTTAGATACCACCTCTTCAAAGAGGG 79230
Qy       2581 CTTTCCTGACCACCTTGGCTGAATTAGCCCTTCACCATCTGATTACTCTCTAGCACATCA 2640
Db      79231 CTTTCCTGACCACCTTGGCTGAATTAGCCCTTCACCATCTGATTACTCTCTAGCACATCA 79290
Qy       2641 CCTGCCCATTTTATTCATGGTACAGGTCAAAATCTGGAATCACCTGATTTGTTTATTTTC 2700
Db      79291 CCTGCCCATTTTATTCATGGTACAGGTCAAAATCTGGAATCACCTGATTTGTTTATTTTC 79350
Qy       2701 TGACTCCTTCTACTGAGATGAAAACTCTACTAGAGCGGAGATTTTATCTGCTTGTATCAG 2760
Db      79351 TGACTCCTTCTACTGAGATGAAAACTCTACTAGAGCGGAGATTTTATCTGCTTGTATCAG 79410
Qy       2761 GTACTGCTTCAAACAGCACCTGATACAGAGTAGGTGGTCAAAAGATATTTCTTAAACAAA 2820
Db      79411 GTACTGCTTCAAACAGCACCTGATACAGAGTAGGTGGTCAAAAGATATTTCTTAAACAAA 79470
Qy       2821 TGAACAAATAAAAAGTAGATCTTTTGAGAGTAAAGCTCTTCCACACTACCAGAGTCATTC 2880
Db      79471 TGAACAAATAAAAAGTAGATCTTTTGAGAGTAAAGCTCTTCCACACTACCAGAGTCATTC 79530
Qy       2881 AGGAATGACAAATCATAGAATAACAGAATTTGATGCTTTGTGCATATCAGAGAAAGAAGG 2940
Db      79531 AGGAATGACAAATCATAGAATAACAGAATTTGATGCTTTGTGCATATCAGAGAAAGAAGG 79590

Qy       2941 TGGAAGGTTGTCAAGGTATCATGATGTACCAGTCCTCGCCTCCTCAAACACAATCTGCAA 3000
Db      79591 TGGAAGGTTGTCAAGGTATCATGATGTACCAGTCCTCGCCTCCTCAAACACAATCTGCAA 79650
Qy       3001 GTCCCACAGTGAAAAAGTAAGTTAACTCATGTGAAGCGTTTTACAAACACTTTTTTAAAA 3060
Db      79651 GTCCCACAGTGAAAAAGTAAGTTAACTCATGTGAAGCGTTTTACAAACACTTTTTTAAAA 79710
Qy       3061 GTCTTAAAACTCCTAAGAAAGCAAGATTTAATAGTCAAAGAAGTGAGTAAACATGAAATG 3120
Db      79711 GTCTTAAAACTCCTAAGAAAGCAAGATTTAATAGTCAAAGAAGTGAGTAAACATGAAATG 79770
Qy       3121 CCTGAACAGAGTAATGAGCTAAGCACAAAGTTAGAGACATGTTAGTTAATATGTCTTGAA 3180
Db      79771 CCTGAACAGAGTAATGAGCTAAGCACAAAGTTAGAGACATGTTAGTTAATATGTCTTGAA 79830
Qy       3181 AGCAGCAGCTCCTGCTTTCAAGGAGCAAGAACAAATTGGGCAAGTGAACACTCCTTGAAT 3240
Db      79831 AGCAGCAGCTCCTGCTTTCAAGGAGCAAGAACAAATTGGGCAAGTGAACACTCCTTGAAT 79890
Qy       3241 AAAATGTGTAAAATTAATTTTGGGTTATGTTCTATACTGTGTATAATAGAATGATAAAAA 3300
Db      79891 AAAATGTGTAAAATTAATTTTGGGTTATGTTCTATACTGTGTATAATAGAATGATAAAAA 79950
Qy       3301 TTATTTGACTAGCACTTTGTAGTTTAGAAATATCTCTATTTACACAGTTTACCTTATTTG 3360
Db      79951 TTATTTGACTAGCACTTTGTAGTTTAGAAATATCTCTATTTACACAGTTTACCTTATTTG 80010
Qy       3361 ATAAGACTGTTGAGTGATGGGATAGCATGGTGGACAATCCACATAACTGAGTATCGAGAC 3420
Db      80011 ATAAGACTGTTGAGTGATGGGATAGCATGGTGGACAATCCACATAACTGAGTATCGAGAC 80070
Qy       3421 ACCTGTATCTGGACCCAGCTCTGTTAGTAAGAAGCTGTAACCTCAGCAAGTCACTTTCTC 3480
Db      80071 ACCTGTATCTGGACCCAGCTCTGTTAGTAAGAAGCTGTAACCTCAGCAAGTCACTTTCTC 80130
Qy       3481 TTTCTGGGTCTCTATTTCCTTTTTGGTGAAATGAGAGTGTTAGGCTAGATTGCCTTTGAA 3540
Db      80131 TTTCTGGGTCTCTATTTCCTTTTTGGTGAAATGAGAGTGTTAGGCTAGATTGCCTTTGAA 80190
Qy       3541 GTCCCATTTTGTCTTTAAAGTCCCATCTATTGCAGTGATTTATATTTAACTCATGACAAA 3600
Db      80191 GTCCCATTTTGTCTTTAAAGTCCCATCTATTGCAGTGATTTATATTTAACTCATGACAAA 80250
Qy       3601 TCAGGCTTCTCTTATTCTAAGTGCAAGACATAAAACTTTTATTGTGGAATTTCAGGCATC 3660
Db      80251 TCAGGCTTCTCTTATTCTAAGTGCAAGACATAAAACTTTTATTGTGGAATTTCAGGCATC 80310
Qy       3661 AGTAAATCTTTTTGGGTACTCACTTATGTTCCTGAAATCAATCTATTTGAGTGATCACTC 3720
Db      80311 AGTAAATCTTTTTGGGTACTCACTTATGTTCCTGAAATCAATCTATTTGAGTGATCACTC 80370
Qy       3721 TTTTAGGTGCCCAGGTAAACAAAGAAGGCCATGGTCTTTCTTTGAGTGACCTTCTTTCCC 3780
Db      80371 TTTTAGGTGCCCAGGTAAACAAAGAAGGCCATGGTCTTTCTTTGAGTGACCTTCTTTCCC 80430
Qy       3781 TTTTAATTAGTCTGACCTCTTTAATGTCAGTTCTGACTGATTCATTTCCCTGGTCCATCT 3840
Db      80431 TTTTAATTAGTCTGACCTCTTTAATGTCAGTTCTGACTGATTCATTTCCCTGGTCCATCT 80490
Qy       3841 TCCTTGGTCTGAGGGCCTTCCTAGTTTCATATTGCACTTCAGTTCCTTCCACACCACCAT 3900
Db      80491 TCCTTGGTCTGAGGGCCTTCCTAGTTTCATATTGCACTTCAGTTCCTTCCACACCACCAT 80550
Qy       3901 CAAGGATGGCTGTCAACATTCATTTGTTCTATGTTATAATTCAAGGAAAAGTTGCCCAGT 3960
Db      80551 CAAGGATGGCTGTCAACATTCATTTGTTCTATGTTATAATTCAAGGAAAAGTTGCCCAGT 80610
Qy       3961 AGCTAATCCAATAAATGCCCTCTTATGGGCGGCTAGAGACTTTTTCCTATAATTTAAATG 4020
Db      80611 AGCTAATCCAATAAATGCCCTCTTATGGGCGGCTAGAGACTTTTTCCTATAATTTAAATG 80670
```

```
Qy       4021 CATCTTCTGTAGATTATGGTCCCTCCACCACTTTACATTTGTCTGCTGTCTCCTTGCTCT 4080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      80671 CATCTTCTGTAGATTATGGTCCCTCCACCACTTTACATTTGTCTGCTGTCTCCTTGCTCT 80730

Qy       4081 GCTAGTCATGGAACGTGTTGGTAGTGGGGGCAGTGTGGGATGTTCAAGGGCACGTATTGG 4140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      80731 GCTAGTCATGGAACGTGTTGGTAGTGGGGGCAGTGTGGGATGTTCAAGGGCACGTATTGG 80790

Qy       4141 GTAGGGCCACATATGGGCATTGCTTTGTGCCATTCTTTCTATATTTTTGGTATTTTGCAT 4200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      80791 GTAGGGCCACATATGGGCATTGCTTTGTGCCATTCTTTCTATATTTTTGGTATTTTGCAT 80850

Qy       4201 CTCACTGGAACCCAACTATTTTTCATCTCTTCCACCTAAACTATTTGATGCCTCTGTTTC 4260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      80851 CTCACTGGAACCCAACTATTTTTCATCTCTTCCACCTAAACTATTTGATGCCTCTGTTTC 80910

Qy       4261 TTATATATAAAGTATAGCTCACTGTAGCCTATGATCAGGAACCTATCTGCTTTCTAAATG 4320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      80911 TTATATATAAAGTATAGCTCACTGTAGCCTATGATCAGGAACCTATCTGCTTTCTAAATG 80970

Qy       4321 AAAGCTGTTTTGGTCAGATCTAGCAATTAATTCCCTTCTTCCACTTATAGCTTTCCTCTG 4380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      80971 AAAGCTGTTTTGGTCAGATCTAGCAATTAATTCCCTTCTTCCACTTATAGCTTTCCTCTG 81030

Qy       4381 TAACTCTGGTGTAGGTATTTGGTTTATGGCTATAAGATGTGAAACACCTGAATGATTCTG 4440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81031 TAACTCTGGTGTAGGTATTTGGTTTATGGCTATAAGATGTGAAACACCTGAATGATTCTG 81090

Qy       4441 TCCATGCAGGCATTTCAGTTCATGATATTGTATGTAAAAGATACTGATTGTCTAGGTGTT 4500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81091 TCCATGCAGGCATTTCAGTTCATGATATTGTATGTAAAAGATACTGATTGTCTAGGTGTT 81150

Qy       4501 CAGAAACACCTATAGGGCTTAATATTCTTACAATCAGTTTGAAGGCTGGTGATACGCAAA 4560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81151 CAGAAACACCTATAGGGCTTAATATTCTTACAATCAGTTTGAAGGCTGGTGATACGCAAA 81210

Qy       4561 GCAAACTACATATTTTTCTGCCTGCTCTCTCTCTTTCTCTCTACATCTCTCTTTCTTTAT 4620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81211 GCAAACTACATATTTTTCTGCCTGCTCTCTCTCTTTCTCTCTACATCTCTCTTTCTTTAT 81270

Qy       4621 CTTTTGAAATATCAGTTTGGAGACTTAGAATTACATAAGACATAAACCCATTTGATATAA 4680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81271 CTTTTGAAATATCAGTTTGGAGACTTAGAATTACATAAGACATAAACCCATTTGATATAA 81330

Qy       4681 GAATTGCTGTGTATATTTGCTCATCTACTCCCTCCTTTGGTCCTCGAGCTGCCGGTTTAG 4740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81331 GAATTGCTGTGTATATTTGCTCATCTACTCCCTCCTTTGGTCCTCGAGCTGCCGGTTTAG 81390

Qy       4741 ACTTTTTACAGGACGCAGGCATGTGAAGGAGAAACTGTCAGTGCTAGGCTGAATTCTGTT 4800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81391 ACTTTTTACAGGACGCAGGCATGTGAAGGAGAAACTGTCAGTGCTAGGCTGAATTCTGTT 81450

Qy       4801 GTTACCAAGATTTCTAGAAAAGTATTCCTCAGTCAGGTTGATTACAGATATAGCAAATCT 4860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81451 GTTACCAAGATTTCTAGAAAAGTATTCCTCAGTCAGGTTGATTACAGATATAGCAAATCT 81510

Qy       4861 ATTTTTCCTAGGGTAGTTTCTGTATGCTGCCGGGCTTATAACTGTCTGTCATCCAGCTAT 4920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81511 ATTTTTCCTAGGGTAGTTTCTGTATGCTGCCGGGCTTATAACTGTCTGTCATCCAGCTAT 81570

Qy       4921 TTCTCTCCACCTTCTTGTTTGCATAACAACCAAGGCAACTTCCGCAAATCACTGCGTGGA 4980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81571 TTCTCTCCACCTTCTTGTTTGCATAACAACCAAGGCAACTTCCGCAAATCACTGCGTGGA 81630

Qy       4981 GACGATGATCCTGCCAGCTCCCTTTTGGAAATCGTGAGGATCAGATCTTGGACCATGTAT 5040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81631 GACGATGATCCTGCCAGCTCCCTTTTGGAAATCGTGAGGATCAGATCTTGGACCATGTAT 81690

Qy       5041 AATATGATGCTTCTAATCCAAAAGAGGAAAGGCATTGGGAGTCAGCTCCTAAGTAAGCTC 5100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81691 AATATGATGCTTCTAATCCAAAAGAGGAAAGGCATTGGGAGTCAGCTCCTAAGTAAGCTC 81750

Qy       5101 CAGAATTCCTGCTGGTACTTTTCCTTCCAGGAAGCAACTTCCTTGATATTTTTTTTTTAC 5160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81751 CAGAATTCCTGCTGGTACTTTTCCTTCCAGGAAGCAACTTCCTTGATATTTTTTTTTTAC 81810

Qy       5161 AGGCATATGAATAAAAACTATATTTTGCAGCATTGTACACTTTTTTTCCTTTTCTAGAAA 5220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81811 AGGCATATGAATAAAAACTATATTTTGCAGCATTGTACACTTTTTTTCCTTTTCTAGAAA 81870

Qy       5221 TTCTAAACCTCTGACATTGGTGGAGACATTGAGTACATTTTTTCCCATATCCCTACTTTT 5280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81871 TTCTAAACCTCTGACATTGGTGGAGACATTGAGTACATTTTTTCCCATATCCCTACTTTT 81930

Qy       5281 CAGAAGGATTTTCTCTGCTCGTTCACTTAACATTGCTGATGCGTCAGTCTTTTCTTCCTC 5340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81931 CAGAAGGATTTTCTCTGCTCGTTCACTTAACATTGCTGATGCGTCAGTCTTTTCTTCCTC 81990

Qy       5341 ATCTCTTTCAGGGGCTGGAGAGGCAGAGGGAGACAGAGGAGCTGGTACTGCAGAGCGGTC 5400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81991 ATCTCTTTCAGGGGCTGGAGAGGCAGAGGGAGACAGAGGAGCTGGTACTGCAGAGCGGTC 82050

Qy       5401 GTCTGATTGGCTGGACGGTCGTAGCTGGGCTATAAAAGAGACCCCTACAGGCTTAGCAGG 5460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      82051 GTCTGATTGGCTGGACGGTCGTAGCTGGGCTATAAAAGAGACCCCTACAGGCTTAGCAGG 82110

Qy       5461 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 5520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      82111 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 82170

Qy       5521 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 5580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      82171 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 82230

Qy       5581 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAGGTAAGATTGCTT 5640
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      82231 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAGGTAAGATTGCTT 82290

Qy       5641 TCAGCCATTAACCATATTAAACTTTTGGCTAGACTTTCTCAGTTATTTACATGTTGTACT 5700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      82291 TCAGCCATTAACCATATTAAACTTTTGGCTAGACTTTCTCAGTTATTTACATGTTGTACT 82350

Qy       5701 TACTAACCTAGTTCTGTGCAATTAGAAACAGTGTGGTCAGGAGAGCACGACTTTCTAACT 5760
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      82351 TACTAACCTAGTTCTGTGCAATTAGAAACAGTGTGGTCAGGAGAGCACGACTTTCTAACT 82410

Qy       5761 TTCCTCCAAGACTAGCTAGATATTGTGACTTAAGACATGTGCTCCCCAAATTTCAGCCCT 5820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      82411 TTCCTCCAAGACTAGCTAGATATTGTGACTTAAGACATGTGCTCCCCAAATTTCAGCCCT 82470

Qy       5821 TATGTGTTGTTTTGTGTGACCTCAGTTTTGAGAACTGTTCTATTCTTTAAGCCAGGTCTA 5880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      82471 TATGTGTTGTTTTGTGTGACCTCAGTTTTGAGAACTGTTCTATTCTTTAAGCCAGGTCTA 82530

Qy       5881 AGAAAGCTAGTTTTAATTAAGAAGCGAGATGAGGTTTGAGGCTATGTACAGTGATCTGTA 5940
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      82531 AGAAAGCTAGTTTTAATTAAGAAGCGAGATGAGGTTTGAGGCTATGTACAGTGATCTGTA 82590

Qy       5941 ATATCTCCATCTGTGATTACTACTGCTATTTGAGCATCCCTGGAGTACATAGAAGCCTGG 6000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      82591 ATATCTCCATCTGTGATTACTACTGCTATTTGAGCATCCCTGGAGTACATAGAAGCCTGG 82650

Qy       6001 CTCTGGGCTTTCTGATTGTATGCTACAACTTGTTTCAGGAAAGGTACCCCAGAATGAGGT 6060
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      82651 CTCTGGGCTTTCTGATTGTATGCTACAACTTGTTTCAGGAAAGGTACCCCAGAATGAGGT 82710

Qy       6061 TTGGCTCCATCATCAGAAAGGCACTAT 6087
             |||||||||||||||||||||||||||
Db      82711 TTGGCTCCATCATCAGAAAGGCACTAT 82737
```

RESULT 3
AC031977/c
LOCUS AC031977 191699 bp DNA linear HTG 12-APR-2001
DEFINITION Homo sapiens chromosome 1 clone RP11-288O18, WORKING DRAFT SEQUENCE, 3 unordered pieces.

```
ACCESSION   AC031977
VERSION     AC031977.7  GI:13194952
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 191699)
  AUTHORS   Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
            Federspiel,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,
            Mao,J., Komp,C., Kottler,S., Lam,B., Marathe,R., Miranda,M.,
            Morehouse,A.J., Nguyen,M., Oefner,P., Palm,C.J., Ramirez,D.,
            Southwick,A.M., Webb,C., Wilhelmy,J., Yu,S. and Davis,R.W.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 191699)
  AUTHORS   Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
            Federspiel,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,
            Mao,J., Marathe,R., Morehouse,A.J., Oefner,P., Palm,C.J.,
            Ramirez,D., Wilhelmy,J., Yu,S. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-APR-2000) DNA Sequencing and Technology Center,
            Stanford University, 855 California Avenue, Palo Alto, CA 94304,
            USA
COMMENT     On Mar 4, 2001 this sequence version replaced gi:9665085.
            -------------- Genome Center
                 Center: Stanford DNA Sequencing and Technology Development
            Center
                 Center code: SDSTDC
                 Web site: http://sequence-www.stanford.edu/group/human/
                 Contact: hum-info@sequence.stanford.edu
            -------------- Project Information
                 Center project name: 880
                 Center clone name: RP11-288O18
            -------------- Summary Statistics
                 Sequencing Vector: M13mp18; X02513; 100% of reads
                 Sequencing Vector: plasmid; plasmid_accession; 0% of reads
                 Chemistry: Dye-primer; 1% of reads
                 Chemistry: Dye-terminator Big Dye; 99% of reads
                 Assembly program: Phrap; version 0.990319
                 Consensus quality: 190680 bases at least Q40
                 Consensus quality: 191287 bases at least Q30
                 Consensus quality: 191336 bases at least Q20
                 Insert size: 195548; agarose-fp
                 Insert size: 191499; sum-of-contigs
                 Quality coverage:  7.9x in Q20 bases; agarose-fp
                 Quality coverage:  8.1x in Q20 bases; sum-of-contigs.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    12646: contig of 12646 bp in length
            *   12647    12746: gap of unknown length
            *   12747    94961: contig of 82215 bp in length
            *   94962    95061: gap of unknown length
            *   95062   191699: contig of 96638 bp in length.
FEATURES             Location/Qualifiers
     source          1. .191699
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-288O18"
                     /clone_lib="RPCI human BAC library 11"
     misc_feature    1. .12646
                     /note="assembly_name:Contig5"
     misc_feature    12747. .94961
                     /note="assembly_name:Contig6
                     clone_end:T7"
     misc_feature    95062. .191699
                     /note="assembly_name:Contig7
                     clone_end:SP6"
ORIGIN

  Query Match             99.7%;  Score 6068.6;  DB 2;  Length 191699;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 6082;  Conservative    0;  Mismatches    4;  Indels    1;  Gaps    1;

Qy          1 CCAGTAGGAAACAGTTCTTCTAAAGCCTATTGTCAACAATACCTTCCAGATGTTAGCATT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7968 CCAGTAGGAAACAGTTCTTCTAAAGCCTATTGTCAACAATACCTTCCAGATGTTAGCATT 7909

Qy         61 TTACAATTTAAGGAACTTAAAATAGCCTTCAAACTTTTTGCCAGTTTCTCTGATATCCAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7908 TTACAATTTAAGGAACTTAAAATAGCCTTCAAACTTTTTGCCAGTTTCTCTGATATCCAA 7849

Qy        121 TCTATTCTTTTACTCTGCCTCCCAAGCTTTCTTTCTAGAATGCTAACCTGATCGGCTTAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7848 TCTATTCTTTTACTCTGCCTCCCAAGCTTTCTTTCTAGAATGCTAACCTGATCGGCTTAA 7789

Qy        181 GTACTTGAACTACCTCTTCTCCTCCATTAACTACAGAGTAAATTCTGGTCTTCAGAGTAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7788 GTACTTGAACTACCTCTTCTCCTCCATTAACTACAGAGTAAATTCTGGTCTTCAGAGTAA 7729

Qy        241 CAAGAAACACCCTTTAGTTCTCAGCATATTCGTGCACCTTCATTTATCTCTCCTTCTCTC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7728 CAAGAAACACCCTTTAGTTCTCAGCATATTCGTGCACCTTCATTTATCTCTCCTTCTCTC 7669

Qy        301 TCAAAGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7668 TCAAAGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGC 7609

Qy        361 AACCAAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7608 AACCAAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTA 7549

Qy        421 ATCAAAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7548 ATCAAAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAA 7489

Qy        481 CTGTCTGTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7488 CTGTCTGTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTA 7429

Qy        541 GACAAAAGGATATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7428 GACAAAAGGATATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTA 7369

Qy        601 TTCAGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7368 TTCAGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGG 7309

Qy        661 GCAGGACTCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7308 GCAGGACTCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGAT 7249

Qy        721 TTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7248 TTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGG 7189

Qy        781 AATTCTGGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7188 AATTCTGGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGG 7129

Qy        841 AGATGGTCAGAAAGAGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAGT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7128 AGATGGTCAGAAAGAGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAGT 7069

Qy        901 ACTTCTTAGCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAACACTGC 960
              |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7068 ACTTCTTAGCATGCCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAACACTGC 7009

Qy        961 AATCATGCTAAACTCCTCTATGACCTTCAAACATTCCACTTGCTTTTATTCTTTATGGTT 1020
```

```
Db      7008 AATCATGCTAAACTCCTCTATGACCTTCAAACATTCCACTTGCTTTTATTCTTTATGGTT 6949

Qy      1021 GTGATGGCATAGAGGTCAATAGCAAAGACCCTGGAGTCCCACTGTCTGAGCTGGCATAAC 1080

Db      6948 GTGATGGCATAGAGGTCAATAGCAAAGACCCTGGAGTCCCACTGTCTGAGCTGGCATAAC 6889

Qy      1081 ATTACTACCACTTAATCAATGTGTAAGCTCAGGTAAGTACTTAAGTCCTCTATGCTTCAT 1140

Db      6888 ATTACTACCACTTAATCAATGTGTAAGCTCAGGTAAGTACTTAAGTCCTCTATGCTTCAT 6829

Qy      1141 CTGTAAAATGAGAATCATTGAAGAACATTCTCTCAGGATGGATCATGAGGAATAAGTGAA 1200

Db      6828 CTGTAAAATGAGAATCATTGAAGAACATTCTCTCAGGATGGATCATGAGGAATAAGTGAA 6769

Qy      1201 TTAACTGGCATATAGTGCTTAAACCAGTGCCTTGCTCAGTTAGTGACAGATAAAATCATC 1260

Db      6768 TTAACTGGCATATAGTGCTTAAACCAGTGCCTTGCTCAGTTAGTGACAGATAAAATCATC 6709

Qy      1261 TGTTATTACTGTGCCCACTATTGTGATGCTCTTCTCTTCTTTGTACAACGACTACATCTC 1320

Db      6708 TGTTATTACTGTGCCCACTATTGTGATGCTCTTCTCTTCTTTGTACAACGACTACATCTC 6649

Qy      1321 TATTTATCATTTTAGGGTCTCCTTGTGAAAAACCACTCCAGATTCAAAAGATTGAGTTTA 1380

Db      6648 TATTTATCATTTTAGGGTCTCCTTGTGAAAAACCACTCCAGATTCAAAAGATTGAGTTTA 6589

Qy      1381 ATCTCTATCCTCTGTGCTTTCCTGGAGTTTTGTAAAGTAAATCTTCACTTGACATCATGG 1440

Db      6588 ATCTCTATCCTCTGTGCTTTCCTGGAGTTTTGTAAAGTAAATCTTCACTTGACATCATGG 6529

Qy      1441 ATAGGTTCTTGGAAACTACAACTTCAAGTGAAAGGACATAACTAAACCAATTTTTTTCTC 1500

Db      6528 ATAGGTTCTTGGAAACTACAACTTCAAGTGAAAGGACATAACTAAACCAATTTTTTTCTC 6469

Qy      1501 ATCAACGTTATAATGAAATGGCATTGATGAAATGATGGCATTCAAGGACCTGCTGTACCT 1560

Db      6468 ATCAACGTTATAATGAAATGGCATTGATGAAATGATGGCATTCAAGGACCTGCTGTACCT 6409

Qy      1561 TGTTTCACTTAAAGTCACTGTTTCCAATAATCTATTGATGACATTGAGGACTTACTATAT 1620

Db      6408 TGTTTCACTTAAAGTCACTGTTTCCAATAATCTATTGATGACATTGAGGACTTACTATAT 6349

Qy      1621 AATAATAAATATATATATAATCGACGAAACAGGAATCAAACTGCTAACTCTGCTAACTGG 1680

Db      6348 AATAATAAATATATATATAATCGACGAAACAGGAATCAAACTGCTAACTCTGCTAACTGG 6289

Qy      1681 TCTCCCTGCTTCCACACTCTGCCCACTCATCTCAGTCTTTCTTTCACAAGAGTCAGAATG 1740

Db      6288 TCTCCCTGCTTCCACACTCTGCCCACTCATCTCAGTCTTTCTTTCACAAGAGTCAGAATG 6229

Qy      1741 ATCAGATGAGACCCCTCCTCTGCTTCTGTTTCTTCCATGGATTTCCACTGCACTCTGATA 1800

Db      6228 ATCAGATGAGACCCCTCCTCTGCTTCTGTTTCTTCCATGGATTTCCACTGCACTCTGATA 6169

Qy      1801 AAGTCCAGCCTCTTGACCACAGCCTACAAATCCTTGCACGATCTATCGTTTACTTTTCCA 1860

Db      6168 AAGTCCAGCCTCTTGACCACAGCCTACAAATCCTTGCACGCTCTATCGTTTACTTTTCCA 6109

Qy      1861 TCTCCTTTTATGCTACTTTCATCTTGTTCTCAATTCTCTAGCTATGCTGGCCCCTTCTTG 1920

Db      6108 TCTCCTTTTATGCTACTTTCATCTTGTTCTCAATTCTCTAGCTATGCTGGCCCCTTCTTG 6049

Qy      1921 TTCTTTCCCA-TTTTTTTTTAATTTTTAAAATTTGTATATATTTATGGGTTATAAGTGAA 1979

Db      6048 TTCTTTCCCATTTTTTTTTTAATTTTTAAAATTTGTATATATTTATGGGTTATAAGTGAA 5989

Qy      1980 ATCTTTTTAGATGCATAGGTTGTATAGTGATAAAATCAGGGCTTTTAGGGTATTCATCAC 2039

Db      5988 ATCTTTTTAGATGCATAGGTTGTATAGTGATAAAATCAGGGCTTTTAGGGTATTCATCAC 5929

Qy      2040 CTGAATGATGTACATTGTACCCCTTAAGTAATTTCTCACCATCCGCTGACTTCTTGCCCC 2099

Db      5928 CTGAATGATGTACATTGTACCCCTTAAGTAATTTCTCACCATCCGCTGACTTCTTGCCCC 5869

Qy      2100 CTGGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCG 2159

Db      5868 CTGGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCG 5809

Qy      2160 CTGACTTCTTGCCCCCTGGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTA 2219

Db      5808 CTGACTTCTTGCCCCCTGGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTA 5749

Qy      2220 ATTTCTCACCATCCGCTGACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCA 2279

Db      5748 ATTTCTCACCATCCGCTGACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCA 5689

Qy      2280 TTCCACACTCTACATCTATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATG 2339

Db      5688 TTCCACACTCTACATCTATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATG 5629

Qy      2340 CAATATTTGTCTTTCTGTGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATC 2399

Db      5628 CAATATTTGTCTTTCTGTGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATC 5569

Qy      2400 TAGGCTGCTGCAAAAGGCATGATTTCATTCTTTTTTATGGCTATGTTCTTTCCCAATTTA 2459

Db      5568 TAGGCTGCTGCAAAAGGCATGATTTCATTCTTTTTTATGGCTATGTTCTTTCCCAATTTA 5509

Qy      2460 GATAAAGAACACTCGCACTTGCTCTTACTTCTATTTGGAATACTAATTCCTAGGCTTCTT 2519

Db      5508 GATAAAGAACACTCGCACTTGCTCTTACTTCTATTTGGAATACTAATTCCTAGGCTTCTT 5449

Qy      2520 GCATTGCTTTCTCCTTCTCACCCATCAAATCTCATTTTAGATACCACCTCTTCAAAGAGG 2579

Db      5448 GCATTGCTTTCTCCTTCTCACCCATCAAATCTCATTTTAGATACCACCTCTTCAAAGAGG 5389

Qy      2580 GCTTTCCTGACCACCTTGGCTGAATTAGCCCTTCACCATCTGATTACTCTCTAGCACATC 2639

Db      5388 GCTTTCCTGACCACCTTGGCTGAATTAGCCCTTCACCATCTGATTACTCTCTAGCACATC 5329

Qy      2640 ACCTGCCCATTTTATTCATGGTACAGGTCAAAATCTGGAATCACCTGATTTGTTTATTTT 2699

Db      5328 ACCTGCCCATTTTATTCATGGTACAGGTCAAAATCTGGAATCACCTGATTTGTTTATTTT 5269

Qy      2700 CTGACTCCTTCTACTGAGATGAAAACTCTACTAGAGCGGAGATTTTATCTGCTTGTATCA 2759

Db      5268 CTGACTCCTTCTACTGAGATGAAAACTCTACTAGAGCGGAGATTTTATCTGCTTGTATCA 5209

Qy      2760 GGTACTGCTTCAAACAGCACCTGATACAGAGTAGGTGGTCAAAAGATATTTCTTAAACAA 2819

Db      5208 GGTACTGCTTCAAACAGCACCTGATACAGAGTAGGTGGTCAAAAGATATTTCTTAAACAA 5149

Qy      2820 ATGAACAAATAAAAAGTAGATCTTTTGAGAGTAAAGCTCTTCCACACTACCAGAGTCATT 2879

Db      5148 ATGAATAAATAAAAAGTAGATCTTTTGAGAGTAAAGCTCTTCCACACTACCAGAGTCATT 5089

Qy      2880 CAGGAATGACAAATCATAGAATAACAGAATTTGATGCTTTGTGCATATCAGAGAAAGAAG 2939

Db      5088 CAGGAATGACAAATCATAGAATAACAGAATTTGATGCTTTGTGCATATCAGAGAAAGAAG 5029

Qy      2940 GTGGAAGGTTGTCAAGGTATCATGATGTACCAGTCCTCGCCTCCTCAAACACAATCTGCA 2999

Db      5028 GTGGAAGGTTGTCAAGGTATCATGATGTACCAGTCCTCGCCTCCTCAAACACAATCTGCA 4969

Qy      3000 AGTCCCACAGTGAAAAAGTAAGTTAACTCATGTGAAGCGTTTTACAAACACTTTTTTAAA 3059

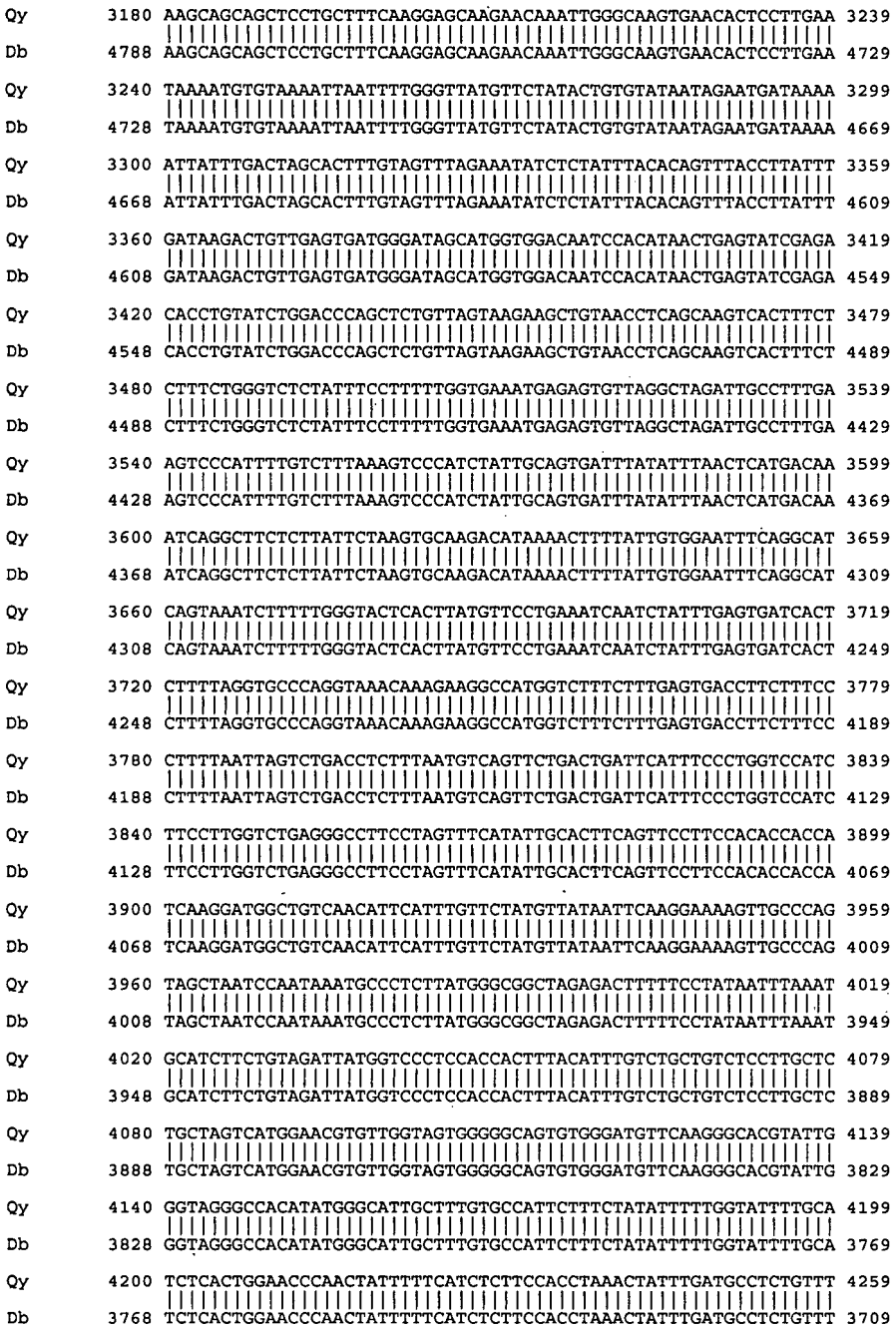
Db      4968 AGTCCCACAGTGAAAAAGTAAGTTAACTCATGTGAAGCGTTTTACAAACACTTTTTTAAA 4909

Qy      3060 AGTCTTAAAACTCCTAAGAAAGCAAGATTTAATAGTCAAAGAAGTGAGTAAACATGAAAT 3119

Db      4908 AGTCTTAAAACTCCTAAGAAAGCAAGATTTAATAGTCAAAGAAGTGAGTAAACATGAAAT 4849

Qy      3120 GCCTGAACAGAGTAATGAGCTAAGCACAAAGTTAGAGACATGTTAGTTAATATGTCTTGA 3179

Db      4848 GCCTGAACAGAGTAATGAGCTAAGCACAAAGTTAGAGACATGTTAGTTAATATGTCTTGA 4789
```


Qy 3180 AAGCAGCAGCTCCTGCTTTCAAGGAGCAAGAACAAATTGGGCAAGTGAACACTCCTTGAA 3239
Db 4788 AAGCAGCAGCTCCTGCTTTCAAGGAGCAAGAACAAATTGGGCAAGTGAACACTCCTTGAA 4729
Qy 3240 TAAAATGTGTAAAATTAATTTTGGGTTATGTTCTATACTGTGTATAATAGAATGATAAAA 3299
Db 4728 TAAAATGTGTAAAATTAATTTTGGGTTATGTTCTATACTGTGTATAATAGAATGATAAAA 4669
Qy 3300 ATTATTTGACTAGCACTTTGTAGTTTAGAAATATCTCTATTTACACAGTTTACCTTATTT 3359
Db 4668 ATTATTTGACTAGCACTTTGTAGTTTAGAAATATCTCTATTTACACAGTTTACCTTATTT 4609
Qy 3360 GATAAGACTGTTGAGTGATGGGATAGCATGGTGGACAATCCACATAACTGAGTATCGAGA 3419
Db 4608 GATAAGACTGTTGAGTGATGGGATAGCATGGTGGACAATCCACATAACTGAGTATCGAGA 4549
Qy 3420 CACCTGTATCTGGACCCAGCTCTGTTAGTAAGAAGCTGTAACCTCAGCAAGTCACTTTCT 3479
Db 4548 CACCTGTATCTGGACCCAGCTCTGTTAGTAAGAAGCTGTAACCTCAGCAAGTCACTTTCT 4489
Qy 3480 CTTTCTGGGTCTCTATTTCCTTTTTGGTGAAATGAGAGTGTTAGGCTAGATTGCCTTTGA 3539
Db 4488 CTTTCTGGGTCTCTATTTCCTTTTTGGTGAAATGAGAGTGTTAGGCTAGATTGCCTTTGA 4429
Qy 3540 AGTCCCATTTTGTCTTTAAAGTCCCATCTATTGCAGTGATTTATATTTAACTCATGACAA 3599
Db 4428 AGTCCCATTTTGTCTTTAAAGTCCCATCTATTGCAGTGATTTATATTTAACTCATGACAA 4369
Qy 3600 ATCAGGCTTCTCTTATTCTAAGTGCAAGACATAAAACTTTTATTGTGGAATTTCAGGCAT 3659
Db 4368 ATCAGGCTTCTCTTATTCTAAGTGCAAGACATAAAACTTTTATTGTGGAATTTCAGGCAT 4309
Qy 3660 CAGTAAATCTTTTTGGGTACTCACTTATGTTCCTGAAATCAATCTATTTGAGTGATCACT 3719
Db 4308 CAGTAAATCTTTTTGGGTACTCACTTATGTTCCTGAAATCAATCTATTTGAGTGATCACT 4249
Qy 3720 CTTTTAGGTGCCCAGGTAAACAAAGAAGGCCATGGTCTTTCTTTGAGTGACCTTCTTTCC 3779
Db 4248 CTTTTAGGTGCCCAGGTAAACAAAGAAGGCCATGGTCTTTCTTTGAGTGACCTTCTTTCC 4189
Qy 3780 CTTTTAATTAGTCTGACCTCTTTAATGTCAGTTCTGACTGATTCATTTCCCTGGTCCATC 3839
Db 4188 CTTTTAATTAGTCTGACCTCTTTAATGTCAGTTCTGACTGATTCATTTCCCTGGTCCATC 4129
Qy 3840 TTCCTTGGTCTGAGGGCCTTCCTAGTTTCATATTGCACTTCAGTTCCTTCCACACCACCA 3899
Db 4128 TTCCTTGGTCTGAGGGCCTTCCTAGTTTCATATTGCACTTCAGTTCCTTCCACACCACCA 4069
Qy 3900 TCAAGGATGGCTGTCAACATTCATTTGTTCTATGTTATAATTCAAGGAAAAGTTGCCCAG 3959
Db 4068 TCAAGGATGGCTGTCAACATTCATTTGTTCTATGTTATAATTCAAGGAAAAGTTGCCCAG 4009
Qy 3960 TAGCTAATCCAATAAATGCCCTCTTATGGGCGGCTAGAGACTTTTTCCTATAATTTAAAT 4019
Db 4008 TAGCTAATCCAATAAATGCCCTCTTATGGGCGGCTAGAGACTTTTTCCTATAATTTAAAT 3949
Qy 4020 GCATCTTCTGTAGATTATGGTCCCTCCACCACTTTACATTTGTCTGCTGTCTCCTTGCTC 4079
Db 3948 GCATCTTCTGTAGATTATGGTCCCTCCACCACTTTACATTTGTCTGCTGTCTCCTTGCTC 3889
Qy 4080 TGCTAGTCATGGAACGTGTTGGTAGTGGGGGCAGTGTGGGATGTTCAAGGGCACGTATTG 4139
Db 3888 TGCTAGTCATGGAACGTGTTGGTAGTGGGGGCAGTGTGGGATGTTCAAGGGCACGTATTG 3829
Qy 4140 GGTAGGGCCACATATGGGCATTGCTTTGTGCCATTCTTTCTATATTTTTGGTATTTTGCA 4199
Db 3828 GGTAGGGCCACATATGGGCATTGCTTTGTGCCATTCTTTCTATATTTTTGGTATTTTGCA 3769
Qy 4200 TCTCACTGGAACCCAACTATTTTTCATCTCTTCCACCTAAACTATTTGATGCCTCTGTTT 4259
Db 3768 TCTCACTGGAACCCAACTATTTTTCATCTCTTCCACCTAAACTATTTGATGCCTCTGTTT 3709
Qy 4260 CTTATATATAAAGTATAGCTCACTGTAGCCTATGATCAGGAACCTATCTGCTTTCTAAAT 4319
Db 3708 CTTATATATAAAGTATAGCTCACTGTAGCCTATGATCAGGAACCTATCTGCTTTCTAAAT 3649
Qy 4320 GAAAGCTGTTTTGGTCAGATCTAGCAATTAATTCCCTTCTTCCACTTATAGCTTTCCTCT 4379
Db 3648 GAAAGCTGTTTTGGTCAGATCTAGCAATTAATTCCCTTCTTCCACTTATAGCTTTCCTCT 3589
Qy 4380 GTAACTCTGGTGTAGGTATTTGGTTTATGGCTATAAGATGTGAAACACCTGAATGATTCT 4439
Db 3588 GTAACTCTGGTGTAGGTATTTGGTATATGGCTATAAGATGTGAAACACCTGAATGATTCT 3529
Qy 4440 GTCCATGCAGGCATTTCAGTTCATGATATTGTATGTAAAAGATACTGATTGTCTAGGTGT 4499
Db 3528 GTCCATGCAGGCATTTCAGTTCATGATATTGTATGTAAAAGATACTGATTGTCTAGGTGT 3469
Qy 4500 TCAGAAACACCTATAGGGCTTAATATTCTTACAATCAGTTTGAAGGCTGGTGATACGCAA 4559
Db 3468 TCAGAAACACCTATAGGGCTTAATATTCTTACAATCAGTTTGAAGGCTGGTGATACGCAA 3409
Qy 4560 AGCAAACTACATATTTTTCTGCCTGCTCTCTCTCTTTCTCTCTACATCTCTCTTTCTTTA 4619
Db 3408 AGCAAACTACATATTTTTCTGCCTGCTCTCTCTCTTTCTCTCTACATCTCTCTTTCTTTA 3349
Qy 4620 TCTTTTGAAATATCAGTTTGGAGACTTAGAATTACATAAGACATAAACCCATTTGATATA 4679
Db 3348 TCTTTTGAAATATCAGTTTGGAGACTTAGAATTACATAAGACATAAACCCATTTGATATA 3289
Qy 4680 AGAATTGCTGTGTATATTTGCTCATCTACTCCCTCCTTTGGTCCTCGAGCTGCCGGTTTA 4739
Db 3288 AGAATTGCTGTGTATATTTGCTCATCTACTCCCTCCTTTGGTCCTCGAGCTGCCGGTTTA 3229
Qy 4740 GACTTTTTACAGGACGCAGGCATGTGAAGGAGAAACTGTCAGTGCTAGGCTGAATTCTGT 4799
Db 3228 GACTTTTTACAGGACGCAGGCATGTGAAGGAGAAACTGTCAGTGCTAGGCTGAATTCTGT 3169
Qy 4800 TGTTACCAAGATTTCTAGAAAAGTATTCCTCAGTCAGGTTGATTACAGATATAGCAAATC 4859
Db 3168 TGTTACCAAGATTTCTAGAAAAGTATTCCTCAGTCAGGTTGATTACAGATATAGCAAATC 3109
Qy 4860 TATTTTTCCTAGGGTAGTTTCTGTATGCTGCCGGGCTTATAACTGTCTGTCATCCAGCTA 4919
Db 3108 TATTTTTCCTAGGGTAGTTTCTGTATGCTGCCGGGCTTATAACTGTCTGTCATCCAGCTA 3049
Qy 4920 TTTCTCTCCACCTTCTTGTTTGCATAACAACCAAGGCAACTTCCGCAAATCACTGCGTGG 4979
Db 3048 TTTCTCTCCACCTTCTTGTTTGCATAACAACCAAGGCAACTTCCGCAAATCACTGCGTGG 2989
Qy 4980 AGACGATGATCCTGCCAGCTCCCTTTTGGAAATCGTGAGGATCAGATCTTGGACCATGTA 5039
Db 2988 AGACGATGATCCTGCCAGCTCCCTTTTGGAAATCGTGAGGATCAGATCTTGGACCATGTA 2929
Qy 5040 TAATATGATGCTTCTAATCCAAAAGAGGAAAGGCATTGGGAGTCAGCTCCTAAGTAAGCT 5099
Db 2928 TAATATGATGCTTCTAATCCAAAAGAGGAAAGGCATTGGGAGTCAGCTCCTAAGTAAGCT 2869
Qy 5100 CCAGAATTCCTGCTGGTACTTTTCCTTCCAGGAAGCAACTTCCTTGATATTTTTTTTTTA 5159
Db 2868 CCAGAATTCCTGCTGGTACTTTTCCTTCCAGGAAGCAACTTCCTTGATATTTTTTTTTTA 2809
Qy 5160 CAGGCATATGAATAAAAACTATATTTTGCAGCATTGTACACTTTTTTTCCTTTTCTAGAA 5219
Db 2808 CAGGCATATGAATAAAAACTATATTTTGCAGCATTGTACACTTTTTTTCCTTTTCTAGAA 2749
Qy 5220 ATTCTAAACCTCTGACATTGGTGGAGACATTGAGTACATTTTTTCCCATATCCCTACTTT 5279
Db 2748 ATTCTAAACCTCTGACATTGGTGGAGACATTGAGTACATTTTTTCCCATATCCCTACTTT 2689
Qy 5280 TCAGAAGGATTTTCTCTGCTCGTTCACTTAACATTGCTGATGCGTCAGTCTTTTCTTCCT 5339
Db 2688 TCAGAAGGATTTTCTCTGCTCGTTCACTTAACATTGCTGATGCGTCAGTCTTTTCTTCCT 2629
Qy 5340 CATCTCTTTCAGGGGCTGGAGAGGCAGAGGGAGACAGAGGAGCTGGTACTGCAGAGCGGT 5399

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2628 CATCTCTTTCAGGGGCTGGAGAGGCAGAGGGAGACAGAGGAGCTGGTACTGCAGAGCGGT 2569

Qy       5400 CGTCTGATTGGCTGGACGGTCGTAGCTGGGCTATAAAAGAGACCCCTACAGGCTTAGCAG 5459
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2568 CGTCTGATTGGCTGGACGGTCGTAGCTGGGCTATAAAAGAGACCCCTACAGGCTTAGCAG 2509

Qy       5460 GAAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAA 5519
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2508 GAAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAA 2449

Qy       5520 GCCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAA 5579
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2448 GCCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAA 2389

Qy       5580 ATAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAGGTAAGATTGCT 5639
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2388 ATAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAGGTAAGATTGCT 2329

Qy       5640 TTCAGCCATTAACCATATTAAACTTTTGGCTAGACTTTCTCAGTTATTTACATGTTGTAC 5699
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2328 TTCAGCCATTAACCATATTAAACTTTTGGCTAGACTTTCTCAGTTATTTACATGTTGTAC 2269

Qy       5700 TTACTAACCTAGTTCTGTGCAATTAGAAACAGTGTGGTCAGGAGAGCACGACTTTCTAAC 5759
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2268 TTACTAACCTAGTTCTGTGCAATTAGAAACAGTGTGGTCAGGAGAGCACGACTTTCTAAC 2209

Qy       5760 TTTCCTCCAAGACTAGCTAGATATTGTGACTTAAGACATGTGCTCCCCAAATTTCAGCCC 5819
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2208 TTTCCTCCAAGACTAGCTAGATATTGTGACTTAAGACATGTGCTCCCCAAATTTCAGCCC 2149

Qy       5820 TTATGTGTTGTTTTGTGTGACCTCAGTTTTGAGAACTGTTCTATTCTTTAAGCCAGGTCT 5879
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2148 TTATGTGTTGTTTTGTGTGACCTCAGTTTTGAGAACTGTTCTATTCTTTAAGCCAGGTCT 2089

Qy       5880 AAGAAAGCTAGTTTTAATTAAGAAGCGAGATGAGGTTTGAGGCTATGTACAGTGATCTGT 5939
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2088 AAGAAAGCTAGTTTTAATTAAGAAGCGAGATGAGGTTTGAGGCTATGTACAGTGATCTGT 2029

Qy       5940 AATATCTCCATCTGTGATTACTACTGCTATTTGAGCATCCCTGGAGTACATAGAAGCCTG 5999
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2028 AATATCTCCATCTGTGATTACTACTGCTATTTGAGCATCCCTGGAGTACATAGAAGCCTG 1969

Qy       6000 GCTCTGGGCTTTCTGATTGTATGCTACAACTTGTTTCAGGAAAGGTACCCCAGAATGAGG 6059
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1968 GCTCTGGGCTTTCTGATTGTATGCTACAACTTGTTTCAGGAAAGGTACCCCAGAATGAGG 1909

Qy       6060 TTTGGCTCCATCATCAGAAAGGCACTA 6086
                ||||||||||||||||||||||||||||
Db       1908 TTTGGCTCCATCATCAGAAAGGCACTA 1882


RESULT 4
AC125563/c
LOCUS       AC125563              275631 bp    DNA     linear   HTG 09-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-9B12, WORKING DRAFT SEQUENCE, 4
            unordered pieces.
ACCESSION   AC125563
VERSION     AC125563.4  GI:24817906
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 275631)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 275631)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 275631)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 9, 2002 this sequence version replaced gi:23096552.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
```

```
            Contact: hgsc-help@bcm.tmc.edu
          ------------- Project Information
            Center project name: GDBK
            Center clone name: CH230-9B12
          ------------- Summary Statistics
            Assembly program: Phrap; version 0.990329
            Consensus quality: 228152 bases at least Q40
            Consensus quality: 230310 bases at least Q30
            Consensus quality: 231872 bases at least Q20
            Estimated insert size: 235029; sum-of-contigs estimation
            Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
          -------------
          * NOTE: Estimated insert size may differ from sequence length
          *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 4 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *       1   192878: contig of 192878 bp in length
          *  192879   192978: gap of unknown length
          *  192979   270295: contig of 77317 bp in length
          *  270296   270395: gap of unknown length
          *  270396   272405: contig of 2010 bp in length
          *  272406   272505: gap of unknown length
          *  272506   275631: contig of 3126 bp in length.
FEATURES             Location/Qualifiers
     source          1. .275631
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-9B12"
     misc_feature    30410. .30884
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH339524"
     misc_feature    complement(56153. .56817)
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH339528"
     misc_feature    57750. .59243
                     /note="wgs_end_extension
                     clone_end:Sp6"
ORIGIN

  Query Match             6.7%;  Score 405.4;  DB 2;  Length 275631;
  Best Local Similarity   55.3%;  Pred. No. 7.6e-79;
  Matches 1792;  Conservative    0;  Mismatches 1231;  Indels  218;  Gaps   43;

Qy       2812 TTAAACAAATGAACAAATAAAAAGTAGATCTTTTGAGAGTAAAGCTCTTCCACACTACCA 2871
Db     145145 TGAACGAAGTGAACGTGTGCACAAAAGACCCTTTAAGAGCAAATCCCATCCAC---ATAA 145089

Qy       2872 GAGTCATTCAGGAATGACAAATCATAGAATAACAGAATTTGATGCTTTGTGCATATCAGA 2931
Db     145088 GAATCATGCAGGAATGACAAACCAGCAGATAATAGGAGTTGATAATCTGTGCGTATCAGA 145029

Qy       2932 GAAAGAAGGTGGAAGGTTGTCAAGGTATCATGATGTACCAGTCCTCGCCTCCTCAAACAC 2991
Db     145028 GAGGGGAGGTGGGCTGTTGTCAGGGTATCATGCTGTACTGGT-CTCGCCTCCTC-AACAC 144971

Qy       2992 AATCTGCAAGTCCCACAGTGAAAAAGTAAGTTAACTCATGTGAAGCGTTTTACAAACACT 3051
Db     144970 AAGCAGCCAGTCGCA-TTAAAAAAAAAAGGCTAATTTATGAGAAGCTTTTCAAACTATCA 144912

Qy       3052 TTTTT------AAAAGTCTTAAAACTCCTAAGAAAGCAAGATTTAATAGTCAAAGAAGTG 3105
Db     144911 GGAGTGCAAGGGACAGGCAAAGAAGTGAGTTAAAGAACTGGGTTTCTTGGGGAAGCAGTC 144852

Qy       3106 AGTAAACATGAAATGCCTGAACAGAGTAATGA--GCTAAGCACAAAGTTAGAGACATGTT 3163
Db     144851 ATGCTGTTAGCTATAGCCCCTGCTTGCATGGAATAAGAGGAATCCAGGCAGAGAGAGAGA 144792

Qy       3164 AGTTAATATGTCTTGAAAGCAGCAGCTCCTGCTTTCAAGGAGCAAGAACAAATTGGGCAA 3223
Db     144791 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAAAGAGAGAGAGA 144732

Qy       3224 GTGAACACTCCTTGAATAAAATGTGTAAAATTAATTTTGGGTTATGTTCTATACTGTGTA 3283
Db     144731 GAGATCCTTGACTGAAATGAATGAAATTAATTAATATTGGGCTTTATTATCTGCTATGTA 144672

Qy       3284 TAATAGAATGATAAAAATTATTTGACTAGCACTTTGTAGTTTAGAAATATCTCTATTTAC 3343
Db     144671 TATTCTGACAATACAAA-TATTTGCCGAGCACTTTACAGTTTAGGAATATCTCTATGTAC 144613

Qy       3344 ACAGTTTACCTTATTTGATAAGACTGTTGAGTGATGGGATAGCATGGTGGACAATCCACA 3403
Db     144612 ACAGCGTATCTCCCTTGATAAGAGTACTGAATGACAGGGCAGCTCTGTGGACAGTCTACA 144553

Qy       3404 TAACTGAGTATCGAGACACCTGTATCTGGACCCAGCTCTGTTAGTAAGAAGCT-GTAACC 3462
Db     144552 CAGCCCAGTGTCAAGACGCCTGGGTCCGCATCCTGTCCGGTTAACAGAAAACTCCGTGTC 144493

Qy       3463 TCAGCAAGTCACTTTCTCTTTCTGGGTCTCTATTTCCTTTTTGGTGAAATGAGAGTGTTA 3522
Db     144492 CTCGGGAGGGACGGTTGCTTTCAGTCTCTGTAATTCTATTTATCATCGGATGAAACGAAG 144433

Qy       3523 GGCTAGATTGCCTTTGAAGTCCCATTTTGTCTTTAAAGTCCCATCTATTGCAGTGATTTA 3582
Db     144432 GCATAGA----------CCGAGTGGTTTTCAAAGTCCATTTTAGCGC--------A 144395

Qy       3583 TATTTAACTCATGACAAATCAGGCTTCTCTTATTCTAAGTGCAAGACATAAAACTTTTAT 3642
Db     144394 TTTAGAACTGATGCCACACCTCTTTTCTTCTAAATCCACGATGAAATGTTCCTGCTCTGC 144335

Qy       3643 TGTGGAATTTCAGGCATCAGTAAATCTTTTTGGGTACT-CACTTATGTTCCTGAAATCAA 3701
Db     144334 AATTGTTTTTCAAGCACCAGTACATCTTTTTGAGTCTTCCACTGATGTCCCCTGAATCAG 144275

Qy       3702 TCTATTTGAGTGATCACTCTTTTAGGTGCCCAGGTAAACAAAGAAGGCCATGGTCTTTCT 3761
Db     144274 ATTGTCTGAGTGATCCCTGTCTTAGGTGCTCGGGTAAACAAAGATTACTCTAGAAG---- 144219

Qy       3762 TTGAGTGACCTTCTTTCCCTTTTAATTAGTCTGACCTCTTTAATGTCAGTTCTGACTGAT 3821
Db     144218 ----GTAACAGTCTCTCCCTTTCAGTTGATCGC----CTTTGATGCAAACTCTGCCCGAA 144167

Qy       3822 TCATTTCCCTGGTCCATCTTCCTTGGTCTGAGGGCCTTCCTAGTTTCATATTGCACTTCA 3881
Db     144166 CCTCTAGTCC--------------------AAGGTCTCTCTCACTTCATCCTGCACGGTG 144127

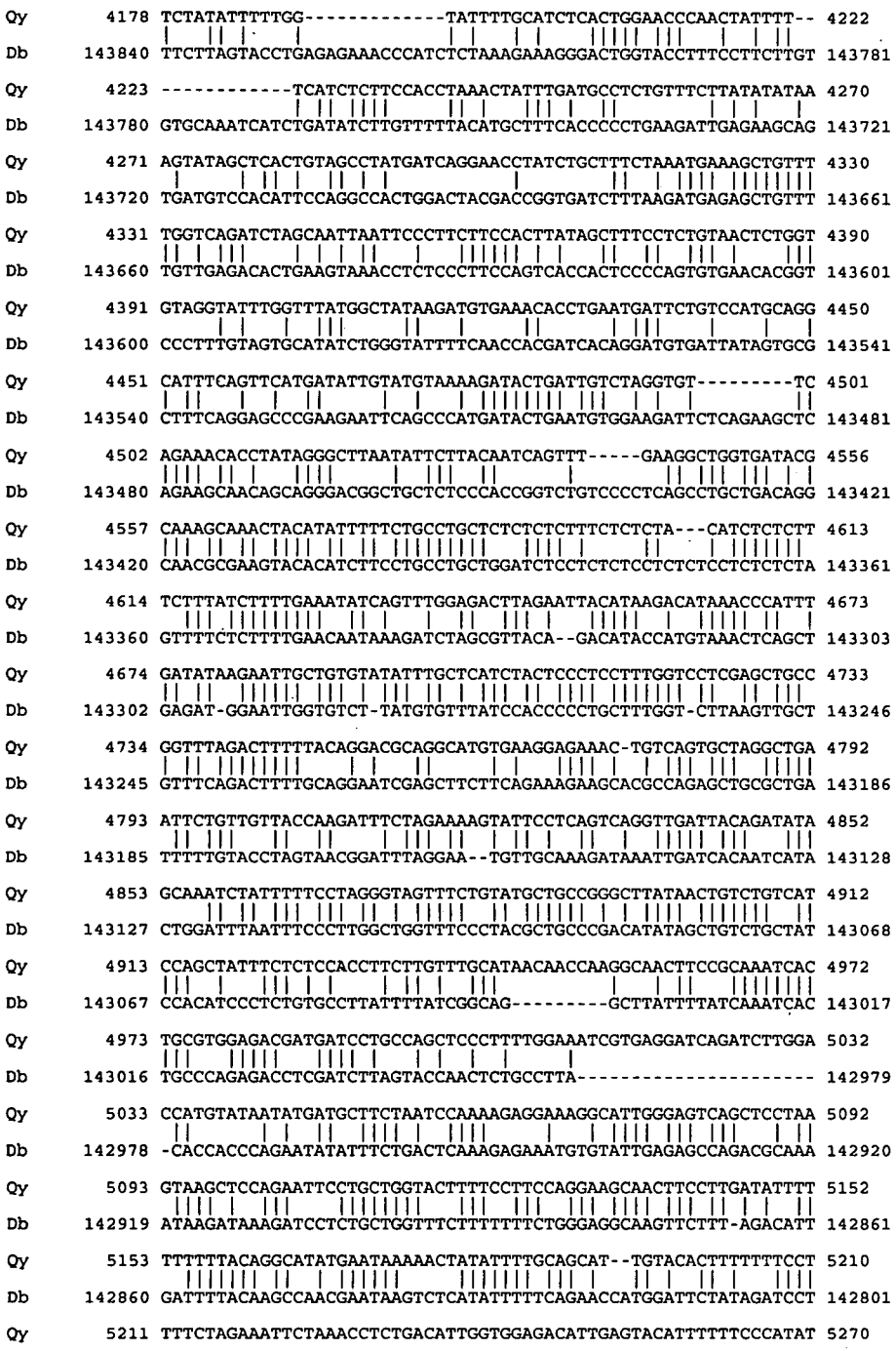
Qy       3882 GTTCCTTCCACACCACCATCAAGGATGGCTGTCAACATTCATTTGTTCTATGTTATAATT 3941
Db     144126 ATCCCTTCCACTCCACCACCAGGGACAGCTATCATCTTTCCTAGGTTTCATGTTAGAAAG 144067

Qy       3942 CAAGGAAAAGTTGCCCAGTAGCTAATCCAATAAATGCCCTCTTATGGGCGGCTAGAGACT 4001
Db     144066 CA--GGAGAGGTTTTCAGTGGCTGATGTAATAGGTGATTTTAGTCTTTACAGGAGAGAGC 144009

Qy       4002 TTTTCCTATAATTTAAATGCATCTTCTGTAGATTATGGTCCCTCCACCACTTTACATTTG 4061
Db     144008 TGGGGACTTGCCCTTAACTGACATACCT------------ATTCCCCTCTTTAACATTTG 143961

Qy       4062 TCTGCTGTCTCCTTGCTCTGCTAGTCATGGAACGTGTTG----GTAGTGGGGGCAGTGTG 4117
Db     143960 TCTACTGTCTGCATACTCAGCTAGCCAGGGGAAGATGTGAATAGTAAGGTGGGTGGTTAG 143901

Qy       4118 GGATGTTCAAGGGCACGTATTGGGTAGGGCCACATATGGGCATTGCTTTGTGCCATTCTT 4177
Db     143900 GGCAATAGTCTTCCTGGTCCAGAGCATAGAACATTGCATGTTTCTATTTCTCCTTTTATA 143841
```




```
RESULT 5
AC115766
LOCUS       AC115766              225600 bp    DNA     linear   ROD 31-JAN-2004
DEFINITION  Mus musculus chromosome 1, clone RP23-82I24, complete sequence.
ACCESSION   AC115766
VERSION     AC115766.9  GI:41470217
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 225600)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus chromosome 1, clone RP23-82I24
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 225600)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
```

Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE Direct Submission

JOURNAL Submitted (22-MAR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE 3 (bases 1 to 225600)

AUTHORS Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N., Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE Direct Submission

JOURNAL Submitted (31-DEC-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE 4 (bases 1 to 225600)

AUTHORS Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N., Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., MacLean,C., Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE Direct Submission

JOURNAL Submitted (31-JAN-2004) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT On Jan 31, 2004 this sequence version replaced gi:40445419.

All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
-------------- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
-------------- Project Information
Center project name: L23317
Center clone name: 82_I_24
--------------.

FEATURES	Location/Qualifiers
source	1. .225600
	/organism="Mus musculus"
	/mol_type="genomic DNA"
	/db_xref="taxon:10090"
	/chromosome="1"
	/map="1"
	/clone="RP23-82I24"
	/clone_lib="RPCI-23 Female Mouse BAC"
misc_feature	1. .10056
	/note="wgs_end_extension
	clone_end:SP6"
repeat_region	2. .58
	/rpt_family="(TCC)n"
repeat_region	complement(61. .205)
	/rpt_family="B1_MM"
repeat_region	complement(210. .297)
	/rpt_family="MIR"
repeat_region	1115. .1173
	/rpt_family="(TTTA)n"
repeat_region	1653. .1689
	/rpt_family="(TG)n"
repeat_region	3086. .3215
	/rpt_family="B1_MM"
repeat_region	3216. .3263
	/rpt_family="(CAA)n"
repeat_region	4885. .4984
	/rpt_family="PB1D9"
repeat_region	4993. .5023
	/rpt_family="(CACCAT)n"
repeat_region	5024. .5064
	/rpt_family="(CAA)n"
repeat_region	7312. .7368
	/rpt_family="GA-rich"
repeat_region	complement(8192. .8990)
	/rpt_family="RMER13A"
unsure	complement(9153. .9174)
	/note="single clone coverage"
repeat_region	9731. .9796
	/rpt_family="(TG)n"
repeat_region	9802. .9981
	/rpt_family="B2_Mm2"
misc_feature	10057. .10062
	/note="clone_boundary
	clone_end:SP6
	site:EcoRI"
repeat_region	complement(11356. .11527)
	/rpt_family="L2"
repeat_region	11528. .11613
	/rpt_family="(TTATA)n"
repeat_region	complement(11614. .11763)
	/rpt_family="L2"
repeat_region	complement(11832. .12036)
	/rpt_family="MSTD"
repeat_region	complement(12839. .16086)
	/rpt_family="Lx2"
repeat_region	17100. .17149
	/rpt_family="(TG)n"
repeat_region	complement(18967. .19081)
	/rpt_family="ID_B1"
repeat_region	19301. .19350

```
                     /rpt_family="(TAA)n"
     repeat_region   complement(20733. .21013)
                     /rpt_family="B4A"
     repeat_region   complement(22840. .23104)
                     /rpt_family="MLT1J"
     repeat_region   26941. .26982
                     /rpt_family="(TTTTG)n"
     repeat_region   complement(28528. .29162)
                     /rpt_family="Lx"
     repeat_region   29150. .29445
                     /rpt_family="L1_MM"
     repeat_region   complement(30205. .30499)
                     /rpt_family="B4"
     repeat_region   32068. .32318
                     /rpt_family="ORR1D"
     repeat_region   32329. .32734
                     /rpt_family="ORR1B-int"
     repeat_region   33123. .33759
                     /rpt_family="Lx"
     repeat_region   33793. .34312
                     /rpt_family="ORR1B-int"
     repeat_region   34313. .34627
                     /rpt_family="ORR1D"
     repeat_region   34628. .34805
                     /rpt_family="(TTTC)n"
     repeat_region   complement(34808. .34947)
                     /rpt_family="B1_MM"
     repeat_region   complement(35046. .35421)
                     /rpt_family="L1MA6"
     repeat_region   35468. .35533
                     /rpt_family="L2"
     repeat_region   complement(35744. .35834)
                     /rpt_family="B1_MM"
     repeat_region   35841. .35977
                     /rpt_family="B3"
     repeat_region   35979. .36012
                     /rpt_family="(CA)n"
     repeat_region   complement(36706. .36743)
                     /rpt_family="U2"

Query Match             6.6%;  Score 402.6;  DB 10;  Length 225600;
Best Local Similarity   54.1%;  Pred. No. 3.2e-78;
Matches 1738;  Conservative    0;  Mismatches 1174;  Indels  301;  Gaps   33;

Qy       2701 TGACTCCTTCTACTGAGATGAAAACTCTACTAGAGCGGAGATTTTATCTGCTTGTATCAG 2760
             || |||| | | ||  | |      |    ||||   | |  |    | ||||  |   |
Db     129273 TGCCTCCCTGTTCTCCGTTCCCTAGCAAACTACCTCCCATCTCACTTTTGCTCATTGTAA 129332

Qy       2761 GTACTGCTTCAAACAGCACCTGATACAGAGTAGGTGGTCAAAAGATATTTCTTAAACAAA 2820
              |     |  || | |  |     |  | |  |||     |     |  | |      ||
Db     129333 GCTTAAGTAAAATCTGTGCTATTGATTCACTGTGTGAGAGATGCTCAGGTGTGTCTGAAG 129392

Qy       2821 TGAACAAATAAAAAGTAGATCTTTTGAGAGTAAAGCTCTTCCACACTACCAGAGTCATTC 2880
              ||| |    | | |  ||| |||| ||||| ||| | | | |||   |  |||  ||  |
Db     129393 TGAGCTTGCACAGAAGAGACCTTTGGAGAGCAAATCCCATTCAC---ATAAGAAGCAAGC 129449

Qy       2881 AGGAATGACAAATCATAGAATAACAGAATTTGATGCTTTGTGCATATCAGAGAAAGAAGG 2940
              |||||||||||| |||   ||||  |   ||| |  ||||||| |||||| ||  | |||
Db     129450 AGGAATGACAAACCATCAGATAATGGGCATTGGTAATTTGTGCCTATCAGGGAGGGGAGG 129509

Qy       2941 TGGAAGGTTGTCAAGGTATCATGATGTACCAGTCCTCGCCTCCTCAAACA-----CAATC 2995
              ||     |||||| |||||| ||||||||||||     |       | |||     | |
Db     129510 TGAGCCATTGTCAGGGTATCCTGATGTACCAGTCTCGCCTCTTCAACACAAACAGCCAGT 129569

Qy       2996 TGCAAGTCCCACAGTGAAAAAGTAAGTTAACTCATGTGAAGCGTTTTACAAACACTTTTT 3055
              | || ||   | |   ||| |  ||| ||| | ||| ||||| ||| | | |   ||
Db     129570 TTCATGTAAAAAATAAAAATAAAAAGCTAATTTATGAGAAGCTTTTCAAATA---TTCGG 129626

Qy       3056 TAAAAGTCTTAAAACTCCTAAGAAAGCAAGATTTAATAGTCAAAGAAGTGAGTAAAC--- 3112
              |  |||| ||| |||| |  ||| |||||    | |   ||||       || ||
Db     129627 TGGAAGTATTAGAACTGTTGGGAATGCAAGGGACAGTCAGTAAAGTGAGTGATAGACTTG 129686

Qy       3113 --------ATGAAATGCCTGAACAGAGTAATGAGCTAAGCACAAAGTTAGAGACATGTTA 3164
```

```
                                | ||  ||    |  |  |  |  |  |||||   ||  |      | |
Db     129687 GGCAGATGAAGAGTTGAAGAACTGGTTTTCTTGGGGAAGCAATCAGGCAATCTGCTATAG 129746

Qy       3165 GTTAATATGTCTTGAAAGCAGCAGCTCCTGCTTTCAAGGAGCAAGA---------- 3210
                |  | || ||  ||  |  |     |  |||  |||
Db     129747 CTCCTGTTTGCCTGGAACAAGAGGAATCCATGCAGAGAGAGAGAGAGAGAGCCAGAGAGA 129806

Qy       3211 ---------ACAAATTGGGCAAGTGAACACTCCTTGAATAAAATGTGTAAAATTAATTTT 3261
                      | | ||      || ||     |  ||||    |||     ||||||| ||
Db     129807 CAGAGAGACAGAGATACTCACAGAGATTCTTGGCTGAAATGTATGAAATTAATTAATATT 129866

Qy       3262 GGGTTATGTTCTATACTGTGTATAATAGAATGATAAAAATTATTTGACTAGCACTTTGTA 3321
              | | | | || | | || | |||   |    |  |||||||  ||||||  | |||||||  | |
Db     129867 GAGCTTTATTATCTGCTATCTATGTTCTGACAATAAAAAATATTTGCCAAGCACTTCGCA 129926

Qy       3322 GTTTAGAAATATCTCTATTTACACAGTTTACCTTATTTGATAAGACTGTTGAGTGATGGG 3381
              |||||||||||||||||  |||||||   || ||  |||||||||    ||||  |||| |
Db     129927 GTTTAGAAATATCTCTATGTACACAGCATATCTCCTTTGATAAGAACATTGAATGATAGA 129986

Qy       3382 ATAGCATGGTGGACAATCCACATAACTGAGTATCGAGACACCTGTATCTGGACCCAGCTC 3441
              | |||   ||||||| || ||| |||  ||| || |||  |||||  ||||  | || |
Db     129987 ACAGCTCTGTGGACAGTCTACACAACCCAGTGTCAAGAAACCTGGGTCTGCATCCTGTCG 130046

Qy       3442 TGTTAGTAAGAAGCTGTAA---CCTCAGCAAGTCACTTTCTCTTTCTGGGTCTCTATTTC 3498
              ||||   |  ||  | |     | |  |    |  |||||  | ||||||  |||||||
Db     130047 GGTTAACAGAAAATTCTGTATCCTTGGGAGGCTGGCTTTCAGTCTCTGGGATTCTATTTA 130106

Qy       3499 CTTTTTGGTGAAATGAGAGTGTTAGGCTAGATTGCCTTTGAAGTCCCATTTTGTCTTTAA 3558
              ||  |||||||||| ||      ||                          ||  |||||
Db     130107 TCTTCGGGTGAAATGAGGGTACAGATTGAG---------------------TGGTTTTAA 130145

Qy       3559 AGTCCCATCTATTGCAGTGATTTATATTTAACTCATGACAAATCAGGCTTCTCTTATTCT 3618
              |  |||| ||  |||| ||  || |        | |  ||  | |     || |
Db     130146 AAGTCCATTTAGCTCAGTTATAGCTAATGC------TATACCTCTTTCCTTCTAGATCCG 130199

Qy       3619 AAGTGCAAGACATAAAACTTTTATTGTGGAATTTCAGGCATCAGTAAATCTTTTTGGGTA 3678
              |  | ||              | |  ||||| ||| | ||| || | |||| ||
Db     130200 AGATAAAATGTTCTTCCTCTGGGATTTCCTTTTTCAAGCAGCGGTACATTTCTTTGAGTC 130259

Qy       3679 CTC-ACTTATGTTCCTGAAATCAATCTATTTGAGTGATCACTCTTTTAGGTGCCCAGGTA 3737
              || ||| |||||||| ||| ||   | | |||||| || || |  | |||||  ||||||
Db     130260 TTCTACTGATGTTCCTAAAACCAGATTGTCTGAGTGCTCCCTGTCGTGGGTGCTCAGGTA 130319

Qy       3738 AACAAAGAAGGCCATGGTCTTTCTTTGAGTGACCTTCTTTCCCTTTTAATTAGTCTGACC 3797
              | |||||||  |  | |        | |||  ||   || |||||||||  ||  |||
Db     130320 AGCAAAGAACACTCTAGGTCACTGTGGAGAAACAGGCTCTCCCTTTTAGTTGATCT---- 130375

Qy       3798 TCTTTAATGTCAGTTCTGACTGATTCATTTCCCTGGTCCATCTTCCTTGGTCTGAGGGCC 3857
              |||||||||  | ||||| |||                  | |       |||| |
Db     130376 TCTTTAATGCAAATTCTGCCTG-------------------CACATCCAGTCCAGGGTC 130415

Qy       3858 TTCCTAGTTTCATATTGCACTTCAGTTCCTTCCACACCACCATCAAGGATGGCTGTCAAC 3917
              |  ||  |||||  ||||||    | ||| |||| |||||| ||||||   ||| ||| |
Db     130416 TCTCTCACTTCATCCTGCACTGTCATCCCTCCCACTCCACCACCAAGGACAGCTCTCATC 130475

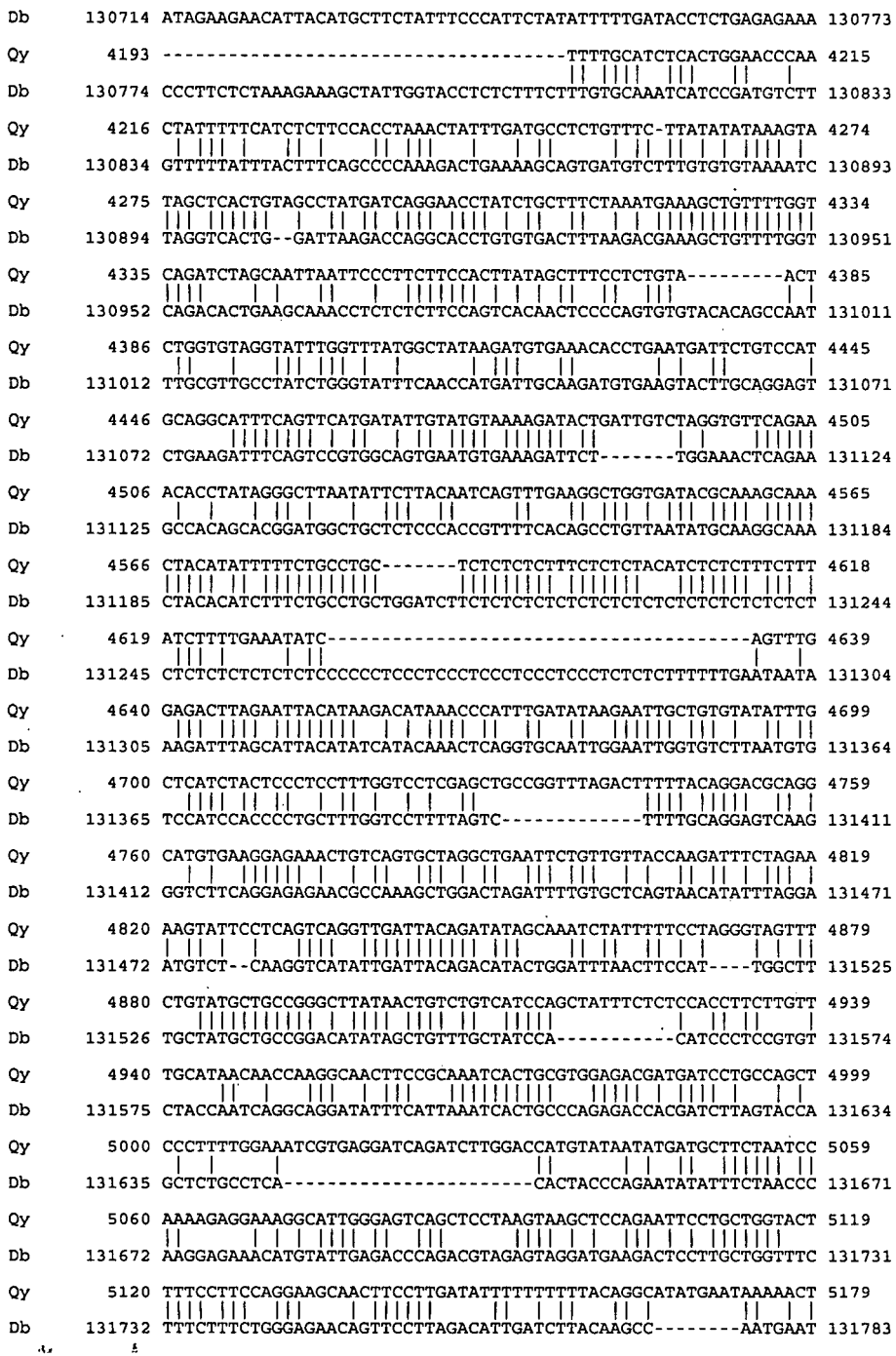
Qy       3918 ATTC----ATTTGTTCTATGTTATAATTCAAGGAAAAGTTGCCCAGTAGCTAATCCAATA 3973
              |||      |  |||  |||||| ||  ||| || | |||  |   |||||  |   ||||
Db     130476 TTTCTCATCTAGGTTTCATGTTAGAAAGCAAAGAGAGGTTTTC--ATAGCTGAATTAATA 130533

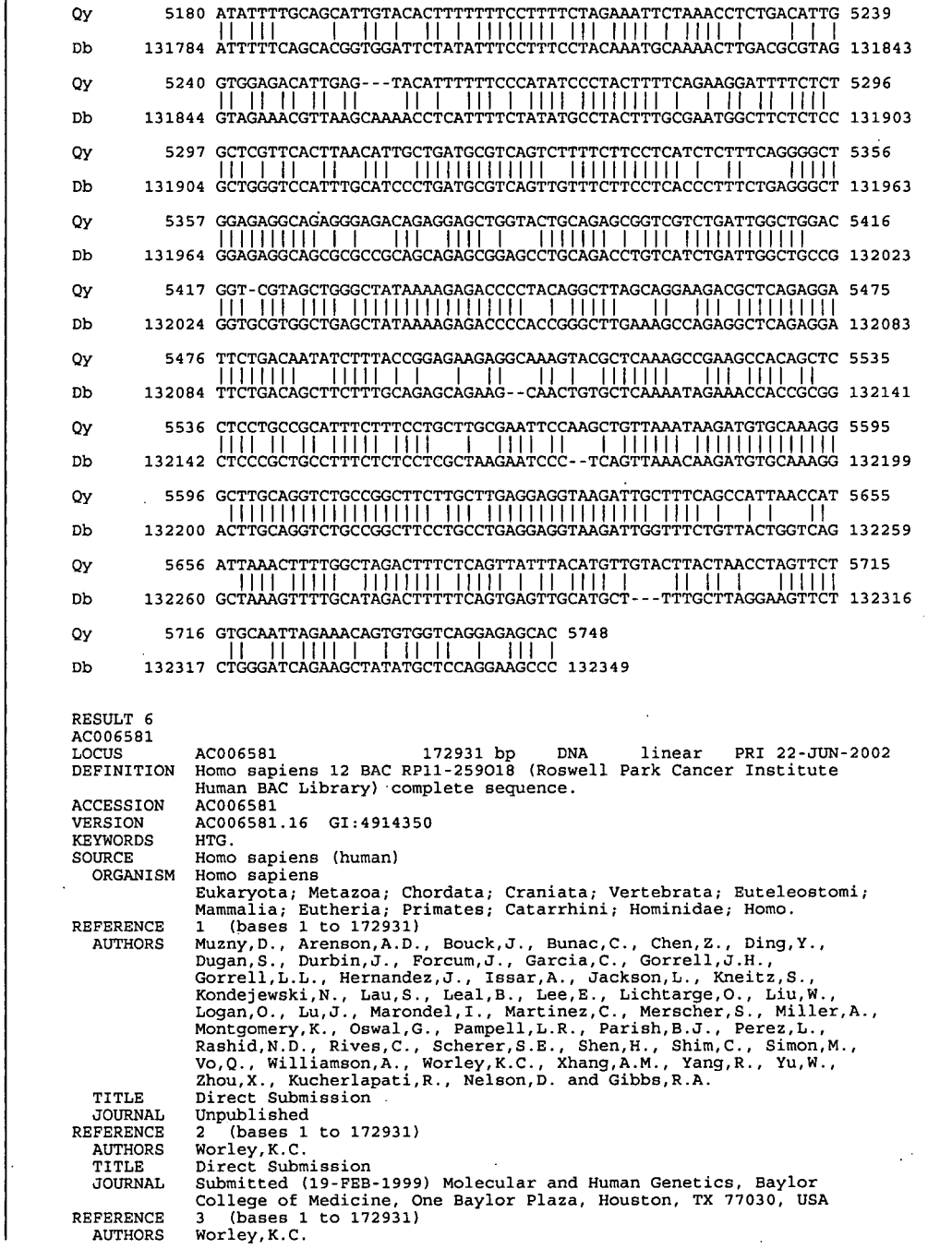
Qy       3974 AATGCCCTCTTATGGGCGGCTAGAGACTTTTTCCTATAATTTAAATGCATCTTCTGTAGA 4033
              | |    | ||| |        |    ||||   | || |||          |    |
Db     130534 AGTTTTTTGTTGTTTTTTTGTTTTTGTTTTTGTTTTTATTTTTGTCTTTACAGGAGAGAG 130593

Qy       4034 TTATGGTCCCTCCACCACTTTACAT-----TTGTCTGCTGTCTCCTTGCTCTGCTAGTCA 4088
              |  || |  |||      | ||||      ||| |  || |  | ||  |||||||
Db     130594 CTGGGGACTTTCCCTATACTGACATACCTATTGCCCTCTTTTACGTTTGTCTGCTATCTC 130653

Qy       4089 TGGAACGTGTTGGTAGTGGGGGCAGTGTGGGATGTTCAAGGGCACGTATTGGGTAGGGCC 4148
              |    | | |   ||  | ||||  |||   | ||       ||||    | |
Db     130654 CATACTCAGCTAGCCAAGGAAGATGTGTACGATAAAGAGGGCAGGTGTTTGGTCCAGAGC 130713

Qy       4149 ACATATGGGCATTGCTTTGTGCCATT---CTTTCTATATTTTTGGTA---------- 4192
              | | | |  |||| | |  | | |||   | ||||||||||||| ||
```





RESULT 6
AC006581
LOCUS AC006581 172931 bp DNA linear PRI 22-JUN-2002
DEFINITION Homo sapiens 12 BAC RP11-259O18 (Roswell Park Cancer Institute Human BAC Library) complete sequence.
ACCESSION AC006581
VERSION AC006581.16 GI:4914350
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 172931)
AUTHORS Muzny,D., Arenson,A.D., Bouck,J., Bunac,C., Chen,Z., Ding,Y., Dugan,S., Durbin,J., Forcum,J., Garcia,C., Gorrell,J.H., Gorrell,L.L., Hernandez,J., Issar,A., Jackson,L., Kneitz,S., Kondejewski,N., Lau,S., Leal,B., Lee,E., Lichtarge,O., Liu,W., Logan,O., Lu,J., Marondel,I., Martinez,C., Merscher,S., Miller,A., Montgomery,K., Oswal,G., Pampell,L.R., Parish,B.J., Perez,L., Rashid,N.D., Rives,C., Scherer,S.E., Shen,H., Shim,C., Simon,M., Vo,Q., Williamson,A., Worley,K.C., Xhang,A.M., Yang,R., Yu,W., Zhou,X., Kucherlapati,R., Nelson,D. and Gibbs,R.A.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 172931)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (19-FEB-1999) Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 172931)
AUTHORS Worley,K.C.

```
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAY-1999) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   4  (bases 1 to 172931)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUN-1999) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   5  (bases 1 to 172931)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   6  (bases 1 to 172931)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JUN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 28, 1999 this sequence version replaced gi:4884002.
            INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
            gc-help@bcm.tmc.edu

            CLONE LENGTH: This sequence does not necessarily represent the
            entire insert of this clone. Overlapping regions of clones are only
            sequenced and submitted once, so the sequence for the remainder of
            the insert may be found in the record for the adjacent clones.
            Overlapping clones are noted at the beginning and end of the
            Features listing.

            ANNOTATION OF FEATURES:
               STSs are identified using ePCR (Genome Res. 7:541-550) searches
            of a local database that includes entries from dbSTS, GDB, and
            local mapping efforts.
               Repeats are identified using RepeatMasker (A. Smit and P. Green,
            unpublished.) for Human and Mouse sequences.
               Genes and Region of sequence similarity are identified by BLAST
            (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
            EST and cDNA sequences. Genes demonstrate at least two exons
            flanked by consensus splice sites that maintained sequence
            continuity across the splice junctions. Sequences that are not
            identical matches are annotated as similar.

            SEQUENCING READ COVERAGE:Sequencing is completed to a minimum
            standard of double strand coverage with a minimum of 2 clones and 2
            reads with no ambiguities or 2 chemistries with a minimum of 2
            clones and 3 reads with no ambiguities. If the sequence quality for
            a region does not meet this standard, it will be indicated in the
            annotation as Low Coverage.

            QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality
            standards - estimated error rate less than 1 per 10,000 bases.
            Reports of lowest quality individual bases and measures of base
            quality are listed below. Description of the metrics can be found
            at URL:
            http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

            QUALSTAT-REPORT.
FEATURES             Location/Qualifiers
     source          1. .172931
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="12"
                     /clone="RP11-259O18"
     misc_feature    1. .2001
                     /note="overlaps bases 32746. .34746 of clone AC006513"
                     /function="clone overlap"
     repeat_region   complement(371. .658)
                     /rpt_family="AluJo"
     repeat_region   1461. .1537
                     /rpt_family="MER94"
     repeat_region   complement(1585. .1632)
                     /rpt_family="MIR"
     repeat_region   complement(1648. .1952)
                     /rpt_family="AluSx"
     repeat_region   complement(1953. .2243)
                     /rpt_family="AluJo"
     repeat_region   complement(2332. .2887)
                     /rpt_family="MLT1F"
     repeat_region   2888. .2968
                     /rpt_family="MER5A"
     repeat_region   3377. .3424
                     /rpt_family="MIR"
     repeat_region   complement(3843. .4148)
                     /rpt_family="AluSg"
     repeat_region   4988. .5161
                     /rpt_family="MIR"
     repeat_region   5283. .5404
                     /rpt_family="L2"
     repeat_region   complement(5870. .6002)
                     /rpt_family="AluSp"
     repeat_region   7182. .7249
                     /rpt_family="MIR"
     repeat_region   7260. .7558
                     /rpt_family="AluJb"
     repeat_region   7573. .7756
                     /rpt_family="MER5A"
     repeat_region   complement(7840. .8077)
                     /rpt_family="MIR"
     repeat_region   8363. .8472
                     /rpt_family="AluSg"
     repeat_region   complement(8523. .8823)
                     /rpt_family="AluSx"
     repeat_region   8889. .8986
                     /rpt_family="MIR"
     repeat_region   complement(8987. .9299)
                     /rpt_family="AluJb"
     repeat_region   9300. .9358
                     /rpt_family="MIR"
     repeat_region   complement(10435. .10740)
                     /rpt_family="AluY"
     repeat_region   complement(10741. .11016)
                     /rpt_family="AluSq"
     repeat_region   complement(11152. .11457)
                     /rpt_family="AluSx"
     repeat_region   complement(12062. .12361)
                     /rpt_family="AluSp"
     repeat_region   13188. .13492
                     /rpt_family="AluSq"
     repeat_region   13841. .14180
                     /rpt_family="AluSx"
     repeat_region   14215. .14496
                     /rpt_family="AluSp"
     repeat_region   14497. .14528
                     /rpt_family="(CAA)n"
     repeat_region   14993. .15288
                     /rpt_family="AluY"
     misc_feature    15260. .15320
                     /function="Low coverage"
     repeat_region   15413. .15440
                     /rpt_family="AT_rich"
     repeat_region   16323. .16617
                     /rpt_family="AluSp"
     repeat_region   17380. .17432
                     /rpt_family="(TG)n"
     repeat_region   17697. .17735
                     /rpt_family="(TA)n"
     repeat_region   17796. .17884
                     /rpt_family="(TA)n"
     repeat_region   complement(18345. .18642)
                     /rpt_family="MER33"
     repeat_region   18656. .18877
```

```
                     /rpt_family="MIR"
     repeat_region   complement(19222. .19523)
                     /rpt_family="AluY"
     repeat_region   complement(19973. .20434)
                     /rpt_family="L1M4"
     repeat_region   complement(20445. .20569)
                     /rpt_family="AluSq/x"
     repeat_region   20855. .21159
                     /rpt_family="AluY"
     repeat_region   complement(21237. .21456)
                     /rpt_family="Charlie1b"
     repeat_region   21457. .21770
                     /rpt_family="AluSg"
     repeat_region   complement(21771. .21938)
                     /rpt_family="Charlie1b"
     repeat_region   complement(21975. .22129)
                     /rpt_family="Charlie1b"
     repeat_region   complement(22200. .22335)
                     /rpt_family="AluSg/x"
     repeat_region   complement(22390. .22725)
                     /rpt_family="L2"
     repeat_region   22956. .22976
                     /rpt_family="AT_rich"
     repeat_region   complement(22977. .23271)
                     /rpt_family="AluSx"

  Query Match             5.4%;  Score 326.2;  DB 9;  Length 172931;
  Best Local Similarity   75.6%;  Pred. No. 2.2e-61;
  Matches  495;  Conservative    0;  Mismatches  128;  Indels   32;  Gaps    6;

Qy        309 GCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAA 368
              || ||| |||| |||| |||||||| |||| |||  ||||||||||||| |   |    |
Db      48094 GCTGTAAGGGTAAAAAAGTGTGATATATTTCCTCACCCATCATAAGGGTTGTGGCTGGCA 48153

Qy        369 CTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGT 428
              |||||||| ||||||||||||||||||||| ||| | |||||||| || || ||||||||
Db      48154 CTCCTATAATAAAAGACAGGTTAATAAGAGAAAAGCATAACAAATGTACTTCATCAAAGT 48213

Qy        429 TTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGT 488
              ||||  |||||   | |||||||||||||||||| |||||||||| | | |||||
Db      48214 TTTAGGTGACACCAGTGTCTTCAGAAATGAAGACCAAAAGACCCAGAGAACACTGT---G 48270

Qy        489 TTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAG 548
              | | | |||| ||||  |||||||||||  |||         |    |||| ||||||||
Db      48271 TCTCTGTGCTTAGGTCTGATGAAGAATGGGCAGCGGTGTAGGAATGTGATTGGACAAAAG 48330

Qy        549 GATATGATCTAGTGGTAAAGGACTCAG-GGGGAAACACAGCAAGGCCTGTCTATTCAGAT 607
              | |||||||||| |||||| ||||| || |||||||| |||||||| || ||| |||||||
Db      48331 GGTATGATCTAATGGTAATGGACTAAGCGGGGAAACCCAGCAAGGTCTCTCTGTTCAGAT 48390

Qy        608 TCTTCTTGATCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGAC 667
              |||||||  |||||        |||  |||||||||||||||| || ||||||||||
Db      48391 CCTTCTTGGCCTCTC--------TGTGCAGCATTCTTTCCTCCTGGGTTTGGGGCAGGAC 48442

Qy        668 TCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATGG 727
              |||||  | |||||||||||||| ||| ||  |||||||||| ||||| ||| |||||||||
Db      48443 CCTTCTGGAGTGAGGGTCTTCAATGGACAACAGAGAAAGTGACCTTTCTAGGTTTTATGG 48502

Qy        728 CTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCTG 787
              |||| ||  ||| |||| |||||||||||||| ||||| | |||||||||||||||| ||
Db      48503 CTTGTTTGTGGGGAGAGAAGTTCTAGTTTCTCTGACCAACCTTGGGGAAGAGGAATTATG 48562

Qy        788 GTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGGT 847
              ||||||||||||  | |    | ||||||||  || |||| ||||||||||||||| |||
Db      48563 GTTTCTGTGACT--CCTCAGGGGAGAAAGAGAGGTGAGAGACAGGAGGGCAGGAGAAGGT 48620

Qy        848 CAGAAAGAGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAGTACTTCTT 907
              |||| |          ||||||||||           | |||   ||     | || ||
Db      48621 CAGAGA-------CTTTGCTTCTGAGG-----------CTTTTTAGTCTCTAAAGTACTC 48662

Qy        908 AGCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAACACTGCAA 962
              ||||| ||||||| | |||||||||  |||  |||||||||||| ||||  || |
Db      48663 AGCATGCCAAAGCGCCATACTTTGGGGTATCATTTTCTGAGCTCCAACAAAGCTA 48717
```

```
RESULT 7
AC119619/c
LOCUS       AC119619               94158 bp    DNA     linear   PRI 28-SEP-2002
DEFINITION  Homo sapiens X BAC RP11-1J4 (Roswell Park Cancer Institute Human
            BAC Library) complete sequence.
ACCESSION   AC119619 AC012577
VERSION     AC119619.5  GI:23343650
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 94158)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 94158)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-APR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 94158)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-AUG-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   4  (bases 1 to 94158)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-SEP-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
```

```
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Sep 28, 2002 this sequence version replaced gi:22297319.
            INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
            gc-help@bcm.tmc.edu

            CLONE LENGTH: This sequence does not necessarily represent the
            entire insert of this clone. Overlapping regions of clones are only
            sequenced and submitted once, so the sequence for the remainder of
            the insert may be found in the record for the adjacent clones.
            Overlapping clones are noted at the beginning and end of the
            Features listing.

            ANNOTATION OF FEATURES:
               STSs are identified using ePCR (Genome Res. 7:541-550) searches
            of a local database that includes entries from dbSTS, GDB, and
            local mapping efforts.
               Repeats are identified using RepeatMasker (A. Smit and P. Green,
            unpublished.) for Human and Mouse sequences.
               Genes and Region of sequence similarity are identified by BLAST
            (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
            EST and cDNA sequences. Genes demonstrate at least two exons
            flanked by consensus splice sites that maintained sequence
            continuity across the splice junctions. Sequences that are not
            identical matches are annotated as similar.

            SEQUENCING READ COVERAGE:Sequencing is completed to a minimum
            standard of double strand coverage with a minimum of 2 clones and 2
            reads with no ambiguities or 2 chemistries with a minimum of 2
            clones and 3 reads with no ambiguities. If the sequence quality for
            a region does not meet this standard, it will be indicated in the
            annotation as Low Coverage.

            QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality
            standards - estimated error rate less than 1 per 10,000 bases.
            Reports of lowest quality individual bases and measures of base
            quality are listed below. Description of the metrics can be found
            at URL:
            http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annotation.ht
            ml.
FEATURES                Location/Qualifiers
     source             1. .94158
                        /organism="Homo sapiens"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9606"
                        /chromosome="X"
                        /clone="RP11-1J4"
     misc_feature       complement(1. .2004)
                        /note="overlaps bases 1. .2004 of clone AC112499"
                        /function="clone overlap"
     repeat_region      4. .92
                        /rpt_family="L1MB2"
     repeat_region      complement(479. .781)
                        /rpt_family="MER58B"
     STS                1591. .1925
                        /standard_name="SHGC-111221"
     STS                1695. .1925
                        /standard_name="SHGC-68548"
     repeat_region      complement(1803. .1861)
                        /rpt_family="MER91A"
     repeat_region      complement(2158. .2465)
                        /rpt_family="AluSx"
     repeat_region      3075. .3245
                        /rpt_family="MIR"
     repeat_region      complement(3246. .3436)
                        /rpt_family="LTR54"
     repeat_region      3437. .3739
                        /rpt_family="AluJo"
     repeat_region      complement(3740. .4007)
                        /rpt_family="LTR54"
     repeat_region      4008. .4046
                        /rpt_family="MIR"
     repeat_region      complement(4405. .4640)
                        /rpt_family="MER30"
     repeat_region      4834. .5153
                        /rpt_family="MER4D"
     repeat_region      complement(5161. .5592)
                        /rpt_family="L1M4"
     repeat_region      complement(7670. .7719)
                        /rpt_family="L2"
     repeat_region      complement(7872. .8182)
                        /rpt_family="AluSg"
     repeat_region      complement(8525. .9084)
                        /rpt_family="MLT1E"
     repeat_region      complement(9362. .9650)
                        /rpt_family="MLT1A1"
     repeat_region      10392. .10501
                        /rpt_family="(TGGA)n"
     repeat_region      complement(11028. .11174)
                        /rpt_family="MER5A"
     repeat_region      11717. .11808
                        /rpt_family="MIR"
     repeat_region      11937. .12248
                        /rpt_family="Charlie4a"
     repeat_region      12249. .12463
                        /rpt_family="MER20"
     repeat_region      12464. .12610
                        /rpt_family="Charlie4a"
     repeat_region      12611. .12947
                        /rpt_family="THE1B"
     repeat_region      14522. .14913
                        /rpt_family="THE1B"
     repeat_region      14914. .14947
                        /rpt_family="Charlie4a"
     repeat_region      15035. .15204
                        /rpt_family="MIR"
     repeat_region      15205. .15423
                        /rpt_family="MER58A"
     repeat_region      15424. .15519
                        /rpt_family="MIR"
     repeat_region      15620. .15661
                        /rpt_family="(TA)n"
     repeat_region      complement(16448. .16619)
                        /rpt_family="MER96B"
     repeat_region      16639. .16659
                        /rpt_family="AT_rich"
     repeat_region      complement(17007. .17274)
                        /rpt_family="AluJb"
     repeat_region      17441. .19201
                        /rpt_family="L1PA5"
     repeat_region      complement(19202. .19496)
                        /rpt_family="AluSp"
     repeat_region      19497. .23797
                        /rpt_family="L1PA5"
     STS                22116. .22396
                        /standard_name="D11S3206"
     STS                22258. .22436
                        /standard_name="D11S2442"
     repeat_region      complement(24720. .24929)
                        /rpt_family="MER30"
     repeat_region      complement(25741. .26040)
                        /rpt_family="AluSx"

  Query Match             5.2%;  Score 317;  DB 9;  Length 94158;
  Best Local Similarity   75.0%;  Pred. No. 2.5e-59;
  Matches  482;  Conservative    0;  Mismatches  140;  Indels   21;  Gaps    6;

Qy        315 GGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTCCTA 374
              |||  ||||     |||||| ||| |||  ||||||||||||||    |  | | |||||
Db      72007 GGGCAGAAAGGTAATGATACCTTTCCTCGCCCATCATAAGGGTCATGGCTGACATTCCTA 71948

Qy        375 TAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTACA 434
              ||   ||||| |||||||||||||| |||||   | |||| ||||||||||  ||||| |
Db      71947 TAACCAAAGATAGGTTAATAAGAGAAAAAC---ATAAATGTATTTAATCAACATTTTATA 71891

Qy        435 TGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTTTTT 494
              | ||||||||  |||||  |||||||||||||||||||||||| | |||||||  |||||
```

```
Db      71890 TTACATGGGAACCTTCATTAATGAAGACCCAAAGACCCAGGGAATACTGTCT-ATTTTTA 71832

Qy        495 TGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGATATG 554

Db      71831 TGATTAGGTTCGACGAAGAATAGAGAGTCATGTAGCAATGTGATTAGACAAAAGGGTATG 71772

Qy        555 ATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTTCTT 614

Db      71771 ATCTAATGGTAATAGACTAAGGGAAAAACCCACCAAGGCCTGTCTGTCCAGATACTTCTT 71712

Qy        615 GATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTTCTT 674

Db      71711 GACCTGTCTC--------TGTAGCTTTCCTTCCTTTGGAGTATGGGGCAGGACTCCTCCA 71660

Qy        675 CAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATGGCTTGCTT 734

Db      71659 AAATGAAGGTCTTCAAGGGAGAAGGGAGACAGTGACCTTTCTAGGTTTTATGGCTTGCAT 71600

Qy        735 CGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCTGGTTTCTG 794

Db      71599 TGTGGGAGATGGGTTCTGGTTCCTACGACCTGTCTTGGAGAAGAGGAAATCTGGTTTCTA 71540

Qy        795 TGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGGTCAGAAAG 854

Db      71539 TGACTCTCTTCGGGACAGAAAGAGAGGCAGGACACAGAAGGGCAGGATAAAGTCAGAGAG 71480

Qy        855 AGACTTGGCTGCTTCTGAGG-GCTTCCGCTCTCCTTTAGTTCCAAGTACTTCTTAGCATA 913

Db      71479 A-----CCTTGCTTCTGAGGTCCTTCCAATCTTCTTTAGTTCAAAGTA---CTCAGCATG 71428

Qy        914 CCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAACA 956

Db      71427 CCAAGGCACCATACTTTGAGGTATCAAGTCCTGAGCCTCAACA 71385
```

RESULT 8
AC137626/c

```
LOCUS       AC137626               175440 bp    DNA     linear   HTG 27-MAR-2003
DEFINITION  Homo sapiens clone CTD-2145O22, WORKING DRAFT SEQUENCE, 19
            unordered pieces.
ACCESSION   AC137626
VERSION     AC137626.2  GI:29293979
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 175440)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 175440)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 175440)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAR-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Mar 27, 2003 this sequence version replaced gi:25696694.
            ---------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Drafting Center Code: BCM
                Contact: hgsc-help@bcm.tmc.edu
            ---------- Project Information
                Center project name: HERR
                Center clone name: CTD-2145O22
            ---------- Summary Statistics
                Sequencing vector: Plasmid;
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 172016 bases at least Q40
                Consensus quality: 174734 bases at least Q30
                Consensus quality: 176344 bases at least Q20
                Estimated insert size: 168157; sum-of-contigs estimation
                Quality coverage: 14x in Q20 bases; sum-of-contigs estimation
            ----------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 19 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1362: contig of 1362 bp in length
            *     1363     1462: gap of unknown length
            *     1463     3030: contig of 1568 bp in length
            *     3031     3130: gap of unknown length
            *     3131     4717: contig of 1587 bp in length
            *     4718     4817: gap of unknown length
            *     4818     6158: contig of 1341 bp in length
            *     6159     6258: gap of unknown length
            *     6259     7733: contig of 1475 bp in length
            *     7734     7833: gap of unknown length
            *     7834     9184: contig of 1351 bp in length
            *     9185     9284: gap of unknown length
            *     9285    11256: contig of 1972 bp in length
            *    11257    11356: gap of unknown length
```

```
            *    11357    12606: contig of 1250 bp in length
            *    12607    12706: gap of unknown length
            *    12707    14194: contig of 1488 bp in length
            *    14195    14294: gap of unknown length
            *    14295    15764: contig of 1470 bp in length
            *    15765    15864: gap of unknown length
            *    15865    17682: contig of 1818 bp in length
            *    17683    17782: gap of unknown length
            *    17783    18813: contig of 1031 bp in length
            *    18814    18913: gap of unknown length
            *    18914    20063: contig of 1150 bp in length
            *    20064    20163: gap of unknown length
            *    20164    21187: contig of 1024 bp in length
            *    21188    21287: gap of unknown length
            *    21288    22341: contig of 1054 bp in length
            *    22342    22441: gap of unknown length
            *    22442    23780: contig of 1339 bp in length
            *    23781    23880: gap of unknown length
            *    23881    25373: contig of 1493 bp in length
            *    25374    25473: gap of unknown length
            *    25474    27379: contig of 1906 bp in length
            *    27380    27479: gap of unknown length
            *    27480   175440: contig of 147961 bp in length.
FEATURES             Location/Qualifiers
     source          1. .175440
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="CTD-2145O22"
ORIGIN

  Query Match             5.2%;  Score 317;  DB 2;  Length 175440;
  Best Local Similarity   75.0%;  Pred. No. 2.4e-59;
  Matches  482;  Conservative    0;  Mismatches  140;  Indels   21;  Gaps    6;

Qy        315 GGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTCCTA 374
              |||  ||||     |||||| ||| |||  ||||||||||||||     | | | |||||
Db     114048 GGGCAGAAAGGTAATGATACCTTTCCTCGCCCATCATAAGGGTCATGGCTGACATTCCTA 113989

Qy        375 TAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTACA 434
              ||   ||||| |||||||||||||| |||||   | |||| ||||||||||  ||||| |
Db     113988 TAACCAAAGATAGGTTAATAAGAGAAAAAC---ATAAATGTATTTAATCAACATTTTATA 113932

Qy        435 TGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTTTTT 494
              | ||||||||  |||||  |||||||||||||||||||||| | |||||||  |||||
Db     113931 TTACATGGGAACCTTCATTAATGAAGACCCAAAGACCCAGGGAATACTGTCT-ATTTTTA 113873

Qy        495 TGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGATATG 554
              || | |||||||| ||||||| || ||         ||     |||||||||||||| ||||
Db     113872 TGATTAGGTTCGACGAAGAATAGAGAGTCATGTAGCAATGTGATTAGACAAAAGGGTATG 113813

Qy        555 ATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTTCTT 614
              |||||  ||||||  |||| ||||  |||| || |||||||||||| | ||||| ||||||
Db     113812 ATCTAATGGTAATAGACTAAGGGAAAAACCCACCAAGGCCTGTCTGTCCAGATACTTCTT 113753

Qy        615 GATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTTCTT 674
              || || ||||         | |||| ||| |||||   ||||||||||||||||| ||
Db     113752 GACCTGTCTC--------TGTAGCTTTCCTTCCTTTGGAGTATGGGGCAGGACTCCTCCA 113701

Qy        675 CAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATGGCTTGCTT 734
              |||||  |||||||||||||||||||||| |||| ||||| ||| |||||||||||| |
Db     113700 AAATGAAGGTCTTCAAGGGAGAAGGGAGACAGTGACCTTTCTAGGTTTTATGGCTTGCAT 113641

Qy        735 CGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCTGGTTTCTG 794
              | || ||| | ||||| ||| ||| ||||  |||||| |||||||||  ||||||||||
Db     113640 TGTGGGAGATGGGTTCTGGTTCCTACGACCTGTCTTGGAGAAGAGGAAATCTGGTTTCTA 113581

Qy        795 TGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGGTCAGAAAG 854
              |||||  |||      ||||||||  | | ||  ||| ||||||||| |  |||||| ||
Db     113580 TGACTCTCTTCGGGACAGAAAGAGAGGCAGGACACAGAAGGGCAGGATAAAGTCAGAGAG 113521

Qy        855 AGACTTGGCTGCTTCTGAGG-GCTTCCGCTCTCCTTTAGTTCCAAGTACTTCTTAGCATA 913
              |       ||||||||||| |||||  ||| ||||||||| |||||   || |||||
```

```
Db     113520 A-----CCTTGCTTCTGAGGTCCTTCCAATCTTCTTTAGTTCAAAGTA---CTCAGCATG 113469

Qy        914 CCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAACA 956
              |||| |||| ||||||||   |||    | ||||||   ||||
Db     113468 CCAAGGCACCATACTTTGAGGTATCAAGTCCTGAGCCTCAACA 113426


RESULT 9
AC073227/c
LOCUS       AC073227              154125 bp    DNA     linear   HTG 12-AUG-2000
DEFINITION  Homo sapiens chromosome 11 clone RP11-652L13, WORKING DRAFT
            SEQUENCE, 45 unordered pieces.
ACCESSION   AC073227
VERSION     AC073227.1  GI:8440039
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 154125)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 154125)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUN-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT
            ---------------- Genome Center ----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            ------------- Project Information -------------
            Center project name: H_NH0652L13
            ------------- Summary Statistics  -------------
            Sequencing vector: M13; 100%
            Sequencing vector: plasmid; 0%
            Chemistry: Dye-primer ET; 100% of reads
            Chemistry: Dye-terminator Big Dye; 0% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 131306 bases at least Q40
            Consensus quality: 137026 bases at least Q30
            Consensus quality: 139583 bases at least Q20
            Insert size: 124000; agarose-fp
            Insert size: 149725; sum-of-contigs
            Quality coverage: 3.61 in Q20 bases; agarose-fp
            Quality coverage: 3.15 in Q20 bases; sum-of-contigs
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 45 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1169: contig of 1169 bp in length
            *     1170     1269: gap of unknown length
            *     1270     2769: contig of 1500 bp in length
            *     2770     2869: gap of unknown length
            *     2870     4193: contig of 1324 bp in length
            *     4194     4293: gap of unknown length
            *     4294     5710: contig of 1417 bp in length
            *     5711     5810: gap of unknown length
            *     5811     6984: contig of 1174 bp in length
            *     6985     7084: gap of unknown length
            *     7085     8592: contig of 1508 bp in length
            *     8593     8692: gap of unknown length
            *     8693     9797: contig of 1105 bp in length
            *     9798     9897: gap of unknown length
            *     9898    11428: contig of 1531 bp in length
```

```
*    11429    11528: gap of unknown length
*    11529    12866: contig of 1338 bp in length
*    12867    12966: gap of unknown length
*    12967    14965: contig of 1999 bp in length
*    14966    15065: gap of unknown length
*    15066    16845: contig of 1780 bp in length
*    16846    16945: gap of unknown length
*    16946    19643: contig of 2698 bp in length
*    19644    19743: gap of unknown length
*    19744    21949: contig of 2206 bp in length
*    21950    22049: gap of unknown length
*    22050    24411: contig of 2362 bp in length
*    24412    24511: gap of unknown length
*    24512    27410: contig of 2899 bp in length
*    27411    27510: gap of unknown length
*    27511    30054: contig of 2544 bp in length
*    30055    30154: gap of unknown length
*    30155    32447: contig of 2293 bp in length
*    32448    32547: gap of unknown length
*    32548    34490: contig of 1943 bp in length
*    34491    34590: gap of unknown length
*    34591    36874: contig of 2284 bp in length
*    36875    36974: gap of unknown length
*    36975    39829: contig of 2855 bp in length
*    39830    39929: gap of unknown length
*    39930    42666: contig of 2737 bp in length
*    42667    42766: gap of unknown length
*    42767    46079: contig of 3313 bp in length
*    46080    46179: gap of unknown length
*    46180    49377: contig of 3198 bp in length
*    49378    49477: gap of unknown length
*    49478    52066: contig of 2589 bp in length
*    52067    52166: gap of unknown length
*    52167    55853: contig of 3687 bp in length
*    55854    55953: gap of unknown length
*    55954    58804: contig of 2851 bp in length
*    58805    58904: gap of unknown length
*    58905    62776: contig of 3872 bp in length
*    62777    62876: gap of unknown length
*    62877    66204: contig of 3328 bp in length
*    66205    66304: gap of unknown length
*    66305    69872: contig of 3568 bp in length
*    69873    69972: gap of unknown length
*    69973    73632: contig of 3660 bp in length
*    73633    73732: gap of unknown length
*    73733    76814: contig of 3082 bp in length
*    76815    76914: gap of unknown length
*    76915    80767: contig of 3853 bp in length
*    80768    80867: gap of unknown length
*    80868    85695: contig of 4828 bp in length
*    85696    85795: gap of unknown length
*    85796    90246: contig of 4451 bp in length
*    90247    90346: gap of unknown length
*    90347    94050: contig of 3704 bp in length
*    94051    94150: gap of unknown length
*    94151    99671: contig of 5521 bp in length
*    99672    99771: gap of unknown length
*    99772   105092: contig of 5321 bp in length
*   105093   105192: gap of unknown length
*   105193   109346: contig of 4154 bp in length
*   109347   109446: gap of unknown length
*   109447   115145: contig of 5699 bp in length
*   115146   115245: gap of unknown length
*   115246   121014: contig of 5769 bp in length
*   121015   121114: gap of unknown length
*   121115   126886: contig of 5772 bp in length
*   126887   126986: gap of unknown length
*   126987   132609: contig of 5623 bp in length
*   132610   132709: gap of unknown length
*   132710   138881: contig of 6172 bp in length
*   138882   138981: gap of unknown length
*   138982   146419: contig of 7438 bp in length
*   146420   146519: gap of unknown length
*   146520   154125: contig of 7606 bp in length.
```

```
FEATURES             Location/Qualifiers
     source          1. .154125
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /clone="RP11-652L13"
     misc_feature    1. .1169
                     /note="assembly_name:Contig11"
     misc_feature    1270. .2769
                     /note="assembly_name:Contig15"
     misc_feature    2870. .4193
                     /note="assembly_name:Contig16"
     misc_feature    4294. .5710
                     /note="assembly_name:Contig18"
     misc_feature    5811. .6984
                     /note="assembly_name:Contig19"
     misc_feature    7085. .8592
                     /note="assembly_name:Contig20"
     misc_feature    8693. .9797
                     /note="assembly_name:Contig21"
     misc_feature    9898. .11428
                     /note="assembly_name:Contig22"
     misc_feature    11529. .12866
                     /note="assembly_name:Contig23"
     misc_feature    12967. .14965
                     /note="assembly_name:Contig24"
     misc_feature    15066. .16845
                     /note="assembly_name:Contig25"
     misc_feature    16946. .19643
                     /note="assembly_name:Contig26"
     misc_feature    19744. .21949
                     /note="assembly_name:Contig27"
     misc_feature    22050. .24411
                     /note="assembly_name:Contig28"
     misc_feature    24512. .27410
                     /note="assembly_name:Contig29"
     misc_feature    27511. .30054
                     /note="assembly_name:Contig30"
     misc_feature    30155. .32447
                     /note="assembly_name:Contig31"
     misc_feature    32548. .34490
                     /note="assembly_name:Contig32"
     misc_feature    34591. .36874
                     /note="assembly_name:Contig33"
     misc_feature    36975. .39829
                     /note="assembly_name:Contig34"
     misc_feature    39930. .42666
                     /note="assembly_name:Contig35"
     misc_feature    42767. .46079
                     /note="assembly_name:Contig36"
     misc_feature    46180. .49377
                     /note="assembly_name:Contig37"
     misc_feature    49478. .52066
                     /note="assembly_name:Contig38"
     misc_feature    52167. .55853
                     /note="assembly_name:Contig39"
     misc_feature    55954. .58804
                     /note="assembly_name:Contig40"
     misc_feature    58905. .62776
                     /note="assembly_name:Contig41"
     misc_feature    62877. .66204

  Query Match             5.1%;  Score 311.2;  DB 2;  Length 154125;
  Best Local Similarity   70.5%;  Pred. No. 4.5e-58;
  Matches  509;  Conservative    0;  Mismatches  188;  Indels   25;  Gaps    6;

Qy        317 GGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTCCTATA 376
              ||   | ||  ||||| ||| ||||||   |||||||||||||| || |    ||||||||
Db      73372 GGGTAGAAAGTGTGGTACGTTTTCTTGCCCATCATAAGGGTCACAGCTGGCACTCCTATC 73313

Qy        377 GTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTACATG 436
```

```
Db      73312 ATAAAAGACAGGTTAACAAGAGAGAAACATAACAAATTTATTTAATCAAAGTTTTGTGCG 73253

Qy        437 ACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTTTTTTG 496

Db      73252 ATACAGGAGCCTTCAGAAATGAAGATCCAAAGACCTAGAGAAAACCGTCTG-CTTTTATG 73194

Qy        497 CTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGATATGAT 556

Db      73193 CTTAGTTTCAACGAAGAATGGACAACCATGTAGCAATGTGATT-GACAAAAGGGTCTGAT 73135

Qy        557 CTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTTCTTGA 616

Db      73134 CTAATGATAATAGACT----GAGAAACCCAGCAAGGCTTGTATGTCCAGATTCTTCTCGG 73079

Qy        617 TCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTTCTTCA 676

Db      73078 TCTCTC--------TGCATAGCATTCCTCCCTCCTAGGTATAGGGCAGAACCCCTGTGGA 73027

Qy        677 ATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATGGCTTGCTTCG 736

Db      73026 ATGAGGGTCTTCAAGGGAGAAGGGAGAGAATGACCTTGCTAGGTTTTATGTCTTGCTTTA 72967

Qy        737 GGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCTGGTTTCTGTG 796

Db      72966 GGGGAGAGGGGTTCTAGTATCTGTGACCCAACTTGGAGGAGAGGAATTCTGGTTTCTATG 72907

Qy        797 ACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGGTCAGAAAGAG 856

Db      72906 ACTCACTTTAGGGGAGATGGAGGGACGGGAAAATGGAGGGCCAGAGAAGGTCAGAGAGAG 72847

Qy        857 ACTTGGCTGCT--------TCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAGTACTTCTTA 908

Db      72846 ACTTTGTTTCTGAGGCTACTTCTGGACCTTCCAATCTCCTTTTGTTCAAAGTA---CTCA 72790

Qy        909 GCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAACACTGCAATCATGC 968

Db      72789 GCATGCCAAAGTGTCATACTTTGCAGTATCATTTTCTGAGCCCCAGAAAGAGAAAAGCTA 72730

Qy        969 TAAACTCCTCTATGACCTTCAAACATTCCACTTGCTTTTATTCTTTATGGTTGTGATGGC 1028

Db      72729 AAATGGCTCATAATCCCACCAACCAGACATTAAAGGTTTATACATTCAAGTTGAGGAGAT 72670

Qy       1029 AT 1030

Db      72669 AT 72668
```

RESULT 10
AC018694
LOCUS AC018694 196832 bp DNA linear PRI 09-MAY-2001
DEFINITION Homo sapiens BAC clone RP11-563P16 from 11, complete sequence.
ACCESSION AC018694
VERSION AC018694.4 GI:9665213
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 196832)
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
PUBMED 9847074
REFERENCE 2 (bases 1 to 196832)
AUTHORS Ali,J., Maupin,R. and Lehnert,L.
TITLE The sequence of Homo sapiens BAC clone RP11-563P16
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 196832)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (16-DEC-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 196832)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 5 (bases 1 to 196832)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (10-SEP-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 6 (bases 1 to 196832)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (08-NOV-2000) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 7 (bases 1 to 196832)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (09-MAY-2001) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On Aug 3, 2000 this sequence version replaced gi:7631073.
---------- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
---------- Summary Statistics
Center project name: H_NH0563P16
----------.

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (http://bacpac.med.buffalo.edu)
VECTOR: pBACe3.6
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-484L12. The clone sequenced to the right is RP11-652L13. Actual start of this clone is at base position 1 of RP11-563P16; actual end is at base position 196832 of RP11-563P16.

The region from base position 99028 to 99031 is covered by a single m13 subclone.
FEATURES Location/Qualifiers
source 1. .196832

```
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11"
                     /clone="RP11-563P16"
                     /clone_lib="RPCI-11"
     repeat_region   25. .466
                     /rpt_family="L1"
     repeat_region   917. .1135
                     /rpt_family="L2"
     repeat_region   1137. .1361
                     /rpt_family="MIR"
     repeat_region   1386. .2075
                     /rpt_family="L2"
     repeat_region   2132. .4576
                     /rpt_family="L1"
     repeat_region   4616. .4774
                     /rpt_family="L2"
     repeat_region   5092. .5202
                     /rpt_family="MIR"
     repeat_region   5241. .5572
                     /rpt_family="ERV"
     repeat_region   5884. .6021
                     /rpt_family="L2"
     repeat_region   6078. .6151
                     /rpt_family="L1"
     repeat_region   6301. .6477
                     /rpt_family="MER1_type"
     repeat_region   6597. .6704
                     /rpt_family="L1"
     repeat_region   8852. .8954
                     /rpt_family="MIR"
     repeat_region   8954. .9034
                     /rpt_family="L2"
     repeat_region   9411. .9560
                     /rpt_family="Alu"
     repeat_region   10259. .10313
                     /rpt_family="L2"
     repeat_region   10395. .10599
                     /rpt_family="MIR"
     repeat_region   10743. .11056
                     /rpt_family="Alu"
     repeat_region   11522. .11673
                     /rpt_family="L1"
     repeat_region   13784. .13928
                     /rpt_family="L2"
     repeat_region   14035. .14188
                     /rpt_family="L2"
     repeat_region   14229. .14659
                     /rpt_family="L1"
     repeat_region   14660. .14751
                     /rpt_family="L2"
     repeat_region   14802. .15365
                     /rpt_family="MER2_type"
     repeat_region   16007. .16389
                     /rpt_family="MER1_type"
     repeat_region   17025. .17123
                     /rpt_family="ERV1"
     misc_feature    17033. .17299
                     /note="similar to  EST BE005163 (NID:g8265396)"
     misc_feature    17050. .17299
                     /note="similar to  EST AA528257 (NID:g2270326) nh26a02.s1"
     misc_feature    17053. .17299
                     /note="similar to  EST AI350387 (NID:g4087593) qo39d10.x1"
     misc_feature    17054. .17299
                     /note="similar to  EST N57929 (NID:g1201819) yv61e06.s1"
     repeat_region   17648. .17735
                     /rpt_family="MIR"
     repeat_region   19011. .19227
                     /rpt_family="MIR"
     repeat_region   21465. .21556
                     /rpt_family="MIR"
     repeat_region   21718. .21927
                     /rpt_family="Alu"
     repeat_region   21961. .22078
                     /rpt_family="MER1_type"
     repeat_region   22071. .22284
                     /rpt_family="MER1_type"
     repeat_region   22285. .22589
                     /rpt_family="Alu"
     repeat_region   22590. .22729
                     /rpt_family="MER1_type"
     repeat_region   22745. .22943
                     /rpt_family="MIR"
     repeat_region   23888. .23989
                     /rpt_family="MIR"
     repeat_region   24058. .24464
                     /rpt_family="L1"
     repeat_region   24466. .24582
                     /rpt_family="L1"
     repeat_region   24583. .24880
                     /rpt_family="Alu"
     repeat_region   24881. .25141
                     /rpt_family="L1"
     repeat_region   25149. .25336
                     /rpt_family="MaLR"
     repeat_region   25948. .26580
                     /rpt_family="ERV1"
     repeat_region   27227. .27361
                     /rpt_family="MIR"
     repeat_region   28278. .28567
                     /rpt_family="Alu"
     repeat_region   28661. .29094

  Query Match             5.1%;  Score 311.2;  DB 9;  Length 196832;
  Best Local Similarity   70.5%;  Pred. No. 4.4e-58;
  Matches  509;  Conservative    0;  Mismatches  188;  Indels   25;  Gaps    6;

Qy        317 GGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTCCTATA 376
              ||   | || ||||| ||| ||||||   ||||||||||||| || |    ||||||||
Db      25953 GGGTAGAAAGTGTGGTACGTTTTCTTGCCCATCATAAGGGTCACAGCTGGCACTCCTATC 26012

Qy        377 GTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTACATG 436
              |||||||||||||||| |||||  |||| |||||||||||||||||||||||||    |
Db      26013 ATAAAAGACAGGTTAACAAGAGAGAAACATAACAAATTTATTTAATCAAAGTTTTGTGCG 26072

Qy        437 ACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTTTTTTG 496
              | |  |||| |||||||||||||| |||||||| || | |||| |||||  |||| ||
Db      26073 ATACAGGAGCCTTCAGAAATGAAGATCCAAAGACCTAGAGAAAACCGTCTG-CTTTTATG 26131

Qy        497 CTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGATATGAT 556
              || || ||| | |||||||||| | |        ||    |||| |||||||||  | ||||
Db      26132 CTTAGTTTCAACGAAGAATGGACAACCATGTAGCAATGTGATT-GACAAAAGGGTCTGAT 26190

Qy        557 CTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTTCTTGA 616
              ||| || |||  ||||     | ||||| ||||||||| ||| | | ||||||||||| |
Db      26191 CTAATGATAATAGACT----GAGAAACCCAGCAAGGCTTGTATGTCCAGATTCTTCTCGG 26246

Qy        617 TCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTTCTTCA 676
              ||||||        || |||||||| | ||||||  |||| |||||| || | | | |
Db      26247 TCTCTC--------TGCATAGCATTCCTCCCTCCTAGGTATAGGGCAGAACCCCTGTGGA 26298

Qy        677 ATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATGGCTTGCTTCG 736
              |||||||||||||||||||||||||| | || |||| ||| |||||||  |||||||
Db      26299 ATGAGGGTCTTCAAGGGAGAAGGGAGAGAATGACCTTGCTAGGTTTTATGTCTTGCTTTA 26358

Qy        737 GGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCTGGTTTCTGTG 796
              ||| ||||| |||||||| ||| ||||||| ||||| | ||||||||||||||||||| ||
Db      26359 GGGGAGAGGGGTTCTAGTATCTGTGACCCAACTTGGAGGAGAGGAATTCTGGTTTCTATG 26418

Qy        797 ACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGGTCAGAAAGAG 856
              |||  ||||  | |||  ||||      ||    |||||||  |||| |||||||  ||||
Db      26419 ACTCACTTTAGGGGAGATGGAGGGACGGGAAAATGGAGGGCCAGAGAAGGTCAGAGAGAG 26478

Qy        857 ACTTGGCTGCT--------TCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAGTACTTCTTA 908
```

```
            |||| | | ||        |    | |||||  |||||||| |||| |||||   || |
Db      26479 ACTTTGTTTCTGAGGCTACTTCTGGACCTTCCAATCTCCTTTTGTTCAAAGTA---CTCA 26535

Qy        909 GCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAACACTGCAATCATGC 968
            |||| ||||||    ||||||||   |||  |||||||||| | |  |    ||
Db      26536 GCATGCCAAAGTGTCATACTTTGCAGTATCATTTTCTGAGCCCCAGAAAGAGAAAAGCTA 26595

Qy        969 TAAACTCCTCTATGACCTTCAAACATTCCACTTGCTTTTATTCTTTATGGTTGTGATGGC 1028
             ||   |   ||   ||  ||| ||  |        |||||  | ||   |||| |  |
Db      26596 AAATGGCTCATAATCCCACCAACCAGACATTAAAGGTTTATACATTCAAGTTGAGGAGAT 26655

Qy       1029 AT 1030
              ||
Db      26656 AT 26657
```

```
RESULT 11
AC009867
LOCUS       AC009867               176773 bp    DNA     linear   HTG 22-NOV-2000
DEFINITION  Homo sapiens chromosome 11 clone RP11-76K11 map 11, WORKING DRAFT
            SEQUENCE, 12 unordered pieces.
ACCESSION   AC009867
VERSION     AC009867.2  GI:11276242
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 176773)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 11, clone RP11-76K11
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 176773)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-SEP-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Nov 22, 2000 this sequence version replaced gi:5822691.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L2148
                Center clone name: 76_K_11
            -------------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 169932 bases at least Q40
                Consensus quality: 173382 bases at least Q30
                Consensus quality: 174685 bases at least Q20
                Insert size: 182000; agarose-fp
                Insert size: 175673; sum-of-contigs
                Quality coverage: 5.4 in Q20 bases; agarose-fp
                Quality coverage: 5.6 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 12 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *      1    18082: contig of 18082 bp in length
            *  18083    18182: gap of 100 bp
            *  18183    19240: contig of 1058 bp in length
            *  19241    19340: gap of 100 bp
            *  19341    69145: contig of 49805 bp in length
            *  69146    69245: gap of 100 bp
            *  69246    71797: contig of 2552 bp in length
            *  71798    71897: gap of 100 bp
            *  71898    76451: contig of 4554 bp in length
            *  76452    76551: gap of 100 bp
            *  76552    84099: contig of 7548 bp in length
            *  84100    84199: gap of 100 bp
            *  84200    94128: contig of 9929 bp in length
            *  94129    94228: gap of 100 bp
            *  94229   102601: contig of 8373 bp in length
            * 102602   102701: gap of 100 bp
            * 102702   115119: contig of 12418 bp in length
            * 115120   115219: gap of 100 bp
            * 115220   136142: contig of 20923 bp in length
            * 136143   136242: gap of 100 bp
            * 136243   175508: contig of 39266 bp in length
            * 175509   175608: gap of 100 bp
            * 175609   176773: contig of 1165 bp in length.
FEATURES             Location/Qualifiers
     source          1. .176773
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11"
                     /clone="RP11-76K11"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .18082
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    18183. .19240
                     /note="assembly_fragment"
     misc_feature    19341. .69145
                     /note="assembly_fragment"
     misc_feature    69246. .71797
                     /note="assembly_fragment"
     misc_feature    71898. .76451
                     /note="assembly_fragment"
     misc_feature    76552. .84099
                     /note="assembly_fragment"
     misc_feature    84200. .94128
                     /note="assembly_fragment"
     misc_feature    94229. .102601
                     /note="assembly_fragment"
     misc_feature    102702. .115119
                     /note="assembly_fragment"
     misc_feature    115220. .136142
                     /note="assembly_fragment"
     misc_feature    136243. .175508
                     /note="assembly_fragment"
     misc_feature    175609. .176773
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
ORIGIN

  Query Match             5.0%;  Score 307.2;  DB 2;  Length 176773;
```

```
  Best Local Similarity   71.5%;  Pred. No. 3.4e-57;
  Matches  495;  Conservative    0;  Mismatches  173;  Indels   24;  Gaps    6;

Qy        293 CTTCTCTCTCAAAGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATA 352
              |||||  |  |    |  |  ||||||     || ||| ||||| |||  ||  |||| |||
Db     104797 CTTCTGTACCTTGGAGGGTGTAGGGACTGGAAAGTGGGATACCTTTCTTCACCCATGATA 104856

Qy        353 AGGGTCGCAACCAAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAA 412
              |||  | || |||  ||||| |||  |||||| ||||||| ||||| ||| | |||||||
Db     104857 AGGACCACAGCCATTACTCC-ATAACAAAAGAGAGGTTAAAAAGAGAAAAGCATAACAAA 104915

Qy        413 TTTATTTAATCAAAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCC 472
              | |||||||||| |||||||  || |||||||| |||||||||||||||| |||||||||
Db     104916 TGTATTTAATCACAGTTTTATGTGTCATGGGAGCCTTCAGAAATGAAGACTCAAAGACCC 104975

Qy        473 AGGGGAAACTGTCTGTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCAT 532
              || | |||||||     ||||| |||| |||||  |||||||||||| | |         |
Db     104976 AGAGAAAACTGT-GCATTTTTATGCTTAGGTTTGATGAAGAATGGACAACCATGTAGAAA 105034

Qy        533 GTAGATTAGACAAAAGGATATGATCTAGTGGTAAAGGACTCA---GGGGGAAACACAGCA 589
                |||| | ||||||||||||||||||    | |   |||| |   || |||   | |
Db     105035 TGTGATTGGGCAAAAGGATATGATCTAACAGAAGTAGACTAAGAGGGAGGAGGGAAACTC 105094

Qy        590 AGGCCTGTCTATTCAGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTC 649
              ||      || | |||||| ||||||  |||||         ||  |||||||| || |||
Db     105095 AGCAAGAGCTGTCCAGATTTTTCTTGGCCTCTC--------TGGGTAGCATTCGTTTCTC 105146

Qy        650 CTGAGTATGGGGCAGGACTCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGG 709
                | |||||||||||| | |||  | ||| |||||| ||||||||||||| | ||||
Db     105147 TGTGGCATGGGGCAGGACCCCTCTGGATTGATGGTCTTTAAGGGAGAAGGGACAGAGTGC 105206

Qy        710 CCTTTTTAGATTTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCT 769
               ||||  |  |||||| |||||||| |  |||||||||||||||||||| ||||||||||
Db     105207 TCTTTCCATCTTTTATAGCTTGCTTTGCAGAAGAGGAGTTCTAGTTTCCATGACCCATCT 105266

Qy        770 TGGGGAAGAGGAATTCTGGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGC 829
              || |||||||||||||| ||||  ||| |||||||     |||||||| ||||  |||||
Db     105267 TGAGGAAGAGGAATTCTGATTTCAATGATTTGCTTTGGAAGAGAAAGAAGAGTGGGAGGC 105326

Qy        830 AGGAGGGCAGGAGATGGTCAGAAAGAGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTT 889
              ||||||||||  | | |||||||||         |||||||| |||||  |||||||
Db     105327 AGGAGGGCAGGCAAAGCTCAGAAAGA--------CGCTTCTGAGGCCTTCCAATCTCCTT 105378

Qy        890 TAGTTCCAAGTACTTCTTAGCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGC 949
              |||||||||| |   || ||||| ||||||  | |||||||  |   ||  ||||||  |||
Db     105379 TAGTTCCAAGAA---CTCAGCATGCCAAAGTGCCATACTTTAGGGTACTGTTTTCCAAGC 105435

Qy        950 TCTAACACTGCAATCATGCTAAACTCCTCTAT 981
                |||||  | ||  | |   |  | | |||
Db     105436 CTTAACAAGGGAAAGAAGTACATTTTGTTTAT 105467


RESULT 12
AP003059
LOCUS       AP003059               89000 bp    DNA     linear   PRI 15-MAR-2003
DEFINITION  Homo sapiens genomic DNA, chromosome 11q, clone:CTD-2005H7,
            complete sequence.
ACCESSION   AP003059
VERSION     AP003059.2  GI:17425250
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens genomic DNA
  JOURNAL   Published Only in Database (2000)
REFERENCE   2  (bases 1 to 89000)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-DEC-2000) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     On Dec 7, 2001 this sequence version replaced gi:11994966.
FEATURES             Location/Qualifiers
     source          1. .89000
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q"
                     /clone="CTD-2005H7"
ORIGIN

  Query Match             5.0%;  Score 305.6;  DB 9;  Length 89000;
  Best Local Similarity   71.4%;  Pred. No. 8.1e-57;
  Matches  494;  Conservative    0;  Mismatches  174;  Indels   24;  Gaps    6;

Qy        293 CTTCTCTCTCAAAGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATA 352
              |||||  |  |    |  |  ||||||     || ||| ||||| |||  ||  |||| |||
Db      28212 CTTCTGTACCTTGGAGGGTGTAGGGACTGGAAAGTGGGATACCTTTCTTCACCCATGATA 28271

Qy        353 AGGGTCGCAACCAAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAA 412
              |||  | || |||  ||||| |||  |||||| ||||||| ||||| ||| | |||||||
Db      28272 AGGACCACAGCCATTACTCC-ATAACAAAAGAGAGGTTAAAAAGAGAAAAGCATAACAAA 28330

Qy        413 TTTATTTAATCAAAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCC 472
              | |||||||||| |||||||  || |||||||| |||||||||||||||| |||||||||
Db      28331 TGTATTTAATCACAGTTTTATGTGTCATGGGAGCCTTCAGAAATGAAGACTCAAAGACCC 28390

Qy        473 AGGGGAAACTGTCTGTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCAT 532
              || | |||||||     ||||| |||| |||||  |||||||||||| | |         |
Db      28391 AGAGAAAACTGT-GCATTTTTATGCTTAGGTTTGATGAAGAATGGACAACCATGTAGAAA 28449

Qy        533 GTAGATTAGACAAAAGGATATGATCTAGTGGTAAAGGACTCA---GGGGGAAACACAGCA 589
                |||| | ||||||||||||||||||    | |   |||| |   || |||   | |
Db      28450 TGTGATTGGGCAAAAGGATATGATCTAACAGAAGTAGACTAAGAGGGAGGAGGGAAACTC 28509

Qy        590 AGGCCTGTCTATTCAGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTC 649
              ||      || | |||||| ||||||  |||||         ||  |||||||| || |||
Db      28510 AGCAAGAGCTGTCCAGATTTTTCTTGGCCTCTC--------TGGGTAGCATTCGTTTCTC 28561

Qy        650 CTGAGTATGGGGCAGGACTCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGG 709
                | |||||||||||| | |||  | ||| |||||| ||||||||||||| | ||||
Db      28562 TGTGGCATGGGGCAGGACCCCTCTGGATTGATGGTCTTTAAGGGAGAAGGGACAGAGTGC 28621

Qy        710 CCTTTTTAGATTTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCT 769
               ||||  |  |||||| |||||||| |  |||||||||||||||||||| ||||||||||
Db      28622 TCTTTCCATCTTTTATAGCTTGCTTTGCAGAAGAGGAGTTCTAGTTTCCATGACCCATCT 28681

Qy        770 TGGGGAAGAGGAATTCTGGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGC 829
              || |||||||||| |||| ||||  ||| |||||||     ||||||| ||||  |||||
Db      28682 TGAGGAAGAGGAAATCTGATTTCAATGATTTGCTTTGGAAGAGAAAGAAGAGTGGGAGGC 28741

Qy        830 AGGAGGGCAGGAGATGGTCAGAAAGAGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTT 889
              ||||||||||  | | |||||||||         |||||||| |||||  |||||||
Db      28742 AGGAGGGCAGGCAAAGCTCAGAAAGA--------CGCTTCTGAGGCCTTCCAATCTCCTT 28793

Qy        890 TAGTTCCAAGTACTTCTTAGCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGC 949
              |||||||||| |   || ||||| ||||||  | |||||||  |   ||  ||||||  |||
Db      28794 TAGTTCCAAGAA---CTCAGCATGCCAAAGTGCCATACTTTAGGGTACTGTTTTCCAAGC 28850

Qy        950 TCTAACACTGCAATCATGCTAAACTCCTCTAT 981
                |||||  | ||  | |   |  | | |||
Db      28851 CTTAACAAGGGAAAGAAGTACATTTTGTTTAT 28882


RESULT 13
AL513284/c
LOCUS       AL513284               169972 bp    DNA     linear   PRI 11-JUN-2002
DEFINITION  Human DNA sequence from clone RP11-518D3 on chromosome 1, complete
```

```
            sequence.
ACCESSION   AL513284
VERSION     AL513284.12  GI:21425224
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 169972)
  AUTHORS   White,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUN-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jun 13, 2002 this sequence version replaced gi:18873617.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            RP11-518D3 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6.
FEATURES             Location/Qualifiers
     source          1. .169972
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-518D3"
                     /clone_lib="RPCI-11.2"
ORIGIN

  Query Match             4.9%;  Score 299.4;  DB 9;  Length 169972;
  Best Local Similarity   69.8%;  Pred. No. 1.8e-55;
  Matches  463;  Conservative    0;  Mismatches  191;  Indels    9;  Gaps    4;

Qy        301 TCAAAGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGC 360
              || |  || | ||||  | | ||| |||| |||  ||||||   |||||||||||| ||
Db     168548 TCTACCCTCCTGTAGAAGAGGAAAGGTGTAATAACTTTTCTTACCCATCATAAGGATCAT 168489

Qy        361 AACCAAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTA 420
                |||| |||||||||  ||||| || |  || ||||| ||| | | |||||| ||||||
Db     168488 GGCCAACACTCCTATAACAAAAGGCAAGAGAACAAGAGAAAAGCATGACAAATGTATTTA 168429

Qy        421 ATCAAAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAA 480
              | ||||||||||  |||||||||| ||| |||||| |||| |||||||||||||| |||
Db     168428 ACCAAAGTTTTATGTGACATGGGAGCCTTGAGAAATAAAGAGCCAAAGACCCAGGGAAAA 168369

Qy        481 CTGTCTGTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTA 540
              |||||| ||||| |||| ||||| |||||||||  || |||        |    |||||
Db     168368 TTGTCTG-TTTTTATGCTTAGGTTTGATGAAGAACTGACAGCCATATAGAAATGTGATTA 168310

Qy        541 GACAAAAGGATATGATCTAGTGGTAAAGGACTCA-GGGGGAAACACAGCAAGGCCTGTCT 599
              ||||||| | |||||||||| || |||  |||| |  || ||||| |||| |||||||| ||

Db     168309 GACAAAAAGGTATGATCTAATGCTAACAGACTGACAGGAGAAACCCAGCCAGGCCTGCCT 168250

Qy        600 ATTCAGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGG 659
                |    |  |          ||||   |||  | | || ||||| |||||||||| |||||
Db     168249 TT----CTGTTCAGATTCTTCTCGGCCTCCCTATGTAACATTCCTTCCTCCTGGGTATGA 168194

Qy        660 GGCAGGACTCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGA 719
              ||| |||||| |||  ||||||  |||||||  ||||||    | |||||  |||  |||
Db     168193 GGCTGGACTCCTCTGGAATGAGATTCTTCAAGACAGAAGGAGAAGAGTGGACTTCCTAGG 168134

Qy        720 TTTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAG 779
              ||||||||| ||||| | |||||||||||||| |||||||||||||| ||||||||||||
Db     168133 TTTTATGGCATGCTTGGAGGAAGAGGAGTTCTGGTTTCTATGACCCACCTTGGGGAAGAG 168074

Qy        780 GAATTCTGGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAG 839
              || |||||||||| |||||   || | |      ||||  |  |    | ||   || |
Db     168073 GAGTTCTGGTTTCTATGACTCATTTACTTAGGCGAAGAAAAAGAGACAGGAGATAGGAGG 168014

Qy        840 GAGATGGTCAGAAAGAGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAG 899
              |          | |||||          |     |  |||||  | ||||| ||||| |||
Db     168013 GTAGGAAGAGGTCAGAGAAACCCTGTTTCTGAGGTCCTTCCAATGTCCTTCAGTTCAAAG 167954

Qy        900 TACTTCTTAGCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAACACTG 959
              ||   || ||||| |||| |  | |||||||||  |||    |||||||| | ||||||
Db     167953 TA---CTCAGCATGCCAAGGTGCCATACTTTGGGGTATCAGATTCTGAGCCCCAACACTC 167897

Qy        960 CAA 962
              |||
Db     167896 CTA 167894


RESULT 14
AL591491
LOCUS       AL591491              121720 bp    DNA     linear   PRI 06-DEC-2001
DEFINITION  Human DNA sequence from clone RP11-435K1 on chromosome Xp15.3-15.5
            Contains the 5' end of a novel gene, complete sequence.
ACCESSION   AL591491
VERSION     AL591491.5  GI:15149577
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 121720)
  AUTHORS   Lawlor,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-DEC-2001) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 13, 2001 this sequence version replaced gi:15020561.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome X, constructed by the Sanger Centre Chromosome X Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/ChrX
            RP11-435K1 is from the library RPCI-11.2 constructed by the group
```

```
          of Pieter de Jong. For further details see
          http://www.chori.org/bacpac/home.htm
          VECTOR: pBACe3.6
          IMPORTANT: This sequence is not the entire insert of clone
          RP11-435K1 It may be shorter because we sequence overlapping
          sections only once, except for a short overlap.
          The true left end of clone RP11-435K1 is at 1 in this sequence. The
          true left end of clone RP11-342D14 is at 119721 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .121720
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="p15.3-15.5"
                     /clone="RP11-435K1"
                     /clone_lib="RPCI-11.2"
     repeat_region   17. .90
                     /note="MIR repeat: matches 28. .111 of consensus"
     repeat_region   443. .559
                     /note="MER5A repeat: matches 59. .189 of consensus"
     misc_feature    1986. .2375
                     /note="match: GSS: Em:AQ725819"
     misc_feature    2022. .2428
                     /note="match: GSS: Em:AQ621701"
     repeat_region   2544. .2635
                     /note="46 copies 2 mer tt 73% conserved"
     repeat_region   2638. .2932
                     /note="AluJo repeat: matches 1. .312 of consensus"
     repeat_region   3343. .3705
                     /note="L2 repeat: matches 2370. .2746 of consensus"
     repeat_region   4022. .4217
                     /note="MIR repeat: matches 8. .208 of consensus"
     repeat_region   4624. .4797
                     /note="87 copies 2 mer ta 69% conserved"
     misc_feature    5082. .5534
                     /note="match: GSS: Em:AQ442500"
     misc_feature    5095. .5533
                     /note="match: GSS: Em:AQ442489"
     repeat_region   6039. .6101
                     /note="MSTB repeat: matches 1. .64 of consensus"
     repeat_region   6102. .6395
                     /note="AluJo repeat: matches 25. .309 of consensus"
     repeat_region   6396. .6768
                     /note="MSTB repeat: matches 64. .426 of consensus"
     repeat_region   6979. .7100
                     /note="L2 repeat: matches 2575. .2709 of consensus"
     repeat_region   8099. .8302
                     /note="MER20 repeat: matches 1. .216 of consensus"
     repeat_region   8324. .8823
                     /note="L1ME repeat: matches 943. .1433 of consensus"
     repeat_region   8829. .9032
                     /note="L1ME3 repeat: matches 3815. .4031 of consensus"
     repeat_region   9478. .9746
                     /note="L1ME3 repeat: matches 4058. .4325 of consensus"
     repeat_region   9747. .10084
                     /note="MER44A repeat: matches 1. .332 of consensus"
     repeat_region   10085. .10943
                     /note="L1ME3 repeat: matches 4325. .5188 of consensus"
     repeat_region   10947. .11151
                     /note="AluJo repeat: matches 86. .297 of consensus"
     repeat_region   11153. .11917
                     /note="L1ME3 repeat: matches 5182. .6002 of consensus"
     repeat_region   11918. .12011
                     /note="MER81 repeat: matches 2. .114 of consensus"
     repeat_region   12012. .12175
                     /note="L1ME3 repeat: matches 6002. .6156 of consensus"
     repeat_region   14703. .15013
                     /note="AluY repeat: matches 1. .311 of consensus"
     repeat_region   15834. .15923
                     /note="L2 repeat: matches 2574. .2666 of consensus"
     repeat_region   16531. .16566
                     /note="18 copies 2 mer ca 100% conserved"
     repeat_region   17622. .18127
                     /note="L1MB5 repeat: matches 5679. .6174 of consensus"
     repeat_region   18145. .18248
                     /note="FLAM_C repeat: matches 5. .108 of consensus"
     repeat_region   18392. .18695
                     /note="AluSq repeat: matches 1. .305 of consensus"
     repeat_region   19551. .19659
                     /note="L2 repeat: matches 2549. .2662 of consensus"
     gene            complement(20811. .29961)
                     /gene="bA435K1.1"
     mRNA            complement(join(<20811. .21075,29654. .29961))
                     /gene="bA435K1.1"
                     /product="bA435K1.1 (novel protein)"
                     /note="match:, cDNAs: Em:AK024165 Em:AK016298
                     match: ESTs: Em:AA420823 Em:AA455585 Em:AI796127
                     Em:AU121832"
                     /evidence=not_experimental
     CDS             complement(join(<20811. .21075,29654. .29886))
                     /gene="bA435K1.1"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="bA435K1.1 (novel protein)"
                     /protein_id="CAD13520.1"
                     /db_xref="GI:17426506"
                     /db_xref="UniProt/TrEMBL:Q8WX00"
                     /translation="MEPQLGPEAAALRPGWLALLLWVSALSCSFSLPASSLSSLVPQV
                     RTSYNFGRTFLGLDKCNACIGTSICKKFFKEEIRSDNWLASHLGLPPDSLLSYPANYS
                     DDSKIWRPVEIFRLVSKYQNEISDRRICASASAPKTCSIERVLRKTERFQKWLQAKRL
                     TPDLVQ"
     repeat_region   21413. .21515
                     /note="MIR repeat: matches 42. .147 of consensus"
     repeat_region   21840. .22137
                     /note="AluSx repeat: matches 2. .299 of consensus"
     repeat_region   23050. .23260
                     /note="MIR repeat: matches 47. .262 of consensus"
     repeat_region   23316. .23609
                     /note="AluSx repeat: matches 1. .294 of consensus"
     repeat_region   24752. .25050
                     /note="L2 repeat: matches 2245. .2543 of consensus"
     misc_feature    complement(25215. .25383)
                     /gene="bA435K1.1"
                     /note="match: GSS: Em:AL013940"
     misc_feature    complement(25215. .25382)
                     /gene="bA435K1.1"
                     /note="match: GSS: Em:AL013938"
     repeat_region   26303. .26595
                     /note="AluSg repeat: matches 1. .293 of consensus"
     repeat_region   26651. .26750
                     /note="L1M4 repeat: matches 5302. .5399 of consensus"
     repeat_region   29031. .29118
                     /note="MIR repeat: matches 25. .114 of consensus"
     repeat_region   31059. .31310
                     /note="L1M4 repeat: matches 5099. .5385 of consensus"
     repeat_region   31315. .31499
                     /note="MER93 repeat: matches 205. .397 of consensus"
     repeat_region   31565. .31643
                     /note="MER93 repeat: matches 3. .82 of consensus"
     repeat_region   31649. .31998
                     /note="MER57-internal repeat: matches 6808. .7145 of
                     consensus"
     repeat_region   31998. .32402
                     /note="MER65-internal repeat: matches 3648. .4734 of
                     consensus"
     repeat_region   32415. .33449
                     /note="MER31-internal repeat: matches 1721. .2860 of
                     consensus"
     repeat_region   33450. .33759
                     /note="AluSx repeat: matches 1. .309 of consensus"
     repeat_region   33888. .35557
                     /note="L1MA2 repeat: matches 4674. .6308 of consensus"
     repeat_region   35554. .37204
                     /note="L1M1 repeat: matches 830. .2496 of consensus"
     repeat_region   37217. .37364
```

```
                     /note="L1M1 repeat: matches -1264. .-1116 of consensus"
     repeat_region   37382. .39552
                     /note="L1M1 repeat: matches -1389. .828 of consensus"
     repeat_region   39638. .39764
                     /note="MER34 repeat: matches 418. .543 of consensus"
     repeat_region   40368. .40646
                     /note="MER31-internal repeat: matches 1117. .1393 of
                     consensus"
     repeat_region   41671. .41824
                     /note="MER93 repeat: matches 248. .394 of consensus"
     misc_feature    41957. .42120
                     /note="match: GSS: Em:AQ170009"
     repeat_region   42164. .42825
                     /note="L1M4 repeat: matches 4435. .5094 of consensus"

  Query Match             4.9%;  Score 297.6;  DB 9;  Length 121720;
  Best Local Similarity   68.6%;  Pred. No. 4.6e-55;
  Matches  495;  Conservative    0;  Mismatches  199;  Indels   28;  Gaps    5;

Qy        313 TAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTCC 372
              || ||  |  ||| ||||| | | ||| ||   ||| ||| ||||||    || | || ||
Db     102399 TATGGATAGAAAGGTGTGGTGCCTTTCCTTACCCAGCATTAGGGTCATGGCCGACACCCC 102458

Qy        373 TATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTA 432
              ||||  ||  || |||||  ||||| ||| | ||||||||||||| ||||||||||||
Db     102459 TATAACAAGGGAAAGGTTAGCAAGAGAAAAGCATAACAAATTTATTTGATCAAAGTTTTA 102518

Qy        433 CATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTTT 492
              |||||  ||||| |||||||||||||||||||||| ||||||  |||    |  ||||
Db     102519 TGTGACACGGGAGCCTTCAGAAATGAAGACCCAAAGGCCCAGGTCAAA--TGATTATTTT 102576

Qy        493 TTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGATA 552
              | |||| ||||| |||||||||||   ||      |  |    |||| |||||||| |
Db     102577 TGTGCTTAGGTTTGATGAAGAATGGGCAGGTATATGGAAATGGGATTGGACAAAAGGGTG 102636

Qy        553 TGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTTC 612
              ||||||  ||||||   ||| |||| ||||| || ||||  |||||| ||||||||| |
Db     102637 TGATCTGTTGGTAATAAACTGAGGGAGAAACCCAACAAGTGCTGTCTGTTCAGATTCTCC 102696

Qy        613 TTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTTC 672
              |||   ||||       ||| ||||||| |||||||  ||||||||||||    |||
Db     102697 TTGGCATCTC--------TGTGTAGCATTCCTTCCTCCTAGGTATGGGGCAGGATCTTTC 102748

Qy        673 TTCAATGAGGGTCTTCAAGGGAGAAGGGAGAA--------AGTGGCCTTTTTAGATTTTA 724
              |  ||||||||||||||||||||||||||||        ||||  ||||  |  |||||
Db     102749 TGGAATGAGGGTCTTCAAGGGAGAAGGGAGAAGACTAGAGAGTGATCTTTCCAAGTTTTA 102808

Qy        725 TGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATT 784
              | |||||||| | ||||| |||||  |||||||||||||  |   | || ||||||||||
Db     102809 TAGCTTGCTTTGCGGAAGTGGAGTCTTAGTTTCTATGACTCGCCATGAGGAAGAGGAATT 102868

Qy        785 CTGGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGAT 844
              ||||||||| ||||||||||   | ||| | || || | ||| ||||||||||||||
Db     102869 CTGGTTTCTATGACTTGCTTT--GGGAGAGAAAGAAGGATGAGACAGGAGGGCAGGAGGA 102926

Qy        845 GGTCAGAAAGAGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAGTACTT 904
              ||||||| |||||| |  || ||   ||         ||| | |        | |||
Db     102927 GGTCAGAGAGAGACATTACTTCTGAGGA--------CTTCTAATCTGTCAGTTCGAACTA 102978

Qy        905 CTTAGCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAACACTGCAATC 964
              || ||||||| || |  |  || |||||  ||| || |||| ||| | |    |  | |
Db     102979 CTCAGCATACTAAGGTGCCGTATTTTGGGGTATTGTGTTCTAAGCCCCAGTGTTATATTA 103038

Qy        965 ATGCTAAACTCCTCTATGACCTTCAAACATTCCACTTGCTTTTATTCTTTATGGTTGTGA 1024
              |||  ||    | | |  |   |||  |    |   ||| |      | |||||
Db     103039 TGGCTGAAGATTTTTTTCTCTGCAAAAATATGTTTATCTATTTTTATAAATCAGATGTGA 103098

Qy       1025 TG 1026
              ||
Db     103099 TG 103100


RESULT 15
```

```
AC010244/c
LOCUS       AC010244              134937 bp    DNA     linear   PRI 15-AUG-2001
DEFINITION  Homo sapiens chromosome 5 clone CTC-364B20, complete sequence.
ACCESSION   AC010244
VERSION     AC010244.6  GI:15187208
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 134937)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 134937)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 134937)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JAN-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
REFERENCE   4  (bases 1 to 134937)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-AUG-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Aug 15, 2001 this sequence version replaced gi:12545284.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 99.6% of Sequence;
            Estimated Total Number of Errors is 0.4.
            STS Content:
            SHGC-79386 G52178.
FEATURES             Location/Qualifiers
     source          1. .134937
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTC-364B20"
ORIGIN

  Query Match             4.8%;  Score 293.8;  DB 9;  Length 134937;
  Best Local Similarity   72.8%;  Pred. No. 3.1e-54;
  Matches  481;  Conservative    0;  Mismatches  157;  Indels   23;  Gaps    7;

Qy        302 CAAAGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCA 361
              | |  | || |||  ||| |||| |||||||  |||||||  | || ||||| ||| ||
Db      22463 CTATCCAGCTGTACAGGTAAAAAGGTGTGATATGTTTTCTCACCTATAATAAGAGTCACA 22404

Qy        362 ACCAAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAA 421
              |  | | |||||||  ||||||  |||||||||||||| ||| | |||||  |||||||||
Db      22403 GCTGACATTCCTATAACAAAAGATAGGTTAATAAGAGGAAAGCATAACACTTTTATTTAA 22344

Qy        422 TCAAAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAAC 481
              ||||||| |||| ||| |   ||| |||||||||||||||||||| ||||||| | ||||
Db      22343 TCAAAGTGTTACTTGATACAAGAGCCTTCAGAAATGAAGACCCAAGGACCCAGAGAAAAC 22284

Qy        482 TGTCTGTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAG 541
              || |  |||||| |||| | |||   |||||||||| |||  |  |  |   |||| |
Db      22283 TG-CCCTTTTTTATGCTTATGTTTAGTGAAGAATGGACAGCCAGGTGGAAATGTGATTGG 22225

Qy        542 ACAAAAGGATATGATCTAGTGGTAAAGGACTCAGGGGG-AAACACAGCAAGGCCTGTCTA 600
              |||||||||||| || ||  ||| |  || | |||||| |||| ||||||||||||  ||
Db      22224 ACAAAAGGATATAATTTAAAGGTGAGAGAATGAGGGGGAAAACCCAGCAAGGCCTGCCTG 22165

Qy        601 TTCAGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGG 660
              ||||||||||||||||  ||| ||| |        | ||||| ||| | ||| |||| ||
```

```
Db      22164 TTCAGATTCTTCTTGACTTCTTTCTGT--------AACATTCCTTCTTTCTGGGTATAGG 22113

Qy        661 GCAGGACTCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGAT 720
              |||||| |  |||  ||||||||||||| || ||||||| |||||   |    |||| |
Db      22112 GCAGGATTACTCTGGAATGAGGGTCTTGAAAGGAGAAGAGAGAA---GCGACTTTTGGC 22056

Qy        721 TTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGG 780
              |||||||||  |||||||| |||| |||||| |||||||||  |    ||||  ||||||
Db      22055 TTTATGGCTGACTTCGGGGGAGAGGGGTTCTAATTTCTATGAACTGCCTTGCAGAAGAGT 21996

Qy        781 AATTCTGGTTTCTGTGACTTGCTT-TCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAG 839
              |||| ||||||  |||| ||||||     | |||||| |  |  ||| ||| ||  |||
Db      21995 AATTTTGGTTTTGGTGAATTGCTTCAGGGGGAGAAAGATGGTTGGGAGACAGAAGAACAG 21936

Qy        840 GAGATGGTCAGAAAGAGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAG 899
              |||| | ||||| |||        ||||| ||||| |||||  |||||||  ||
Db      21935 GAGAAGATCAGACAGA-----CTTTGCTTTTGAGGCCTTCCAATCTCCTTAGGT----CA 21885

Qy        900 TACTTCTTAGCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAACACTG 959
              | |  | ||||||||||||| | | | ||||||  ||| || |||||||| | ||||| |
Db      21884 AAGTAATCAGCATACCAAAGTAATGTTCTTTGGGCTATCGTATTCTGAGCCCCAACACAG 21825

Qy        960 C 960
              |
Db      21824 C 21824
```

```
RESULT 16
AC091991
LOCUS       AC091991              170219 bp    DNA     linear   PRI 30-JUL-2002
DEFINITION  Homo sapiens chromosome 5 clone RP11-86C20, complete sequence.
ACCESSION   AC091991
VERSION     AC091991.2  GI:22004175
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 170219)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 170219)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUN-2001) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 170219)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JUL-2002) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Jul 30, 2002 this sequence version replaced gi:14333927.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 100% of Sequence;
            Estimated Total Number of Errors is 0.1.
FEATURES             Location/Qualifiers
     source          1. .170219
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="RP11-86C20"
ORIGIN

  Query Match             4.8%;  Score 293.8;  DB 9;  Length 170219;
  Best Local Similarity   72.8%;  Pred. No. 3.1e-54;
  Matches  481;  Conservative    0;  Mismatches  157;  Indels   23;  Gaps    7;

Qy        302 CAAAGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCA 361
```

```
              | |  | ||  |||  ||| |||| ||||||||  |||||||  |  || |||||  ||| ||
Db     121792 CTATCCAGCTGTACAGGTAAAAAGGTGTGATATGTTTTCTCACCTATAATAAGAGTCACA 121851

Qy        362 ACCAAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAA 421
              |  | | |||||||  ||||||  |||||||||||||| ||| | |||||  |||||||||
Db     121852 GCTGACATTCCTATAACAAAAGATAGGTTAATAAGAGGAAAGCATAACACTTTTATTTAA 121911

Qy        422 TCAAAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAAC 481
              ||||||| |||| ||| |   ||| ||||||||||||||||||||| ||||||| | ||||
Db     121912 TCAAAGTGTTACTTGATACAAGAGCCTTCAGAAATGAAGACCCAAGGACCCAGAGAAAAC 121971

Qy        482 TGTCTGTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAG 541
              || |  |||||| |||| | |||   |||||||||||| |||  |  |  |    |||| |
Db     121972 TG-CCCTTTTTTATGCTTATGTTTAGTGAAGAATGGACAGCCAGGTGGAAATGTGATTGG 122030

Qy        542 ACAAAAGGATATGATCTAGTGGTAAAGGACTCAGGGGG-AAACACAGCAAGGCCTGTCTA 600
              |||||||||||| || ||  ||| |  || | |||||| |||| ||||||||||| ||
Db     122031 ACAAAAGGATATAATTTAAAGGTGAGAGAATGAGGGGGAAAACCCAGCAAGGCCTGCCTG 122090

Qy        601 TTCAGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGG 660
              |||||||||||||||  ||| ||| |       | ||||| ||| | ||| |||| ||
Db     122091 TTCAGATTCTTCTTGACTTCTTTCTGT--------AACATTCCTTCTTTCTGGGTATAGG 122142

Qy        661 GCAGGACTCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGAT 720
              |||||| |  |||  |||||||||||| || ||||||| |||||   |    |||| |
Db     122143 GCAGGATTACTCTGGAATGAGGGTCTTGAAAGGAGAAGAGAGAA---GCGACTTTTGGC 122199

Qy        721 TTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGG 780
              |||||||||  |||||||| |||| |||||| |||||||||  |    ||||  ||||||
Db     122200 TTTATGGCTGACTTCGGGGGAGAGGGGTTCTAATTTCTATGAACTGCCTTGCAGAAGAGT 122259

Qy        781 AATTCTGGTTTCTGTGACTTGCTT-TCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAG 839
              |||| ||||||  |||| ||||||     | |||||| |  |  ||| ||| ||  |||
Db     122260 AATTTTGGTTTTGGTGAATTGCTTCAGGGGGAGAAAGATGGTTGGGAGACAGAAGAACAG 122319

Qy        840 GAGATGGTCAGAAAGAGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAG 899
              |||| | ||||| |||        ||||| ||||| |||||  |||||||  ||
Db     122320 GAGAAGATCAGACAGA-----CTTTGCTTTTGAGGCCTTCCAATCTCCTTAGGT----CA 122370

Qy        900 TACTTCTTAGCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAACACTG 959
              | |  | ||||||||||||| | | | ||||||  ||| || |||||||| | ||||| |
Db     122371 AAGTAATCAGCATACCAAAGTAATGTTCTTTGGGCTATCGTATTCTGAGCCCCAACACAG 122430

Qy        960 C 960
              |
Db     122431 C 122431
```

```
RESULT 17
AC021239/c
LOCUS       AC021239              182049 bp    DNA     linear   HTG 24-AUG-2002
DEFINITION  Homo sapiens chromosome 5 clone RP11-297I14 map 5, WORKING DRAFT
            SEQUENCE, 9 unordered pieces.
ACCESSION   AC021239
VERSION     AC021239.3  GI:7341773
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 182049)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 5, clone RP11-297I14
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 182049)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
```

```
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 182049)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 30, 2000 this sequence version replaced gi:6940111.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L2368
                Center clone name: 297_I_14
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 174800 bases at least Q40
                Consensus quality: 178733 bases at least Q30
                Consensus quality: 179947 bases at least Q20
                Insert size: 182000; agarose-fp
                Insert size: 181249; sum-of-contigs
                Quality coverage: 4.8 in Q20 bases; agarose-fp
                Quality coverage: 4.8 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     4146: contig of 4146 bp in length
            *     4147     4246: gap of 100 bp
            *     4247     9349: contig of 5103 bp in length
            *     9350     9449: gap of 100 bp
            *     9450    19658: contig of 10209 bp in length
            *    19659    19758: gap of 100 bp
            *    19759    27772: contig of 8014 bp in length
            *    27773    27872: gap of 100 bp
            *    27873    47101: contig of 19229 bp in length
            *    47102    47201: gap of 100 bp
            *    47202    74998: contig of 27797 bp in length
            *    74999    75098: gap of 100 bp
            *    75099   100337: contig of 25239 bp in length
            *   100338   100437: gap of 100 bp
            *   100438   143242: contig of 42805 bp in length
            *   143243   143342: gap of 100 bp
            *   143343   182049: contig of 38707 bp in length.
FEATURES             Location/Qualifiers
     source          1. .182049
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /map="5"
                     /clone="RP11-297I14"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .4146
                     /note="assembly_fragment"
     misc_feature    4247. .9349
                     /note="assembly_fragment"
     misc_feature    9450. .19658
                     /note="assembly_fragment"
     misc_feature    19759. .27772
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    27873. .47101
                     /note="assembly_fragment"
     misc_feature    47202. .74998
                     /note="assembly_fragment"
     misc_feature    75099. .100337
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     misc_feature    100438. .143242
                     /note="assembly_fragment"
     misc_feature    143343. .182049
                     /note="assembly_fragment"
ORIGIN

  Query Match             4.8%;  Score 293.8;  DB 2;  Length 182049;
  Best Local Similarity   72.8%;  Pred. No. 3e-54;
  Matches  481;  Conservative    0;  Mismatches  157;  Indels   23;  Gaps    7;

Qy        302 CAAAGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCA 361
              | |  | || |||  ||| |||| ||||||||  |||||||  | || |||| ||| ||
Db     161391 CTATCCAGCTGTACAGGTAAAAAGGTGTGATATGTTTTCTCACCTATAATAAGAGTCACA 161332

Qy        362 ACCAAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAA 421
              | | | |||||||  |||||| ||||||||||||| ||| | |||||  |||||||||
Db     161331 GCTGACATTCCTATAACAAAAGATAGGTTAATAAGAGGAAAGCATAACACTTTTATTTAA 161272

Qy        422 TCAAAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAAC 481
              ||||||| |||| ||| |   ||| ||||||||||||||||||| ||||||| | ||||
Db     161271 TCAAAGTGTTACTTGATACAAGAGCCTTCAGAAATGAAGACCCAAGGACCCAGAGAAAAC 161212

Qy        482 TGTCTGTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAG 541
              || |  |||||| |||| | |||   |||||||||| |||  |  | |    |||| |
Db     161211 TG-CCCTTTTTTATGCTTATGTTTAGTGAAGAATGGACAGCCAGGTGGAAATGTGATTGG 161153

Qy        542 ACAAAAGGATATGATCTAGTGGTAAAGGACTCAGGGGG-AAACACAGCAAGGCCTGTCTA 600
              |||||||||||| || ||  ||| |  || | |||||| |||| ||||||||||| ||
Db     161152 ACAAAAGGATATAATTTAAAGGTGAGAGAATGAGGGGGAAAACCCAGCAAGGCCTGCCTG 161093

Qy        601 TTCAGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGG 660
              |||||||||||||||  ||| ||| |        | ||||| ||| | ||| |||| ||
Db     161092 TTCAGATTCTTCTTGACTTCTTTCTGT--------AACATTCCTTCTTTCTGGGTATAGG 161041
```

```
Qy        661 GCAGGACTCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGAT 720
              |||||| |  |||  ||||||||||| || ||||||| |||||   |    |||| |
Db     161040 GCAGGATTACTCTGGAATGAGGGTCTTGAAAGGAGAAGAGAGAA---GCGACTTTTTGGC 160984

Qy        721 TTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGG 780
              |||||||||  |||||||| ||||| |||||| ||||||||| |   ||||  ||||||
Db     160983 TTTATGGCTGACTTCGGGGGAGAGGGGTTCTAATTTCTATGAACTGCCTTGCAGAAGAGT 160924

Qy        781 AATTCTGGTTTCTGTGACTTGCTT-TCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAG 839
              |||| ||||||  |||| ||||||     | ||||||| |  |  ||| ||| ||  |||
Db     160923 AATTTTGGTTTTGGTGAATTGCTTCAGGGGGAGAAAGATGGTTGGGAGACAGAAGAACAG 160864

Qy        840 GAGATGGTCAGAAAGAGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAG 899
              |||| | ||||| |||        ||||| ||||| |||||  |||||||  ||
Db     160863 GAGAAGATCAGACAGA-----CTTTGCTTTTGAGGCCTTCCAATCTCCTTAGGT----CA 160813

Qy        900 TACTTCTTAGCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAACACTG 959
               | |  | ||||||||||||| | | | ||||||  ||| || |||||||| | ||||| |
Db     160812 AAGTAATCAGCATACCAAAGTAATGTTCTTTGGGCTATCGTATTCTGAGCCCCAACACAG 160753

Qy        960 C 960
              |
Db     160752 C 160752
```

```
RESULT 18
AC021800/c
LOCUS       AC021800              188207 bp    DNA     linear   HTG 05-APR-2000
DEFINITION  Homo sapiens clone RP11-26B2, WORKING DRAFT SEQUENCE, 13 unordered
            pieces.
ACCESSION   AC021800
VERSION     AC021800.3  GI:7417767
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 188207)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-26B2
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 188207)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Apr 5, 2000 this sequence version replaced gi:6939798.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L4621
                Center clone name: 26_B_2
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 181424 bases at least Q40
                Consensus quality: 184545 bases at least Q30
                Consensus quality: 185748 bases at least Q20
                Insert size: 187007; sum-of-contigs
                Quality coverage: 5.2 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 13 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1465: contig of 1465 bp in length
            *     1466     1565: gap of 100 bp
            *     1566     5618: contig of 4053 bp in length
            *     5619     5718: gap of 100 bp
            *     5719     9391: contig of 3673 bp in length
            *     9392     9491: gap of 100 bp
            *     9492    14782: contig of 5291 bp in length
            *    14783    14882: gap of 100 bp
            *    14883    20606: contig of 5724 bp in length
            *    20607    20706: gap of 100 bp
            *    20707    30108: contig of 9402 bp in length
            *    30109    30208: gap of 100 bp
            *    30209    37414: contig of 7206 bp in length
            *    37415    37514: gap of 100 bp
            *    37515    45315: contig of 7801 bp in length
            *    45316    45415: gap of 100 bp
            *    45416    64824: contig of 19409 bp in length
            *    64825    64924: gap of 100 bp
            *    64925    90998: contig of 26074 bp in length
            *    90999    91098: gap of 100 bp
            *    91099   117243: contig of 26145 bp in length
            *   117244   117343: gap of 100 bp
            *   117344   146290: contig of 28947 bp in length
            *   146291   146390: gap of 100 bp
            *   146391   188207: contig of 41817 bp in length.
FEATURES             Location/Qualifiers
     source          1. .188207
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="RP11-26B2"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1465
                     /note="assembly_fragment"
     misc_feature    1566. .5618
                     /note="assembly_fragment"
     misc_feature    5719. .9391
                     /note="assembly_fragment"
     misc_feature    9492. .14782
                     /note="assembly_fragment"
     misc_feature    14883. .20606
                     /note="assembly_fragment"
     misc_feature    20707. .30108
                     /note="assembly_fragment"
     misc_feature    30209. .37414
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    37515. .45315
                     /note="assembly_fragment"
     misc_feature    45416. .64824
                     /note="assembly_fragment"
     misc_feature    64925. .90998
```

```
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    91099. .117243
                     /note="assembly_fragment"
     misc_feature    117344. .146290
                     /note="assembly_fragment"
     misc_feature    146391. .188207
                     /note="assembly_fragment"
ORIGIN

  Query Match             4.8%;  Score 293.8;  DB 2;  Length 188207;
  Best Local Similarity   72.8%;  Pred. No. 3e-54;
  Matches  481;  Conservative    0;  Mismatches  157;  Indels   23;  Gaps    7;

Qy        302 CAAAGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCA 361
              | |  | || |||  ||| |||| ||||||||  |||||||  | || ||||| ||| ||
Db     134957 CTATCCAGCTGTACAGGTAAAAAGGTGTGATATGTTTTCTCACCTATAATAAGAGTCACA 134898

Qy        362 ACCAAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAA 421
              |  | | |||||||  |||||| |||||||||||||| ||| | |||||  |||||||||
Db     134897 GCTGACATTCCTATAACAAAAGATAGGTTAATAAGAGGAAAGCATAACACTTTTATTTAA 134838

Qy        422 TCAAAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAAC 481
              ||||||| |||| ||| |   ||| ||||||||||||||||||||| ||||||| | ||||
Db     134837 TCAAAGTGTTACTTGATACAAGAGCCTTCAGAAATGAAGACCCAAGGACCCAGAGAAAAC 134778

Qy        482 TGTCTGTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAG 541
              || |  |||||| |||| | |||   |||||||||| ||| |  |  |    |||| |
Db     134777 TG-CCCTTTTTTATGCTTATGTTTAGTGAAGAATGGACAGCCAGGTGGAAATGTGATTGG 134719

Qy        542 ACAAAAGGATATGATCTAGTGGTAAAGGACTCAGGGGG-AAACACAGCAAGGCCTGTCTA 600
              |||||||||||| || ||  ||| |  || | |||||| |||| ||||||||||||| ||
Db     134718 ACAAAAGGATATAATTTAAAGGTGAGAGAATGAGGGGGAAAACCCAGCAAGGCCTGCCTG 134659

Qy        601 TTCAGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGG 660
              |||||||||||||||  ||| ||| |         | ||||| ||| | ||| |||| ||
Db     134658 TTCAGATTCTTCTTGACTTCTTTCTGT--------AACATTCCTTCTTTCTGGGTATAGG 134607

Qy        661 GCAGGACTCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGAT 720
              |||||| |  |||  |||||||||||| || ||||||| |||||    |    |||| |
Db     134606 GCAGGATTACTCTGGAATGAGGGTCTTGAAAGGAGAAGAGAGAA---GCGACTTTTTGGC 134550

Qy        721 TTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGG 780
              |||||||||  |||||||| ||||| |||||| ||||||||| |    ||||  ||||||
Db     134549 TTTATGGCTGACTTCGGGGGAGAGGGGTTCTAATTTCTATGAACTGCCTTGCAGAAGAGT 134490

Qy        781 AATTCTGGTTTCTGTGACTTGCTT-TCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAG 839
              |||| ||||||  |||| ||||||      | ||||||| |  |  ||| ||| ||  |||
Db     134489 AATTTTGGTTTTGGTGAATTGCTTCAGGGGGAGAAAGATGGTTGGGAGACAGAAGAACAG 134430

Qy        840 GAGATGGTCAGAAAGAGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAG 899
              |||| | ||||| |||        ||||| ||||| |||||  |||||||  ||
Db     134429 GAGAAGATCAGACAGA-----CTTTGCTTTTGAGGCCTTCCAATCTCCTTAGGT----CA 134379

Qy        900 TACTTCTTAGCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAACACTG 959
              | |  | |||||||||||| | | | ||||||  ||| || |||||||| | ||||| |
Db     134378 AAGTAATCAGCATACCAAAGTAATGTTCTTTGGGCTATCGTATTCTGAGCCCCAACACAG 134319

Qy        960 C 960
              |
Db     134318 C 134318


RESULT 19
AL139275/c
LOCUS       AL139275              193126 bp    DNA     linear   PRI 04-SEP-2001
DEFINITION  Human DNA sequence from clone RP11-552E20 on chromosome
            6p12.3-21.2, complete sequence.
ACCESSION   AL139275
VERSION     AL139275.30  GI:15485119
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 193126)
  AUTHORS   Smith,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 6, 2001 this sequence version replaced gi:14800199.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            RP11-552E20 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            This sequence is the entire insert of clone RP11-552E20 The true
            left end of clone RP11-535K1 is at 161600 in this sequence. The
            true right end of clone RP11-200E19 is at 35574 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .193126
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="p12.3-21.2"
                     /clone="RP11-552E20"
                     /clone_lib="RPCI-11.2"
     misc_feature    12. .455
                     /note="match: GSS: Em:AQ807014"
     misc_feature    4253. .4780
                     /note="match: GSS: Em:B71338"
     misc_feature    4256. .4678
                     /note="match: GSS: Em:AQ437688"
     misc_feature    complement(8999. .9566)
                     /note="match: STS: Em:G63453"
     misc_feature    9829. .10237
                     /note="match: GSS: Em:AQ030932"
     misc_feature    complement(10116. .10617)
                     /note="match: STS: Em:G56076
                     match: GSS: Em:AQ318725"
     misc_feature    13203. .13321
                     /note="match: GSS: Em:AQ792274"
     misc_feature    13238. .13428
                     /note="match: GSS: Em:AQ042370"
     misc_feature    complement(13332. .13469)
                     /note="match: GSS: Em:AQ504765"
     misc_feature    15417. .15794
                     /note="match: GSS: Em:AQ138303"
     misc_feature    15426. .15971
                     /note="match: GSS: Em:AQ384582"
     misc_feature    15435. .16067
                     /note="match: GSS: Em:AQ198235"
```

```
misc_feature    20235. .20758
                /note="match: GSS: Em:AQ480601"
misc_feature    20275. .20620
                /note="match: GSS: Em:AQ097268"
misc_feature    33115. .33591
                /note="match: GSS: Em:AQ613320"
misc_feature    complement(35079. .35571)
                /note="match: GSS: Em:AQ692857"
misc_feature    complement(38680. .39467)
                /note="match: GSS: Em:AQ892878"
misc_feature    41966. .42582
                /note="match: GSS: Em:AQ308209"
misc_feature    41976. .42381
                /note="match: GSS: Em:AQ124355"
misc_feature    46566. .46984
                /note="match: GSS: Em:AQ882156"
misc_feature    47399. .47848
                /note="match: GSS: Em:AQ811844"
misc_feature    64782. .65002
                /note="match: GSS: Em:AQ772609"
misc_feature    64783. .65002
                /note="match: GSS: Em:AQ683445"
misc_feature    complement(64835. .64983)
                /note="match: STS: Em:G08189"
misc_feature    complement(64864. .64983)
                /note="match: GSS: Em:AL328221"
misc_feature    complement(64864. .64975)
                /note="match: GSS: Em:AL324222"
misc_feature    complement(64888. .64983)
                /note="match: GSS: Em:AQ932064"
misc_feature    64898. .65015
                /note="match: STS: Em:G09742"
misc_feature    complement(64898. .64990)
                /note="match: GSS: Em:AZ257084"
misc_feature    complement(64898. .64989)
                /note="match: GSS: Em:AQ674454"
misc_feature    complement(64900. .64983)
                /note="match: GSS: Em:AZ237938"
misc_feature    64903. .65013
                /note="match: GSS: Em:AL191898"
misc_feature    64903. .64991
                /note="match: GSS: Em:AL309316 Em:AL316981 Em:AL346021"
misc_feature    complement(64903. .64991)
                /note="match: GSS: Em:AL284370"
misc_feature    64903. .64983
                /note="match: GSS: Em:AL224697"
misc_feature    complement(64903. .64983)
                /note="match: STS: Em:G10109"
misc_feature    64905. .65002
                /note="match: GSS: Em:AQ284087"
misc_feature    complement(74424. .74751)
                /note="match: GSS: Em:AQ151384"
misc_feature    complement(77154. .77547)
                /note="match: GSS: Em:AQ192713"
misc_feature    80797. .81303
                /note="match: GSS: Em:AQ606690"
misc_feature    complement(80886. .81417)
                /note="match: GSS: Em:AQ602297"
misc_feature    complement(91202. .91756)
                /note="match: GSS: Em:AQ063130"
misc_feature    complement(102120. .102591)
                /note="match: GSS: Em:AQ177699"
misc_feature    107123. .107721
                /note="match: GSS: Em:AQ524997"
misc_feature    complement(115935. .116552)
                /note="match: GSS: Em:AQ080194"
misc_feature    complement(116067. .116478)
                /note="match: GSS: Em:AQ031004"
misc_feature    complement(116195. .116555)
                /note="match: GSS: Em:B32716"
misc_feature    complement(128470. .128806)
                /note="match: GSS: Em:B36077"
misc_feature    complement(128811. .129176)
                /note="match: STS: Em:HS164XF4"
misc_feature    139541. .139825
                /note="match: GSS: Em:AQ278260"
misc_feature    165778. .165946
                /note="match: STS: Em:G26303"
misc_feature    171874. .172092
                /note="Single clone region. Sequence from reads from a
                short insert library derived from a clone PCR. Restriction
                digest data confirm the assembly."
misc_feature    complement(175807. .176203)
                /note="match: GSS: Em:AQ882642"
misc_feature    176202. .176743
                /note="match: GSS: Em:AQ374550"
misc_feature    complement(177911. .178307)
                /note="match: GSS: Em:AQ131194"
misc_feature    complement(177997. .178301)
                /note="match: GSS: Em:AQ148743"
misc_feature    178325. .178821
                /note="match: GSS: Em:AQ420001"
misc_feature    complement(179120. .179217)
                /note="match: GSS: Em:AQ545214"
ORIGIN

  Query Match             4.8%;  Score 292;  DB 9;  Length 193126;
  Best Local Similarity   74.3%;  Pred. No. 7.6e-54;
  Matches  507;  Conservative    0;  Mismatches  125;  Indels   50;  Gaps    9;

Qy        308 TGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAA 367
              ||| |||||||   ||| ||||||| | | | |   |||||||||| || || |
Db       2999 TGCTGTAGGGGCAGAAAGGTGTGATGCCTATCCACACTCATCATAAGGATCACAGCTGCC 2940

Qy        368 ACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAG 427
              |||||||||  |||||||| ||||| ||||| ||  | ||||| | ||||||||||||||
Db       2939 ACTCCTATAACAAAAGACAAGTTAACAAGAGAAATTCATAACACAATTATTTAATCAAAG 2880

Qy        428 TTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTG 487
              |||||  |||||||| || |||||||||||||| ||||         ||||| ||
Db       2879 TTTTATGTGACATGGCAGCCTTCAGAAATGAAGATCCAA---------GGAAAATGGTCT 2829

Qy        488 TTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAA 547
              |||| || | || || |||||||||||| |  |||||||  |||| |||| |||||||
Db       2828 GATTTTATGTTTAGTTTAGATGAAGAATGGCTTCCATGTAGAAATGT-GATTGGACAAAA 2770

Qy        548 GGATATGATCTAGTGGTAAAGGACTCAGG----GGGAAACACAGCAAGGCCTGTCTATTC 603
              ||  |||||||| |||||| || ||  ||     ||||||| ||||||| | | ||| |||
Db       2769 GG-CATGATCTAATGGTAATGGGCTGGGGAGTTGGGAAACCCAGCAAGTCTTATCTGTTC 2711

Qy        604 AGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCA 663
              ||||||||||||||||||         || |||||||| |||||||| |||||||||
Db       2710 AGATTCTTCTTGATCTCTC--------TGAATAGCATTCCTTCCTCCTGGGTATGGGGCA 2659

Qy        664 GGACTCTTCTTCAATGAGGGTCTTCAAGGG-------AGAAGGGAGAAAGTGGCCTTTTT 716
              ||| | |||| ||||| ||||||||||||       |||||| ||| ||| ||||| |
Db       2658 AGACCCCTCTTGAATGAAGGTCTTCAAGGGAAATGGAAGAAGGCAGAGAGTAACCTTTCT 2599

Qy        717 AGATTTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAA 776
              || |||||    || ||| | || ||||| ||||||||||||||||||||||| |  |||
Db       2598 AGGTTTTACAAATTCCTTTGTGGGAGAGGGGTTCTAGTTTCTATGACCCATCTGGAAGAA 2539

Qy        777 GAGGAATTCTGGTTTCTGTGACTT--------GCTTTCATGAAGAAAGAGGAGTAAGAGG 828
              |||||||||||||||| ||||||         |  |   ||  |  ||||| | ||||
Db       2538 GAGGAATTCTGGTTTCTATGACTTACGTCGGGGGCTGGTTGGGGGGAGAGGGGCGAGAGA 2479

Qy        829 CAGGAGGGCAGGAGATGGTCAGAAAGAGACTT---------GGCTGCTTCTGAGGGCTTC 879
              |||||||| |||||| ||||||| ||| ||||         ||||||||||||  ||||
Db       2478 CAGGAGGGTAGGAGAAGGTCAGAGAGAAACTTTGATTCTGAGGCTGCTTCTGAGACCTTC 2419

Qy        880 CGCTCTCCTTTAGTTCCAAGTACTTCTTAGCATACCAAAGCACTATACTTTGGCATATGG 939
              |  |||||||||||||| |||||    ||||||| ||||||  | |||||||||  ||| |
Db       2418 CAGTCTCCTTTAGTTCAAAGTA---TTTAGCATGCCAAAGTGCCATACTTTGGGGTATTG 2362

Qy        940 TTTTCTGAGCTCTAACACTGCA 961
              |||| ||||| | |||| ||||
```

```
Db       2361 TTTTTTGAGCCCCAACAATGCA 2340


RESULT 20
AC073070
LOCUS       AC073070              158983 bp    DNA     linear   HTG 07-JUL-2000
DEFINITION  Homo sapiens chromosome 6 clone RP11-200E19, WORKING DRAFT
            SEQUENCE, 16 unordered pieces.
ACCESSION   AC073070
VERSION     AC073070.1  GI:8348041
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 158983)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 158983)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JUN-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT
            ---------------- Genome Center ----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0200E19
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 100%
            Sequencing vector: plasmid; 0%
            Chemistry: Dye-primer ET; 100% of reads
            Chemistry: Dye-terminator Big Dye; 0% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 150411 bases at least Q40
            Consensus quality: 153191 bases at least Q30
            Consensus quality: 154640 bases at least Q20
            Insert size: 141000; agarose-fp
            Insert size: 157483; sum-of-contigs
            Quality coverage: 4.79 in Q20 bases; agarose-fp
            Quality coverage: 4.34 in Q20 bases; sum-of-contigs
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 16 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2233: contig of 2233 bp in length
            *     2234     2333: gap of unknown length
            *     2334     5177: contig of 2844 bp in length
            *     5178     5277: gap of unknown length
            *     5278     9015: contig of 3738 bp in length
            *     9016     9115: gap of unknown length
            *     9116    13714: contig of 4599 bp in length
            *    13715    13814: gap of unknown length
            *    13815    18427: contig of 4613 bp in length
            *    18428    18527: gap of unknown length
            *    18528    22634: contig of 4107 bp in length
            *    22635    22734: gap of unknown length
            *    22735    29174: contig of 6440 bp in length
            *    29175    29274: gap of unknown length
            *    29275    35652: contig of 6378 bp in length
            *    35653    35752: gap of unknown length
            *    35753    42585: contig of 6833 bp in length
            *    42586    42685: gap of unknown length
            *    42686    53206: contig of 10521 bp in length
            *    53207    53306: gap of unknown length
            *    53307    61368: contig of 8062 bp in length
            *    61369    61468: gap of unknown length
            *    61469    72869: contig of 11401 bp in length
            *    72870    72969: gap of unknown length
            *    72970    87473: contig of 14504 bp in length
            *    87474    87573: gap of unknown length
            *    87574   101199: contig of 13626 bp in length
            *   101200   101299: gap of unknown length
            *   101300   120299: contig of 19000 bp in length
            *   120300   120399: gap of unknown length
            *   120400   158983: contig of 38584 bp in length.
FEATURES             Location/Qualifiers
     source          1. .158983
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP11-200E19"
     misc_feature    1. .2233
                     /note="assembly_name:Contig6"
     misc_feature    2334. .5177
                     /note="assembly_name:Contig7"
     misc_feature    5278. .9015
                     /note="assembly_name:Contig8"
     misc_feature    9116. .13714
                     /note="assembly_name:Contig9"
     misc_feature    13815. .18427
                     /note="assembly_name:Contig10"
     misc_feature    18528. .22634
                     /note="assembly_name:Contig11"
     misc_feature    22735. .29174
                     /note="assembly_name:Contig12"
     misc_feature    29275. .35652
                     /note="assembly_name:Contig13"
     misc_feature    35753. .42585
                     /note="assembly_name:Contig14"
     misc_feature    42686. .53206
                     /note="assembly_name:Contig15"
     misc_feature    53307. .61368
                     /note="assembly_name:Contig16"
     misc_feature    61469. .72869
                     /note="assembly_name:Contig17"
     misc_feature    72970. .87473
                     /note="assembly_name:Contig18"
     misc_feature    87574. .101199
                     /note="assembly_name:Contig19
                     clone_end:SP6
                     vector_side:right"
     misc_feature    101300. .120299
                     /note="assembly_name:Contig20"
     misc_feature    120400. .158983
                     /note="assembly_name:Contig21"
ORIGIN

  Query Match             4.8%;  Score 290.4;  DB 2;  Length 158983;
  Best Local Similarity   74.2%;  Pred. No. 1.7e-53;
  Matches  506;  Conservative    0;  Mismatches  126;  Indels   50;  Gaps    9;

Qy        308 TGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAA 367
              ||| |||||||   ||| ||||||| | | | | |   |||||||||| || || |
Db       6575 TGCTGTAGGGGCAGAAAGGTGTGATGCCTATCCACACTCATCATAAGGATCACAGCTGCC 6634

Qy        368 ACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAG 427
              |||||||||  |||||||| ||||| ||||| ||  | ||||| | ||||||||||||||
Db       6635 ACTCCTATAACAAAAGACAAGTTAACAAGAGAAATTCATAACACAATTATTTAATCAAAG 6694

Qy        428 TTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTG 487
              |||||  |||||||| || ||||||||||||||| ||||          ||||| ||
Db       6695 TTTTATGTGACATGGCAGCCTTCAGAAATGAAGATCCAA---------GGAAAATGGTCT 6745

Qy        488 TTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAA 547
                  |||| || | || || ||||||||||  |  |||||||  |||| |||| |||||||
```

```
Db       6746 GATTTTATGTTTAGTTTAGATGAAGAATGACTTCCATGTAGAAATGT-GATTGGACAAAA 6804

Qy        548 GGATATGATCTAGTGGTAAAGGACTCAGG----GGGAAACACAGCAAGGCCTGTCTATTC 603
              ||  |||||||| |||||| || ||  ||    ||||||| ||||||| | | ||| |||
Db       6805 GG-CATGATCTAATGGTAATGGGCTGGGGAGTTGGGAAACCCAGCAAGTCTTATCTGTTC 6863

Qy        604 AGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCA 663
              |||||||||||||||||||         || |||||||| |||||||| ||||||||||
Db       6864 AGATTCTTCTTGATCTCTC--------TGAATAGCATTCCTTCCTCCTGGGTATGGGGCA 6915

Qy        664 GGACTCTTCTTCAATGAGGGTCTTCAAGGG-------AGAAGGGAGAAAGTGGCCTTTTT 716
               ||| | |||| ||||| ||||||||||||       |||||| ||| |||  ||||| |
Db       6916 AGACCCCTCTTGAATGAAGGTCTTCAAGGGAAATGGAAGAAGGCAGAGAGTAACCTTTCT 6975

Qy        717 AGATTTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAA 776
              || |||||    || ||| | || |||||||||||||||||||||||||||| |  |||
Db       6976 AGGTTTTACAAATTCCTTTGTGGGAGAGGGGTTCTAGTTTCTATGACCCATCTGGAAGAA 7035

Qy        777 GAGGAATTCTGGTTTCTGTGACTT--------GCTTTCATGAAGAAAGAGGAGTAAGAGG 828
              ||||||||||||||||| ||||||         | |   ||  |  |||||  |  ||||
Db       7036 GAGGAATTCTGGTTTCTATGACTTACGTCGGGGGCTGGTTGGGGGGAGAGGGGCGAGAGA 7095

Qy        829 CAGGAGGGCAGGAGATGGTCAGAAAGAGACTT---------GGCTGCTTCTGAGGGCTTC 879
              |||||||| |||||| ||||||| ||| ||||          ||||||||||||  ||||
Db       7096 CAGGAGGGTAGGAGAAGGTCAGAGAGAAACTTTGATTCTGAGGCTGCTTCTGAGACCTTC 7155

Qy        880 CGCTCTCCTTTAGTTCCAAGTACTTCTTAGCATACCAAAGCACTATACTTTGGCATATGG 939
              |  |||||||||||||| |||||    ||||||| ||||||  | |||||||||  ||| |
Db       7156 CAGTCTCCTTTAGTTCAAAGTA---TTTAGCATGCCAAAGTGCCATACTTTGGGGTATTG 7212

Qy        940 TTTTCTGAGCTCTAACACTGCA 961
              |||| |||||  | |||| ||||
Db       7213 TTTTTTGAGCCCCAACAATGCA 7234
```

RESULT 21
AC069216/c
LOCUS AC069216 150145 bp DNA linear PRI 07-MAY-2002
DEFINITION Homo sapiens 3 BAC RP11-621D6 (Roswell Park Cancer Institute Human BAC Library) complete sequence.
ACCESSION AC069216
VERSION AC069216.13 GI:20467989
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 150145)
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T., Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., He,X., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M., Neal,D., Nelson,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Naylor,S.L., Weinstock,G. and Gibbs,R.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 150145)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (22-MAY-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 150145)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (01-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 4 (bases 1 to 150145)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (07-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On May 7, 2002 this sequence version replaced gi:20336722.
INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE:Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL:
http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

```
          QUALSTAT-REPORT.
FEATURES             Location/Qualifiers
     source          1. .150145
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-621D6"
     repeat_region   1. .68
                     /rpt_family="MER41B"
     repeat_region   69. .369
                     /rpt_family="MER41A"
     repeat_region   370. .481
                     /rpt_family="L1"
     repeat_region   533. .707
                     /rpt_family="AluJo"
     repeat_region   826. .852
                     /rpt_family="(TA)n"
     repeat_region   1159. .1194
                     /rpt_family="AT_rich"
     repeat_region   complement(1687. .2098)
                     /rpt_family="MSTB"
     repeat_region   2142. .2173
                     /rpt_family="(A)n"
     repeat_region   2505. .2802
                     /rpt_family="AluSx"
     repeat_region   complement(3288. .3429)
                     /rpt_family="L1ME"
     repeat_region   3497. .3569
                     /rpt_family="L1ME"
     repeat_region   complement(4252. .5813)
                     /rpt_family="L1MA4A"
     repeat_region   5829. .5910
                     /rpt_family="(TA)n"
     repeat_region   complement(5911. .6195)
                     /rpt_family="AluSx"
     repeat_region   complement(6198. .6460)
                     /rpt_family="L1MA4A"
     repeat_region   complement(7908. .8336)
                     /rpt_family="MSTB"
     repeat_region   9136. .9427
                     /rpt_family="AluY"
     repeat_region   9440. .9737
                     /rpt_family="AluJo"
     repeat_region   10250. .10529
                     /rpt_family="L1ME"
     repeat_region   10545. .10695
                     /rpt_family="(TA)n"
     repeat_region   complement(10845. .11140)
                     /rpt_family="AluJb"
     misc_feature    11693
                     /note="G/T"
                     /function="polymorphic site"
     misc_feature    11737
                     /note="C/T"
                     /function="polymorphic site"
     repeat_region   11949. .11973
                     /rpt_family="AT_rich"
     repeat_region   12047. .12353
                     /rpt_family="AluJb"
     repeat_region   12681. .12740
                     /rpt_family="(TC)n"
     repeat_region   12789. .13036
                     /rpt_family="L1ME"
     repeat_region   13256. .13312
                     /rpt_family="(TA)n"
     repeat_region   13668. .13689
                     /rpt_family="AT_rich"
     repeat_region   13702. .13758
                     /rpt_family="AT_rich"
     repeat_region   15903. .16199
                     /rpt_family="AluSx"
     repeat_region   complement(16357. .16599)
                     /rpt_family="MIR"
     repeat_region   complement(16611. .16745)
                     /rpt_family="MIR"
     repeat_region   complement(17201. .17297)
                     /rpt_family="L2"
     repeat_region   17425. .17466
                     /rpt_family="(CAAAA)n"
     repeat_region   complement(18079. .18242)
                     /rpt_family="MIR"
     repeat_region   18310. .18405
                     /rpt_family="GA-rich"
     repeat_region   complement(18934. .19002)
                     /rpt_family="MER47"
     repeat_region   19017. .19084
                     /rpt_family="MER47"
     repeat_region   complement(19069. .19166)
                     /rpt_family="MER47"

  Query Match             4.8%;  Score 289.6;  DB 9;  Length 150145;
  Best Local Similarity   71.6%;  Pred. No. 2.6e-53;
  Matches  484;  Conservative    0;  Mismatches  169;  Indels   23;  Gaps    7;

Qy        284 TTATCTCTCCTTCTCTCTCAAAGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCT 343
              |||  ||    |  ||  ||||  |       |  |  | |||  ||||  ||   ||  ||
Db      69138 TTAATTCATTTATTCATTCAATACATGCACACTGTAGGAAAGTTGTGGTATCCTTCTTCA 69079

Qy        344 TCCATCATAAGGGTCGCAACCAAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAA 403
              ||||||||||||||    |||| |||||||||   |||||||| |||||||||| |||
Db      69078 TCCATCATAAGGGTCATGGCCAACACTCCTATAAC-AAAGACAGATTAATAAGAGAAAAG 69020

Qy        404 CCTAACAAATTTATTTAATCAAAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACC 463
              | || ||||| ||||||||||||||||||  |||||||||   ||||||||||||||||
Db      69019 CATAGCAAATGTATTTAATCAAAGTTTTA-GTGACATGGGGAACTTCAGAAATGAAGATT 68961

Qy        464 CAAAGACCCAGGGGAAACTGTCTGTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCA 523
              |||||||||||| |||||| |   ||||  |||| || || ||||||||||| | |||
Db      68960 CAAAGACCCAGGGAAAACTGAC-CATTTTAATGCTTAGATTTGATGAAGAATGAACAGCC 68902

Qy        524 TGTAGCCATGTAGATTAGACAAAAGGATATGATCTAGTGGTAAAGGACTCAGGGGG-AAA 582
              |   | ||| |||||| || ||| ||||| ||||||  |||| |||||| |||
Db      68901 ATGTAGAAATGAAATTGGACAAAGGGGTATAATCTAATGGTAATTGACTGAGGGGGAAAA 68842

Qy        583 CACAGCAAGGCCTGTCTATTCAGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTC 642
              | |||||||||||||||| |||| | |||||||  || ||         ||| |||||  |
Db      68841 CTCAGCAAGGCCTGTCTGTTCATAGTCTTCTTGGCCTGTC--------TGTGTAGCACCC 68790

Qy        643 TTTCCTCCTGAGTATGGGGCAGGACTCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAG 702
              |||||||   |||| || |||||| | |||  ||||||||||| |||||||| ||||
Db      68789 CTTCCTCCCAGGTATAGGACAGGACCCCTCTGGAATGAGGGTCTTTAAGGGAGAAAGGAG 68730

Qy        703 AAAGTGGCCTTTTTAGATTTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGA 762
              | |||  ||||| |||||||| | |||| | |  | || ||||||||||||||| ||||
Db      68729 AGAGTCACCTTTCTAGATTTTTTAGCTTACATTAGAGATGAGGAGTTCTAGTTTTTATGT 68670

Qy        763 CCCATCTTGGGGAAGAGGAATTCTGGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGT 822
              ||   ||||| |||||||||||||||||||||| |||||  |||    | || | || | ||
Db      68669 CCTGCCTTGGAGAAGAGGAATTCTGGTTTCTTTGACTCCCTTCAGGGGAGGAGGAAGTGT 68610

Qy        823 AAGAGGCAGGAGGGCAGGAGATGGTCAGAAAGAGACTT--------GGCTGCTTCTGAGG 874
              ||   | ||||||||||||| ||||    |       ||         ||||||||||||
Db      68609 GGGAAACCAGAGGGCAGGAGAAGGTCGAGAGAGACTTTGCTTCTGAGGCTGCTTCTGAGG 68550

Qy        875 GCTTCCGCTCTCCTTTAGTTCCAAGTACTTCTTAGCATACCAAAGCACTATACTTTGGCA 934
              |||||  | | |||||||| |||||   || || |  ||||||    ||||||| |
Db      68549 CCTTCCAGTTTTCTTTAGTTCAAAGTA---CTCAGTACGCCAAAGTGACATACTTTTGAG 68493

Qy        935 TATGGTTTTCTGAGCT 950
              |||  |||||| ||||
Db      68492 TATCATTTTCTAAGCT 68477


RESULT 22
```

```
AC024412
LOCUS       AC024412              178168 bp    DNA     linear   HTG 25-MAY-2000
DEFINITION  Homo sapiens chromosome 3 clone RP11-740D17 map 3, WORKING DRAFT
            SEQUENCE, 33 unordered pieces.
ACCESSION   AC024412
VERSION     AC024412.3  GI:8076909
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 178168)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 3, clone RP11-740D17
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 178168)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M.,
            Fenestor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., Landers,T., Largocque,K., Lehoczky,J., Levine,R.,
            Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
            McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivar,T.M.,
            Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
            Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
            Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
            Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
            Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 25, 2000 this sequence version replaced gi:7212066.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L6889
                Center clone name: 740_D_17
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 157219 bases at least Q40
                Consensus quality: 168334 bases at least Q30
                Consensus quality: 172739 bases at least Q20
                Insert size: 182000; agarose-fp
                Insert size: 174968; sum-of-contigs
                Quality coverage: 3.5 in Q20 bases; agarose-fp
                Quality coverage: 3.7 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 33 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1025: contig of 1025 bp in length
            *     1026     1125: gap of 100 bp
            *     1126     2296: contig of 1171 bp in length
            *     2297     2396: gap of 100 bp
            *     2397     4307: contig of 1911 bp in length
            *     4308     4407: gap of 100 bp
            *     4408     6169: contig of 1762 bp in length
            *     6170     6269: gap of 100 bp
            *     6270     7785: contig of 1516 bp in length
            *     7786     7885: gap of 100 bp
            *     7886    10094: contig of 2209 bp in length
            *    10095    10194: gap of 100 bp
            *    10195    12891: contig of 2697 bp in length
            *    12892    12991: gap of 100 bp
            *    12992    16495: contig of 3504 bp in length
            *    16496    16595: gap of 100 bp
            *    16596    18954: contig of 2359 bp in length
            *    18955    19054: gap of 100 bp
            *    19055    21997: contig of 2943 bp in length
            *    21998    22097: gap of 100 bp
            *    22098    26348: contig of 4251 bp in length
            *    26349    26448: gap of 100 bp
            *    26449    29582: contig of 3134 bp in length
            *    29583    29682: gap of 100 bp
            *    29683    32837: contig of 3155 bp in length
            *    32838    32937: gap of 100 bp
            *    32938    36977: contig of 4040 bp in length
            *    36978    37077: gap of 100 bp
            *    37078    40554: contig of 3477 bp in length
            *    40555    40654: gap of 100 bp
            *    40655    44062: contig of 3408 bp in length
            *    44063    44162: gap of 100 bp
            *    44163    47890: contig of 3728 bp in length
            *    47891    47990: gap of 100 bp
            *    47991    53061: contig of 5071 bp in length
            *    53062    53161: gap of 100 bp
            *    53162    57537: contig of 4376 bp in length
            *    57538    57637: gap of 100 bp
            *    57638    62270: contig of 4633 bp in length
            *    62271    62370: gap of 100 bp
            *    62371    67026: contig of 4656 bp in length
            *    67027    67126: gap of 100 bp
            *    67127    72268: contig of 5142 bp in length
            *    72269    72368: gap of 100 bp
            *    72369    76670: contig of 4302 bp in length
            *    76671    76770: gap of 100 bp
            *    76771    82195: contig of 5425 bp in length
            *    82196    82295: gap of 100 bp
            *    82296    87658: contig of 5363 bp in length
            *    87659    87758: gap of 100 bp
            *    87759    95525: contig of 7767 bp in length
            *    95526    95625: gap of 100 bp
            *    95626   103428: contig of 7803 bp in length
            *   103429   103528: gap of 100 bp
            *   103529   112551: contig of 9023 bp in length
            *   112552   112651: gap of 100 bp
            *   112652   123241: contig of 10590 bp in length
            *   123242   123341: gap of 100 bp
            *   123342   133097: contig of 9756 bp in length
            *   133098   133197: gap of 100 bp
            *   133198   142314: contig of 9117 bp in length
            *   142315   142414: gap of 100 bp
            *   142415   158518: contig of 16104 bp in length
            *   158519   158618: gap of 100 bp
            *   158619   178168: contig of 19550 bp in length.
FEATURES             Location/Qualifiers
     source          1. .178168
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /map="3"
                     /clone="RP11-740D17"
                     /clone_lib="RPCI-11 Human Male BAC"
```

```
     misc_feature    1. .1025
                     /note="assembly_fragment"
     misc_feature    1126. .2296
                     /note="assembly_fragment"
     misc_feature    2397. .4307
                     /note="assembly_fragment"
     misc_feature    4408. .6169
                     /note="assembly_fragment"
     misc_feature    6270. .7785
                     /note="assembly_fragment"
     misc_feature    7886. .10094
                     /note="assembly_fragment"
     misc_feature    10195. .12891
                     /note="assembly_fragment"
     misc_feature    12992. .16495
                     /note="assembly_fragment"
     misc_feature    16596. .18954
                     /note="assembly_fragment"
     misc_feature    19055. .21997
                     /note="assembly_fragment"
     misc_feature    22098. .26348
                     /note="assembly_fragment"
     misc_feature    26449. .29582
                     /note="assembly_fragment"
     misc_feature    29683. .32837
                     /note="assembly_fragment"
     misc_feature    32938. .36977
                     /note="assembly_fragment"
     misc_feature    37078. .40554
                     /note="assembly_fragment"
     misc_feature    40655. .44062
                     /note="assembly_fragment"
     misc_feature    44163. .47890
                     /note="assembly_fragment"
     misc_feature    47991. .53061
                     /note="assembly_fragment"
     misc_feature    53162. .57537
                     /note="assembly_fragment"
     misc_feature    57638. .62270
                     /note="assembly_fragment"
     misc_feature    62371. .67026
                     /note="assembly_fragment"
     misc_feature    67127. .72268
                     /note="assembly_fragment"
     misc_feature    72369. .76670
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    76771. .82195
                     /note="assembly_fragment"
     misc_feature    82296. .87658
                     /note="assembly_fragment"
     misc_feature    87759. .95525
                     /note="assembly_fragment"
     misc_feature    95626. .103428

  Query Match             4.8%;  Score 289.6;  DB 2;  Length 178168;
  Best Local Similarity   71.6%;  Pred. No. 2.6e-53;
  Matches  484;  Conservative    0;  Mismatches  169;  Indels   23;  Gaps    7;

Qy        284 TTATCTCTCCTTCTCTCTCAAAGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCT 343
              |||  ||   |  ||  |||| |      |  |  | |||  |||| ||   ||  ||
Db     150065 TTAATTCATTTATTCATTCAATACATGCACACTGTAGGAAAGTTGTGGTATCCTTCTTCA 150124

Qy        344 TCCATCATAAGGGTCGCAACCAAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAA 403
              ||||||||||||||    |||| |||||||||   |||||||| |||||||||| |||
Db     150125 TCCATCATAAGGGTCATGGCCAACACTCCTATAAC-AAAGACAGATTAATAAGAGAAAAG 150183

Qy        404 CCTAACAAATTTATTTAATCAAAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACC 463
              | || ||||| ||||||||||||||||||  |||||||||   ||||||||||||||
Db     150184 CATAGCAAATGTATTTAATCAAAGTTTTA-GTGACATGGGGAACTTCAGAAATGAAGATT 150242

Qy        464 CAAAGACCCAGGGGAAACTGTCTGTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCA 523
              |||||||||||||| |||||| |   ||||  |||| || || |||||||||| | |||
Db     150243 CAAAGACCCAGGGAAAACTGAC-CATTTTAATGCTTAGATTTGATGAAGAATGAACAGCC 150301

Qy        524 TGTAGCCATGTAGATTAGACAAAAGGATATGATCTAGTGGTAAAGGACTCAGGGGG-AAA 582
                 |  | ||| |||||| || ||| ||||| ||||||  |||| |||||| |||
Db     150302 ATGTAGAAATGAAATTGGACAAAGGGGTATAATCTAATGGTAATTGACTGAGGGGGAAAA 150361

Qy        583 CACAGCAAGGCCTGTCTATTCAGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTC 642
              | |||||||||||||| |||| | |||||||  || ||        ||| |||||  |
Db     150362 CTCAGCAAGGCCTGTCTGTTCATAGTCTTCTTGGCCTGTC--------TGTGTAGCACCC 150413

Qy        643 TTTCCTCCTGAGTATGGGGCAGGACTCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAG 702
              |||||||   |||| || |||||| | |||  ||||||||||||| |||||||| |||| .
Db     150414 CTTCCTCCCAGGTATAGGACAGGACCCCTCTGGAATGAGGGTCTTTAAGGGAGAAAGGAG 150473

Qy        703 AAAGTGGCCTTTTTAGATTTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGA 762
              | |||  |||| |||||||| | |||| | |  | || ||||||||||||||| ||||
Db     150474 AGAGTCACCTTTCTAGATTTTTTAGCTTACATTAGAGATGAGGAGTTCTAGTTTTTATGT 150533

Qy        763 CCCATCTTGGGGAAGAGGAATTCTGGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGT 822
              ||   |||| ||||||||||||||||||||| |||||  |||    | || | || | ||
Db     150534 CCTGCCTTGGAGAAGAGGAATTCTGGTTTCTTTGACTCCCTTCAGGGGAGGAGGAAGTGT 150593

Qy        823 AAGAGGCAGGAGGGCAGGAGATGGTCAGAAAGAGACTT--------GGCTGCTTCTGAGG 874
              ||  |  ||||||||||||| ||||   |       ||        ||||||||||||
Db     150594 GGGAAACCAGAGGGCAGGAGAAGGTCGAGAGAGACTTTGCTTCTGAGGCTGCTTCTGAGG 150653

Qy        875 GCTTCCGCTCTCCTTTAGTTCCAAGTACTTCTTAGCATACCAAAGCACTATACTTTGGCA 934
              |||||  | | |||||||| |||||   || || |  ||||||    ||||||| |
Db     150654 CCTTCCAGTTTTCTTTAGTTCAAAGTA---CTCAGTACGCCAAAGTGACATACTTTTGAG 150710

Qy        935 TATGGTTTTCTGAGCT 950
              |||  |||||| ||||
Db     150711 TATCATTTTCTAAGCT 150726


RESULT 23
AC109998/c
LOCUS       AC109998              126394 bp    DNA     linear   PRI 27-MAR-2002
DEFINITION  Homo sapiens chromosome 5 clone CTD-2203A3, complete sequence.
ACCESSION   AC109998
VERSION     AC109998.2  GI:19747167
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 126394)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 126394)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-FEB-2002) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 126394)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAR-2002) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Mar 27, 2002 this sequence version replaced gi:18642722.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 100% of Sequence;
            Estimated Total Number of Errors is 0.
            NOTE:  BACTERIAL TRANSPOSON excised at 118365.
FEATURES             Location/Qualifiers
     source          1. .126394
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
```

```
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTD-2203A3"
ORIGIN

  Query Match             4.7%;  Score 289;  DB 9;  Length 126394;
  Best Local Similarity   71.8%;  Pred. No. 3.6e-53;
  Matches  469;  Conservative    0;  Mismatches  160;  Indels   24;  Gaps    6;

Qy        313 TAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTCC 372
              ||| ||| ||| ||||||||| ||| || |||||||||||||| |||| | |||||
Db      45396 TAGAAGTGGAAAGGTGTGATACCTTTCATCACCCATCATAAGGGTCACAACTGATACTCC 45337

Qy        373 TATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTA 432
              |||   ||||||    |||||||||| | |   |||||||| |||||| ||||||||||
Db      45336 TATGACAAAAGA----TTAATAAGAGAAGACTGTAACAAATGTATTTATTCAAAGTTTTA 45281

Qy        433 CATGACATGGGAGTCTTCAGAAATG-AAGACCCAAAGACCCAGGGGAAACTGTCTGTTTT 491
              |||||||||||||| ||||||||| | |||||||||||||||   | |||     |  |||
Db      45280 CATGACATGGGAGCCTTCAGAAAGGAAAGACCCAAAGACCCAAAGAAAAACTGCCCTTTG 45221

Qy        492 TTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGAT 551
              || |||| |||||| |||| |||||||||          |    ||||||||||||||
Db      45220 TTATGCTTAGGTTCAATGACAAATGGATAGTCATGTAGAAATGTGATTAGACAAAAGGAA 45161

Qy        552 ATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTT 611
              || ||| | ||||||  |||| |||||| |||    |||| ||||| |||||||  |||||
Db      45160 ATAATCGAATGGTAATAGACTGAGGGGGGAACCAGGCAAAGCCTGGCTATTCATGTTCTT 45101

Qy        612 CTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTT 671
              ||||  |||||| | |    | ||  | | |   |        |||||| |||||   | |
Db      45100 CTTGGCCTCTCTGTGT---AGCATTCCTTCCCCCCACATATGGTATGGAGCAGGGTCCCT 45044

Qy        672 CTTCAATGAGGGTCTTCAAGGG-------AGAAGGGAGAAAGTGGCCTTTTTAGATTTTA 724
              ||  ||  ||| |||||||||        |||||||| ||||| ||||| ||| |||||
Db      45043 CTGGAACAAGGATCTTCAAGGGAGAAGAAAGAAGGGAGGAAGTGACCTTTCTAGGTTTTA 44984

Qy        725 TGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATT 784
              ||| |||||| |||||| | || ||||| |||||||| ||  || ||||||||||||||
Db      44983 TGGATTGCTTTGGGGAAAAAGACTTCTAATTTCTATGATCCCCCTAGGGGAAGAGGAATT 44924

Qy        785 CTGGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGAT 844
              ||||||||  |||||  |||        ||||||    | ||| || || |||||||||
Db      44923 CTGGTTTCCATGACTCTCTTCTGATGGAAAAGAGAGACAGGAGACAAGAAGGCAGGAGAA 44864

Qy        845 GGTCAGAAAGAGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAGTACTT 904
              || ||  |  |          |||||||||| ||||   |||||||||||| |||||
Db      44863 GGCCAAGAGAA------CTTGCTTCTGAGGCCTTCTAATCTCCTTTAGTTCAAAGTA--- 44813

Qy        905 CTTAGCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAACAC 957
              || ||||| |||| ||   ||||||||  |||  | |||||||| | | |||
Db      44812 CTCAGCATGCCAAGACAACGTACTTTGGGGTATCATGTTCTGAGCCCCACCAC 44760


RESULT 24
AC091903
LOCUS       AC091903              194860 bp    DNA     linear   PRI 20-OCT-2001
DEFINITION  Homo sapiens chromosome 5 clone RP11-182I21, complete sequence.
ACCESSION   AC091903
VERSION     AC091903.2  GI:16303417
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 194860)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 194860)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUN-2001) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 194860)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-OCT-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Oct 20, 2001 this sequence version replaced gi:14333839.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 99.9% of Sequence;
            Estimated Total Number of Errors is 0.1.
FEATURES             Location/Qualifiers
     source          1. .194860
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="RP11-182I21"
ORIGIN

  Query Match             4.7%;  Score 289;  DB 9;  Length 194860;
  Best Local Similarity   71.8%;  Pred. No. 3.5e-53;
  Matches  469;  Conservative    0;  Mismatches  160;  Indels   24;  Gaps    6;

Qy        313 TAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTCC 372
              ||| ||| ||| ||||||||| ||| || |||||||||||||| |||| | |||||
Db       9170 TAGAAGTGGAAAGGTGTGATACCTTTCATCACCCATCATAAGGGTCACAACTGATACTCC 9229

Qy        373 TATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTA 432
              |||   ||||||    |||||||||| | |   |||||||| |||||| ||||||||||
Db       9230 TATGACAAAAGA----TTAATAAGAGAAGACTGTAACAAATGTATTTATTCAAAGTTTTA 9285

Qy        433 CATGACATGGGAGTCTTCAGAAATG-AAGACCCAAAGACCCAGGGGAAACTGTCTGTTTT 491
              |||||||||||||| ||||||||| | |||||||||||||||   | |||     |  |||
Db       9286 CATGACATGGGAGCCTTCAGAAAGGAAAGACCCAAAGACCCAAAGAAAAACTGCCCTTTG 9345

Qy        492 TTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGAT 551
              || |||| |||||| |||| |||||||||          |    ||||||||||||||
Db       9346 TTATGCTTAGGTTCAATGACAAATGGATAGTCATGTAGAAATGTGATTAGACAAAAGGAA 9405

Qy        552 ATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTT 611
              || ||| | ||||||  |||| |||||| |||    |||| ||||| |||||||  |||||
Db       9406 ATAATCGAATGGTAATAGACTGAGGGGGGAACCAGGCAAAGCCTGGCTATTCATGTTCTT 9465

Qy        612 CTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTT 671
              ||||  |||||| | |    | ||  | | |   |        |||||| |||||   | |
Db       9466 CTTGGCCTCTCTGTGT---AGCATTCCTTCCCCCCACATATGGTATGGAGCAGGGTCCCT 9522

Qy        672 CTTCAATGAGGGTCTTCAAGGG-------AGAAGGGAGAAAGTGGCCTTTTTAGATTTTA 724
              ||  ||  ||| |||||||||        |||||||| ||||| ||||| ||| |||||
Db       9523 CTGGAACAAGGATCTTCAAGGGAGAAGAAAGAAGGGAGGAAGTGACCTTTCTAGGTTTTA 9582

Qy        725 TGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATT 784
              ||| |||||| |||||| | || ||||| |||||||| ||  || ||||||||||||||
Db       9583 TGGATTGCTTTGGGGAAAAAGACTTCTAATTTCTATGATCCCCCTAGGGGAAGAGGAATT 9642

Qy        785 CTGGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGAT 844
              ||||||||  |||||  |||        ||||||    | ||| || || |||||||||
Db       9643 CTGGTTTCCATGACTCTCTTCTGATGGAAAAGAGAGACAGGAGACAAGAAGGCAGGAGAA 9702

Qy        845 GGTCAGAAAGAGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAGTACTT 904
              || ||  |  |          |||||||||| ||||   |||||||||||| |||||
Db       9703 GGCCAAGAGAA------CTTGCTTCTGAGGCCTTCTAATCTCCTTTAGTTCAAAGTA--- 9753

Qy        905 CTTAGCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAACAC 957
              || ||||| |||| ||   ||||||||  |||  | |||||||| | | |||
Db       9754 CTCAGCATGCCAAGACAACGTACTTTGGGGTATCATGTTCTGAGCCCCACCAC 9806


RESULT 25
AL450243
```

```
LOCUS       AL450243              65608 bp    DNA     linear   PRI 11-SEP-2001
DEFINITION  Human DNA sequence from clone RP11-567E16 on chromosome 9, complete
            sequence.
ACCESSION   AL450243
VERSION     AL450243.13  GI:15591207
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 65608)
  AUTHORS   Babbage,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 12, 2001 this sequence version replaced gi:14272321.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr9
            RP11-567E16 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-567E16 It may be shorter because we sequence overlapping
            sections only once, except for a short overlap.
            The true left end of clone RP11-375O18 is at 65509 in this
            sequence. The true right end of clone RP11-159N19 is at 100 in this
            sequence.
FEATURES             Location/Qualifiers
     source          1. .65608
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /clone="RP11-567E16"
                     /clone_lib="RPCI-11.2"
ORIGIN

  Query Match             4.7%;  Score 288.2;  DB 9;  Length 65608;
  Best Local Similarity   71.2%;  Pred. No. 5.6e-53;
  Matches  482;  Conservative    0;  Mismatches  178;  Indels   17;  Gaps    7;

Qy        313 TAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTCC 372
              | |||    |||| ||||||||| |||||||  ||||||||| ||||    |  | |||||
Db      16581 TGGGGAGAAAAATGTGTGATACCTTTTCTCACCCATCATAAAGGTCATGGCTGACACTCC 16640

Qy        373 TATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTA 432
              ||||  |||||||| ||||| ||||| ||| | |||||| |||| | ||||||||||||
Db      16641 TATAACAAAAGACAAGTTAACAAGAGAAAAGCATAACAAGTTTACTCAATCAAAGTTTTA 16700

Qy        433 CATGAC--ATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTT 490
                ||||  |  ||||  |||| ||||||||||| ||| | |||||||| |||||||||  ||

Db      16701 AGTGACAAACAGGAGATTTCAAAAATGAAGACTCAATGGCCCAGGGAAAACTGTCT-ATT 16759

Qy        491 TTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGA 550
              ||| |||| || |||  |||||||||| |||          |   | || ||||   ||
Db      16760 TTTATGCTTAGCTTCAGTGAAGAATGGACAGCCGTAGAGAAATAAAATCAGACGGCAGAG 16819

Qy        551 TATGATCTAGTGGTAAAGGACTCAGGGGGA-AACACAGCAAGGCCTGTCTATTCAGATTC 609
              |||||||||  | |||| |||| | ||||| ||| ||||||| ||||||| ||||||| |
Db      16820 TATGATCTACGGCTAAAAGACTGAAGGGGAGAACTCAGCAAGACCTGTCTGTTCAGATCC 16879

Qy        610 TTCT---TGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGA 666
              |  |        ||||  ||||| ||| ||||||||  ||| || |  |   ||||||||
Db      16880 TGTTCAGATCCTTCTCCATCTCTCTGTGTAGCATTCCCTCCCCCAG-CTGCAGGGCAGGA 16938

Qy        667 CTCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATG 726
              | | |||  |||  | ||||||||  |||||||||| |||  ||||| ||| ||||||||
Db      16939 CCCCTCTGGAATACGAGTCTTCAAAAGAGAAGGGAGAGAGTATCCTTTCTAGGTTTTATG 16998

Qy        727 GCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCT 786
              ||    || | |  |||||||| ||| |||| ||| ||| ||||||||||||||||||
Db      16999 GCCATTTTTGAGAGAGAGGAGTTTTAGCTTCTGTGATCCACCTTGGGGAAGAGGAATTCT 17058

Qy        787 GGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGG 846
              ||||||| |||||  |||| || |||||||||      |  |  | |||   |   ||
Db      17059 GGTTTCTATGACTCACTTTGGTGGAGAAAGAGGGAGGGGGAGAGGAAGGCAGGAGAATTC 17118

Qy        847 TCAGAAAGAGACT------TGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAGT 900
              | ||| |     |       |||||||||||||| |||||  | ||||  |||||  ||||
Db      17119 TGAGAGAACTTTTCTTCTGAGGCTGCTTCTGAGGCCTTCCAATGTCCTAAAGTTCAAAGT 17178

Qy        901 ACTTCTTAGCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAACACTGC 960
              |   || ||||  |||||||| ||| |||||  | | |||||||||| | ||||  |
Db      17179 A---CTCAGCACGCCAAAGCACCATATTTTGGGGTGTTGTTTTCTGAGCCCCAACAAGGG 17235

Qy        961 AATCATGCTAAACTCCT 977
                | |||  ||||  | |
Db      17236 GAACATTTTAAAAGCTT 17252


RESULT 26
AL442129
LOCUS       AL442129              181618 bp    DNA     linear   HTG 02-OCT-2000
DEFINITION  Homo sapiens chromosome 9 clone RP11-2C8, WORKING DRAFT SEQUENCE.
ACCESSION   AL442129
VERSION     AL442129.2  GI:10635948
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Plumb,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 4, 2000 this sequence version replaced gi:10278027.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bA2C8
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 181423 bases at least Q40
            Consensus quality: 181504 bases at least Q30
            Consensus quality: 181552 bases at least Q20
            Insert size: 181618; sum-of-contigs
            Insert size: 190546; agarose-fp
```

```
          Quality coverage: 8.11x in Q20 bases; sum-of-contigs Quality
          coverage: 7.97x in Q20 bases; agarose-fp
          --------------.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 1 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1   181618: contig of 181618 bp in length.
FEATURES             Location/Qualifiers
     source          1. .181618
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /clone="RP11-2C8"
                     /clone_lib="RPCI-11.1"
     misc_feature    1. .181618
                     /note="assembly_fragment:04053"
ORIGIN

  Query Match             4.7%;  Score 288.2;  DB 2;  Length 181618;
  Best Local Similarity   71.2%;  Pred. No. 5.2e-53;
  Matches  482;  Conservative    0;  Mismatches  178;  Indels   17;  Gaps    7;

Qy        313 TAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTCC 372
              | |||   |||| ||||||||| |||||||  ||||||||| ||||    |  | |||||
Db      24785 TGGGGAGAAAAATGTGTGATACCTTTTCTCACCCATCATAAAGGTCATGGCTGACACTCC 24844

Qy        373 TATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTA 432
              ||||  |||||||| ||||| ||||| ||| | |||||| |||| | ||||||||||||||
Db      24845 TATAACAAAAGACAAGTTAACAAGAGAAAAGCATAACAAGTTTACTCAATCAAAGTTTTA 24904

Qy        433 CATGAC--ATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTT 490
                ||||  |  ||||  |||| |||||||||| ||| | ||||||| |||||||||  ||
Db      24905 AGTGACAAACAGGAGATTTCAAAAATGAAGACTCAATGGCCCAGGGAAAACTGTCT-ATT 24963

Qy        491 TTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGA 550
              ||| |||| || |||  |||||||||| |||         |   | || ||||   ||
Db      24964 TTTATGCTTAGCTTCAGTGAAGAATGGACAGCCGTAGAGAAATAAAATCAGACGGCAGAG 25023

Qy        551 TATGATCTAGTGGTAAAGGACTCAGGGGGA-AACACAGCAAGGCCTGTCTATTCAGATTC 609
              |||||||||  | |||| |||| | ||||| ||| ||||||| ||||||| ||||||| |
Db      25024 TATGATCTACGGCTAAAAGACTGAAGGGGAGAACTCAGCAAGACCTGTCTGTTCAGATCC 25083

Qy        610 TTCT---TGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGA 666
              |   |        ||||  ||||| ||| ||||||||  ||| || |  |   |||||||||
Db      25084 TGTTCAGATCCTTCTCCATCTCTCTGTGTAGCATTCCCTCCCCCAG-CTGCAGGGCAGGA 25142

Qy        667 CTCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATG 726
              | | |||  |||  | ||||||||  |||||||||| |||  ||||| ||| |||||||
Db      25143 CCCCTCTGGAATACGAGTCTTCAAAAGAGAAGGGAGAGAGTATCCTTTCTAGGTTTTATG 25202

Qy        727 GCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCT 786
              ||    || | |  |||||||||| ||| |||| ||| ||| ||||||||||||||||||
Db      25203 GCCATTTTTGAGAGAGAGGAGTTTTAGCTTCTGTGATCCACCTTGGGGAAGAGGAATTCT 25262

Qy        787 GGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGG 846
              ||||||| |||||  ||||  || |||||||||      |  |  | |||   |    ||
Db      25263 GGTTTCTATGACTCACTTTGGTGGAGAAAGAGGGAGGGGGAGAGGAAGGCAGGAGAATTC 25322

Qy        847 TCAGAAAGAGACT------TGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAGT 900
              | ||| |     |       |||||||||||||| |||||  | ||||  ||||| ||||
Db      25323 TGAGAGAACTTTTCTTCTGAGGCTGCTTCTGAGGCCTTCCAATGTCCTAAAGTTCAAAGT 25382

Qy        901 ACTTCTTAGCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAACACTGC 960
              |   || ||||  ||||||||| ||| |||||  | | |||||||||| | ||||  |
Db      25383 A---CTCAGCACGCCAAAGCACCATATTTTGGGGTGTTGTTTTCTGAGCCCCAACAAGGG 25439

Qy        961 AATCATGCTAAACTCCT 977
```

```
              | |||  ||||  | |
Db      25440 GAACATTTTAAAAGCTT 25456


RESULT 27
AC120502
LOCUS       AC120502               190092 bp    DNA     linear   HTG 07-MAY-2002
DEFINITION  Lemur catta clone LB2-287D5, WORKING DRAFT SEQUENCE, 13 unordered
            pieces.
ACCESSION   AC120502
VERSION     AC120502.1  GI:20486402
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Lemur catta (ring-tailed lemur)
  ORGANISM  Lemur catta
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Strepsirhini; Lemuridae; Lemur.
REFERENCE   1  (bases 1 to 190092)
  AUTHORS   Martin,J., Hosseini,R., Peng,Y., Peng,Z., Rubin,E.M. and
            Cheng,J.-F.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 190092)
  AUTHORS   Martin,J., Hosseini,R., Peng,Y., Peng,Z., Rubin,E.M. and
            Cheng,J.-F.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAY-2002) Genome Sciences, Lawrence Berkeley National
            Laboratory, 1 Cyclotron Rd., Berkeley, CA 94720, USA
COMMENT     Draft Sequence Produced by Berkeley PGA
            Web site: http://pga.lbl.gov
            Center Code: PGABERK
            Center Project Name: L017-287D5
            Bac Clone Name: LB2-287D5

            Additional information on comparative analysis and ordering are
            available at:
            http://pga.lbl.gov/cgi-bin/search_cvcgd?type=n&value=SFTPD
            Funding agent: Programs for Genomic Applications (NHLBI)
            Contact: 'Jody Schwartz' jrschwartz@lbl.gov
            if library name is LB1 to LB4, please see website
            for a description: http://www-gsd.lbl.gov/cheng/BAC.html
            Summary Statistics:
            Sequencing vector: Plasmid; pUC18
            Chemistry: Dye-terminator Big Dye
            Assembly program: Phrap version 0.990329.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 13 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2537: contig of 2537 bp in length
            *     2538     2637: gap of unknown length
            *     2638     6734: contig of 4097 bp in length
            *     6735     6834: gap of unknown length
            *     6835    14153: contig of 7319 bp in length
            *    14154    14253: gap of unknown length
            *    14254    21025: contig of 6772 bp in length
            *    21026    21125: gap of unknown length
            *    21126    33188: contig of 12063 bp in length
            *    33189    33288: gap of unknown length
            *    33289    43658: contig of 10370 bp in length
            *    43659    43758: gap of unknown length
            *    43759    59694: contig of 15936 bp in length
            *    59695    59794: gap of unknown length
            *    59795    75288: contig of 15494 bp in length
            *    75289    75388: gap of unknown length
            *    75389    93753: contig of 18365 bp in length
            *    93754    93853: gap of unknown length
            *    93854   112133: contig of 18280 bp in length
            *   112134   112233: gap of unknown length
            *   112234   136450: contig of 24217 bp in length
```

```
           *   136451   136550: gap of unknown length
           *   136551   161260: contig of 24710 bp in length
           *   161261   161360: gap of unknown length
           *   161361   190092: contig of 28732 bp in length.
FEATURES             Location/Qualifiers
     source          1. .190092
                     /organism="Lemur catta"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9447"
                     /clone="LB2-287D5"
ORIGIN

  Query Match             4.7%;  Score 288.2;  DB 2;  Length 190092;
  Best Local Similarity   70.5%;  Pred. No. 5.2e-53;
  Matches  455;  Conservative    0;  Mismatches  158;  Indels   32;  Gaps    4;

Qy        312 GTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTC 371
              ||||||| ||| ||||||| ||| ||| | || | |||||||  |||| ||||
Db     126207 GTAGGGGCAGAAAGTTGTGATATCTTTCCTCACCTATTACAAGGGTCATGGCCAACACTC 126266

Qy        372 CTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTT 431
              |||||  |||||||| | || ||||| ||| | |||||||||| ||||||||||||||
Db     126267 CTATAACAAAAGACAGATGAACAAGAGAAAAGCATAACAAATTTACTTAATCAAAGTTTT 126326

Qy        432 ACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTT 491
              |||||||||||||| |||||||||||   ||||||||||||||| |||||||||| ||
Db     126327 ACATGACATGGGAGCCTTCAGAAATGGGAACCCAAAGACCCAGGGAAAACTGTCTGCTTC 126386

Qy        492 TTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGAT 551
              | |   | | || |   |     |     || |||| ||||   |     |   | ||
Db     126387 TATGCTTAGGATTAGTGCAGACTGGACAGCCACGTAGACATGTGATTGGACAAGGGGTAT 126446

Qy        552 ATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTT 611
                ||  | ||  || ||        |||| || ||| ||||||| | | |||| ||||||
Db     126447 GGAATAGACTGAGAAGGG-----CAGGGAGACCCAGTAAGGCCTTTGTGTTCAAATTCTT 126501

Qy        612 CTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTT 671
              ||||  |||||          |  || |||| ||| || || | |||  ||||||| | |
Db     126502 CTTGGCCTCTC----------TGCAGTATTCCTTCTTCTTGGGCATGATGCAGGACCCCT 126551

Qy        672 CTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATGGCTTG 731
              |  |||||||||||||||||||| ||||| |||||||||| ||  |||||||| |||
Db     126552 ATGGGATGAGGGTCTTCAAGGGAGAAAGGAGAGAGTGGCCTTTCTATGTTTTATGGTTTG 126611

Qy        732 CTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCTGGTTT 791
              ||||   | |||| |||||||||||||| ||||||| |||||||||||||||| ||||||
Db     126612 CTTCAAAGGAGAGAAGTTCTAGTTTCTACGACCCATTTTGGGGAAGAGGAATTTTGGTTT 126671

Qy        792 CTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGGTCAGA 851
              || |||||  ||||    |||||||||  || ||||   |||||||  ||| | |||||
Db     126672 CTATGACTCACTTTGGGTAAGAAAGAGAGGTGAGAGAGTGGAGGGCAAAAGAGGATCAGA 126731

Qy        852 AAGAGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAGTACTTCTTAGCA 911
              |||              |||||| |||||  ||||||| ||||  ||    | || ||||
Db     126732 GAGA-------------CTGAGGCCTTCCAGTCTCCTTCAGTTAAAA----TACTCAGCA 126774

Qy        912 TACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAACA 956
              | |||| |      |||||||| || |  ||||||||||  | ||||
Db     126775 TGCCAAGGTGACGTACTTTGGGATGTCATTTTCTGAGCCCCAACA 126819


RESULT 28
HS372K1/c
LOCUS       HS372K1               147055 bp    DNA     linear   PRI 05-MAR-2003
DEFINITION  Human DNA sequence from clone RP3-372K1 on chromosome 6q24,
            complete sequence.
ACCESSION   AL023580
VERSION     AL023580.1  GI:3702133
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 147055)
  AUTHORS   Spragon,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 6, 1998 this sequence version replaced gi:3646075.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            RP3-372K1 is from the library RPCI-3 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pCYPAC2.
FEATURES             Location/Qualifiers
     source          1. .147055
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="q24"
                     /clone="RP3-372K1"
                     /clone_lib="RPCI-3"
ORIGIN

  Query Match             4.7%;  Score 284.4;  DB 9;  Length 147055;
  Best Local Similarity   71.9%;  Pred. No. 3.6e-52;
  Matches  466;  Conservative    0;  Mismatches  151;  Indels   31;  Gaps    6;

Qy        298 CTCTCAAAGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGT 357
              ||   | | ||   | |||||   ||| ||||| ||| |||||||  |||||||||| ||
Db      26216 CTTCTATATCTATTGAAGGGGCAGAAAAGTGTGGTACCTTTTCTCACCCATCATAAGAGT 26157

Qy        358 CGCAACCAAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTAT 417
              | |   ||| |||||| ||  |||||| |||||| ||||| |||    ||||||||||
Db      26156 CACCGTCAACACTCCTGTAACAAAAGATAGGTTAACAAGAGAAAAGTACAACAAATTTAT 26097

Qy        418 TTAATCAAAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGG 477
              ||||||||||||||  ||||||| ||  | ||||||||||||||||||||| |||||
Db      26096 TTAATCAAAGTTTTATGTGACATGAGAAACATCAGAAATGAAGACCCAAAGATCCAGGAA 26037

Qy        478 AAACTGTCTGTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGA 537
              |||||| ||  ||||| |||| ||| || |||||||||| | |||         |     |
Db      26036 AAACTGCCT-ATTTTTATGCTTAGGATCAATGAAGAATGAACAGCCATGTAGAAATGTCA 25978

Qy        538 TTAGACAAAAGGATATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGT 597
                | ||| ||||| ||||||||| |||| |  | || |||||||||| ||||||||||||
```

```
Db      25977 CTGGACCAAAGGGTATGATCTAATGGTGATAGGCTGAGGGGGAAACCCAGCAAGGCCTGC 25918

Qy        598 CTATTCAGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTAT 657
              || ||||||||||||||| ||||||       ||| |||  | | ||| |||||  |  |
Db      25917 CTGTTCAGATTCTTCTTGGTCTCTC--------TGTGTAGTCTGCCTTCATCCTG-GGCT 25867

Qy        658 GGGGCAGGACTCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTA 717
              ||| ||||| | | |  ||||| ||||||| ||| ||||             |      |||
Db      25866 TGGGTAGGACCCCTGTGGAATGAAGGTCTTTAAGAGAGA-----------GTGACCTTTC 25818

Qy        718 GATTTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAG 777
              |  |||||||||||||| |||||| ||  |||||||||||||||||||| |||||| |||
Db      25817 TAGGTTATGGCTTGCTTGGGGGAAAAGAGGTTCTAGTTTCTATGACCCACCTTGGGAAAG 25758

Qy        778 AGGAATTCTGGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGC 837
              |||||||||||||||| |||||  ||||   | ||||||||  |  |||  |||||||
Db      25757 AGGAATTCTGGTTTCTATGACTCACTTTATGGGAGAAAGAGAGGCGAGAAACAGGAGGAT 25698

Qy        838 AGGAGATGGTCAGAAAGAGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCA 897
              |||||  ||||||| |||       ||| |||||||  ||| |   |||||||  |||| |
Db      25697 GGGAGAGGGTCAGAGAGA-------CTGTTTCTGAGACCTTTCAACCTCCTTTTGTTCAA 25645

Qy        898 AGTACTTCTTAGCATACCAAAGCACTATACTTTGGCATATGGTTTTCT 945
              || |    | ||||| |||| |  | |||||||||  |||  |  |||
Db      25644 AGCA---TTCAGCATGCCAAGGTGCCATACTTTGGGGTATCATGCTCT 25600
```

RESULT 29
AC024153/c

```
LOCUS       AC024153               178905 bp    DNA     linear   PRI 01-APR-2003
DEFINITION  Homo sapiens 12 BAC RP13-359K18 (Roswell Park Cancer Institute
            Human BAC Library) complete sequence.
ACCESSION   AC024153
VERSION     AC024153.22  GI:10190646
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 178905)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C., Cleveland,C.D.,
            Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Emerling,S., Escotto,M., Falls,T., Ferraguto,D.,
            Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J.,
            Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W.,
            Gunaratne,P., Hale,S., Hamilton,K., Han,J., Harris,C., Harris,K.,
            Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O.,
            Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F.,
            Howard,S., Huber,J., Hulyk,S., Hume,J., Ioshikhes,I., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lee,E., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Marondel,I., Martin,R.,
            Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P.,
            Meador,M., Mei,G., Merscher,S., Metzker,M., Miller,A., Miner,G.,
            Miner,Z., Mitchell,T., Mohabbat,K., Montgomery,K.T., Morgan,M.,
            Morris,S., Moser,M., Neal,D., Nelson,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shim,C.,
            Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A.,
            Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A.,
            Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B.,
            Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D.,
            Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R.,
            Washington,C., Watlington,S., Williams,G., Williamson,A.,
            Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J.,
            Zorrilla,S., Kucherlapati,R. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 178905)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-FEB-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 178905)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-SEP-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   4  (bases 1 to 178905)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-JAN-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   5  (bases 1 to 178905)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-FEB-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   6  (bases 1 to 178905)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-APR-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Sep 19, 2000 this sequence version replaced gi:9966020.
            INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
            gc-help@bcm.tmc.edu

            CLONE LENGTH: This sequence does not necessarily represent the
            entire insert of this clone. Overlapping regions of clones are only
            sequenced and submitted once, so the sequence for the remainder of
            the insert may be found in the record for the adjacent clones.
            Overlapping clones are noted at the beginning and end of the
            Features listing.

            ANNOTATION OF FEATURES:
               STSs are identified using ePCR (Genome Res. 7:541-550) searches
            of a local database that includes entries from dbSTS, GDB, and
            local mapping efforts.
               Repeats are identified using RepeatMasker (A. Smit and P. Green,
            unpublished.) for Human and Mouse sequences.
               Genes and Region of sequence similarity are identified by BLAST
            (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
            EST and cDNA sequences. Genes demonstrate at least two exons
            flanked by consensus splice sites that maintained sequence
            continuity across the splice junctions. Sequences that are not
            identical matches are annotated as similar.

            SEQUENCING READ COVERAGE:Sequencing is completed to a minimum
            standard of double strand coverage with a minimum of 2 clones and 2
            reads with no ambiguities or 2 chemistries with a minimum of 2
            clones and 3 reads with no ambiguities. If the sequence quality for
            a region does not meet this standard, it will be indicated in the
            annotation as Low Coverage.

            QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality
            standards - estimated error rate less than 1 per 10,000 bases.
```

Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL:
http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

QUALSTAT-REPORT----------

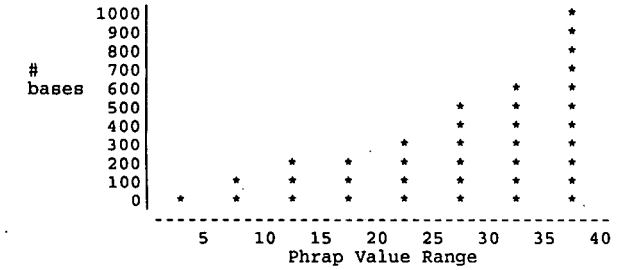
---------- Summary Statistics ----------

Contig length:	178905
Phrap values in estimate:	177537
Average error rate (BCM-Phrap estimate):	0.000263085
Fraction of Phrap values less than 40 :	0.022277
Number of consensus changing edits:	27
Number of N's in consensus :	0

---------- Consensus changing edits ----------

Position	Original+Context	Edited+Context
4748	aaaaagaaaa(n)cttttaacca	aaaaagaaaa(g)cttttaacca
5704	ccatgcatcc(t)ttacttctca	ccatgcatcc(a)tcacttctca
5706	atgcatcctt(t)acttctcata	atgcatccat(c)acttctcata
11559	aatttttactt(n)ttcttccagt	aatttttactt(c)ttcttccagt
17872	ttactattca(n)taggtaaatg	ttactattca(t)taggtaaatg
40216	tcagacaaag(n)ctattggtct	tcagacaaag(t)ctattggtct
41700	gagcatcata(n)tacatgaaac	gagcatcata(g)tacatgaaac
43693	ggctgtggat(n)tntnccaacc	ggctgtggat(a)tctaccaacc
43695	ctgtggatnt(n)tnccaacctc	ctgtggatat(c)taccaacctc
43697	gtggatntnt(n)ccaacctcaa	gtggatatct(a)ccaacctcaa
43811	cttctgcaat(n)ctggtttttt	cttctgcaat(t)ctggtttttt
43852	ttttttttt(n)gagacggagt	ttttttttt(t)gagacggagt
44158	tggaatattc(n)ttttggtatt	tggaatattc(t)ttttggtatt
54827	tttattcatt(n)acaaggatat	tttattcatt(c)acaaggatat
54874	aagggaagca(n)ancataaaag	aagggaagca(g)aacataaaag
54876	gggaagcana(n)cataaaagtt	gggaagcaga(a)cataaaagtt
74216	tcctattcag(n)ttttttttt	tcctattcag(t)ttttttttt
74241	ttttttttt(a)agttttaagg	ttttttttt(t)agttttaagg
76761	tatataacct(n)tgtaggataa	tatataacct(c)tgtaggataa
93171	ttttttaat(a)acagttctca	ttttttaat(t)acagttctca
93815	ctttcaatat(n)taatgaaaga	ctttcaatat(c)taatgaaaga
94701	caccctcgg(n)gcaaaaataa	caccctcgg(t)gcaaaaataa
97007	tcgggaggct(n)aggcaggaga	tcgggaggct(g)aggcaggaga
116441	ttaaggctta(n)aagtgtaaca	ttaaggctta(g)aagtgtaaca
141823	gatctttgtc(n)tgtgatattt	gatctttgtc(t)tgtgatattt
150571	aaaaaagtgt(n)ttctggcttg	aaaaaagtgt(c)ttctggcttg
166877	aacagaaaag(n)tctccaaatc	aacagaaaag(t)tctccaaatc

--------- Distribution of Quality < 40 Bases ----------



Version: 1.01 qxfo.

```
FEATURES             Location/Qualifiers
     source          1. .178905
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="12"
                     /clone="RP13-359K18"
     misc_feature    complement(1. .2042)
                     /note="overlaps bases 1. .2042 of clone AC129028"
                     /function="clone overlap"
     misc_feature    complement(1. .1083)
                     /note="May overlap bases 1. .1083 of clone AC016956"
                     /function="Overlaps adjacent clone AC016956"
     repeat_region   1. .210
                     /rpt_family="L1M2"
```

Query Match 4.6%; Score 282.2; DB 9; Length 178905;
Best Local Similarity 71.2%; Pred. No. 1.1e-51;
Matches 476; Conservative 0; Mismatches 173; Indels 20; Gaps 7;

```
Qy      315 GGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTCCTA 374
            |||||| ||| ||||||||| ||| |||  ||||||| || |||    | || |||||||
Db    76407 GGGGTGGAAATGTGTGATACCTTTCCTCACCCATCATGAGTGTCATGGCTAACACTCCTA 76348

Qy      375 TAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTACA 434
            ||| |||||||||||| || || |  | | | |||||||||||| ||||||||| |||| ||
Db    76347 TAGGAAAAGACAGGTGAACAAAAAAATAGCATAACAAATTTACTTAATCAAAATTTTGCA 76288

Qy      435 TGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTTTTT 494
            |||||  |||  |||||| |||||| | | ||| |||||||  |  ||||||  ||||
Db    76287 TGACACAGGAAACTTCAGCAATGAAAATCTAAAAACCCAGGTAACGCTGTCT-ATTTTCA 76229

Qy      495 TGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGATATG 554
            |||| ||||| |||||||||||| |||   |    |  | |||| ||| ||| | | ||
Db    76228 TGCTTAGGTTTGATGAAGAATGGACAGC-CATGCAGAAATGGATTGGACCAAAAGGTCTG 76170

Qy      555 ATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTTCTT 614
            ||||| || |||   |  || ||||  || |||||||||||| ||| ||   |||||||||
Db    76169 ATCTAATGATAA--TAAACAAGGGGGTACCCAGCAAGGCCTCTCTGTTTGAATTCTTCTT 76112

Qy      615 GATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTTCTT 674
            ||          ||||| |||||||||||  ||  |||    ||||||||||||| || | |
Db    76111 GA--------CCTCTAGGTATAGCATTCCTTTTTCCCAGATATGGGGCAGGATTCCTTTA 76060

Qy      675 CAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATGGCTTGCTT 734
            |||||||||||||| |||||||| || | |||| ||||| ||| ||||||| ||||||||
Db    76059 GAATGAGGGTCTTCATGGGAGAAGAGACAGAGTGACCTTTCTAGGTTTTATGACTTGCTT 76000

Qy      735 CGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCTGGTTTCTG 794
            |||| ||||| |||||||||||||||  ||    |||||||||| ||||||||||||||||
Db    75999 TGGGGGAGAGGGGTTCTAGTTTCTATGCCCTGCCTTGGGGAAGGGGAATTCTGGTTTCTA 75940

Qy      795 TGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGGTCAGAAAG 854
            | | |  ||||   | |||||||||| ||||| || || |    |||| ||||||||| |
Db    75939 TCATTAACTTTAGGGGAGAAAGAGGAGCAAGAGACAAGAAGATGAGAGAAGGTCAGAACG 75880

Qy      855 AGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAGTACTTCTTAGCATAC 914
            |  |       |   || |  | | |||||| || |||||   || |||||  | |
Db    75879 ACCTT----GTTTGTGAGGCTTTTTCCATTTTCTTTAGATCAAAGTA---CTCAGCATGC 75827

Qy      915 CAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAACACTGCAATCATGCTAAACT 974
            ||| |  | || ||||||  |||    || ||||| ||| ||  |  || ||  | | |
Db    75826 CAAGGTGCCATTCTTTGGGGTATCACATT-TGAGCCCTAGCAAAGGTATTATTATCACCA 75768

Qy      975 CCTCTATGA 983
              | | ||
Db    75767 TATTGAAGA 75759
```

```
RESULT 30
AC012271
LOCUS       AC012271               98232 bp    DNA     linear   PRI 04-NOV-2001
DEFINITION  Homo sapiens chromosome 15, clone RP11-64O3, complete sequence.
ACCESSION   AC012271
VERSION     AC012271.11  GI:16646932
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
```

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 98232)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 15, clone RP11-6403
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 98232)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-OCT-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 98232)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Nov 4, 2001 this sequence version replaced gi:16596589.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L1395
                Center clone name: 64_O_3
            --------------

            Only the first 98.2 kilobases of this clone are being submitted.
            The remainder overlaps accession number AC087457 [WICGR project
            L11876].
FEATURES             Location/Qualifiers
     source          1. .98232
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15"
                     /clone="RP11-6403"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   187. .255
                     /rpt_family="AT_rich"
     repeat_region   456. .524
                     /rpt_family="(TA)n"
     repeat_region   621. .652
                     /rpt_family="AT_rich"
     repeat_region   1878. .1902
                     /rpt_family="(TTTTG)n"
     repeat_region   2969. .3011
                     /rpt_family="AT_rich"
     repeat_region   3634. .3654
                     /rpt_family="AT_rich"
     repeat_region   3666. .3814
                     /rpt_family="L2"
     repeat_region   4496. .4526
                     /rpt_family="AT_rich"
     repeat_region   5168. .5232
                     /rpt_family="GA-rich"
     repeat_region   complement(5510. .5720)
                     /rpt_family="AluJo"
     repeat_region   complement(5728. .5929)
                     /rpt_family="LTR37A"
     repeat_region   complement(5981. .6276)
                     /rpt_family="AluJb"
     repeat_region   complement(6424. .6574)
                     /rpt_family="AluJo"
     repeat_region   6575. .6600
                     /rpt_family="(CA)n"
     repeat_region   complement(6601. .6750)
                     /rpt_family="AluJo"
     repeat_region   7910. .7960
                     /rpt_family="AT_rich"
     repeat_region   complement(8020. .8316)
                     /rpt_family="AluSx"
     repeat_region   8326. .8369
                     /rpt_family="AT_rich"
     repeat_region   complement(8864. .9479)
                     /rpt_family="LTR26"
     repeat_region   10221. .10488
                     /rpt_family="AluY"
     repeat_region   complement(10650. .10825)
                     /rpt_family="MER5B"
     repeat_region   complement(10983. .11214)
                     /rpt_family="MER20"
     repeat_region   complement(12031. .12211)
                     /rpt_family="MIR"
     repeat_region   complement(12228. .12743)
                     /rpt_family="MER119"
     repeat_region   12834. .12915
                     /rpt_family="L2"
     repeat_region   13594. .13779
                     /rpt_family="L1MC/D"
     repeat_region   complement(13782. .13912)
                     /rpt_family="FLAM_A"
     repeat_region   complement(13931. .14148)
                     /rpt_family="MER58A"
     repeat_region   complement(15644. .15843)
                     /rpt_family="MIR3"
     repeat_region   complement(16217. .16531)
                     /rpt_family="AluY"
     repeat_region   16610. .16634
                     /rpt_family="(ATG)n"
     repeat_region   complement(16640. .16792)
                     /rpt_family="MIR"
     repeat_region   complement(16883. .16998)
                     /rpt_family="MIR3"
     repeat_region   17150. .17235
                     /rpt_family="MIR"
     repeat_region   complement(17968. .18557)
                     /rpt_family="L1PA5"
```

```
     repeat_region   19205. .19225
                     /rpt_family="(TAAA)n"
     repeat_region   complement(19341. .19543)
                     /rpt_family="MIR"
     repeat_region   19693. .19853
                     /rpt_family="MIR"
     repeat_region   20164. .20297
                     /rpt_family="LTR37A"
     repeat_region   20976. .21633
                     /rpt_family="MER41B"
     repeat_region   complement(21974. .22229)
                     /rpt_family="L2"
     unsure          22097. .22271
                     /note="PCR only region"
     repeat_region   complement(22275. .22939)
                     /rpt_family="L2"
     unsure          complement(22534. .22538)
                     /note="<30 qual SNGL region"
     unsure          complement(22648. .22653)
                     /note="<30 qual SNGL region"
     unsure          complement(22671. .22677)
                     /note="<30 qual SNGL region"
     unsure          complement(22708. .22713)
                     /note="<30 qual SNGL region"
     unsure          complement(22753. .22757)
                     /note="<30 qual SNGL region"
     repeat_region   complement(23044. .23343)
                     /rpt_family="MLT2D"
     repeat_region   complement(23377. .23472)
                     /rpt_family="MLT2D"
     repeat_region   complement(23474. .23561)
                     /rpt_family="L2"
     repeat_region   23579. .23696
                     /rpt_family="MER5B"
     repeat_region   complement(23763. .23852)
                     /rpt_family="MER5A"
     repeat_region   complement(23847. .24110)
                     /rpt_family="L2"
     repeat_region   complement(24146. .24512)
                     /rpt_family="L2"
     repeat_region   24691. .24732
                     /rpt_family="(GA)n"
     repeat_region   25102. .25277
                     /rpt_family="MIR"
     repeat_region   25787. .26059
                     /rpt_family="L1MD2"
     repeat_region   26062. .26132

  Query Match             4.6%;  Score 280.2;  DB 9;  Length 98232;
  Best Local Similarity   71.5%;  Pred. No. 3.2e-51;
  Matches  460;  Conservative    0;  Mismatches  158;  Indels   25;  Gaps    6;

Qy        312 GTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTC 371
              |||||      ||  | ||||||  || ||   ||||||||||||||    |||| | ||
Db      56369 GTAGGAACAGGAAATTATGATACTATTCCTAACCCATCATAAGGGTCATGGCCAATATTC 56428

Qy        372 CTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTT 431
              ||||  |||| |||||||||  |||| ||| | ||| | ||||||||| |||||||||
Db      56429 TTATAAAAAAATACAGGTTAACTAGAGAAAATCATAATACATTTATTTAACCAAAGTTTT 56488

Qy        432 ACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTT 491
              ||| ||||| |||| ||| ||||||||||||||||  ||    ||||    | | ||||
Db      56489 ACACGACATAGGAGCCTTGAGAAATGAAGACCCAAAAGCC---AGGAAGAACTTTCTTTT 56545

Qy        492 TTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACA-AAAGGA 550
              || ||||  |||| ||||||||||||| |||         |    |||| |||| |||||
Db      56546 TTATGCTTGGGTTTGATGAAGAATGGACAGCTGTATAAAAATGTGATTGGACAGAAAGGG 56605

Qy        551 TATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCT 610
              ||| |||      | ||  |||| |||||||||   |||||| ||||||| ||||| ||||
Db      56606 TATAATCCGAGAGCAACAGACTGAGGGGGAAAGCAAGCAAGACCTGTCTGTTCAGCTTCT 56665

Qy        611 TCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCT 670
              |||||| |||||        ||| |||| ||  |||||||  | | ||| |      |
Db      56666 TCTTGACCTCTC--------TGTGTAGCGTTTCTTCCTCCTAGGAAGGGGTC------CC 56711

Qy        671 TCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATGGCTT 730
              |||  |||||||||||||||||||||||||| | || ||||| ||| ||||| || |
Db      56712 TCTGGAATGAGGGTCTTCAAGGGAGAAGGGAGAGAATGACCTTTCTAGGTTTTACGGGTG 56771

Qy        731 GCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCTGGTT 790
              |||| | | |||||||||||||||||||||||||| |||||||||| |||| | ||| |
Db      56772 GCTTTGAGAAAGAGGAGTTCTAGTTTCTATGACCCACCTTGGGGAAGGGGAAGTGTGGCT 56831

Qy        791 TCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGGTCAG 850
              ||| |||||  ||||  |||||||||| || | ||| |||||||   ||||| | | |
Db      56832 TCTATGACTCACTTTGGGGAAGAAAGAGAAGCAGGAGACAGGAGGATGGGAGAAGATGGG 56891

Qy        851 AAAGAGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAGTACTTCTTAGC 910
              |||||         || |||||||| ||| |     |   ||   | |||  | | || |||
Db      56892 AAAGA-----CCTTGGTTCTGAGGCCTTTCCAATTTTCTTCAGTTCAA--AGTACTCAGC 56944

Qy        911 ATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTA 953
              | |||||||| | || ||||||  |||  | ||||||||  | |
Db      56945 ACACCAAAGCGCCATGCTTTGGAGTATCATATTCTGAGCCCAA 56987


RESULT 31
AC100834
LOCUS       AC100834               152005 bp    DNA     linear   PRI 30-AUG-2002
DEFINITION  Homo sapiens chromosome 15, clone CTD-2125J1, complete sequence.
ACCESSION   AC100834
VERSION     AC100834.3  GI:22549868
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 152005)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 15, clone CTD-2125J1
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 152005)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 152005)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
```

```
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 152005)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-AUG-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Aug 30, 2002 this sequence version replaced gi:22381992.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L21556
                Center clone name: 2125_J_1
            --------------.
FEATURES             Location/Qualifiers
     source          1. .152005
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15"
                     /clone="CTD-2125J1"
                     /clone_lib="CITD1 Human BAC"
     repeat_region   1992. .2041
                     /rpt_family="GC_rich"
     repeat_region   2116. .2169
                     /rpt_family="(CCG)n"
     repeat_region   3339. .3361
                     /rpt_family="(G)n"
     repeat_region   4178. .5489
                     /rpt_family="L1PA5"
     repeat_region   complement(5821. .5991)
                     /rpt_family="MIR"
     repeat_region   complement(5996. .6302)
                     /rpt_family="AluY"
     repeat_region   complement(6861. .7237)
                     /rpt_family="THE1C"
     repeat_region   complement(7310. .7464)
                     /rpt_family="MIR"
     repeat_region   complement(7747. .8057)
                     /rpt_family="AluSx"
     repeat_region   8529. .8705
                     /rpt_family="L1PA6"
     repeat_region   9084. .9410
                     /rpt_family="L2"
     repeat_region   complement(9411. .9638)
                     /rpt_family="MER20"
     repeat_region   9639. .10125
                     /rpt_family="L2"
     repeat_region   10220. .10254
                     /rpt_family="(CCTG)n"
     repeat_region   complement(10338. .10647)
                     /rpt_family="AluSg"
     repeat_region   10665. .10774
                     /rpt_family="L2"
     repeat_region   complement(10993. .11209)
                     /rpt_family="AluSp"
     repeat_region   complement(11351. .11385)
                     /rpt_family="L1M4"
     repeat_region   complement(11386. .11691)
                     /rpt_family="AluY"
     repeat_region   complement(11692. .12404)
                     /rpt_family="L1M4"
     repeat_region   12445. .12579
                     /rpt_family="L1MC/D"
     repeat_region   complement(12953. .13326)
                     /rpt_family="MLT1A2"
     repeat_region   13821. .14133
                     /rpt_family="AluSx"
     repeat_region   complement(14593. .15041)
                     /rpt_family="MSTB2"
     repeat_region   15071. .15309
                     /rpt_family="(TA)n"
     repeat_region   complement(15322. .15601)
                     /rpt_family="AluSx"
     repeat_region   16005. .16025
                     /rpt_family="AT_rich"
     repeat_region   complement(16166. .16331)
                     /rpt_family="L1MA8"
     repeat_region   16332. .16351
                     /rpt_family="(TTTTA)n"
     repeat_region   complement(16356. .16637)
                     /rpt_family="AluSx"
     repeat_region   complement(16638. .17414)
                     /rpt_family="L1MA8"
     repeat_region   complement(17533. .17649)
                     /rpt_family="L2"
     repeat_region   complement(18410. .18635)
                     /rpt_family="LTR33A"
     repeat_region   complement(19292. .19857)
                     /rpt_family="ERVL"
     repeat_region   19951. .19981
                     /rpt_family="AT_rich"
     repeat_region   complement(20154. .20380)
                     /rpt_family="HAL1"
     repeat_region   21493. .21515
                     /rpt_family="AT_rich"
     repeat_region   21543. .21586
                     /rpt_family="AT_rich"
     repeat_region   complement(22291. .23134)
                     /rpt_family="L1PA10"
     repeat_region   23135. .23607
                     /rpt_family="L1PA16"
     repeat_region   24001. .24031
                     /rpt_family="AT_rich"
     repeat_region   25070. .25122
                     /rpt_family="AT_rich"
     repeat_region   complement(25773. .25828)
                     /rpt_family="MIR"
     repeat_region   complement(25882. .25953)
                     /rpt_family="MIR"
```

```
     repeat_region   26198. .26457
                     /rpt_family="MIR"
     repeat_region   complement(26716. .26770)
                     /rpt_family="MER112"
     repeat_region   26906. .26928
                     /rpt_family="(TATG)n"
     repeat_region   27039. .27071

  Query Match             4.6%;  Score 280.2;  DB 9;  Length 152005;
  Best Local Similarity   71.5%;  Pred. No. 3.1e-51;
  Matches  460;  Conservative    0;  Mismatches  158;  Indels   25;  Gaps    6;

Qy        312 GTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTC 371
              |||||      ||  | ||||||  || ||   ||||||||||||||    |||| | ||
Db     110131 GTAGGAACAGGAAATTATGATACTATTCCTAACCCATCATAAGGGTCATGGCCAATATTC 110190

Qy        372 CTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTT 431
              ||||  |||| |||||||||  |||| ||| | ||| | ||||||||||| |||||||||
Db     110191 TTATAAAAAAATACAGGTTAACTAGAGAAAATCATAATACATTTATTTAACCAAAGTTTT 110250

Qy        432 ACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTT 491
              ||| ||||| |||| ||| ||||||||||||||||  ||    ||||    | | ||||
Db     110251 ACACGACATAGGAGCCTTGAGAAATGAAGACCCAAAAGCC---AGGAAGAACTTTCTTTT 110307

Qy        492 TTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACA-AAAGGA 550
              || ||||  |||| |||||||||||| |||       |    |||| |||| |||||
Db     110308 TTATGCTTGGGTTTGATGAAGAATGGACAGCTGTATAAAAATGTGATTGGACAGAAAGGG 110367

Qy        551 TATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCT 610
              ||| |||     | ||  |||| |||||||||   |||||| ||||||| ||||| ||||
Db     110368 TATAATCCGAGAGCAACAGACTGAGGGGGAAAGCAAGCAAGACCTGTCTGTTCAGCTTCT 110427

Qy        611 TCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCT 670
              |||||| |||||         ||| |||| ||  ||||||||  | | ||| |      |
Db     110428 TCTTGACCTCTC--------TGTGTAGCGTTTCTTCCTCCTAGGAAGGGGTC------CC 110473

Qy        671 TCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATGGCTT 730
              |||  |||||||||||||||||||||||||||| | || ||||| ||| ||||| || |
Db     110474 TCTGGAATGAGGGTCTTCAAGGGAGAAGGGAGAGAATGACCTTTCTAGGTTTTACGGGTG 110533

Qy        731 GCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCTGGTT 790
              |||| | | |||||||||||||||||||||||||||| |||||||||| |||| | ||| |
Db     110534 GCTTTGAGAAAGAGGAGTTCTAGTTTCTATGACCCACCTTGGGGAAGGGGAAGTGTGGCT 110593

Qy        791 TCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGGTCAG 850
              ||| |||||  ||||   |||||||||| || | ||| |||||||   ||||| | |  |
Db     110594 TCTATGACTCACTTTGGGGAAGAAAGAGAAGCAGGAGACAGGAGGATGGGAGAAGATGGG 110653

Qy        851 AAAGAGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAGTACTTCTTAGC 910
              |||||        || |||||||| ||| |    |   ||   | |||  | | || |||
Db     110654 AAAGA-----CCTTGGTTCTGAGGCCTTTCCAATTTTCTTCAGTTCAA--AGTACTCAGC 110706

Qy        911 ATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTA 953
              | ||||||||  | ||  ||||||  |||  | |||||||| | |
Db     110707 ACACCAAAGCGCCATGCTTTGGAGTATCATATTCTGAGCCCAA 110749


RESULT 32
BX284678/c
LOCUS       BX284678               27724 bp    DNA     linear   PRI 03-APR-2003
DEFINITION  Human DNA sequence from clone RP4-740O15 on chromosome 1, complete
            sequence.
ACCESSION   BX284678
VERSION     BX284678.3  GI:29539190
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 27724)
  AUTHORS   Harrison,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-APR-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Apr 3, 2003 this sequence version replaced gi:28881719.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            RP4-740O15 is from the library RPCI-4 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pCYPAC2.
FEATURES             Location/Qualifiers
     source          1. .27724
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP4-740O15"
                     /clone_lib="RPCI-4"
ORIGIN

  Query Match             4.6%;  Score 280;  DB 9;  Length 27724;
  Best Local Similarity   73.4%;  Pred. No. 3.9e-51;
  Matches  414;  Conservative    0;  Mismatches  140;  Indels   10;  Gaps    4;

Qy        301 TCAAAGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGC 360
              || |  || | ||||  | | ||| |||| |||  ||||||   |||||||||| ||
Db        576 TCTACCCTCCTGTAGAAGAGGAAAGGTGTAATAACTTTTCTTACCCATCATAAGGATCAT 517

Qy        361 AACCAAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTA 420
              ||||  |||||||||  ||||| || |  || ||||| ||| | | ||||||  ||||||
Db        516 GGCCAACACTCCTATAACAAAAGGCAAGAGAACAAGAGAAAAGCATGACAAATGTATTTA 457

Qy        421 ATCAAAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAA 480
              | ||||||||||  |||||||||| ||| |||||| |||| ||||||||||||||| |||
Db        456 ACCAAAGTTTTATGTGACATGGGAGCCTTGAGAAATAAAGAGCCAAAGACCCAGGGAAAA 397

Qy        481 CTGTCTGTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTA 540
              ||||||  |||||  |||| ||||| |||||||||  || |||         |    |||||
Db        396 TTGTCTG-TTTTTATGCTTAGGTTTGATGAAGAACTGACAGCCATATAGAAATGTGATTA 338

Qy        541 GACAAAAGGATATGATCTAGTGGTAAAGGACTCA-GGGGGAAACACAGCAAGGCCTGTCT 599
              ||||||| | |||||||| || |||  |||| |  || ||||| |||| ||||||| ||
Db        337 GACAAAAAGGTATGATCTAATGCTAACAGACTGACAGGAGAAACCCAGCCAGGCCTGCCT 278

Qy        600 ATTCAGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGG 659
              |    | |         ||||   |||  | | || ||||| |||||||| |||||
Db        277 TT----CTGTTCAGATTCTTCTCGGCCTCCCTATGTAACATTCCTTCCTCCTGGGTATGA 222
```

```
Qy        660 GGCAGGACTCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGA 719
              ||| |||||| ||| |||||| |||||||| |||||| | ||||| ||| |||
Db        221 GGCTGGACTCCTCTGGAATGAGATTCTTCAAGACAGAAGGAGAAGAGTGGACTTCCTAGG 162

Qy        720 TTTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAG 779
              ||||||||| ||||| | |||||||||||||| |||||||||||||| ||||||||||||
Db        161 TTTTATGGCATGCTTGGAGGAAGAGGAGTTCTGGTTTCTATGACCCACCTTGGGGAAGAG 102

Qy        780 GAATTCTGGTTTCTGTGAC----TTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGG 835
              || |||||||||| ||||    || |||    ||||||| ||    | |||  ||||||
Db        101 GAGTTCTGGTTTCTATGACTCATTTACTTAGGCGAAGAAAAAGAGACAGGAGATAGGAGG 42

Qy        836 GCAGGAGATGGTCAGAAAGAGACT 859
              | ||||   ||||||| | |  ||
Db         41 GTAGGAAGAGGTCAGAGAAACCCT 18


RESULT 33
CQ434203
LOCUS       CQ434203                1238 bp    DNA     linear   PAT 28-JAN-2004
DEFINITION  Sequence 19237 from Patent WO0151628.
ACCESSION   CQ434203
VERSION     CQ434203.1  GI:41386432
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Lillie,J., Xu,Y., Wang,Y. and Steinmann,K.
  TITLE     Novel genes, compositions, kits, and methods for identification,
            assessment, prevention, and therapy of breast cancer
  JOURNAL   Patent: WO 0151628-A 19237 19-JUL-2001;
            Millennium Pharmaceuticals, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .1238
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

  Query Match             4.6%;  Score 277.6;  DB 6;  Length 1238;
  Best Local Similarity   98.6%;  Pred. No. 1.6e-50;
  Matches  280;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       5345 CTTTCAGGGGCTGGAGAGGCAGAGGGAGACAGAGGAGCTGGTACTGCAGAGCGGTCGTCT 5404
              ||   | ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         58 CTCCTAAGGGCTGGAGAGGCAGAGGGAGACAGAGGAGCTGGTACTGCAGAGCGGTCGTCT 117

Qy       5405 GATTGGCTGGACGGTCGTAGCTGGGCTATAAAAGAGACCCCTACAGGCTTAGCAGGAAGA 5464
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        118 GATTGGCTGGACGGTCGTAGCTGGGCTATAAAAGAGACCCCTACAGGCTTAGCAGGAAGA 177

Qy       5465 CGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGCCGA 5524
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        178 CGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGCCGA 237

Qy       5525 AGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAATAAG 5584
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        238 AGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAATAAG 297

Qy       5585 ATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
              |||||||||||||||||||||||||||||||||||||||||||
Db        298 ATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 341


RESULT 34
AL590392
LOCUS       AL590392              104011 bp    DNA     linear   PRI 03-JUL-2001
DEFINITION  Human DNA sequence from clone RP11-453D21 on chromosome 6, complete
            sequence.
ACCESSION   AL590392
VERSION     AL590392.5  GI:14596401
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 104011)
  AUTHORS   Almeida,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 4, 2001 this sequence version replaced gi:14455961.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            RP11-453D21 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-453D21 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true left end of clone RP1-249N8 is at 102012 in this sequence.
            The true right end of clone RP3-492L20 is at 2000 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .104011
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP11-453D21"
                     /clone_lib="RPCI-11.2"
     repeat_region   402. .445
                     /note="22 copies 2 mer tt 75% conserved"
     repeat_region   692. .987
                     /note="AluSx repeat: matches 1. .301 of consensus"
     repeat_region   1153. .1446
                     /note="AluSq repeat: matches 1. .293 of consensus"
     repeat_region   1554. .1754
                     /note="MER53 repeat: matches 2. .189 of consensus"
     repeat_region   2442. .2473
                     /note="16 copies 2 mer tg 90% conserved"
     repeat_region   2669. .2968
                     /note="AluJb repeat: matches 1. .299 of consensus"
     repeat_region   2970. .3009
                     /note="20 copies 2 mer ac 100% conserved"
     repeat_region   3044. .3155
                     /note="L2 repeat: matches 2370. .2485 of consensus"
     repeat_region   4811. .5165
                     /note="MLT1A2 repeat: matches 1. .360 of consensus"
     repeat_region   5236. .5528
                     /note="AluY repeat: matches 3. .292 of consensus"
     repeat_region   5534. .5676
```

```
                     /note="L1ME3A repeat: matches 5927. .6096 of consensus"
  repeat_region      5690. .5734
                     /note="L1ME repeat: matches 5691. .5735 of consensus"
  repeat_region      5739. .6036
                     /note="L1ME repeat: matches 5376. .5681 of consensus"
  repeat_region      6688. .7187
                     /note="LTR7 repeat: matches 1. .448 of consensus"
  repeat_region      8526. .8557
                     /note="16 copies 2 mer tt 84% conserved"
  repeat_region      8729. .8833
                     /note="MER5A repeat: matches 84. .188 of consensus"
  repeat_region      11073. .11225
                     /note="LTR39 repeat: matches 630. .780 of consensus"
  repeat_region      11226. .11517
                     /note="AluSp repeat: matches 1. .295 of consensus"
  repeat_region      11518. .12133
                     /note="LTR39 repeat: matches 1. .630 of consensus"
  repeat_region      14251. .14389
                     /note="L2 repeat: matches 2601. .2744 of consensus"
  repeat_region      14700. .14991
                     /note="AluSx repeat: matches 3. .293 of consensus"
  repeat_region      17500. .17675
                     /note="FRAM repeat: matches -3. .171 of consensus"
  repeat_region      17681. .17980
                     /note="HAL1 repeat: matches 1211. .1531 of consensus"
  repeat_region      18192. .18239
                     /note="2 copies 24 mer 100% conserved"
  repeat_region      18242. .18577
                     /note="HAL1 repeat: matches 572. .931 of consensus"
  repeat_region      18580. .18945
                     /note="MER39 repeat: matches 12. .380 of consensus"
  repeat_region      18954. .19136
                     /note="MER39b repeat: matches 356. .549 of consensus"
  repeat_region      19539. .19842
                     /note="AluSc repeat: matches 1. .303 of consensus"
  repeat_region      20829. .20917
                     /note="MIR repeat: matches 56. .139 of consensus"
  repeat_region      21625. .21664
                     /note="MIR repeat: matches 104. .144 of consensus"
  repeat_region      22562. .22772
                     /note="L1M4 repeat: matches 3813. .4025 of consensus"
  repeat_region      26833. .26895
                     /note="L2 repeat: matches 2698. .2750 of consensus"
  repeat_region      26947. .27435
                     /note="LOR1a repeat: matches 1. .497 of consensus"
  repeat_region      28184. .28631
                     /note="MLT1C repeat: matches 11. .466 of consensus"
  repeat_region      29097. .29124
                     /note="7 copies 4 mer tcat 92% conserved"
  repeat_region      29181. .29491
                     /note="L2 repeat: matches 2340. .2614 of consensus"
  repeat_region      30030. .30190
                     /note="L2 repeat: matches 1693. .1853 of consensus"
  repeat_region      30219. .30404
                     /note="MIR repeat: matches 13. .255 of consensus"
  repeat_region      31530. .31628
                     /note="MER5A repeat: matches 9. .109 of consensus"
  repeat_region      32698. .33016
                     /note="AluSx repeat: matches 1. .303 of consensus"
  repeat_region      33915. .34079
                     /note="LTR41 repeat: matches 597. .767 of consensus"
  repeat_region      34139. .34260
                     /note="L2 repeat: matches 2360. .2484 of consensus"
  repeat_region      35696. .36022
                     /note="AluSx repeat: matches 1. .309 of consensus"
  repeat_region      36904. .37000
                     /note="L1M4c repeat: matches 1494. .1590 of consensus"
  repeat_region      37629. .37923
                     /note="AluJb repeat: matches 1. .299 of consensus"
  repeat_region      37925. .38219
                     /note="AluSx repeat: matches 1. .303 of consensus"
  repeat_region      39379. .40277
                     /note="L1M4 repeat: matches 4661. .5649 of consensus"
  repeat_region      40855. .40967
                     /note="MER20 repeat: matches 53. .179 of consensus"
  repeat_region      40968. .41018
                     /note="MLT2B repeat: matches 394. .446 of consensus"
  repeat_region      41086. .41481
                     /note="MLT2B repeat: matches 1. .403 of consensus"
  repeat_region      42872. .42918
                     /note="L2 repeat: matches 2701. .2747 of consensus"
  repeat_region      43005. .43129
                     /note="FLAM_C repeat: matches 1. .125 of consensus"
  repeat_region      43452. .43492
                     /note="L2 repeat: matches 2358. .2398 of consensus"
  repeat_region      43746. .44022
                     /note="AluSq repeat: matches 6. .283 of consensus"
  repeat_region      45239. .45679
                     /note="LTR1 repeat: matches 294. .738 of consensus"
  repeat_region      45762. .45945
                     /note="LTR1 repeat: matches 1. .194 of consensus"
  repeat_region      46327. .46901
                     /note="LTR8 repeat: matches 109. .691 of consensus"
  repeat_region      46902. .47082
                     /note="AluSg repeat: matches 1. .181 of consensus"
  repeat_region      47083. .47192
                     /note="LTR8 repeat: matches 1. .109 of consensus"
  repeat_region      48024. .48170
                     /note="MER5A repeat: matches 35. .189 of consensus"
  repeat_region      48573. .48694
                     /note="MIR repeat: matches 20. .143 of consensus"
  repeat_region      51579. .51668
                     /note="MIR repeat: matches 157. .256 of consensus"
  repeat_region      51675. .51763
                     /note="MIR repeat: matches 168. .262 of consensus"
  repeat_region      56121. .56259
                     /note="L1MB7 repeat: matches 6033. .6171 of consensus"
  repeat_region      56693. .56730
                     /note="19 copies 2 mer tt 78% conserved"
  repeat_region      57084. .57129
                     /note="23 copies 2 mer aa 78% conserved"
  repeat_region      57091. .57130
                     /note="10 copies 4 mer aaat 97% conserved"
  repeat_region      57818. .58123
                     /note="AluSx repeat: matches 1. .307 of consensus"
  repeat_region      60495. .60615
                     /note="L2 repeat: matches 2625. .2750 of consensus"
  repeat_region      61012. .61271
                     /note="LTR1 repeat: matches 1. .260 of consensus"
  repeat_region      61272. .61668
                     /note="LTR28 repeat: matches 351. .817 of consensus"
  repeat_region      63738. .63868
                     /note="MIR repeat: matches 98. .241 of consensus"

 Query Match             4.5%;  Score 276.2;  DB 9;  Length 104011;
 Best Local Similarity   71.8%;  Pred. No. 2.4e-50;
 Matches  453;  Conservative    0;  Mismatches  153;  Indels   25;  Gaps    6;

Qy        285 TATCTCTCCTTCTCTCTCAAAGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTC-TCT 343
              ||| |     | || |   ||    | | |||||   ||| |||||||||| | ||| ||
Db      18555 TATTTTCAAATTTCACAACAATGCCCTGCAGGGGAAGAAAGGTGTGATACCTGTTCGTCA 18614

Qy        344 TCCATCATAAGGGTCGCAACCAAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAA 403
              ||||| || | ||     ||||| | |||||||  ||||||  ||||||  |||   |||
Db      18615 CCCATCCTAGGAGTTATGACCAACATTCCTATAACAAAAGATGGGTTAACAAGCAAAAAG 18674

Qy        404 CCTAACAAATTTATTTAATCAAAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACC 463
              ||||||||||||||||||||||||||||  |||||  |||   ||||||||||||||||
Db      18675 CCTAACAAATTTATTTAATCAAAGTTTTATGTGACAGAGGAACTTTCAGAAATGAAGACC 18734

Qy        464 CAAAGACCCAGGGGAAACTGTCTGTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCA 523
              |||||||||||| | ||||     ||||| |||| |||||| |||||||| || |||
Db      18735 GAAAGACCCAGGGAACACTG-----TTTTTATGCTTAGGTTCAATGAAGAATAGACAGCT 18789

Qy        524 TGTAGCCATGTAGATTAGACAAAAGGATATGATCTAGTGGTAAAGGACTCAG-GGGGAAA 582
                  |    |||| ||||||||  || ||||||  ||||||  |||| || ||||||
```

```
Db      18790 GTGTAGAAATGGGATTGGACAAAAGTGTAAGATCTAATGGTAACAGACTGAGCGGGGAAG 18849

Qy        583 CACAGCAAGGCCTGTCTATTCAGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTC 642

Db      18850 CCCAGCAAAGCCTGTCTGCTCAGATGCTTCTTGGT---------TCTCTGGGTAGCATTC 18900

Qy        643 TTTCCTCCTGAGTATGGGGCAGGACTCTTCTTCAATGAGGGTCTTCAAGGGAGAA----G 698

Db      18901 CTTCCTCCTGGGTATGGGGTGGAACCCCTCTGGAATGAGGGTCTTCAAGGGAGAAGAGGG 18960

Qy        699 GGAGAAAGTGGCCTTTTTAGATTTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCT 758

Db      18961 AGGGAGAGTGAGCTTTCTAGGTTTTATGGCTTGCTTTGGGGGAGGGCAGGTCTAGTTTCT 19020

Qy        759 ATGACCCATCTTGGGGAAGAGGAATTCTGGTTTCTGTGACTTGCTTTCATGAAGAAAGAG 818

Db      19021 ACAACCTGCCTTGAGGAAGAGGAATTGTGGCTTCTGTGATTTATTTCAGGGGAGAAAGGT 19080

Qy        819 GAGTAAGAGGCAGGAGGGCAGGAGATGGTCAGAAAGAGACTTGGCTGCTTCTGAGGGCTT 878

Db      19081 GGCAGGAAGGCAGAAGGGCAGGAGGAGGGCAGATCGA-----CCTTGTTTCTGAGACCTT 19135

Qy        879 CCGCTCTCCTTTAGTTCCAAGTACTTCTTAG 909

Db      19136 CCCACCTCCTTCAGTTCAAAGTACTCAGTAG 19166


RESULT 35
AP001982
LOCUS       AP001982               137996 bp    DNA     linear   PRI 15-MAR-2003
DEFINITION  Homo sapiens genomic DNA, chromosome 11q, clone:RP11-344F5,
            complete sequence.
ACCESSION   AP001982
VERSION     AP001982.4  GI:17425236
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens genomic DNA
  JOURNAL   Published Only in Database (2000)
REFERENCE   2  (bases 1 to 137996)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-MAY-2000) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     On Dec 7, 2001 this sequence version replaced gi:9798619.
FEATURES             Location/Qualifiers
     source          1. .137996
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q"
                     /clone="RP11-344F5"
ORIGIN

  Query Match             4.5%;  Score 275.8;  DB 9;  Length 137996;
  Best Local Similarity   70.1%;  Pred. No. 2.9e-50;
  Matches  452;  Conservative    0;  Mismatches  167;  Indels   26;  Gaps    5;

Qy        312 GTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTC 371

Db      79303 GTAGGAGGAGAACTATGTGATGCCTTTCCTCACCCATCATAAGGGTCACGACTGACATTT 79362

Qy        372 CTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTT 431

Db      79363 TTATAACAAAAGACAGGTTAACCAGAGAAAAGCATAACAATTTTATTTAATTGAAGTTGT 79422

Qy        432 ACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTT 491

Db      79423 ACATGACACAGGAGCCTTTAGGATGAAGACCCAAGA---CCTGGGAAAACT-----ATTT 79474

Qy        492 TTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGAT 551

Db      79475 TTATGCTTAGGTTCAGTGAAGAACAAACGGCCATGTAGAAATGTGACCAAAAAACTGTGT 79534

Qy        552 ATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTT 611

Db      79535 TCTAATGGTAAGAGACCGAGTGGAGAAGGCCAGCCGCCAGGACTGTGTGTTCAGATTCTT 79594

Qy        612 CTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTT 671

Db      79595 CCTGGTTTTTC--------TGTGTAGCATTCTTTCCTCCTGGGTGTCGGGCGGGACCCCT 79646

Qy        672 CTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATGGCTTG 731

Db      79647 CTGGAATGAGGGTCTTCAAAGGAGAAGGGAGAGAGTGATCTTTGTAGGTTTTATGGCTTG 79706

Qy        732 CTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCTGGTTT 791

Db      79707 CTTTGGGGAGGAGAAGTTCTAGTTTCTACGACCTGTGTTGGGGAAGAGGAATTCTGGTTT 79766

Qy        792 CTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGGTCAGA 851

Db      79767 GTAGGACTCGCTTTGGTGGAGAAAGAGGGGCAAGAGACAGGAGGGCAGGAGAAGATCGGA 79826

Qy        852 AAGAGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAGTACTTCTTAGCA 911

Db      79827 AA-------GTTTGCTTCTGAGATCTTCCAGTCTCCTTTAGTTTGAAGTA---TTCGGTG 79876

Qy        912 TACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAACA 956

Db      79877 TATCAAAGCACCAGACTTTCGAGTATTGTCTTCTGAGCTCCAACA 79921


RESULT 36
AP002011
LOCUS       AP002011               160800 bp    DNA     linear   HTG 30-MAY-2000
DEFINITION  Homo sapiens chromosome 11 clone RP11-96N14 map 11q, WORKING DRAFT
            SEQUENCE, 21 unordered pieces.
ACCESSION   AP002011
VERSION     AP002011.2  GI:8117629
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 160800)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens 160,800 genomic DNA of 11q
  JOURNAL   Published Only in DataBase (2000)
REFERENCE   2  (bases 1 to 160800)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-MAY-2000) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
            Japan (E-mail:hattori@gsc.riken.go.jp,
            URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
            Fax:81-42-778-9924)
COMMENT     On May 30, 2000 this sequence version replaced gi:7768936.
            -------------- Genome Center
                Center: RIKEN Genomic Sciences Center(GSC)
                Center code: RIKEN
                Web site: http://hgp.gsc.riken.go.jp/
                Contact: hattori@gsc.riken.go.jp
```

```
------------- Project Information
    Center project name: HumDraft11
    Center clone name: RP11-96N14
------------- Summary Statistics
    Sequencing vector: PCR products; 100% of reads
    Chemistry: Dye-terminator ET-amersham; 100% of reads
    Assembly program: Phrap; version 0.990329
    Consensus quality: 153477 bases at least Q40
    Consensus quality: 156297 bases at least Q30
    Consensus quality: 157655 bases at least Q20
    Insert size: 158800; sum-of-contigs
    Quality coverage: 6.37x in Q20 bases; sum-of-contigs
-------------.
NOTE: This is a 'working draft' sequence. It currently consists of
21 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved
         1     25551 contig of  25551 bp in length
     25652     39527 contig of  13876 bp in length
     39628     54679 contig of  15052 bp in length
     54780     66520 contig of  11741 bp in length
     66621     79378 contig of  12758 bp in length
     79479     89839 contig of  10361 bp in length
     89940     99761 contig of   9822 bp in length
     99862    107801 contig of   7940 bp in length
    107902    114450 contig of   6549 bp in length
    114551    121525 contig of   6975 bp in length
    121626    127312 contig of   5687 bp in length
    127413    132396 contig of   4984 bp in length
    132497    137634 contig of   5138 bp in length
    137735    141989 contig of   4255 bp in length
    142090    146645 contig of   4556 bp in length
    146746    150112 contig of   3367 bp in length
    150213    153851 contig of   3639 bp in length
    153952    156227 contig of   2276 bp in length
    156328    158363 contig of   2036 bp in length
    158464    159557 contig of   1094 bp in length
    159658    160800 contig of   1143 bp in length
Sequence updated (26-May-2000).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*         1     25551: contig of 25551 bp in length
*     25552     25651: gap of 100 bp
*     25652     39527: contig of 13876 bp in length
*     39528     39627: gap of 100 bp
*     39628     54679: contig of 15052 bp in length
*     54680     54779: gap of 100 bp
*     54780     66520: contig of 11741 bp in length
*     66521     66620: gap of 100 bp
*     66621     79378: contig of 12758 bp in length
*     79379     79478: gap of 100 bp
*     79479     89839: contig of 10361 bp in length
*     89840     89939: gap of 100 bp
*     89940     99761: contig of 9822 bp in length
*     99762     99861: gap of 100 bp
*     99862    107801: contig of 7940 bp in length
*    107802    107901: gap of 100 bp
*    107902    114450: contig of 6549 bp in length
*    114451    114550: gap of 100 bp
*    114551    121525: contig of 6975 bp in length
*    121526    121625: gap of 100 bp
*    121626    127312: contig of 5687 bp in length
*    127313    127412: gap of 100 bp
*    127413    132396: contig of 4984 bp in length
*    132397    132496: gap of 100 bp
*    132497    137634: contig of 5138 bp in length
*    137635    137734: gap of 100 bp
*    137735    141989: contig of 4255 bp in length
*    141990    142089: gap of 100 bp
*    142090    146645: contig of 4556 bp in length
*    146646    146745: gap of 100 bp
*    146746    150112: contig of 3367 bp in length
*    150113    150212: gap of 100 bp
*    150213    153851: contig of 3639 bp in length
*    153852    153951: gap of 100 bp
*    153952    156227: contig of 2276 bp in length
*    156228    156327: gap of 100 bp
*    156328    158363: contig of 2036 bp in length
*    158364    158463: gap of 100 bp
*    158464    159557: contig of 1094 bp in length
*    159558    159657: gap of 100 bp
*    159658    160800: contig of 1143 bp in length.
FEATURES             Location/Qualifiers
     source          1. .160800
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q"
                     /clone="RP11-96N14"
     misc_feature    1. .25551
                     /note="assembly_fragment"
     misc_feature    25652. .39527
                     /note="assembly_fragment"
     misc_feature    39628. .54679
                     /note="assembly_fragment"
     misc_feature    54780. .66520
                     /note="assembly_fragment"
     misc_feature    66621. .79378
                     /note="assembly_fragment"
     misc_feature    79479. .89839
                     /note="assembly_fragment"
     misc_feature    89940. .99761
                     /note="assembly_fragment"
     misc_feature    99862. .107801
                     /note="assembly_fragment"
     misc_feature    107902. .114450
                     /note="assembly_fragment"
     misc_feature    114551. .121525
                     /note="assembly_fragment"
     misc_feature    121626. .127312
                     /note="assembly_fragment"
     misc_feature    127413. .132396
                     /note="assembly_fragment"
     misc_feature    132497. .137634
                     /note="assembly_fragment"
     misc_feature    137735. .141989
                     /note="assembly_fragment"
     misc_feature    142090. .146645
                     /note="assembly_fragment"
     misc_feature    146746. .150112
                     /note="assembly_fragment"
     misc_feature    150213. .153851
                     /note="assembly_fragment clone_end:T7 vector_side:right"
     misc_feature    153952. .156227
                     /note="assembly_fragment"
     misc_feature    156328. .158363
                     /note="assembly_fragment clone_end:SP6 vector_side:left"
     misc_feature    158464. .159557
                     /note="assembly_fragment"
     misc_feature    159658. .160800
                     /note="assembly_fragment"
ORIGIN

  Query Match          4.5%;  Score 275.8;  DB 2;  Length 160800;
  Best Local Similarity   70.1%;  Pred. No. 2.8e-50;
  Matches  452;  Conservative    0;  Mismatches  167;  Indels   26;  Gaps    5;
```

```
Qy        312 GTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTC 371
              ||||| |   ||   |||||| | ||| |||  ||||||||||||| | ||  | | |
Db      86525 GTAGGAGGAGAACTATGTGATGCCTTTCCTCACCCATCATAAGGGTCACGACTGACATTT 86584

Qy        372 CTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTT 431
              ||||  |||||||||||||  |||| ||| | ||||||  |||||||||  |||||
Db      86585 TTATAACAAAAGACAGGTTAACCAGAGAAAAGCATAACAATTTTATTTAATTGAAGTTGT 86644

Qy        432 ACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTT 491
              ||||||||  ||||  ||| || |   |   || | |   || ||| |||||      |||
Db      86645 ACATGACACAGGAGCCTTTAGGATGAAGACCCAAGA---CCTGGGAAAACT-----ATTT 86696

Qy        492 TTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGAT 551
              || |||| ||||||  |||||||   |  ||        |   ||  | | ||  |  |
Db      86697 TTATGCTTAGGTTCAGTGAAGAACAAACGGCCATGTAGAAATGTGACCAAAAAACTGTGT 86756

Qy        552 ATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTT 611
                |    |  |  || |   |  |     | || ||| |||| | ||||||||||||
Db      86757 TCTAATGGTAAGAGACCGAGTGGAGAAGGCCAGCCGCCAGGACTGTGTGTTCAGATTCTT 86816

Qy        612 CTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTT 671
              | || | | ||         ||| ||||||||||||||||| || | |||| |||| | |
Db      86817 CCTGGTTTTTC--------TGTGTAGCATTCTTTCCTCCTGGGTGTCGGGCGGGACCCCT 86868

Qy        672 CTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATGGCTTG 731
              ||  |||||||||||||| |||||||||||| ||||  |||| ||| ||||||||||||
Db      86869 CTGGAATGAGGGTCTTCAAAGGAGAAGGGAGAGAGTGATCTTTGTAGGTTTTATGGCTTG 86928

Qy        732 CTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCTGGTTT 791
              ||| ||||| ||| ||||||||||||||| ||||  | ||||||||||||||||||||
Db      86929 CTTTGGGGAGGAGAAGTTCTAGTTTCTACGACCTGTGTTGGGGAAGAGGAATTCTGGTTT 86988

Qy        792 CTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGGTCAGA 851
              | ||||  |||||  || |||||||| | ||||| ||||||||||||||| | || ||
Db      86989 GTAGGACTCGCTTTGGTGGAGAAAGAGGGGCAAGAGACAGGAGGGCAGGAGAAGATCGGA 87048

Qy        852 AAGAGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAGTACTTCTTAGCA 911
              ||        | ||||||||||  |||||  ||||||||||||  |||||    | |
Db      87049 AA-------GTTTGCTTCTGAGATCTTCCAGTCTCCTTTAGTTTGAAGTA---TTCGGTG 87098

Qy        912 TACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAACA 956
              || |||||||| | ||||| |  ||| || |||||||||| ||||
Db      87099 TATCAAAGCACCAGACTTTCGAGTATTGTCTTCTGAGCTCCAACA 87143
```

```
RESULT 37
AP002763/c
LOCUS       AP002763               162584 bp    DNA     linear   HTG 13-JUL-2000
DEFINITION  Homo sapiens chromosome 11 clone RP11-812L16 map 11q, WORKING DRAFT
            SEQUENCE, 23 unordered pieces.
ACCESSION   AP002763
VERSION     AP002763.1  GI:9188601
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 162584)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens 162,584 genomic DNA of 11q
  JOURNAL   Published Only in DataBase (2000)
REFERENCE   2  (bases 1 to 162584)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-42-778-9923, Fax:81-42-778-9924)
COMMENT     -------------- Genome Center
            Center: RIKEN Genomic Sciences Center(GSC)
            Center code: RIKEN
            Web site: http://hgp.gsc.riken.go.jp/
            Contact: hattori@gsc.riken.go.jp
          -------------- Project Information
            Center project name: HumDraft11
            Center clone name: RP11-812L16
          -------------- Summary Statistics
            Sequencing vector: PCR products; 100% of reads
            Chemistry: Dye-terminator ET-amersham; 100% of reads
            Assembly program: Phrap; version 0.990329
            Consensus quality: 148478 bases at least Q40
            Consensus quality: 155428 bases at least Q30
            Consensus quality: 158571 bases at least Q20
            Insert size: 160384; sum-of-contigs
            Quality coverage: 4.77x in Q20 bases; sum-of-contigs
          --------------.
          NOTE: This is a 'working draft' sequence. It currently consists of
          23 contigs. The true order of the pieces is not known and their
          order in this sequence record is arbitrary. Gaps between the
          contigs are represented as runs N, but the exact sizes of the gaps
          are unknown. This record will be updated with the finished sequence
          as soon as it is available and the accession number will be
          preserved
                1     19158 contig of  19158 bp in length
            19259     35737 contig of  16479 bp in length
            35838     46375 contig of  10538 bp in length
            46476     57728 contig of  11253 bp in length
            57829     68212 contig of  10384 bp in length
            68313     78745 contig of  10433 bp in length
            78846     86727 contig of   7882 bp in length
            86828     93406 contig of   6579 bp in length
            93507    100114 contig of   6608 bp in length
           100215    106482 contig of   6268 bp in length
           106583    112453 contig of   5871 bp in length
           112554    117886 contig of   5333 bp in length
           117987    124003 contig of   6017 bp in length
           124104    130126 contig of   6023 bp in length
           130227    136615 contig of   6389 bp in length
           136716    141804 contig of   5089 bp in length
           141905    145754 contig of   3850 bp in length
           145855    148951 contig of   3097 bp in length
           149052    153206 contig of   4155 bp in length
           153307    156779 contig of   3473 bp in length
           156880    158578 contig of   1699 bp in length
           158679    160876 contig of   2198 bp in length
           160977    162584 contig of   1608 bp in length.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 23 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1     19158: contig of 19158 bp in length
          *    19159     19258: gap of 100 bp
          *    19259     35737: contig of 16479 bp in length
          *    35738     35837: gap of 100 bp
          *    35838     46375: contig of 10538 bp in length
          *    46376     46475: gap of 100 bp
          *    46476     57728: contig of 11253 bp in length
          *    57729     57828: gap of 100 bp
          *    57829     68212: contig of 10384 bp in length
          *    68213     68312: gap of 100 bp
          *    68313     78745: contig of 10433 bp in length
          *    78746     78845: gap of 100 bp
          *    78846     86727: contig of 7882 bp in length
          *    86728     86827: gap of 100 bp
          *    86828     93406: contig of 6579 bp in length
          *    93407     93506: gap of 100 bp
          *    93507    100114: contig of 6608 bp in length
          *   100115    100214: gap of 100 bp
```

```
          *   100215   106482: contig of 6268 bp in length
          *   106483   106582: gap of 100 bp
          *   106583   112453: contig of 5871 bp in length
          *   112454   112553: gap of 100 bp
          *   112554   117886: contig of 5333 bp in length
          *   117887   117986: gap of 100 bp
          *   117987   124003: contig of 6017 bp in length
          *   124004   124103: gap of 100 bp
          *   124104   130126: contig of 6023 bp in length
          *   130127   130226: gap of 100 bp
          *   130227   136615: contig of 6389 bp in length
          *   136616   136715: gap of 100 bp
          *   136716   141804: contig of 5089 bp in length
          *   141805   141904: gap of 100 bp
          *   141905   145754: contig of 3850 bp in length
          *   145755   145854: gap of 100 bp
          *   145855   148951: contig of 3097 bp in length
          *   148952   149051: gap of 100 bp
          *   149052   153206: contig of 4155 bp in length
          *   153207   153306: gap of 100 bp
          *   153307   156779: contig of 3473 bp in length
          *   156780   156879: gap of 100 bp
          *   156880   158578: contig of 1699 bp in length
          *   158579   158678: gap of 100 bp
          *   158679   160876: contig of 2198 bp in length
          *   160877   160976: gap of 100 bp
          *   160977   162584: contig of 1608 bp in length.
FEATURES             Location/Qualifiers
     source          1. .162584
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q"
                     /clone="RP11-812L16"
     misc_feature    1. .19158
                     /note="assembly_fragment clone_end:T7 vector_side:left"
     misc_feature    19259. .35737
                     /note="assembly_fragment"
     misc_feature    35838. .46375
                     /note="assembly_fragment"
     misc_feature    46476. .57728
                     /note="assembly_fragment"
     misc_feature    57829. .68212
                     /note="assembly_fragment"
     misc_feature    68313. .78745
                     /note="assembly_fragment"
     misc_feature    78846. .86727
                     /note="assembly_fragment"
     misc_feature    86828. .93406
                     /note="assembly_fragment"
     misc_feature    93507. .100114
                     /note="assembly_fragment"
     misc_feature    100215. .106482
                     /note="assembly_fragment"
     misc_feature    106583. .112453
                     /note="assembly_fragment"
     misc_feature    112554. .117886
                     /note="assembly_fragment"
     misc_feature    117987. .124003
                     /note="assembly_fragment"
     misc_feature    124104. .130126
                     /note="assembly_fragment"
     misc_feature    130227. .136615
                     /note="assembly_fragment"
     misc_feature    136716. .141804
                     /note="assembly_fragment"
     misc_feature    141905. .145754
                     /note="assembly_fragment"
     misc_feature    145855. .148951
                     /note="assembly_fragment"
     misc_feature    149052. .153206
                     /note="assembly_fragment"
     misc_feature    153307. .156779
                     /note="assembly_fragment"
     misc_feature    156880. .158578
                     /note="assembly_fragment clone_end:SP6 vector_side:left"
     misc_feature    158679. .160876
                     /note="assembly_fragment"
     misc_feature    160977. .162584
                     /note="assembly_fragment"
ORIGIN

  Query Match             4.5%;  Score 275.8;  DB 2;  Length 162584;
  Best Local Similarity   70.1%;  Pred. No. 2.8e-50;
  Matches  452;  Conservative    0;  Mismatches  167;  Indels   26;  Gaps    5;

Qy        312 GTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTC 371
              ||||| |   ||   |||||| | ||| |||  |||||||||||||| | ||  | | |
Db      61162 GTAGGAGGAGAACTATGTGATGCCTTTCCTCACCCATCATAAGGGTCACGACTGACATTT 61103

Qy        372 CTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTT 431
              ||||  |||||||||||||||  |||| ||| | |||||| ||||||||||  ||||| |
Db      61102 TTATAACAAAAGACAGGTTAACCAGAGAAAAGCATAACAATTTTATTTAATTGAAGTTGT 61043

Qy        432 ACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTT 491
              ||||||||  |||| ||| || |   |   || | |   || ||| |||||      |||
Db      61042 ACATGACACAGGAGCCTTTAGGATGAAGACCCAAGA---CCTGGGAAAACT-----ATTT 60991

Qy        492 TTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGAT 551
              || |||| ||||||  |||||||   | ||          |    ||  | | ||  | |
Db      60990 TTATGCTTAGGTTCAGTGAAGAACAAACGGCCATGTAGAAATGTGACCAAAAAACTGTGT 60931

Qy        552 ATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTT 611
              |     |  |  || |   |  |    | || ||| |||| | ||||||||||
Db      60930 TCTAATGGTAAGAGACCGAGTGGAGAAGGCCAGCCGCCAGGACTGTGTGTTCAGATTCTT 60871

Qy        612 CTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTT 671
              | || | | ||        ||| ||||||||||||||||||| || | |||| |||| | |
Db      60870 CCTGGTTTTTC--------TGTGTAGCATTCTTTCCTCCTGGGTGTCGGGCGGGACCCCT 60819

Qy        672 CTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATGGCTTG 731
              ||  |||||||||||||| |||||||||||| ||||  |||| ||| ||||||||||||
Db      60818 CTGGAATGAGGGTCTTCAAAGGAGAAGGGAGAGAGTGATCTTTGTAGGTTTTATGGCTTG 60759

Qy        732 CTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCTGGTTT 791
              ||| ||||| ||| |||||||||||||| ||||  | ||||||||||||||||||||||
Db      60758 CTTTGGGGAGGAGAAGTTCTAGTTTCTACGACCTGTGTTGGGGAAGAGGAATTCTGGTTT 60699

Qy        792 CTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGGTCAGA 851
              |  |||| ||||  || |||||||| | ||||| |||||||||||||| | || ||
Db      60698 GTAGGACTCGCTTTGGTGGAGAAAGAGGGGCAAGAGACAGGAGGGCAGGAGAAGATCGGA 60639

Qy        852 AAGAGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAGTACTTCTTAGCA 911
              ||        | ||||||||||  |||||  |||||||||||  |||||    |  |
Db      60638 AA-------GTTTGCTTCTGAGATCTTCCAGTCTCCTTTAGTTTGAAGTA---TTCGGTG 60589

Qy        912 TACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAACA 956
              || |||||||| | ||||| |  ||| || |||||||||| ||||
Db      60588 TATCAAAGCACCAGACTTTCGAGTATTGTCTTCTGAGCTCCAACA 60544


RESULT 38
AC016996/c
LOCUS       AC016996               183674 bp    DNA     linear   HTG 16-JUL-2000
DEFINITION  Homo sapiens chromosome 11 clone RP11-33G24, WORKING DRAFT
            SEQUENCE, 4 unordered pieces.
ACCESSION   AC016996
VERSION     AC016996.3  GI:9211325
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 183674)
  AUTHORS   Waterston,R.H.
```

```
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 183674)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Jul 15, 2000 this sequence version replaced gi:7021767.

            ----------------- Genome Center -----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0033G24
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 82%
            Sequencing vector: plasmid; 18%
            Chemistry: Dye-primer ET; 82% of reads
            Chemistry: Dye-terminator Big Dye; 18% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 181170 bases at least Q40
            Consensus quality: 181898 bases at least Q30
            Consensus quality: 182447 bases at least Q20
            Insert size: 180000; agarose-fp
            Insert size: 183374; sum-of-contigs
            Quality coverage: 6.57 in Q20 bases; agarose-fp
            Quality coverage: 6.47 in Q20 bases; sum-of-contigs
            ----------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    30414: contig of 30414 bp in length
            *   30415    30514: gap of unknown length
            *   30515    74774: contig of 44260 bp in length
            *   74775    74874: gap of unknown length
            *   74875   126758: contig of 51884 bp in length
            *  126759   126858: gap of unknown length
            *  126859   183674: contig of 56816 bp in length.
FEATURES             Location/Qualifiers
     source          1. .183674
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /clone="RP11-33G24"
     misc_feature    1. .30414
                     /note="assembly_name:Contig25"
     misc_feature    30515. .74774
                     /note="assembly_name:Contig26
                     clone_end:T7
                     vector_side:right"
     misc_feature    74875. .126758
                     /note="assembly_name:Contig27
                     clone_end:SP6
                     vector_side:left"
     misc_feature    126859. .183674
                     /note="assembly_name:Contig28"
ORIGIN

  Query Match             4.5%;  Score 275.8;  DB 2;  Length 183674;
  Best Local Similarity   70.1%;  Pred. No. 2.8e-50;
  Matches  452;  Conservative    0;  Mismatches  167;  Indels   26;  Gaps    5;

Qy        312 GTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTC 371
              ||||| |   ||  ||||||| | ||| |||  ||||||||||||||| | ||  | | |
Db       4709 GTAGGAGGAGAACTATGTGATGCCTTTCCTCACCCATCATAAGGGTCACGACTGACATTT 4650

Qy        372 CTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTT 431
              ||||  |||||||||||||||  |||| ||| | |||||| ||||||||||  ||||| |
Db       4649 TTATAACAAAAGACAGGTTAACCAGAGAAAAGCATAACAATTTTATTTAATTGAAGTTGT 4590

Qy        432 ACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTT 491
              ||||||||  |||| ||| || |   |   || | |   || ||| |||||      |||
Db       4589 ACATGACACAGGAGCCTTTAGGATGAAGACCCAAGA---CCTGGGAAAACT-----ATTT 4538

Qy        492 TTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGAT 551
              || |||| ||||||  |||||||   | ||        |    ||  | | ||  |  |
Db       4537 TTATGCTTAGGTTCAGTGAAGAACAAACGGCCATGTAGAAATGTGACCAAAAAACTGTGT 4478

Qy        552 ATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTT 611
              |       | |  || |   | |      | || ||| |||| | ||||||||||
Db       4477 TCTAATGGTAAGAGACCGAGTGGAGAAGGCCAGCCGCCAGGACTGTGTGTTCAGATTCTT 4418

Qy        612 CTTGATCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTT 671
              | || | | ||        ||| ||||||||||||||||| || | |||| |||| | |
Db       4417 CCTGGTTTTTC--------TGTGTAGCATTCTTTCCTCCTGGGTGTCGGGCGGGACCCCT 4366

Qy        672 CTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATGGCTTG 731
              ||  |||||||||||||| |||||||||||| ||||  |||| ||| ||||||||||||
Db       4365 CTGGAATGAGGGTCTTCAAAGGAGAAGGGAGAGAGTGATCTTTGTAGGTTTTATGGCTTG 4306

Qy        732 CTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCTGGTTT 791
              ||| |||| ||| ||||||||||||||| |||| | ||||||||||||||||||||||
Db       4305 CTTTGGGGAGGAGAAGTTCTAGTTTCTACGACCTGTGTTGGGGAAGAGGAATTCTGGTTT 4246

Qy        792 CTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGGTCAGA 851
              | |||| |||||  || |||||||| | ||||| |||||||||||||| | || ||
Db       4245 GTAGGACTCGCTTTGGTGGAGAAAGAGGGGCAAGAGACAGGAGGGCAGGAGAAGATCGGA 4186

Qy        852 AAGAGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAGTACTTCTTAGCA 911
              ||        | ||||||||||  |||||  |||||||||||  |||||    | |
Db       4185 AA-------GTTTGCTTCTGAGATCTTCCAGTCTCCTTTAGTTTGAAGTA---TTCGGTG 4136

Qy        912 TACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAACA 956
              || |||||||| | ||||| |  ||| || |||||||||| ||||
Db       4135 TATCAAAGCACCAGACTTTCGAGTATTGTCTTCTGAGCTCCAACA 4091


RESULT 39
BS000038/c
LOCUS       BS000038              142533 bp    DNA     linear   PRI 12-JUN-2004
DEFINITION  Pan troglodytes chromosome 22 clone:PTB-005J02, map 22, complete
            sequences.
ACCESSION   BS000038 BA000046
VERSION     BS000038.1  GI:37537305
KEYWORDS    HTG.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   The International Chimpanzee Chromosome 22 Consortium.
  TITLE     DNA sequence and comparative analysis of chimpanzee chromosome 22
  JOURNAL   Nature 429, 382-388 (2004)
REFERENCE   2  (bases 1 to 142533)
  AUTHORS   Hattori,M., Toyoda,A., Watanabe,H., Taylor,T.D., Kuroki,Y.,
            Fujiyama,A. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAY-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     The Chimpanzee Chromosome 22 Sequencing Consortium consists of:
            *Chinese National Human Genome Center at Shanghai, Shanghai, China;
            *GBF, Dept. of Genome Analysis, Braunschweig, Germany; *Institute
            of Molecular Biotechnology, Jena, Germany; *KRIBB Genome Research
            Center, Daejeon, Korea;
            *Max-Planck-Institute for Molecular Genetics, Berlin, Germany;
```

*National Institute of Genetics, Mishima, Japan;
*National Yang Ming University Genome Research Center, Taipei, Taiwan;
*RIKEN Genomic Sciences Center, Yokohama, Japan.
---------- Genome Center
Center: RIKEN Genomic Sciences Center
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gsc.riken.go.jp
---------- Project Information
Center project name:The Chimpanzee Chromosome 22 Sequencing Project
Center clone name: PTB-005J02
---------- Summary Statistics
Sequencing vector: pUC18,pUC13,pTZ19R; 100% of reads Chemistry: Dye-terminator Big Dye and ET; 100% of reads Assembly program: Phrap; version 0.990329
Consensus quality: 138,380 bases at least Q40
Consensus quality: 3,766 bases at least Q30
Consensus quality: 375 bases at least Q20

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30);
an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at one plasmid
subclone or more than one M13 subclone;
and the assembly was confirmed by restriction digest.

Source information:
The RPCI-43 chimpanzee BAC library was prepared from DNA isolated from the blood of a single male chimpanzee using published protocols (Osoegawa,K. et al. Genomics 52:1-8). The DNA from the chimpanzee ('Clint') was obtained from the Yerkes Primate Center in Atlanta. The library was prepared by Baoli Zhu, Chung Li Shu, Kazutoyo Osoegawa, Evan Eichler & Pieter J de Jong. The library characteristics are described at
http://www.chori.org/bacpac/mchimp43.htm.
The clone may be obtained from Pieter J. de Jong and coworkers (http://www.chori.org/bacpac).
VECTOR: pBACe3.6
The CHORI-251 chimpanzee BAC library was prepared from DNA isolated from the blood of a single male chimpanzee using published protocols (Osoegawa,K. et al. Genomics 52:1-8). The DNA from the chimpanzee ('Clint') was obtained from the Yerkes Primate Center in Atlanta. The library was prepared by Baoli Zhu, Chung Li Shu, Kazutoyo Osoegawa, Evan Eichler & Pieter J de Jong. The library characteristics are described at
http://www.chori.org/bacpac/chimpanzee251.htm.
The clone may be obtained from Pieter J. de Jong and coworkers (http://www.chori.org/bacpac).
VECTOR: pTARBAC2.1
The PTB1 chimpanzee BAC library was prepared from DNA isolated from cultured cells established from the blood of a single male chimpanzee.
Clones may be obtained from Asao Fujiyama and co-workers (http://www.gsc.riken.go.jp).
VECTOR: pKS145
The PTF22 chimpanzee Fosmid library was prepared from DNA isolated from cultured cells established from the blood of a single male chimpanzee.
Clones may be obtained from Asao Fujiyama and co-workers (http://www.gsc.riken.go.jp).
VECTOR: pKS143

Sequence Quality Assessment:
This entry has been annotated with sequence estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than 1 error in 10,000 bp.

Neighboring clones: PTB-003F15(left) and CH251-530H12(right).

```
FEATURES             Location/Qualifiers
     source          1. .142533
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /chromosome="22"
                     /clone="PTB-005J02"
                     /clone_lib="PTB1 chimpanzee BAC"
ORIGIN

  Query Match             4.5%;  Score 275.4;  DB 9;  Length 142533;
  Best Local Similarity   71.9%;  Pred. No. 3.5e-50;
  Matches  464;  Conservative    0;  Mismatches  156;  Indels   25;  Gaps    7;

Qy        308 TGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAA 367
              ||  ||||| |   ||| || |||||||||||||  | || || |||     | |
Db      32505 TGTTGTAGGAGCAGAAAGGTATGATACATTTTCTCACCAATTATCAGG-------CTGAC 32453

Qy        368 ACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAG 427
              | | |||||  |||||||||||||| |||   ||  | |||||  |  | ||||||||||
Db      32452 ATTTCTATAACAAAAGACAGGTTAACAAGGAAAAGGCATAACAGGTGCA-TTAATCAAAG 32394

Qy        428 TTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTG 487
              ||||| |||||||| |||||||||||||||| |||||||||||||||| ||||||| |
Db      32393 TTTTATATGACATGAGAGTCTTCAGAAATGAAAACCCAAAGACCCAGGGAAAACTGT-TT 32335

Qy        488 TTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAA 547
              ||||| |||| || || |||||||  |||| ||||    |  |    ||||  ||||||
Db      32334 CTTTTTATGCTTAGTTTTGATGAAGTGTGGACAGCAGTGTGGAAATGTGATTGCACAAAA 32275

Qy        548 GGATATGATCTAGTGGTAAAGGACTCAGG-GGGAAACACAGCAAGGCCTGTCTATTCAGA 606
              || |||||| || |||||   ||   |||  ||||||  ||||  |||||||||  | ||||
Db      32274 GGGTATGATTTAATGGTAGTAGATGGAGGCAGGAAACCCAGCCCGGCCTGTCTGTCCAGA 32215

Qy        607 TTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGA 666
              |||||||||  ||||           |||||  ||  ||||||||  ||||||||||||
Db      32214 TTCTTCTTGACTTCTC--------CACGTAGCAATCCATCCTCCTGGGTATGGGGCAGGA 32163

Qy        667 CTCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATG 726
              |  ||||  |||||||||||||| | |||||| || |||||| ||||  |||  |||||||
Db      32162 CATTTCTGGAATGAGGGTCTTCAGGAGAGAAGAGATAAAGTGACCTTCCTAGGTTTTATG 32103

Qy        727 GCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCT 786
              ||||| || ||   ||||| |||||||||||||||| ||  |||||| ||||| ||||||
Db      32102 GCTTGATTTGGAAGAGAGGGGTTCTAGTTTCTATGATCCGCCTTGGGCAAGAGAAATTCT 32043

Qy        787 GGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGG 846
              | || || | |||   |||   | || |||||| ||   || |||||| |||| ||| ||
Db      32042 GATTACTATAACTCAGTTTGGGGGAGGAAGAGGGGTGGAAGACAGGAGAGCAGCAGAGGG 31983

Qy        847 TCAGAAAGAGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAGTACTTCT 906
              ||||| |||   |    | |    |   ||||  | ||||| ||||  |||||     |
Db      31982 TCAGAGAGACCTT----GGTTCTGAGGCCTTTCCAATGTCCTTCAGTTTAAAGTA---TT 31930

Qy        907 TAGCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTC 951
              ||||| ||||| |||  |||||||  ||| || || ||| |||
Db      31929 GAGCATGCCAAAACACCGTACTTTGGGGTATTGTGTTTTGAACTC 31885


RESULT 40
AL445283/c
LOCUS       AL445283              228353 bp    DNA     linear   HTG 01-JUN-2003
DEFINITION  Homo sapiens chromosome 1 clone RP11-355E8, WORKING DRAFT SEQUENCE,
            11 unordered pieces.
ACCESSION   AL445283
VERSION     AL445283.8  GI:31335501
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 228353)
```

```
AUTHORS   McLay,K.
TITLE     Direct Submission
JOURNAL   Submitted (31-MAY-2003) Wellcome Trust Sanger Institute, Hinxton,
          Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
          humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT   On Jun 2, 2003 this sequence version replaced gi:31071434.
          -------------- Genome Center
          Center: Wellcome Trust Sanger Institute
          Center code: SC
          Web site: http://www.sanger.ac.uk
          Contact: humquery@sanger.ac.uk
          -------------- Project Information
          Center project name: bA355E8
          -------------- Summary Statistics
          Assembly program: XGAP4; version 4.5
          Sequencing vector: plasmid; L08752; 100% of reads
          Chemistry: Dye-terminator; 63% of reads
          Chemistry: Dye-terminator Big Dye; 36% of reads
          Consensus quality: 221499 bases at least Q40
          Consensus quality: 223476 bases at least Q30
          Consensus quality: 225347 bases at least Q20
          Insert size: 227353; sum-of-contigs
          Insert size: 212075; 26.8% error; agarose-fp
          Quality coverage: 12.78x in Q20 bases; sum-of-contigs Quality
          coverage: 16.93x in Q20 bases; agarose-fp
          --------------.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 11 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *       1    72361: contig of 72361 bp in length
          *   72362    72461: gap of 100 bp
          *   72462   126279: contig of 53818 bp in length
          *  126280   126379: gap of 100 bp
          *  126380   147834: contig of 21455 bp in length
          *  147835   147934: gap of 100 bp
          *  147935   204926: contig of 56992 bp in length
          *  204927   205026: gap of 100 bp
          *  205027   210877: contig of 5851 bp in length
          *  210878   210977: gap of 100 bp
          *  210978   214639: contig of 3662 bp in length
          *  214640   214739: gap of 100 bp
          *  214740   218225: contig of 3486 bp in length
          *  218226   218325: gap of 100 bp
          *  218326   220847: contig of 2522 bp in length
          *  220848   220947: gap of 100 bp
          *  220948   223136: contig of 2189 bp in length
          *  223137   223236: gap of 100 bp
          *  223237   225413: contig of 2177 bp in length
          *  225414   225513: gap of 100 bp
          *  225514   228353: contig of 2840 bp in length.
FEATURES             Location/Qualifiers
     source          1. .228353
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-355E8"
                     /clone_lib="RPCI-11.2"
     misc_feature    1. .72361
                     /note="assembly_fragment:02601
                     fragment_chain:1"
     misc_feature    72462. .126279
                     /note="assembly_fragment:00152
                     fragment_chain:1"
     misc_feature    126380. .147834
                     /note="assembly_fragment:00908
                     fragment_chain:1"
     misc_feature    147935. .204926
                     /note="assembly_fragment:05322
                     fragment_chain:1"
     misc_feature    205027. .210877
                     /note="assembly_fragment:04697
                     fragment_chain:2"
     misc_feature    210978. .214639
                     /note="assembly_fragment:00846
                     fragment_chain:2"
     misc_feature    214740. .218225
                     /note="assembly_fragment:00550"
     misc_feature    218326. .220847
                     /note="assembly_fragment:00847"
     misc_feature    220948. .223136
                     /note="assembly_fragment:01203"
     misc_feature    223237. .225413
                     /note="assembly_fragment:01239"
     misc_feature    225514. .228353
                     /note="assembly_fragment:02478"
ORIGIN

  Query Match             4.5%;  Score 274;  DB 2;  Length 228353;
  Best Local Similarity   74.8%;  Pred. No. 6.9e-50;
  Matches  485;  Conservative    0;  Mismatches  130;  Indels   33;  Gaps   10;

Qy        313 TAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTCC 372
              || ||||| ||| |||||||||| ||| |||  ||||||| | ||       | || |||||
Db     117756 TACGGGTGGAAAGGTGTGATACCTTTCCTCAGCCATCATTAAGG-----GCTAACACTCC 117702

Qy        373 TATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTA 432
              ||||  |||| ||| ||||| ||||| ||||| ||||| |||||||||||| |||||||
Db     117701 TATAACAAAA-ACAAGTTAACAAGAGAAAAACATAACACATTTATTTAATCAGAGTTTTA 117643

Qy        433 CATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTTT 492
              |||| ||  |||| ||||||||||||||||||||||||||||||| |||||      ||||
Db     117642 CATGGCACAGGAGGCTTCAGAAATGAAGACCCAAAGACCCAGGGAAAACT-----ATTTT 117588

Qy        493 TTTGCTGAGGTTCGATGAAGAATGGATAGC-ATGTAGCCATGTAGATTAGACAAAAGGAT 551
              | |||| || ||  ||| ||||| |  ||| ||| ||  ||||    |      ||| || |
Db     117587 TATGCTCAGATTTAATGCAGAATAGTCAGCTATGAAGGAATGTGATTGGACAAAATGGGT 117528

Qy        552 ATGATCTAGTGGTAAAGGACTCAGGGGGAAAC---ACAGCAAGGCCTGTCTATTCAGATT 608
              ||||||| |||| |  ||||  ||||  | |    ||||||| ||||||| ||||||||
Db     117527 ATGATCTAATGGTTATAGACTGGGGGGTGAGCGGACCAGCAAGCCCTGTCTGTTCAGATT 117468

Qy        609 CTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACT 668
              |||||||  | |||         || |||| |||  |||||| || ||||||||||||||
Db     117467 CTTCTTGGACGCTC--------TGCATAGTATTTATTCCTCTTGGGTATGGGGCAGGACT 117416

Qy        669 CTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATGGC 728
              | |||  |  |||||||||||||||||||| ||| |||| ||||| ||| |||||| |
Db     117415 CCTCTAAATGGAGGGTCTTCAAGGGAGAAGGAAGAGAGTGACCTTTGTAGGTTTTAT-GT 117357

Qy        729 TTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCTGG 788
              |||||| | || |||||||| |||||||||| |||     ||| ||||||||||||||||
Db     117356 TTGCTTTGAGGGAGAGGAGTCCTAGTTTCTAAGACTGGCTTTGAGGAAGAGGAATTCTGG 117297

Qy        789 TTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGGTC 848
              ||||| || ||||| | || | ||||| ||||||  |||   |||||||| |||| ||
Db     117296 TTTCTATGGCTTGCATCCAGGGAGAAAAAGGAGTGTGAGAGGGGAGGGCAAGAGAAGG-- 117239

Qy        849 AGAAAGAGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAGTACTTCTTA 908
              | |||||| |   |||||||||| |||||  |||||||| |||| |||||   || |
Db     117238 -CAGAGAGACCT---TGCTTCTGAGGCCTTCCAATCTCCTTTGGTTCAAAGTA---CTCA 117186

Qy        909 GCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAACA 956
              ||||||||||     | |||| ||  |||  |  | |||||   ||||
Db     117185 GCATACCAAAGTGTCACACTTAGGGGTATCATACTGTGAGCCACAACA 117138


RESULT 41
AC083796/c
LOCUS       AC083796               172461 bp    DNA     linear   HTG 08-JAN-2003
DEFINITION  Homo sapiens chromosome 3 clone RP11-817L6, WORKING DRAFT SEQUENCE,
```

```
            13 unordered pieces.
ACCESSION   AC083796
VERSION     AC083796.6  GI:20335635
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 172461)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 172461)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-OCT-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 172461)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JAN-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Apr 28, 2002 this sequence version replaced gi:11995513.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: HBEW
                Center clone name: RP11-817L6
            -------------- Summary Statistics
                Sequencing vector: M13;
                Chemistry: Dye-primer Bodipy: 50% of reads
                Chemistry: Dye-terminator Big Dye: 50% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 162060 bases at least Q40
                Consensus quality: 165789 bases at least Q30
                Consensus quality: 167723 bases at least Q20
                Estimated insert size: 168513; sum-of-contigs estimation
                Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 13 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2807: contig of 2807 bp in length
            *     2808     2907: gap of unknown length
            *     2908     5767: contig of 2860 bp in length
            *     5768     5867: gap of unknown length
            *     5868    11906: contig of 6039 bp in length
            *    11907    12006: gap of unknown length
            *    12007    17970: contig of 5964 bp in length
            *    17971    18070: gap of unknown length
            *    18071    23748: contig of 5678 bp in length
            *    23749    23848: gap of unknown length
            *    23849    29974: contig of 6126 bp in length
            *    29975    30074: gap of unknown length
            *    30075    38700: contig of 8626 bp in length
            *    38701    38800: gap of unknown length
            *    38801    47979: contig of 9179 bp in length
            *    47980    48079: gap of unknown length
            *    48080    66509: contig of 18430 bp in length
            *    66510    66609: gap of unknown length
            *    66610    89425: contig of 22816 bp in length
            *    89426    89525: gap of unknown length
            *    89526   116375: contig of 26850 bp in length
            *   116376   116475: gap of unknown length
            *   116476   141932: contig of 25457 bp in length
            *   141933   142032: gap of unknown length
            *   142033   172461: contig of 30429 bp in length.
FEATURES             Location/Qualifiers
     source          1. .172461
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-817L6"
ORIGIN

  Query Match             4.5%;  Score 273.8;  DB 2;  Length 172461;
  Best Local Similarity   71.9%;  Pred. No. 7.8e-50;
  Matches  507;  Conservative    0;  Mismatches  162;  Indels   36;  Gaps   10;

Qy        311 AGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACT 370
              ||  |||||| |||  |||||||| ||| |||   |||||| |||||| || |  | ||
Db      92816 AGGTGGGGTGGAAAGATGTGATACCTTTCCTCACTCATCATGAGGGTCACAGCTGACACC 92757

Qy        371 CCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTT 430
              ||| ||  |||||||||||| ||  ||  ||| | |||||||| ||||| ||||||||||
Db      92756 CCTTTAACAAAAGACAGGTTGAT--GATAAAAGCATAACAAATGTATTTCATCAAAGTTT 92699

Qy        431 TACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTT 490
              |||||||| |||||||||||||||| |||||||| ||||||| | |  ||||| ||
Db      92698 TACATGACAAGGGAGTCTTCAGAAATGTAGACCCAAAGGCCCAGGGAAGATGGTCTG-TT 92640

Qy        491 TTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGAC--AAAAG 548
              | | ||||  |||  || |||||||||  |||         |    |||| |||  |||||
Db      92639 TCTATGCTTTGGTCTGAGGAAGAATGGGCAGCTGTGTAGAAATGTGATTGGACAAAAAAG 92580
```

```
Qy        549 GATATGATCTAGTGGTAAAGGACTCAGGGGG-AAACACAGCAAGGCCTGTCTATTCAGAT 607
              | | ||||||| || | |  |||| |||||| | ||| ||||||||||||||| |||||||
Db      92579 GGTCTGATCTAATGATGACAGACTGAGGGGGTAGACAGAGCAAGGCCTGTCTGTTCAGAT 92520

Qy        608 TCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGAC 667
              ||||||||  ||| |           |||||||||| ||||||| | ||| |||| |||||
Db      92519 TCTTCTTGGCCTCCC----------TATAGCATTCCTTCCTCCAG-GTAGGGGGTAGGAC 92471

Qy        668 TCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATGG 727
              | |||  ||||| ||  ||  |||||||||||||| | || ||||| ||| |||||||||
Db      92470 CCCTCTGAAATGAAGG-GTTTTAGGGAGAAGGGAGAGAATGACCTTTCTAGGTTTTATGG 92412

Qy        728 CTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCTG 787
              |||||| |||  ||||| ||||||| |||||||||  ||||| ||||||||||||||||
Db      92411 ATTGCTTTGGGAGAGAGGGGTTCTAGATTCTATGACCTGTCTTGAGGAAGAGGAATTCTG 92352

Qy        788 GTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGGT 847
              |||||||||| |  ||||   | |||  |  ||        |||||||||||||| |||
Db      92351 GTTTCTGTGATTCCCTTTAGGGGAGATGGGAGA-------CAAGGAGGGCAGGAGAAGGT 92299

Qy        848 CAGAAAGAGACTTGGCTGCT--------TCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAG 899
              |||| | |||||| ||| ||        |||||||    ||  | ||||||||||  ||
Db      92298 CAGAGAAAGACTTTGCTTCTGAGGCAGCTCTGAGGCTACCCAATATCCTTTAGTTCAGAG 92239

Qy        900 TACTTCTTAGCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAACACTG 959
              ||   || ||| | ||||||||| ||| ||||| |||  |||||||| || ||  ||||
Db      92238 TA---CTCAGCCTGCCAAAGCACCATAGTTTGGGATACTGTTTTCTGCGCCCTGGCACTT 92182

Qy        960 CAATCATGCTAAACTCCTCTATGACCTTCAAACATTCCACTTGCT 1004
                |    |    |||||  | | |  | |  | |   | | | | |
Db      92181 ACAAGTTAAGCAACTCACCCAGGGTCATAGAGCTAACAAATGGTT 92137


RESULT 42
CNS01DSU/c
LOCUS       CNS01DSU              173010 bp    DNA     linear   PRI 01-MAY-2001
DEFINITION  Human chromosome 14 DNA sequence BAC R-204K16 of library RPCI-11
            from chromosome 14 of Homo sapiens (Human), complete sequence.
ACCESSION   AL122013
VERSION     AL122013.5  GI:13940010
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 173010)
  AUTHORS   Heilig,R., Petit,J.L., Vico,V., Dasilva,C., Robert,C., Wincker,P.,
            Brottier,P., Cattolico,L., Barbe,V., Pelletier,E., Artiguenave,F.,
            Levy,M., Eckenberg,R., Bruls,T., deBerardinis,V., Cruaud,C.,
            Gyapay,G., Saurin,W. and Weissenbach,J.
  TITLE     Sequencing of the human chromosome 14
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 173010)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-APR-2001) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     On May 3, 2001 this sequence version replaced gi:12001715.
            -------------- Genome Center
            Center: Genoscope / Centre National de Sequencage
            Center code: GS
            Web site: http://www.genoscope.cns.fr/
            Contact: SeqRef@genoscope.cns.fr
            --------------
            The following BAC sequence is oriented from the T7 to the SP6 end.
            Upstream BAC (overlapping the T7 end) : R-108B17 (AC=AL589168)
            Downstream BAC (overlapping the SP6 end) : C-2171B16
            -------------- Summary Statistics
            Assembly program: Phrap; version 2.0
            Quality coverage: 7.63x in Q20 bases; sum-of-contigs
            --------------
            ----------------------
            Overall quality chart :
            Range     :   bases.
            0         :
            1  -  9   :
            10 - 19   :
            20 - 29   :      17
            30 - 39   :     186
            40 - 49   :    4404
            50 - 59   :   11880
            60 - 69   :   12401
            70 - 79   :   24582
            80 - 89   :   50572
            90 - 99   :   68968
            ----------------------
            Percentage of bases with a quality value >= 40 : 99 %.
FEATURES              Location/Qualifiers
     source           1. .173010
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:9606"
                      /chromosome="14"
                      /clone="R-204K16"
                      /clone_lib="RPCI-11"
     STS              48796. .49037
                      /note="matching EMBL:AA007615
                      RHdb:RH75459
                      dbSTS:STS52547
                      Identified using the e-PCR software (G. Schuler)"
     STS              48817. .49053
                      /note="matching EMBL:R93167
                      RHdb:RH45281
                      dbSTS:STS38349
                      Identified using the e-PCR software (G. Schuler)"
     STS              55055. .55154
                      /note="matching EMBL:G03442
                      RHdb:RH53695
                      RHdb:RH3796
                      dbSTS:STS4275
                      Identified using the e-PCR software (G. Schuler)"
ORIGIN

  Query Match             4.5%;  Score 273.8;  DB 9;  Length 173010;
  Best Local Similarity   71.9%;  Pred. No. 7.8e-50;
  Matches  507;  Conservative    0;  Mismatches  162;  Indels   36;  Gaps   10;

Qy        311 AGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACT 370
              || |||||| |||  |||||||| ||| |||   |||||| |||||| || |  | ||
Db      98470 AGGTGGGGTGGAAAGATGTGATACCTTTCCTCACTCATCATGAGGGTCACAGCTGACACC 98411

Qy        371 CCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTT 430
              ||| ||  |||||||||||| ||  ||  ||| | |||||||| ||||| ||||||||||
Db      98410 CCTTTAACAAAAGACAGGTTGAT--GATAAAAGCATAACAAATGTATTTCATCAAAGTTT 98353

Qy        431 TACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTT 490
              ||||||||| |||||||||||||||| |||||||||| ||||||| | |  ||||| ||
Db      98352 TACATGACAAGGGAGTCTTCAGAAATGTAGACCCAAAGGCCCAGGGAAGATGGTCTG-TT 98294

Qy        491 TTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGAC--AAAAG 548
              | | ||||  |||  || |||||||||  |||         |    |||| |||  |||||
Db      98293 TCTATGCTTTGGTCTGAGGAAGAATGGGCAGCTGTGTAGAAATGTGATTGGACAAAAAAG 98234

Qy        549 GATATGATCTAGTGGTAAAGGACTCAGGGGG-AAACACAGCAAGGCCTGTCTATTCAGAT 607
              | | ||||||| || | |  |||| |||||| | ||| ||||||||||||||| |||||||
Db      98233 GGTCTGATCTAATGATGACAGACTGAGGGGGTAGACAGAGCAAGGCCTGTCTGTTCAGAT 98174

Qy        608 TCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGAC 667
              ||||||||  ||| |           |||||||||| ||||||| | ||| |||| |||||
Db      98173 TCTTCTTGGCCTCCC----------TATAGCATTCCTTCCTCCAG-GTAGGGGGTAGGAC 98125

Qy        668 TCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATGG 727
              | |||  ||||| ||  ||  |||||||||||||| | || ||||| ||| |||||||||
Db      98124 CCCTCTGAAATGAAGG-GTTTTAGGGAGAAGGGAGAGAATGACCTTTCTAGGTTTTATGG 98066
```

```
Qy        728 CTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCTG 787
              |||||| |||  ||||| ||||||| ||||||||||  ||||| ||||||||||||||||
Db      98065 ATTGCTTTGGGAGAGAGGGGTTCTAGATTCTATGACCTGTCTTGAGGAAGAGGAATTCTG 98006

Qy        788 GTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGGT 847
              |||||||||| |  ||||   | |||  |  ||          ||||||||||||| |||
Db      98005 GTTTCTGTGATTCCCTTTAGGGGAGATGGGAGA-------CAAGGAGGGCAGGAGAAGGT 97953

Qy        848 CAGAAAGAGACTTGGCTGCT--------TCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAG 899
              |||| | |||||| ||| ||         |||||||    ||  | ||||||||||  ||
Db      97952 CAGAGAAAGACTTTGCTTCTGAGGCAGCTCTGAGGCTACCCAATATCCTTTAGTTCAGAG 97893

Qy        900 TACTTCTTAGCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAACACTG 959
              ||   || ||| | ||||||||| ||| ||||| |||  |||||||| || ||  ||||
Db      97892 TA---CTCAGCCTGCCAAAGCACCATAGTTTGGGATACTGTTTTCTGCGCCCTGGCACTT 97836

Qy        960 CAATCATGCTAAACTCCTCTATGACCTTCAAACATTCCACTTGCT 1004
                |    |    |||||  | | |  | |  | |   | | | |
Db      97835 ACAAGTTAAGCAACTCACCCAGGGTCATAGAGCTAACAAATGGTT 97791
```

RESULT 43
AC117409/c
LOCUS AC117409 154336 bp DNA linear PRI 23-OCT-2002
DEFINITION Homo sapiens 3 BAC RP11-64K7 (Roswell Park Cancer Institute Human BAC Library) complete sequence.
ACCESSION AC117409
VERSION AC117409.4 GI:22218445
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 154336)
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T., Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., He,X., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M., Neal,D., Nelson,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Naylor,S.L., Weinstock,G. and Gibbs,R.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 154336)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (10-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 154336)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (04-JUN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 4 (bases 1 to 154336)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (14-AUG-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 5 (bases 1 to 154336)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (23-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Aug 14, 2002 this sequence version replaced gi:21314809.
INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE:Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL:
http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

```
FEATURES             Location/Qualifiers
     source          1. .154336
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="3"
```

```
                     /clone="RP11-64K7"
     misc_feature    1. .2050
                     /note="overlaps bases 58644. .60701 of clone AC076961"
                     /function="clone overlap"
     repeat_region   complement(226. .340)
                     /rpt_family="MLT1G1"
     repeat_region   complement(801. .1177)
                     /rpt_family="MLT1A1"
     repeat_region   complement(1220. .1635)
                     /rpt_family="L2"
     repeat_region   complement(2056. .2447)
                     /rpt_family="L2"
     repeat_region   complement(3510. .3616)
                     /rpt_family="L1ME"
     repeat_region   3695. .3971
                     /rpt_family="AluSx"
     repeat_region   3972. .4038
                     /rpt_family="GA-rich"
     repeat_region   4039. .4222
                     /rpt_family="(GGGGA)n"
     repeat_region   4236. .4281
                     /rpt_family="(GGGGA)n"
     repeat_region   4283. .4338
                     /rpt_family="G-rich"
     misc_feature    4326. .4329
                     /note="QUALITY BELOW PHRED30."
                     /function="low quality"
     misc_feature    4328
                     /note="UNSURE BASE C/CC"
                     /function="unsure"
     repeat_region   4339. .4512
                     /rpt_family="(GGGAA)n"
     misc_feature    4397. .4400
                     /note="QUALITY BELOW PHRED30."
                     /function="low quality"
     misc_feature    4399
                     /note="UNSURE BASE C/CC"
                     /function="unsure"
     repeat_region   4892. .5280
                     /rpt_family="L1MD3"
     repeat_region   5637. .5939
                     /rpt_family="AluY"
     repeat_region   6511. .6994
                     /rpt_family="MLT1H"
     repeat_region   complement(6995. .7110)
                     /rpt_family="MIR"
     repeat_region   7743. .7934
                     /rpt_family="L2"
     repeat_region   complement(7944. .8121)
                     /rpt_family="AluJb"
     repeat_region   8130. .8213
                     /rpt_family="L2"
     repeat_region   8214. .8239
                     /rpt_family="(CA)n"
     repeat_region   8240. .8309
                     /rpt_family="L2"
     repeat_region   complement(8329. .13762)
                     /rpt_family="L1PA5"
     STS             9698. .9875
                     /standard_name="D11S2442"
     repeat_region   13865. .13960
                     /rpt_family="L2"
     repeat_region   14938. .15323
                     /rpt_family="MSTC"
     repeat_region   complement(15946. .16598)
                     /rpt_family="L1PA5"
     repeat_region   complement(17001. .17412)
                     /rpt_family="L2"
     repeat_region   complement(17502. .17702)
                     /rpt_family="L2"
     repeat_region   17946. .18279
                     /rpt_family="LTR16C"
     repeat_region   complement(18872. .19279)
                     /rpt_family="MLT1B"
     repeat_region   19510. .19530
                     /rpt_family="AT_rich"
     repeat_region   complement(19531. .21413)
                     /rpt_family="L1PA14"
```

```
  Query Match             4.5%;  Score 271.6;  DB 9;  Length 154336;
  Best Local Similarity   74.4%;  Pred. No. 2.4e-49;
  Matches  415;  Conservative    0;  Mismatches  124;  Indels   19;  Gaps    5;

Qy        307 CTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAA 366
              | || |||||||   ||| |||||||||| ||  ||| |||||||          || ||||
Db      29974 CCGCTGTAGGGGGAGAAAGGTGTGATACTTTCCCTCATCCATCA---------CAGCCAA 29924

Qy        367 AACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAA 426
              |||||||||  ||||| || ||| | || || | | | ||||||||||||||| || ||
Db      29923 CACTCCTATAACAAAAGGCATGTTGACAAAAGAAGAGCATAACAAATTTATTTACTCTAA 29864

Qy        427 GTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCT 486
              |||||| |||||||||||  |||||| |||||| || |||||||||||||| | |||||||
Db      29863 GTTTTATATGACATGGGACCCTTCAGCAATGAACACTCAAAGACCCAGGGAAGACTGTCT 29804

Qy        487 GTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAA 546
              | ||| |||| ||||||  ||||||||||  |||              |||| ||||||
Db      29803 -ATGTTTATGCTCAGGTTCAGTGAAGAATGGGGAGCCATGGAAACATGTGATTGGACAAA 29745

Qy        547 AGGATATGATCTAGTGGTAAAGGACT-CAGGGGGAAACACAGCAAGGCCTGTCTATTCAG 605
              ||| |||||||| ||||||  ||||  || ||||||||||||||| |||||| |||||
Db      29744 GGGACATGATCTAATGGTAACAGACTGAAGAGGGAAACACAGCAAGACCTGTCTCTTCAG 29685

Qy        606 ATTCTTCTTGATCTCTCTCTCTCTATGTATAGC----ATTCTTTCCTCCTGAGTATGGGG 661
              |||||||||| || ||||| |     | || |       |   |||||     ||| ||
Db      29684 ATTCTTCTTGGCCTTTCTCTGTAGCATTTTACCCCCTCCCCCAGCCTCCCCGCCATGAGG 29625

Qy        662 CAGGACTCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATT 721
              |||||    |||  ||| ||| ||||||||| |||||| ||| | || || || ||| ||
Db      29624 CAGGATCTCTCTGGAATAAGGTTCTTCAAGGCAGAAGGTAGAGAATGACCCTTCTAGGTT 29565

Qy        722 T----TATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAG 777
              |    ||||| |||||| |||||||||  ||||| ||||||| |||||||||| |||||
Db      29564 TTATGTATGGGTTGCTTTGGGGAAGAGGTCTTCTAATTTCTATTACCCATCTTGTGGAAG 29505

Qy        778 AGGAATTCTGGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGC 837
              |||||||||||||||| || ||   |||  ||||||||||| ||  ||| ||||||||||
Db      29504 AGGAATTCTGGTTTCTATGGCTCATTTTGGGGAAGAAAGAGGGGTGGGAGACAGGAGGGC 29445

Qy        838 AGGAGATGGTCAGAAAGA 855
              |||||| ||||||| |||
Db      29444 AGGAGAAGGTCAGAGAGA 29427
```

```
RESULT 44
HS1164K7
LOCUS       HS1164K7              154341 bp    DNA     linear   PRI 21-MAR-2003
DEFINITION  Homo sapiens chromosome 3 from BAC RPCI-11 1164K7 map 3q21 region
            D3S3607-D3S1290, complete sequence.
ACCESSION   AL592209
VERSION     AL592209.3  GI:29150201
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Sudbrak,R., Borzym,K., Kosiura,A., Rak,K., Omran,H., Lehrach,H. and
            Reinhardt,R.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 154341)
  AUTHORS   MOLGENR.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JUN-2001) MPIMG, Abt.Lehrach, Max Planck Institut
            Fuer Molekulare Genetik, Ihnestrasse 73, Berlin, 14195 Germany
COMMENT     On Mar 21, 2003 this sequence version replaced gi:16415917.
```

```
          contig 01          1. .154341
          Clone received from Resource Centre of the Human Genome Project at
          the Max-Planck-Institut for Molecular Genetics.
FEATURES             Location/Qualifiers
     source          1. .154341
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /map="3q21"
                     /clone="BAC RPCI-11 1164K7"
                     /clone_lib="RPCI-11 Human BAC library, originating
                     insitute: Roswell Park Cancer Institute"
                     /note="region between markers D3S3607-D3S1290"
ORIGIN

  Query Match             4.5%;  Score 271.6;  DB 9;  Length 154341;
  Best Local Similarity   74.4%;  Pred. No. 2.4e-49;
  Matches  415;  Conservative    0;  Mismatches  124;  Indels   19;  Gaps    5;

Qy        307 CTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAA 366
              | || |||||||   ||| |||||||||| ||  ||| |||||||         || ||||
Db     124358 CCGCTGTAGGGGGAGAAAGGTGTGATACTTTCCCTCATCCATCA---------CAGCCAA 124408

Qy        367 AACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAA 426
              |||||||||  ||||| || ||| | || || | | | ||||||||||||||| || ||
Db     124409 CACTCCTATAACAAAAGGCATGTTGACAAAAGAAGAGCATAACAAATTTATTTACTCTAA 124468

Qy        427 GTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCT 486
              |||||| ||||||||||  |||||| |||||| || ||||||||||||| | |||||||
Db     124469 GTTTTATATGACATGGGACCCTTCAGCAATGAACACTCAAAGACCCAGGGAAGACTGTCT 124528

Qy        487 GTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAA 546
                | ||| |||| ||||||  ||||||||||  |||              |||| ||||||
Db     124529 -ATGTTTATGCTCAGGTTCAGTGAAGAATGGGGAGCCATGGAAACATGTGATTGGACAAA 124587

Qy        547 AGGATATGATCTAGTGGTAAAGGACT-CAGGGGGAAACACAGCAAGGCCTGTCTATTCAG 605
                ||| |||||||| ||||||  ||||  || ||||||||||||||| ||||||| |||||
Db     124588 GGGACATGATCTAATGGTAACAGACTGAAGAGGGAAACACAGCAAGACCTGTCTCTTCAG 124647

Qy        606 ATTCTTCTTGATCTCTCTCTCTCTATGTATAGC----ATTCTTTCCTCCTGAGTATGGGG 661
              ||||||||||  || ||||| |     | || |       |   |||||     ||| ||
Db     124648 ATTCTTCTTGGCCTTTCTCTGTAGCATTTTACCCCCTCCCCCAGCCTCCCCGCCATGAGG 124707

Qy        662 CAGGACTCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATT 721
              |||||    |||  ||| ||| |||||||||| |||||| ||| | || || || ||| ||
Db     124708 CAGGATCTCTCTGGAATAAGGTTCTTCAAGGCAGAAGGTAGAGAATGACCCTTCTAGGTT 124767

Qy        722 T----TATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAG 777
              |    ||||| |||||| |||||||||  ||||| ||||||| |||||||||| |||||
Db     124768 TTATGTATGGGTTGCTTTGGGGAAGAGGTCTTCTAATTTCTATTACCCATCTTGTGGAAG 124827

Qy        778 AGGAATTCTGGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGC 837
              |||||||||||||||| || ||   |||   |||||||||| ||  ||| |||||||||
Db     124828 AGGAATTCTGGTTTCTATGGCTCATTTTGGGGAAGAAAGAGGGGTGGGAGACAGGAGGGC 124887

Qy        838 AGGAGATGGTCAGAAAGA 855
              |||||| ||||||| |||
Db     124888 AGGAGAAGGTCAGAGAGA 124905


RESULT 45
CNS01RHQ/c
LOCUS       CNS01RHQ              204662 bp    DNA     linear   PRI 30-NOV-2001
DEFINITION  Human chromosome 14 DNA sequence BAC R-349D8 of library RPCI-11
            from chromosome 14 of Homo sapiens (Human), complete sequence.
ACCESSION   AL162191
VERSION     AL162191.6  GI:17383599
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 204662)
  AUTHORS   Heilig,R., Petit,J.L., Vico,V., Dasilva,C., Robert,C., Wincker,P.,
            Brottier,P., Cattolico,L., Barbe,V., Pelletier,E., Artiguenave,F.,
            Levy,M., Eckenberg,R., Bruls,T., deBerardinis,V., Cruaud,C.,
            Gyapay,G., Saurin,W. and Weissenbach,J.
  TITLE     Sequencing of the human chromosome 14
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 204662)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-NOV-2001) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     On Dec 5, 2001 this sequence version replaced gi:12001743.
            ---------- Genome Center
            Center: Genoscope / Centre National de Sequencage
            Center code: GS
            Web site: http://www.genoscope.cns.fr/
            Contact: SeqRef@genoscope.cns.fr
            ----------
            The following BAC sequence is oriented from the T7 to the SP6 end.
            Upstream BAC (overlapping the T7 end) : C-2350E12
            Downstream BAC (overlapping the SP6 end) : C-2174H20 (AC=AL356806)
            ---------- Summary Statistics
            Assembly program: Phrap; version 2.0
            Quality coverage: 7.86x in Q20 bases; sum-of-contigs
            ----------
            ------------------------
            Overall quality chart :
            Range    :   bases
            0        :
            1  -  9  :
            10 - 19  :
            20 - 29  :      23
            30 - 39  :     261
            40 - 49  :    4362
            50 - 59  :   18448
            60 - 69  :   16906
            70 - 79  :   22873
            80 - 89  :   45682
            90 - 99  :   96107
            ------------------------
            Percentage of bases with a quality value >= 40 : 99 %.
FEATURES             Location/Qualifiers
     source          1. .204662
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="14"
                     /clone="R-349D8"
                     /clone_lib="RPCI-11"
     STS             121704. .121966
                     /note="matching EMBL:Z51462
                     RHdb:RH31306
                     dbSTS:STS19836
                     Identified using the e-PCR software (G. Schuler)"
     STS             157692. .157835
                     /note="matching EMBL:G33165
                     RHdb:RH42760
                     dbSTS:STS36190
                     Identified using the e-PCR software (G. Schuler)"
ORIGIN

  Query Match             4.4%;  Score 268.8;  DB 9;  Length 204662;
  Best Local Similarity   70.9%;  Pred. No. 9.7e-49;
  Matches  437;  Conservative    0;  Mismatches  152;  Indels   27;  Gaps    5;

Qy        304 AAGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAAC 363
              || |||| | ||||||  ||| |||||||| ||| ||| |||||||||||||| |  | |
Db      71328 AAACTGCTGCAGGGGTAGAAAGGTGTGATATCTTTCCTCATCCATCATAAGGGACATAGC 71269

Qy        364 CAAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATC 423
              ||| ||||  |||  ||||| |||||||  ||||   || | | ||||||||||||||||
```

```
Db      71268 CAACACTCTGATAACAAAAGGCAGGTTAGCAAGAAAGAAGCATTACAAATTTATTTAATC 71209

Qy        424 AAAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTG 483

Db      71208 AAAGTTTTGTGTGTCTCAGGACCCGTCAGAAATGAAGACCCAAATACCTAGGGAAAAGTG 71149

Qy        484 T---------CTGTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGT 534

Db      71148 TTATGCTTAGGTTACATTTATGCTTAGCTTCGATGAAGAATGGACAGCCATGTAGAAATG 71089

Qy        535 AGATTAGACAAAAGGATATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCC 594

Db      71088 TGATTGGACAAAAGGGTATGATCTAATGGTAATAGACTAAAGGGGAAACCCAGCAAAGCC 71029

Qy        595 TGTCTATTCAGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAG 654

Db      71028 TGTCTGCTCGGATTCTTTTGG-----CCTCTCTGTGTAGCATTCTTCCCCTCCTACCTGG 70974

Qy        655 TATGGGGCAGGACTCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAA-------AGT 707

Db      70973 TATGGGACAGGATCCCTCTGGAATGAGAGTCTTCAGGGGAGAAGGGAGGAAAGAGAGAGC 70914

Qy        708 GGCCTTTTTAGATTTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCAT 767

Db      70913 GACCTTTCTAGGTTTTATGGCTTACTTTGGGGAAGAGGAATTCTGGTTTCTATGACCTGC 70854

Qy        768 CTTGGGGAAGAGGAATTCTGGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAG 827

Db      70853 CTTGGGGAAAATAAATTCTAGTTCTATGATTTGCTTTCAGGGGAGAAAGAGGGTTGGCAG 70794

Qy        828 GCAGGAGGGCAGGAGATGGTCAGAAAGAGACTTGGCTGCTTCTGAGG-GCTTCCGCTCTC 886

Db      70793 ACAGGAAGACATGAGAAGGTCAGAAAGA-----CCTTGCTTCTGAGGTGCTTCCAATCTT 70739

Qy        887 CTTTAGTTCCAAGTAC 902

Db      70738 CTTCAGTTCAAAGTAC 70723
```

```
RESULT 46
AC008885/c
LOCUS       AC008885               127790 bp    DNA     linear   PRI 30-JAN-2002
DEFINITION  Homo sapiens chromosome 5 clone CTD-2218K11, complete sequence.
ACCESSION   AC008885
VERSION     AC008885.5  GI:18425268
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 127790)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 127790)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 127790)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JAN-2002) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Jan 30, 2002 this sequence version replaced gi:9256068.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 99.6% of Sequence;
            Estimated Total Number of Errors is 0.3.
            NOTE:   166 bp break in paired end plasmid links from 103909 to
            104075.
```

```
FEATURES             Location/Qualifiers
     source          1. .127790
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTD-2218K11"
ORIGIN
```

```
  Query Match             4.4%;  Score 267.2;  DB 9;  Length 127790;
  Best Local Similarity   71.9%;  Pred. No. 2.3e-48;
  Matches  411;  Conservative    0;  Mismatches  143;  Indels   18;  Gaps    4;

Qy        283 TTTATCTCTCCTTCTCTCTCAAAGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTC 342

Db     106504 TTAAACTTTCTACACCGATCTTAGATATTATAGGAATGGAAAGGAATGATACGTTTCCTC 106445

Qy        343 TTCCATCATAAGGGTCGCAACCAAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAA 402

Db     106444 ACTCATTATAAGAGTCACAGACAACACTCCTATAACAAAAGACAGGTTAACAAGAAAAAA 106385

Qy        403 ACCTAACAAATTTATTTAATCAAAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGAC 462

Db     106384 GCATAACAAACTTATTTAATCAAAGTTTTATGTGACACAGGAGCCTTCAGAAATGGAGAC 106325

Qy        463 CCAAAGACCCAGGGGAAACTGTCTGTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGC 522

Db     106324 CCAAAGACCCAGGGAAAACTATCT-TTTTTATGCTTACGTTTGATGAAGAATG--TACA 106268

Qy        523 ATGTAGCCATGTAGATTAGACAAAAGGATATGATCTAGTGGTAAAGGACTCAGGGGGAAA 582

Db     106267 ACCATAGAAATGTGATTGGACAAAAGGGTATAATCTAATGGTCATAGACAAAGGGGAGAA 106208

Qy        583 CACAGCAAGGCCTGTCTATTCAGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTC 642

Db     106207 TCCAGCAAGGCCTATCTGTTCAGAGTCCTCTTGGCCTCTC--------TGTGTTGCATTA 106156

Qy        643 TTTCCTCCTGAGTATGGGGCAGGACTCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAG 702

Db     106155 CTTCCGCCTACGTATGGGCCAGGCTTCTTCAGAAATGAGAGTTTTC-------AAGGGAC 106103

Qy        703 AAAGTGGCCTTTTTAGATTTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGA 762

Db     106102 AGAGCAGCATTTCTCAATTTTATGGCTTGCTTTGGGGGAAAGCAGTTCTAGTTTCATTGA 106043

Qy        763 CCCATCTTGGGGAAGAGGAATTCTGGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGT 822

Db     106042 TTTGCCTAAAAGAAGAATAATTCTGGTTTCTATGACTCACTTTACAGAAGAAAGAGGGGT 105983

Qy        823 AAGAGGCAGGAGGGCAGGAGATGGTCAGAAAG 854

Db     105982 GGATGACAGGAGGACAGGAGAAGATCGGGAAG 105951
```

```
RESULT 47
HSQ9D13
LOCUS       HSQ9D13                37067 bp    DNA     linear   PRI 06-MAY-2003
DEFINITION  Homo sapiens chromosome 21 sequence from Cosmid LL21NC02-9D13 map
            21q21 region D21S189-LA297, complete sequence.
ACCESSION   AL109764
VERSION     AL109764.2  GI:5931932
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Ramser,J., Borzym,K., Lehrack,S., Klein,M., Arndt,S., Hildmann,T.,
            Gardiner,K., Yaspo,M.L., Reinhardt,R. and Lehrach,H.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 37067)
  AUTHORS   MPIMG.
  TITLE     Direct Submission
```

```
  JOURNAL   Submitted (06-AUG-1999) MPIMG, Abt.Lehrach, Max Planck Institut
            Fuer Molekulare Genetik, Ihnestrasse 73, Berlin, 14195 Germany
COMMENT     On Sep 28, 1999 this sequence version replaced gi:5725299.
            contig 01         1. .37067.
FEATURES             Location/Qualifiers
     source          1. .37067
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q21"
                     /clone="LL21NC02-9D13"
                     /clone_lib="LL21NC02, chromosome 21 specific cosmid
                     library, creator:Pieter de Jong"
                     /note="region between markers D21S189-LA297"
     repeat_region   complement(2349. .2502)
                     /note="MER4A"
                     /rpt_family="LTR/MER4-group"
     repeat_region   complement(2708. .2826)
                     /note="MER4C"
                     /rpt_family="LTR/MER4-group"
     repeat_region   2864. .3387
                     /note="L1MC4"
                     /rpt_family="LINE/L1"
     repeat_region   3394. .3532
                     /note="L1MC4"
                     /rpt_family="LINE/L1"
     repeat_region   3606. .3907
                     /note="AluSq"
                     /rpt_family="SINE/Alu"
     repeat_region   3962. .4008
                     /note="AT_rich"
                     /rpt_family="Low_complexity"
     repeat_region   complement(4355. .4545)
                     /note="MIR"
                     /rpt_family="SINE/MIR"
     repeat_region   5370. .5409
                     /note="(CATA)n"
                     /rpt_family="Simple_repeat"
     repeat_region   complement(5519. .5549)
                     /note="AT_rich"
                     /rpt_family="Low_complexity"
     repeat_region   complement(5564. .5755)
                     /note="MIR"
                     /rpt_family="SINE/MIR"
     repeat_region   6192. .12630
                     /note="L1PA2"
                     /rpt_family="LINE/L1"
     repeat_region   14989. .15334
                     /note="MLT1A1"
                     /rpt_family="LTR/MaLR"
     repeat_region   complement(15564. .15602)
                     /note="U2"
                     /rpt_family="snRNA"
     repeat_region   15896. .15975
                     /note="AT_rich"
                     /rpt_family="Low_complexity"
     repeat_region   complement(15984. .16285)
                     /note="AluY"
                     /rpt_family="SINE/Alu"
     repeat_region   complement(16570. .16619)
                     /note="MER5A"
                     /rpt_family="DNA/MER1_type"
     repeat_region   complement(16829. .16928)
                     /note="MER45"
                     /rpt_family="DNA/MER1_type"
     repeat_region   complement(16939. .17054)
                     /note="(TAAA)n"
                     /rpt_family="Simple_repeat"
     repeat_region   17502. .17668
                     /note="MER5A"
                     /rpt_family="DNA/MER1_type"
     repeat_region   complement(17824. .17862)
                     /note="(CA)n"
                     /rpt_family="Simple_repeat"
     repeat_region   19589. .19739
                     /note="AluJo/FRAM"
                     /rpt_family="SINE/Alu"
     repeat_region   complement(20209. .20554)
                     /note="MLT1A1"
                     /rpt_family="LTR/MaLR"
     repeat_region   complement(21983. .22680)
                     /note="MER67C"
                     /rpt_family="Other"
     repeat_region   22741. .22784
                     /note="Alu"
                     /rpt_family="SINE/Alu"
     repeat_region   22786. .22847
                     /note="L1MA9"
                     /rpt_family="LINE/L1"
     repeat_region   23016. .23259
                     /note="MER39"
                     /rpt_family="Other/MER21_gro"
     repeat_region   complement(23277. .23670)
                     /note="MSTB"
                     /rpt_family="LTR/MaLR"
     repeat_region   23686. .23787
                     /note="MER39"
                     /rpt_family="Other/MER21_gro"
     repeat_region   23787. .24048
                     /note="MER39b_j"
                     /rpt_family="Other/MER21_gro"
     repeat_region   24167. .24262
                     /note="AluSp"
                     /rpt_family="SINE/Alu"
     repeat_region   24280. .25648
                     /note="pTR5"
                     /rpt_family="LTR/Retroviral"
     repeat_region   25671. .25754
                     /note="AluSg/x"
                     /rpt_family="SINE/Alu"
     repeat_region   26075. .26377
                     /note="AluJb"
                     /rpt_family="SINE/Alu"
     repeat_region   26386. .26670
                     /note="AluJo"
                     /rpt_family="SINE/Alu"
     repeat_region   26721. .26783
                     /note="L1"
                     /rpt_family="LINE/L1"
     repeat_region   26868. .27163
                     /note="AluJb"
                     /rpt_family="SINE/Alu"
     repeat_region   27181. .27210
                     /note="AT_rich"
                     /rpt_family="Low_complexity"
     repeat_region   27477. .27765
                     /note="L1P5"
                     /rpt_family="LINE/L1"
     repeat_region   28126. .28193
                     /note="(CATA)n"
                     /rpt_family="Simple_repeat"
     repeat_region   28196. .28560
                     /note="MER39"
                     /rpt_family="Other/MER21_gro"
     repeat_region   28558. .28741
                     /note="MER39b_j"
                     /rpt_family="Other/MER21_gro"
     repeat_region   28822. .28862
                     /note="(CATA)n"
                     /rpt_family="Simple_repeat"
     repeat_region   complement(28998. .29032)
                     /note="AT_rich"
                     /rpt_family="Low_complexity"
     repeat_region   30128. .34316
                     /note="L1PA2"
```

```
                     /rpt_family="LINE/L1"
     repeat_region   34319. .34374
                     /note="(GAAAA)n"
                     /rpt_family="Simple_repeat"
     repeat_region   complement(35013. .35117)
                     /note="MER34"
                     /rpt_family="Other/MER21_gro"
     repeat_region   35221. .35275
                     /note="MER4D"
                     /rpt_family="LTR/MER4-group"
     repeat_region   complement(35440. .35535)
                     /note="MER34"
                     /rpt_family="Other/MER21_gro"
     repeat_region   complement(36094. .36448)
                     /note="MER39"
                     /rpt_family="Other/MER21_gro"
     repeat_region   complement(36449. .36613)
                     /note="MER39b_j"
                     /rpt_family="Other/MER21_gro"
     repeat_region   36614. .36915
                     /note="AluY"
                     /rpt_family="SINE/Alu"
     repeat_region   complement(36916. .37067)
                     /note="MER39b_j"
                     /rpt_family="Other/MER21_gro"
ORIGIN

  Query Match             4.4%;  Score 266.6;  DB 9;  Length 37067;
  Best Local Similarity   72.1%;  Pred. No. 3.4e-48;
  Matches  465;  Conservative    0;  Mismatches  154;  Indels   26;  Gaps    8;

Qy        308 TGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAA 367
              ||  ||||  |   ||| || ||||||||||||||  | || || |||       | |
Db      28193 TGTTGTAGAAGCAGAAAGGTATGATACATTTTCTCACCAATTATCAGG-------CTGAC 28245

Qy        368 ACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAG 427
              | | |||||  |||||||||||||| |||   ||  | |||||  |  | ||||||||||
Db      28246 ATTTCTATAACAAAAGACAGGTTAACAAGGAAAAGGCATAACAGGTGCA-TTAATCAAAG 28304

Qy        428 TTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTG 487
              ||||| |||||||| |||||||||||||||| |||||||||||||| ||||||| |
Db      28305 TTTTATATGACATGAGAGTCTTCAGAAATGAAAACCCAAAGACCCAGGGAAAACTGT-TT 28363

Qy        488 TTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAA 547
              |||||  |||| || || |||||||  |||| ||||    |  |    ||||  ||||||
Db      28364 CTTTTTATGCTTAGTTTTGATGAAGTGTGGACAGCAGTGTGGAAATGTGATTGCACAAAA 28423

Qy        548 GGATATGATCTAGTGGTAAAGGACTCAGG-GGGAAACACAGCAAGGCCTGTCTATTCAGA 606
              || |||||| || |||||   ||   |||  |||||| |||| |||||||||  | ||||
Db      28424 GGGTATGATTTAATGGTAGTAGATGGAGGAAGGAAACCCAGCCCGGCCTGTCTGTCCAGA 28483

Qy        607 TTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGA 666
              ||||||||||  ||||           ||||| ||  |||||||| ||||||||||||||
Db      28484 TTCTTCTTGACTTCTC--------CACGTAGCAATCCATCCTCCTGGGTATGGGGCAGGA 28535

Qy        667 CTCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATG 726
              |  ||||  |||||||||| ||| | || |||||| |||||| |||   ||| |||||||
Db      28536 CCTTTCTGGAATGAGGGTC-TCAGGAGAAAAGGGATAAAGTGACCTTCCTAGGTTTTATG 28594

Qy        727 GCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCT 786
              ||||| || ||   ||||| |||||||||||||||| ||| |||||| ||||| ||||||
Db      28595 GCTTGATTTGGAAGAGAGGGGTTCTAGTTTCTATGATCCACCTTGGGCAAGAGAAATTCT 28654

Qy        787 GGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGG 846
              |||| || | |||   |||   | || |||||| ||   || |||||| |||| ||| ||
Db      28655 GGTTACTATAACTCAGTTTGGGGGAGGAAGAGGGGTGGAAGACAGGAGAGCAGCAGAGGG 28714

Qy        847 TCAGAAAGAGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAGTACTTCT 906
              ||||| |||   |    | |    |  |||| | ||||| |||| |||||    |
Db      28715 TCAGAGAGACCTT----GGTTCTGAGGCCTTTCCAATGTCCTTCAGTTTAAAGTA---TT 28767

Qy        907 TAGCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTC 951
              ||||| ||||| ||| |||||||||  ||| || || ||| |||
```

```
Db      28768 GAGCATGCCAAAACACCATACTTTGGGGTATTGTGTTTTGAACTC 28812


RESULT 48
HSF1614
LOCUS       HSF1614                78447 bp    DNA     linear   PRI 18-FEB-2000
DEFINITION  Homo sapiens chromosome 21 sequence from PAC RPCI-1 14F16 map 21q21
            region D21S189-LA297, complete sequence.
ACCESSION   AL109761
VERSION     AL109761.3  GI:7018381
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Ramser,J., Borzym,K., Langer,I., Lehrack,S., Klein,M., Arndt,S.,
            Marquardt,I., Noitz,M., Riesselmann,L., Yaspo,M.L., Reinhardt,R.
            and Lehrach,H.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 78447)
  AUTHORS   MPIMG.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-AUG-1999) MPIMG, Abt.Lehrach, Max Planck Institut
            Fuer Molekulare Genetik, Ihnestrasse 73, Berlin, 14195 Germany
COMMENT     On Feb 21, 2000 this sequence version replaced gi:6982091.
            contig 01         1. .78447
            Clone received from Resource Centre of the Human Genome Project at
            the Max-Planck-Institut for Molecular Genetics.
FEATURES             Location/Qualifiers
     source          1. .78447
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q21"
                     /clone="PAC RPCI-1 14F16"
                     /clone_lib="RPCI1,3-5 Human PAC library, originating
                     institute: Roswell Park Cancer Institute, creator: Pieter
                     de Jong, P.Ioannou"
                     /note="region between markers D21S189-LA297"
ORIGIN

  Query Match             4.4%;  Score 266.6;  DB 9;  Length 78447;
  Best Local Similarity   72.1%;  Pred. No. 3.2e-48;
  Matches  465;  Conservative    0;  Mismatches  154;  Indels   26;  Gaps    8;

Qy        308 TGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAA 367
              ||  ||||  |   ||| || ||||||||||||||  | || || |||       | |
Db      68306 TGTTGTAGAAGCAGAAAGGTATGATACATTTTCTCACCAATTATCAGG-------CTGAC 68358

Qy        368 ACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAG 427
              | | |||||  |||||||||||||| |||   ||  | |||||  |  | ||||||||||
Db      68359 ATTTCTATAACAAAAGACAGGTTAACAAGGAAAAGGCATAACAGGTGCA-TTAATCAAAG 68417

Qy        428 TTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTG 487
              ||||| |||||||| |||||||||||||||| |||||||||||||| ||||||| |
Db      68418 TTTTATATGACATGAGAGTCTTCAGAAATGAAAACCCAAAGACCCAGGGAAAACTGT-TT 68476

Qy        488 TTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAA 547
              |||||  |||| || || |||||||  |||| ||||    |  |    ||||  ||||||
Db      68477 CTTTTTATGCTTAGTTTTGATGAAGTGTGGACAGCAGTGTGGAAATGTGATTGCACAAAA 68536

Qy        548 GGATATGATCTAGTGGTAAAGGACTCAGG-GGGAAACACAGCAAGGCCTGTCTATTCAGA 606
              || |||||| || |||||   ||   |||  |||||| |||| |||||||||  | ||||
Db      68537 GGGTATGATTTAATGGTAGTAGATGGAGGAAGGAAACCCAGCCCGGCCTGTCTGTCCAGA 68596

Qy        607 TTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGA 666
              ||||||||||  ||||           ||||| ||  |||||||| ||||||||||||||
Db      68597 TTCTTCTTGACTTCTC--------CACGTAGCAATCCATCCTCCTGGGTATGGGGCAGGA 68648

Qy        667 CTCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATG 726
              |  ||||  |||||||||| ||| | || |||||| |||||| |||   ||| |||||||
```

```
Db      68649 CCTTTCTGGAATGAGGGTC-TCAGGAGAAAAGGGATAAAGTGACCTTCCTAGGTTTTATG 68707

Qy        727 GCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCT 786
              ||||| || ||    ||||| ||||||||||||||| ||| |||||| ||||| ||||||
Db      68708 GCTTGATTTGGAAGAGAGGGGTTCTAGTTTCTATGATCCACCTTGGGCAAGAGAAATTCT 68767

Qy        787 GGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGG 846
              |||| || | |||    |||   | || |||||| ||   || |||||| |||| ||| ||
Db      68768 GGTTACTATAACTCAGTTTGGGGGAGGAAGAGGGGTGGAAGACAGGAGAGCAGCAGAGGG 68827

Qy        847 TCAGAAAGAGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAGTACTTCT 906
              ||||| |||    |     | |     |   ||||  | ||||| |||| |||||    |
Db      68828 TCAGAGAGACCTT----GGTTCTGAGGCCTTTCCAATGTCCTTCAGTTTAAAGTA---TT 68880

Qy        907 TAGCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTC 951
              ||||| ||||| ||| |||||||||  ||| || || ||| |||
Db      68881 GAGCATGCCAAAACACCATACTTTGGGGTATTGTGTTTTGAACTC 68925


RESULT 49
AP001671/c
LOCUS       AP001671               340000 bp    DNA     linear   PRI 21-MAY-2003
DEFINITION  Homo sapiens genomic DNA, chromosome 21q, section 15/105.
ACCESSION   AP001671 AL163216 BA000005
VERSION     AP001671.1  GI:7768665
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
            Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E.,
            Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
            Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
            Patterson,D., Reichwald,K., Rump,A., Schillhabel,M., Schudy,A.,
            Zimmermann,W., Rosenthal,A., Kudoh,J., Shibuya,K., Kawasaki,K.,
            Asakawa,S., Shintani,A., Sasaki,K., Nagamine,K., Mitsuyama,S.,
            Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordsiek,G.,
            Hornischer,K., Barandt,P., Scharfe,M., Schoen,O., Desario,A.,
            Reichelt,J., Kauer,G., Bloecker,H., Ramser,J., Beck,A., Klages,S.,
            Hennig,S., Riesselmann,L., Dagand,E., Wehrmeyer,S., Borzym,K.,
            Gardiner,K., Nizetic,D., Francis,F., Lehrach,H., Reinhardt,R. and
            Yaspo,M.L.
  TITLE     The DNA sequence of human chromosome 21
  JOURNAL   Nature 405 (6784), 311-319 (2000)
  MEDLINE   20289799
   PUBMED   10830953
REFERENCE   2  (bases 1 to 340000)
  AUTHORS   Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
            Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E.,
            Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
            Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
            Patterson,D., Reichwald,K., Rump,A., Schillhabel,M., Schudy,A.,
            Zimmermann,W., Rosenthal,A., Kudoh,J., Shibuya,K., Kawasaki,K.,
            Asakawa,S., Shintani,A., Sasaki,T., Nagamine,K., Mitsuyama,S.,
            Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordsiek,G.,
            Hornischer,K., Barandt,P., Scharfe,M., Schoen,O., Desario,A.,
            Reichelt,J., Kauer,G., Bloecker,H., Ramser,J., Beck,A., Klages,S.,
            Hennig,S., Riesselmann,L., Dagand,E., Wehrmaeyer,S., Borzym,K.,
            Gardiner,K., Nizetic,D., Francis,F., Lehrach,H., Reinhardt,R. and
            Yaspo,M.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-APR-2000) The Chromosome 21 Mapping and Sequencing
            Consortium: * RIKEN Genomic Sciences Center, Human Genome Research
            Group * Institute of Molecular Biotechnology, Genome Analysis *
            Keio University School of Medicine, Dept. of Molecular Biology *
            GBF, Dept. of Genome Analysis * Max-Planck Institute for Molecular
            Genetics (addresses see below)
COMMENT     On May 30, 2000 this sequence version replaced gi:7717265.
            The chromosome 21 mapping and sequencing consortium consisting of
            * RIKEN Genomic Sciences Center, Human Genome Research Group, *
            Sagamihara 228-8555, Japan,
            * e.mail: hattori@gsc.riken.go.jp
            * URL: http://hgp.gsc.riken.go.jp/
            and
            * Institute of Molecular Biotechnology, Genome Analysis, *
            Beutenbergstrasse 11, D-07745 Jena, Germany,
            * e.mail: gscj-submit@genome.imb-jena.de
            * URL: http://genome.imb-jena.de/
            and
            * Keio University School of Medicine, Molecular Biology, * Tokyo
            160-8582, Japan,
            * e.mail: nshimizu@dmb-med.keio.ac.jp
            * URL: http://www.dmb.med.keio.ac.jp/
            and
            * GBF, Dept. of Genome Analysis,
            * Mascheroder Weg 1, D-38124 Braunschweig, Germany, * e.mail:
            info.genome@gbf.de
            * URL: http://genome.gbf.de/
            and
            * Max-Planck Institute for Molecular Genetics,
            * Ihnestrasse 73, D-14195 Berlin, Germany,
            * e.mail: info-chr21@molgen.mpg.de
            * URL: http://chr21.rz-berlin.mpg.de/
            AL163216: Submitted (10-Apr-2000).
FEATURES              Location/Qualifiers
     source           1. .340000
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:9606"
                      /chromosome="21"
     source           <1. .36248
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:9606"
                      /chromosome="21"
                      /map="21q21.1"
                      /clone="R396A17, 5' partial"
                      /clone_lib="RPCI-11 BAC library"
                      /note="Accession No. AP000432"
     source           34783. .113251
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:9606"
                      /chromosome="21"
                      /map="21q21.1"
                      /clone="P14F16"
                      /clone_lib="RPCI1,3-5 PAC library"
                      /note="Accession No. AL109761"
     source           75213. .140592
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:9606"
                      /chromosome="21"
                      /map="21q21.1"
                      /clone="pT1545"
                      /clone_lib="CMP21 P1 library"
                      /note="Accession No. AP000404"
     source           137289. .193160
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:9606"
                      /chromosome="21"
                      /map="21q21.1"
                      /clone="R546A20"
                      /clone_lib="RPCI-11 BAC library"
                      /note="Accession No. AP000745"
     source           192785. .284864
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:9606"
                      /chromosome="21"
                      /map="21q21.1"
                      /clone="P39J11"
                      /clone_lib="RPCI1,3-5 PAC library"
```

```
                /note="Accession No. AF165175"
     source     283283. .>340000
                /organism="Homo sapiens"
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"
                /chromosome="21"
                /map="21q21.1"
                /clone="R304D2, 3' partial"
                /clone_lib="RPCI-11 BAC library"
                /note="Accession No. AP000998"
     repeat_region   complement(3292. .3734)
                /note="L1MA9"
                /rpt_family="LINE/L1"
                /rpt_type=DISPERSED
     repeat_region   complement(3755. .4445)
                /note="THE1B-internal"
                /rpt_family="LTR/MaLR"
                /rpt_type=DISPERSED
     repeat_region   complement(4446. .4744)
                /note="THE1B"
                /rpt_family="LTR/MaLR"
                /rpt_type=DISPERSED
     repeat_region   complement(6107. .6389)
                /note="AluSx"
                /rpt_family="SINE/Alu"
                /rpt_type=DISPERSED
     repeat_region   complement(6590. .6926)
                /note="MLT1C"
                /rpt_family="LTR/MaLR"
                /rpt_type=DISPERSED
     repeat_region   6929. .7566
                /note="L1PB3"
                /rpt_family="LINE/L1"
                /rpt_type=DISPERSED
     repeat_region   complement(7610. .7829)
                /note="AluSx"
                /rpt_family="SINE/Alu"
                /rpt_type=DISPERSED
     repeat_region   7834. .8505
                /note="L1PB3"
                /rpt_family="LINE/L1"
                /rpt_type=DISPERSED
     repeat_region   complement(8506. .8544)
                /note="THE1C"
                /rpt_family="LTR/MaLR"
                /rpt_type=DISPERSED
     repeat_region   8545. .8840
                /note="AluY"
                /rpt_family="SINE/Alu"
                /rpt_type=DISPERSED
     repeat_region   complement(8841. .9188)
                /note="THE1C"
                /rpt_family="LTR/MaLR"
                /rpt_type=DISPERSED
     repeat_region   9189. .9350
                /note="L1PB3"
                /rpt_family="LINE/L1"
                /rpt_type=DISPERSED
     repeat_region   9359. .10350
                /note="SVA"
                /rpt_family="LTR/Retroviral"
                /rpt_type=DISPERSED
     repeat_region   10351. .10596
                /note="SVA"
                /rpt_family="LTR/Retroviral"
                /rpt_type=DISPERSED
     repeat_region   10597. .11085
                /note="SVA"
                /rpt_family="LTR/Retroviral"
                /rpt_type=DISPERSED
     repeat_region   11150. .11325
                /note="SVA"
                /rpt_family="LTR/Retroviral"
                /rpt_type=DISPERSED
     repeat_region   11326. .11828
```

Query Match 4.4%; Score 266.6; DB 9; Length 340000;
Best Local Similarity 72.1%; Pred. No. 2.9e-48;
Matches 465; Conservative 0; Mismatches 154; Indels 26; Gaps 8;

```
Qy        308 TGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAA 367
              ||  ||||  |   ||| || ||||||||||||||  | || || |||       | |
Db      44946 TGTTGTAGAAGCAGAAAGGTATGATACATTTTCTCACCAATTATCAGG-------CTGAC 44894

Qy        368 ACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAG 427
              | | |||||  |||||||||||||| |||   ||  | |||||  |  | ||||||||||
Db      44893 ATTTCTATAACAAAAGACAGGTTAACAAGGAAAAGGCATAACAGGTGCA-TTAATCAAAG 44835

Qy        428 TTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTG 487
              ||||| |||||||| |||||||||||||||| |||||||||||||||| ||||||| |
Db      44834 TTTTATATGACATGAGAGTCTTCAGAAATGAAAACCCAAAGACCCAGGGAAAACTGT-TT 44776

Qy        488 TTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAA 547
              ||||| |||| || || ||||||| |||| ||||    |  |    |||| ||||||
Db      44775 CTTTTTATGCTTAGTTTTGATGAAGTGTGGACAGCAGTGTGGAAATGTGATTGCACAAAA 44716

Qy        548 GGATATGATCTAGTGGTAAAGGACTCAGG-GGGAAACACAGCAAGGCCTGTCTATTCAGA 606
              || |||||| || |||||   ||   ||| |||||| |||| |||||||||  | ||||
Db      44715 GGGTATGATTTAATGGTAGTAGATGGAGGAAGGAAACCCAGCCCGGCCTGTCTGTCCAGA 44656

Qy        607 TTCTTCTTGATCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGA 666
              |||||||||  ||||           ||||| ||  ||||||||  ||||||||||||
Db      44655 TTCTTCTTGACTTCTC--------CACGTAGCAATCCATCCTCCTGGGTATGGGGCAGGA 44604

Qy        667 CTCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATG 726
              |  ||||  |||||||||| ||| | || |||||| |||||| ||||  ||| |||||||
Db      44603 CCTTTCTGGAATGAGGGTC-TCAGGAGAAAAGGGATAAAGTGACCTTCCTAGGTTTTATG 44545

Qy        727 GCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCT 786
              ||||| || ||   ||||| |||||||||||||||| ||| |||||| ||||| ||||||
Db      44544 GCTTGATTTGGAAGAGAGGGGTTCTAGTTTCTATGATCCACCTTGGGCAAGAGAAATTCT 44485

Qy        787 GGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGG 846
              |||| || | |||   |||   | || |||||| ||   || |||||| |||| ||| ||
Db      44484 GGTTACTATAACTCAGTTTGGGGGAGGAAGAGGGGTGGAAGACAGGAGAGCAGCAGAGGG 44425

Qy        847 TCAGAAAGAGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAGTACTTCT 906
              ||||| |||   |     | |     |  |||| | ||||| |||| ||||    |
Db      44424 TCAGAGAGACCTT----GGTTCTGAGGCCTTTCCAATGTCCTTCAGTTTAAAGTA---TT 44372

Qy        907 TAGCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTC 951
               ||||| ||||| ||| |||||||||  ||| || || ||| |||
Db      44371 GAGCATGCCAAAACACCATACTTTGGGGTATTGTGTTTTGAACTC 44327
```

```
RESULT 50
AL354856/c
LOCUS       AL354856              113146 bp    DNA     linear   PRI 19-FEB-2001
DEFINITION  Human DNA sequence from clone RP11-290P3 on chromosome 6 Contains
            STSs and GSSs, complete sequence.
ACCESSION   AL354856
VERSION     AL354856.9  GI:11121034
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 113146)
  AUTHORS   Bagguley,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Nov 8, 2000 this sequence version replaced gi:11043020.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
```

together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr6
IMPORTANT: This sequence is not the entire insert of clone RP11-290P3 It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.
The true left end of clone RP11-290P3 is at 1 in this sequence. The true left end of clone RP3-372K1 is at 113047 in this sequence. The true right end of clone RP11-343K13 is at 38139 in this sequence. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated repeat sequence elements. Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key. RP11-290P3 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6.

FEATURES Location/Qualifiers
source 1. .113146
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="6"
/clone="RP11-290P3"
/clone_lib="RPCI-11.2"
repeat_region 1011. .1325
/note="AluSg repeat: matches 1. .309 of consensus"
repeat_region 1338. .1388
/note="FRAM/FAM repeat: matches 1. .51 of consensus"
repeat_region 1382. .1504
/note="MER46B repeat: matches 84. .210 of consensus"
repeat_region 1505. .1846
/note="THE1B repeat: matches 1. .364 of consensus"
repeat_region 1847. .1916
/note="MER46B repeat: matches 13. .84 of consensus"
repeat_region 1917. .2232
/note="AluJb repeat: matches 1. .302 of consensus"
repeat_region 2233. .2383
/note="MER46B repeat: matches 2. .13 of consensus"
repeat_region 2666. .2957
/note="MLT1E repeat: matches 1. .292 of consensus"
repeat_region 3025. .3113
/note="MER34 repeat: matches 455. .543 of consensus"
repeat_region 3126. .3573
/note="LTR29 repeat: matches 3. .498 of consensus"
repeat_region 3574. .3676
/note="MER4B repeat: matches 1. .105 of consensus"
repeat_region 3672. .4235
/note="MER4D repeat: matches 229. .972 of consensus"
misc_feature 3963. .4105
/note="match: GSS: Em:B49391"
misc_feature complement(3978. .4268)
/note="match: GSS: Em:AQ344420"
misc_feature complement(3986. .4276)
/note="match: GSS: Em:AQ053552 Em:AQ342961"
misc_feature complement(3986. .4273)
/note="match: GSS: Em:AQ545892"
misc_feature 3986. .4102
/note="match: GSS: Em:AQ240745"
misc_feature complement(4006. .4274)
/note="match: GSS: Em:B94003"
misc_feature 4115. .4273
/note="match: GSS: Em:AQ240745"
misc_feature complement(4117. .4275)
/note="match: GSS: Em:AQ268095"
misc_feature 4122. .4273
/note="match: GSS: Em:AQ887080"
misc_feature complement(4135. .4273)
/note="match: GSS: Em:AQ037473"
misc_feature 4150. .4273
/note="match: GSS: Em:B47699"
repeat_region 4332. .4476
/note="MLT1E repeat: matches 372. .557 of consensus"
repeat_region 4878. .5034
/note="MIR repeat: matches 35. .205 of consensus"
repeat_region 5259. .5780
/note="MER4B repeat: matches 1. .539 of consensus"
repeat_region 5984. .6210
/note="MIR repeat: matches 3. .254 of consensus"
repeat_region 7278. .7819
/note="MER39 repeat: matches 18. .546 of consensus"
repeat_region 8154. .8224
/note="MLT1H repeat: matches 104. .172 of consensus"
misc_feature complement(8189. .8649)
/note="match: GSS: Em:AQ154917"
repeat_region 8247. .8361
/note="MER34 repeat: matches 18. .136 of consensus"
repeat_region 9030. .9468
/note="MER39 repeat: matches 13. .477 of consensus"
misc_feature complement(9093. .9684)
/note="match: GSS: Em:AQ063023"
repeat_region 9544. .10244
/note="MER50 repeat: matches 1. .711 of consensus"
repeat_region 10259. .10565
/note="AluY repeat: matches 1. .301 of consensus"
repeat_region 10914. .10981
/note="L2 repeat: matches 2691. .2750 of consensus"
repeat_region 10942. .10987
/note="MIR repeat: matches 203. .248 of consensus"
repeat_region 10988. .11137
/note="MER93 repeat: matches 5. .166 of consensus"
repeat_region 11268. .11464
/note="MER93 repeat: matches 201. .397 of consensus"
repeat_region 11644. .12014
/note="THE1C repeat: matches 1. .368 of consensus"
repeat_region 12610. .12902
/note="LTR40a repeat: matches 1. .272 of consensus"
repeat_region 12903. .13314
/note="LTR38 repeat: matches 149. .556 of consensus"
misc_feature 13036. .13514
/note="match: GSS: Em:AQ558061"
misc_feature 13326. .13801
/note="match: STS: Em:AF192026"
repeat_region 13392. .13631
/note="LTR40a repeat: matches 272. .519 of consensus"
repeat_region 13924. .14430
/note="MER9 repeat: matches 1. .511 of consensus"
repeat_region 14752. .14819
/note="MIR repeat: matches 1. .68 of consensus"
repeat_region 14822. .14863
/note="5S repeat: matches 1. .42 of consensus"
repeat_region 14864. .15067
/note="MIR repeat: matches 58. .262 of consensus"
repeat_region 15328. .15368
/note="L2 repeat: matches 2664. .2710 of consensus"
misc_feature 15441. .15580
/note="match: GSS: Em:AQ553037"
repeat_region 15595. .16074
/note="MER39 repeat: matches 13. .452 of consensus"
misc_feature complement(16914. .17735)
/note="match: GSS: Em:AQ742830"
misc_feature complement(17241. .17735)
/note="match: GSS: Em:AQ513131"
repeat_region 17327. .17382
/note="MIR repeat: matches 86. .140 of consensus"

```
     misc_feature    17764. .18371
                     /note="match: STS: Em:AF192023"
     repeat_region   18532. .19012
                     /note="L2 repeat: matches 2205. .2708 of consensus"
     repeat_region   19352. .19415
                     /note="32 copies 2 mer tt 67% conserved"
     repeat_region   19803. .19834
                     /note="8 copies 4 mer tcat 93% conserved"
     repeat_region   19985. .20103
                     /note="FLAM_C repeat: matches 2. .121 of consensus"
     repeat_region   20265. .20378
                     /note="L2 repeat: matches 2576. .2696 of consensus"
     repeat_region   20849. .20957
                     /note="MER34 repeat: matches 436. .543 of consensus"
     repeat_region   21012. .21263
                     /note="MER39 repeat: matches 199. .410 of consensus"
     repeat_region   21264. .21568
                     /note="AluY repeat: matches 3. .307 of consensus"
     repeat_region   21569. .21656
                     /note="MER39 repeat: matches 104. .199 of consensus"
     repeat_region   21657. .21905
                     /note="AluJ repeat: matches 1. .310 of consensus"
     repeat_region   22238. .22265
                     /note="7 copies 4 mer atga 96% conserved"
     misc_feature    22833. .23316
                     /note="match: GSS: Em:AQ467221"
     repeat_region   22874. .23096
                     /note="MIR repeat: matches 33. .259 of consensus"
     misc_feature    complement(24592. .25037)
                     /note="match: GSS: Em:AQ053426
                     match: STS: Em:G53032"
     repeat_region   25120. .25197
                     /note="2 copies 39 mer 94% conserved"
     misc_feature    complement(25253. .25618)
                     /note="match: GSS: Em:AQ506502"

  Query Match             4.4%;  Score 266.2;  DB 9;  Length 113146;
  Best Local Similarity   69.0%;  Pred. No. 3.8e-48;
  Matches  466;  Conservative    0;  Mismatches  173;  Indels   36;  Gaps    6;

Qy        298 CTCTCAAAGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGT 357
              ||||  | ||    ||| |||   |||  ||||| || ||| ||| |||| || |   |
Db      16087 CTCTTTATGCACTTGTAAGGGCAGAAAGATGTGACACCTTTCCTCATCCAGCACAGCTGA 16028

Qy        358 CGCAACCAAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTAT 417
              |    | | |||      |   |  ||| |||||  ||||| ||| | ||||||||||||
Db      16027 CATTTCTATAAC-----AAAAGACTGACTGGTTAGCAAGAGAAAAGCGTAACAAATTTAT 15973

Qy        418 TTAATCAAAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGG 477
              |||||||||||||||  ||||||||||  |||||||  ||||||||| |||||||| |||
Db      15972 TTAATCAAAGTTTTATGTGACATGGGAAACTTCAGAGATGAAGACCTAAAGACCCTGGGA 15913

Qy        478 AAACTGTCTGTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGA 537
              ||||  |||  ||||| |||| ||| || |||||||||| ||||      |    ||
Db      15912 AAACCATCT-ATTTTTATGCTTAGGATCAATGAAGAATGGGTAGCCGTGTAGAAGTGTGA 15854

Qy        538 TTAGACAAAAGGATATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGT 597
              || ||| ||||              |||||  |||| ||||   ||| |||  ||| |||
Db      15853 TTGGACCAAAG-------------GGTAATAGACTGAGGGCAGAACCCAGTGAGGTCTGA 15807

Qy        598 CTATTCAGATTCTTCTTGATCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTAT 657
              ||  |||||||||| |||  ||| |         ||| |||| |||||||||||||  |||
Db      15806 CTGATCAGATTCTTTTTGGCCTCCC--------TGTGTAGCCTTCTTTCCTCCTGGATAT 15755

Qy        658 GGGGCAGGACTCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTA 717
              |||||||||| | |||  ||||||  | ||||||||||||||||| |||| |||||  ||
Db      15754 GGGGCAGGACCCATCTGGAATGAGAATTTTCAAGGGAGAAGGGAGAGAGTGACCTTTATA 15695

Qy        718 GATTTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAG 777
              |||||||  |||||||  ||   | ||  |||||||||||| |||||||  ||||||  |||
Db      15694 TGTTTTATGACTTGCTTTGGAAGAAAGAGGTTCTAGTTTCTGTGACCCACCTTGGGAAAG 15635

Qy        778 AGGAATTCTGGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGC 837
```

```
              ||||||||||||||| ||||| |||||   | |||||||| | || ||  |||||| |
Db      15634 AGGAATTCTGGTTTCTATGACTCGCTTTGTGGTAGAAAGAGGGGCAACAGATAGGAGGAC 15575

Qy        838 AGGAGATGGTCAGAAAGAGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCA 897
               ||||| |  ||||| ||   |       |       | ||| |  ||||||||||| |
Db      15574 GGGAGAGGAGCAGAAGGACCGTGCCTCTC------AGTCTTTCAATCTCCTTTAGTTCAA 15521

Qy        898 AGTACTTCTTAGCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAACAC 957
              ||||   || ||||| |||| |  |  |||||||  |||  | || | ||| | || ||
Db      15520 AGTA---CTCAGCATGCCAAGGTGCCTTACTTTGGGGTATCATATTGTCAGCCCCAATAC 15464

Qy        958 TGCAATCATGCTAAA 972
                 || |   | ||
Db      15463 CCTCATGACTATCAA 15449


RESULT 51
AC068966/c
LOCUS       AC068966               176042 bp    DNA     linear   HTG 08-JUL-2000
DEFINITION  Homo sapiens chromosome 6 clone RP11-290P3, WORKING DRAFT SEQUENCE,
            5 unordered pieces.
ACCESSION   AC068966
VERSION     AC068966.5  GI:8927682
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 176042)
  AUTHORS   Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
            Federspiel,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,
            Mao,J., Marathe,R., Morehouse,A.J., Oefner,P., Palm,C.J.,
            Ramirez,D., Wilhelmy,J., Yu,S. and Davis,R.W.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 176042)
  AUTHORS   Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
            Federspiel,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,
            Mao,J., Marathe,R., Morehouse,A.J., Oefner,P., Palm,C.J.,
            Ramirez,D., Wilhelmy,J., Yu,S. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-MAY-2000) DNA Sequencing and Technology Center,
            Stanford University, 855 California Avenue, Palo Alto, CA 94304,
            USA
COMMENT     On Jul 6, 2000 this sequence version replaced gi:7960317.
            -------------- Genome Center
                 Center: Stanford DNA Sequencing and Technology Development
            Center
                 Center code: SDSTDC
                 Web site: http://sequence-www.stanford.edu/group/human/
                 Contact: hum-info@sequence.stanford.edu
            -------------- Project Information
                 Center project name: 930
                 Center clone name: RP11-290P3
            -------------- Summary Statistics
                 Sequencing Vector: M13mp18; X02513
                 Chemistry: Dye-primer; 0% of reads
                 Chemistry: Dye-terminator Big Dye; 100% of reads
                 Assembly program: Phrap; version 0.990319
                 Consensus quality: 168170 bases at least Q40
                 Consensus quality: 170491 bases at least Q30
                 Consensus quality: 170900 bases at least Q20
                 Insert size: 166209; agarose-fp
                 Insert size: 175642; sum-of-contigs
                 Quality coverage:  6.4x in Q20 bases; agarose-fp
                 Quality coverage:  6.1x in Q20 bases; sum-of-contigs.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
```

```
            *        1     1648: contig of 1648 bp in length
            *     1649     1748: gap of unknown length
            *     1749     7331: contig of 5583 bp in length
            *     7332     7431: gap of unknown length
            *     7432    24054: contig of 16623 bp in length
            *    24055    24154: gap of unknown length
            *    24155    68765: contig of 44611 bp in length
            *    68766    68865: gap of unknown length
            *    68866   176042: contig of 107177 bp in length.
FEATURES             Location/Qualifiers
     source          1. .176042
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP11-290P3"
                     /clone_lib="RPCI human BAC library 11"
     misc_feature    1. .1648
                     /note="assembly_name:Contig2"
     misc_feature    1749. .7331
                     /note="assembly_name:Contig4
                     clone_end:SP6
                     vector_side:left"
     misc_feature    7432. .24054
                     /note="assembly_name:Contig5
                     clone_end:T7
                     vector_side:left"
     misc_feature    24155. .68765
                     /note="assembly_name:Contig6"
     misc_feature    68866. .176042
                     /note="assembly_name:Contig7"
ORIGIN

  Query Match             4.4%;  Score 266.2;  DB 2;  Length 176042;
  Best Local Similarity   69.0%;  Pred. No. 3.7e-48;
  Matches  466;  Conservative    0;  Mismatches  173;  Indels   36;  Gaps    6;

Qy        298 CTCTCAAAGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGT 357
              ||||  | ||   ||| |||   |||  ||||| || ||| ||| |||| || |  |
Db      23524 CTCTTTATGCACTTGTAAGGGCAGAAAGATGTGACACCTTTCCTCATCCAGCACAGCTGA 23465

Qy        358 CGCAACCAAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTAT 417
              |   | | |||       |   | ||| |||||  ||||| ||| | ||||||||||
Db      23464 CATTTCTATAAC-----AAAAGACTGACTGGTTAGCAAGAGAAAAGCGTAACAAATTTAT 23410

Qy        418 TTAATCAAAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGG 477
              ||||||||||||||  |||||||||  |||||| |||||||| |||||||  |||
Db      23409 TTAATCAAAGTTTTATGTGACATGGGAAACTTCAGAGATGAAGACCTAAAGACCCTGGGA 23350

Qy        478 AAACTGTCTGTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGA 537
              ||||  |||  ||||| |||| ||| || |||||||||| ||||       |   ||
Db      23349 AAACCATCT-ATTTTTATGCTTAGGATCAATGAAGAATGGGTAGCCGTGTAGAAGTGTGA 23291

Qy        538 TTAGACAAAAGGATATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGT 597
              || ||| ||||              |||||  |||| ||||   ||| |||  ||| |||
Db      23290 TTGGACCAAAG------------GGTAATAGACTGAGGGCAGAACCCAGTGAGGTCTGA 23244

Qy        598 CTATTCAGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTAT 657
              ||  |||||||||| |||  ||| |        ||| |||| ||||||||||||  |||
Db      23243 CTGATCAGATTCTTTTTGGCCTCCC--------TGTGTAGCCTTCTTTCCTCCTGGATAT 23192

Qy        658 GGGGCAGGACTCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTA 717
              |||||||||| | |||  ||||||  | ||||||||||||||||||| |||| ||||| ||
Db      23191 GGGGCAGGACCCATCTGGAATGAGAATTTTCAAGGGAGAAGGGAGAGAGTGACCTTTATA 23132

Qy        718 GATTTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAG 777
              |||||||  |||||||  ||    | ||  |||||||||||| |||||| ||||||  |||
Db      23131 TGTTTTATGACTTGCTTTGGAAGAAAGAGGTTCTAGTTTCTGTGACCCACCTTGGGAAAG 23072

Qy        778 AGGAATTCTGGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGC 837
              |||||||||||||||| ||||| |||||   | |||||||| | || ||  |||||| |
Db      23071 AGGAATTCTGGTTTCTATGACTCGCTTTGTGGTAGAAAGAGGGGCAACAGATAGGAGGAC 23012

Qy        838 AGGAGATGGTCAGAAAGAGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCA 897
              |||||  |  ||||| ||   |      |        | |||  |  |||||||||||| |
Db      23011 GGGAGAGGAGCAGAAGGACCGTGCCTCTC------AGTCTTTCAATCTCCTTTAGTTCAA 22958

Qy        898 AGTACTTCTTAGCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAACAC 957
              ||||   || ||||| |||| |  |  |||||||  |||  | || | ||| | || ||
Db      22957 AGTA---CTCAGCATGCCAAGGTGCCTTACTTTGGGGTATCATATTGTCAGCCCCAATAC 22901

Qy        958 TGCAATCATGCTAAA 972
                 || |   | ||
Db      22900 CCTCATGACTATCAA 22886


RESULT 52
CQ714476
LOCUS       CQ714476                 840 bp    DNA     linear   PAT 03-FEB-2004
DEFINITION  Sequence 410 from Patent WO02068579.
ACCESSION   CQ714476
VERSION     CQ714476.1  GI:42275333
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Venter,C.J., Adams,M.C., Li,P.W. and Myers,E.W.
  TITLE     Kits, such as nucleic acid arrays, comprising a majority of
            humanexons or transcripts, for detecting expression and other uses
            thereof
  JOURNAL   Patent: WO 02068579-A 410 06-SEP-2002;
            PE Corporation (NY) (US)
FEATURES             Location/Qualifiers
     source          1. .840
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

  Query Match             4.4%;  Score 266;  DB 6;  Length 840;
  Best Local Similarity   100.0%;  Pred. No. 6.1e-48;
  Matches  266;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5363 GCAGAGGGAGACAGAGGAGCTGGTACTGCAGAGCGGTCGTCTGATTGGCTGGACGGTCGT 5422
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCAGAGGGAGACAGAGGAGCTGGTACTGCAGAGCGGTCGTCTGATTGGCTGGACGGTCGT 60

Qy       5423 AGCTGGGCTATAAAAGAGACCCCTACAGGCTTAGCAGGAAGACGCTCAGAGGATTCTGAC 5482
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AGCTGGGCTATAAAAGAGACCCCTACAGGCTTAGCAGGAAGACGCTCAGAGGATTCTGAC 120

Qy       5483 AATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGCCGAAGCCACAGCTCCTCCTGC 5542
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGCCGAAGCCACAGCTCCTCCTGC 180

Qy       5543 CGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAATAAGATGTGCAAAGGGCTTGCA 5602
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAATAAGATGTGCAAAGGGCTTGCA 240

Qy       5603 GGTCTGCCGGCTTCTTGCTTGAGGAG 5628
              |||||||||||||||||||||||||||
Db        241 GGTCTGCCGGCTTCTTGCTTGAGGAG 266


RESULT 53
CNS01DTT/c
LOCUS       CNS01DTT                196993 bp    DNA     linear   PRI 26-APR-2001
DEFINITION  Human chromosome 14 DNA sequence BAC R-1103G16 of library RPCI-11
            from chromosome 14 of Homo sapiens (Human), complete sequence.
ACCESSION   AL132985
VERSION     AL132985.4  GI:12619148
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
```

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 196993)
  AUTHORS   Heilig,R., Petit,J.L., Vico,V., Dasilva,C., Robert,C., Wincker,P.,
            Brottier,P., Cattolico,L., Barbe,V., Pelletier,E., Artiguenave,F.,
            Levy,M., Eckenberg,R., Bruls,T., deBerardinis,V., Cruaud,C.,
            Gyapay,G., Saurin,W. and Weissenbach,J.
  TITLE     Sequencing of the human chromosome 14
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 196993)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-APR-2001) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     On Jan 30, 2001 this sequence version replaced gi:11611135.
            -------------- Genome Center
            Center: Genoscope / Centre National de Sequencage
            Center code: GS
            Web site: http://www.genoscope.cns.fr/
            Contact: SeqRef@genoscope.cns.fr
            --------------
            The following BAC sequence is oriented from the T7 to the SP6 end.
            Upstream BAC (overlapping the T7 end) : R-354E14 (AC=AL117355)
            Downstream BAC (overlapping the SP6 end) : R-306B9 ----------
            Summary Statistics
            Assembly program: Phrap; version 2.0
            Quality coverage: 6.72x in Q20 bases; sum-of-contigs
            --------------
            ------------------------
            Overall quality chart :
            Range     :   bases
            0         :
            1   -  9  :      16
            10  - 19  :     106
            20  - 29  :     408
            30  - 39  :    2050
            40  - 49  :   14617
            50  - 59  :   10762
            60  - 69  :   12876
            70  - 79  :   31124
            80  - 89  :   68200
            90  - 99  :   56834
            ------------------------
            Percentage of bases with a quality value >= 40 : 98 %.
FEATURES              Location/Qualifiers
     source           1. .196993
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:9606"
                      /chromosome="14"
                      /clone="R-1103G16"
                      /clone_lib="RPCI-11"
     STS              72539. .72684
                      /note="matching EMBL:G07569
                      RHdb:RH32505
                      dbSTS:STS11342
                      Identified using the e-PCR software (G. Schuler)"
ORIGIN

  Query Match            4.4%;  Score 265.8;  DB 9;  Length 196993;
  Best Local Similarity   72.2%;  Pred. No. 4.5e-48;
  Matches  440;  Conservative    0;  Mismatches  137;  Indels   32;  Gaps    6;

Qy        312 GTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTC 371
              ||| |||   ||| |||||||| ||| ||| || | ||||   || || |||| ||||
Db      25502 GTAAGGGCTGAAAGGTGTGATATCTTTCCTCACCCGTTATAATCATCACAGCCAACACTC 25443

Qy        372 CTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTT 431
              |||||  ||||  |||| ||| ||||  ||     |||| |||| |||||||||| ||||
Db      25442 CTATAACAAAAACCAGGATAACAAGAAAAA-----AACATATTTCTTTAATCAAACTTTT 25388

Qy        432 ACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTT 491
              |||||||| |||| |||||||||||| ||| ||| | |||||| |||   | || |||
Db      25387 ACATGACATAGGAGGCTTCAGAAATGAAAACCTAAAAATCCAGGGAAAATAATTTG-TTT 25329

Qy        492 TTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGAT 551
              || | || |||||| |||||||| || |||         |    ||||||||||||| |
Db      25328 TTATACTTAGGTTCAATGAAGAATAGACAGCTGTGTAGAAATGTGATTAGACAAAAGGGT 25269

Qy        552 ATGATCTAGTGGTAAAGGACTCAGGG-GGAAACACAGCAAGGCCTGTCTATTCAGATTCT 610
              |||||||| ||||||  |||| |||| ||| |  |||||| |||||  | ||||||||||
Db      25268 ATGATCTAATGGTAACAGACTGAGGGAGGAGATCCAGCAAAGCCTGCTTGTTCAGATTCT 25209

Qy        611 TCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCT 670
              |||||  |||||| |      |          |||||||    | |||| |||||  ||
Db      25208 TCTTGGCCTCTCTGTGCAACATTC---------TTCCTCCCAGATTTGGGCCAGGATCCT 25158

Qy        671 TCTTCAATGAGGGTCTTCAAG-------GGAGAAGGGAGAAAGTGGCCTTTTTAGATTTT 723
              | |  |||||||| |||||||        ||| |||||| | ||| |||||| |||  ||
Db      25157 TATGGAATGAGGGCCTTCAAGAGAGAATGGAAAAGGGAAAGAGTAGCCTTTCTAGGATTG 25098

Qy        724 ATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAAT 783
              |||||||||| |||| ||| |  |||||||||||||||||  || |||||  |||||
Db      25097 ATGGCTTGCTTTGGGGGAGAAGGATTCTAGTTTCTATGACCCACATTAGGGAAATGGAAT 25038

Qy        784 TCTGGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGA 843
              |||||| ||| | ||| ||||    | |||| |||| | | ||||||||||||||| |||
Db      25037 TCTGGTCTCTATAACTCGCTTCAGGGGAGAAGGAGGGGCAGGAGGCAGGAGGGCAGAAGA 24978

Qy        844 TGGTCAGAAAGAGACT---------TGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTT 894
              || || | |||| ||          ||||| |||||||| |||||  |||||||||| |
Db      24977 AGGGCAAAGAGAGGCTATGCTTCTGGGGCTGTTTCTGAGGCCTTCCAGTCTCCTTTAGCT 24918

Qy        895 CCAAGTACT 903
              | |||||||
Db      24917 CAAAGTACT 24909


RESULT 54
HS315B17
LOCUS       HS315B17               53185 bp    DNA     linear   PRI 05-MAR-2003
DEFINITION  Human DNA sequence from clone RP1-315B17 on chromosome Xq22.2-22.3,
            complete sequence.
ACCESSION   Z73967
VERSION     Z73967.1  GI:1370101
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 53185)
  AUTHORS   Sulston,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAR-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
```

```
          occasion of the clone being a YAC.
          The following abbreviations are used to associate primary accession
          numbers given in the feature table with their source databases:
          Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
          on the WORMPEP database can be found at
          http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
          was generated from part of bacterial clone contigs of human
          chromosome X, constructed by the Sanger Centre Chromosome X Mapping
          Group.  Further information can be found at
          http://www.sanger.ac.uk/HGP/ChrX
          RP1-315B17 is from the library RPCI-1 constructed by the group of
          Pieter de Jong. For further details see
          http://www.chori.org/bacpac/home.htm
          VECTOR: pCYPAC2.
FEATURES             Location/Qualifiers
     source          1. .53185
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="RZPD:RPCIP704B17315"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="q22.2-22.3"
                     /clone="RP1-315B17"
                     /clone_lib="RPCI-1"
ORIGIN

  Query Match             4.3%;  Score 264.2;  DB 9;  Length 53185;
  Best Local Similarity   68.8%;  Pred. No. 1.1e-47;
  Matches  460;  Conservative    0;  Mismatches  183;  Indels   26;  Gaps    6;

Qy        305 AGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACC 364
              || |   |||   ||  ||| ||||||||| ||| |||   | ||||||||||   || ||
Db      49694 AGATAGTGTAAAAGTACAAAAGTGTGATACCTTTCCTCACCAATCATAAGGGTAACAGCC 49753

Qy        365 AAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCA 424
              || |||||||||  |||||| ||||||| |||||  || | | ||||||||||||||||||
Db      49754 AAGACTCCTATAACAAAAGATAGGTTAACAAGAGACAAGCATGACAAATTTATTTAATCA 49813

Qy        425 AAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGT 484
              || ||| ||||||||  | || |||| ||||||| |||||  | | ||  | | ||
Db      49814 AAATTTCACATGACACAGCAGCCTTCGGAAATGAGGACCCCCAAAACCTAGAGGAA---- 49869

Qy        485 CTGTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACA 544
                | | | | || |||||| || ||||||||||||        |    |||||||||
Db      49870 ACCATCTATATACTTAGGTTCAATAAAGAATGGATAGCCACGTAGAAATGTGATTAGACA 49929

Qy        545 AAAGGATATGATCTAGTGGTAAAGGACT-CAGGGGGAAACACAGCAAGGCCTGTCTATTC 603
              ||||  ||||| ||| ||||||  || | |  ||||  |  |||||||||| ||| |||
Db      49930 AAAGAGTATGACCTAATGGTAATAGATTGCGTGGGGGGATCCAGCAAGGCCTATCTGTTC 49989

Qy        604 AGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCA 663
              |||   ||  ||||  ||||         ||| |||||||| ||| |||| ||||||   ||
Db      49990 AGAGACTGGTTGACTTCTC--------TGTGTAGCATTCCTTCTTCCTCAGTATGATACA 50041

Qy        664 GGACTCTTCTTCAATGAGGGTCTTCAAGGGAGA-------AGGGAGAAAGTGGCCTTTTT 716
                |||   |||  ||||| |||||||||||| |       ||||||| |||  ||||| |
Db      50042 CGACCTCTCTGGAATGAAGGTCTTCAAGGGAAAGGGGAACAGGGAGAGAGTTACCTTTCT 50101

Qy        717 AGATTTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAA 776
              || ||| ||||||||||  ||| ||||| ||||||||||  |||||||   |||| |||
Db      50102 AGGTTTAATGGCTTGCTTTTGGGGAGAGGGGTTCTAGTTTTCATGACCCTCTTTGGAGAA 50161

Qy        777 GAGGAATTCTGGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGG 836
              ||||||||||||||   ||||||||||    | ||||  ||||| | ||| ||||||
Db      50162 GAGGAATTCTGGTTTCCATGACTTGCTTTGGGGGAGAACAAGGAGGAGGAGACAGGAGAA 50221

Qy        837 CAGGAGATGGTCAGAAAGAGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCC 896
                  |||  |    ||    ||    |||||||||| ||| |    | | ||   | |
Db      50222 TGTAAGAAAGATCAGAAAGACCT----TGCTTCTGAGGCCTTTCTAGTTTCCTTCAGTTC 50277

Qy        897 AAGTACTTCTTAGCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAACA 956
              | | |  | || || |||| |  | |||||| ||  |||  | |||||||| | ||||
Db      50278 --GAAGTATTCAGTATGCCAAGGTGCCATACTTCGGGGTATCATGTTCTGAGCCCCAACA 50335

Qy        957 CTGCAATCA 965
                |   || |
Db      50336 ATAGCATAA 50344


RESULT 55
AC022962
LOCUS       AC022962               180008 bp    DNA     linear   HTG 26-MAY-2000
DEFINITION  Homo sapiens clone RP11-258L5, WORKING DRAFT SEQUENCE, 6 unordered
            pieces.
ACCESSION   AC022962
VERSION     AC022962.3  GI:8076841
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 180008)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome, clone RP11-258L5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 180008)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 25, 2000 this sequence version replaced gi:7596819.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L6427
                Center clone name: 258_L_5
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 176901 bases at least Q40
                Consensus quality: 178256 bases at least Q30
                Consensus quality: 178929 bases at least Q20
                Insert size: 176000; agarose-fp
                Insert size: 179508; sum-of-contigs
                Quality coverage: 8.7 in Q20 bases; agarose-fp
                Quality coverage: 8.5 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
```

```
           * as soon as it is available and the accession number will
           * be preserved.
           *         1    24517: contig of 24517 bp in length
           *     24518    24617: gap of 100 bp
           *     24618    47524: contig of 22907 bp in length
           *     47525    47624: gap of 100 bp
           *     47625    71798: contig of 24174 bp in length
           *     71799    71898: gap of 100 bp
           *     71899    99611: contig of 27713 bp in length
           *     99612    99711: gap of 100 bp
           *     99712   134522: contig of 34811 bp in length
           *    134523   134622: gap of 100 bp
           *    134623   180008: contig of 45386 bp in length.
FEATURES             Location/Qualifiers
     source          1. .180008
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="RP11-258L5"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .24517
                     /note="assembly_fragment"
     misc_feature    24618. .47524
                     /note="assembly_fragment"
     misc_feature    47625. .71798
                     /note="assembly_fragment"
     misc_feature    71899. .99611
                     /note="assembly_fragment"
     misc_feature    99712. .134522
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     misc_feature    134623. .180008
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
ORIGIN

  Query Match             4.3%;  Score 264.2;  DB 2;  Length 180008;
  Best Local Similarity   68.8%;  Pred. No. 1e-47;
  Matches  460;  Conservative    0;  Mismatches  183;  Indels   26;  Gaps    6;

Qy        305 AGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACC 364
              || |   |||   ||  ||| ||||||||| ||| |||  | ||||||||||  || ||
Db     137881 AGATAGTGTAAAAGTACAAAAGTGTGATACCTTTCCTCACCAATCATAAGGGTAACAGCC 137940

Qy        365 AAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCA 424
              || |||||||||  |||||| ||||||| |||||  || | | ||||||||||||||||||
Db     137941 AAGACTCCTATAACAAAAGATAGGTTAACAAGAGACAAGCATGACAAATTTATTTAATCA 138000

Qy        425 AAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGT 484
              || ||| ||||||||  | || |||| ||||||| |||||  | | ||  | | ||
Db     138001 AAATTTCACATGACACAGCAGCCTTCGGAAATGAGGACCCCCAAAACCTAGAGGAA---- 138056

Qy        485 CTGTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACA 544
              | | | | || |||||| || ||||||||||||||         |     |||||||||
Db     138057 ACCATCTATATACTTAGGTTCAATAAAGAATGGATAGCCACGTAGAAATGTGATTAGACA 138116

Qy        545 AAAGGATATGATCTAGTGGTAAAGGACT-CAGGGGGAAACACAGCAAGGCCTGTCTATTC 603
              ||||  ||||| ||| ||||||  || | |  |||| |  |||||||||| ||| |||
Db     138117 AAAGAGTATGACCTAATGGTAATAGATTGCGTGGGGGGATCCAGCAAGGCCTATCTGTTC 138176

Qy        604 AGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCA 663
              |||  ||  |||| ||||        ||| |||||||| ||| |||| ||||||   ||
Db     138177 AGAGACTGGTTGACTTCTC--------TGTGTAGCATTCCTTCTTCCTCAGTATGATACA 138228

Qy        664 GGACTCTTCTTCAATGAGGGTCTTCAAGGGAGA-------AGGGAGAAAGTGGCCTTTTT 716
               |||   |||  ||||| |||||||||||| |       ||||||| |||  ||||| |
Db     138229 CGACCTCTCTGGAATGAAGGTCTTCAAGGGAAAGGGGAACAGGGAGAGAGTTACCTTTCT 138288

Qy        717 AGATTTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAA 776
              || ||| |||||||||||  ||| |||| ||||||||||  |||||||  |||| |||
Db     138289 AGGTTTAATGGCTTGCTTTTGGGGAGAGGGGTTCTAGTTTTCATGACCCTCTTTGGAGAA 138348

Qy        777 GAGGAATTCTGGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGG 836
              ||||||||||||||||  ||||||||||    | ||||  |||||  | |||  ||||||
Db     138349 GAGGAATTCTGGTTTCCATGACTTGCTTTGGGGGAGAACAAGGAGGAGGAGACAGGAGAA 138408

Qy        837 CAGGAGATGGTCAGAAAGAGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCC 896
                |||  |     ||    ||    |||||||||||  |||  |    | | ||   | |
Db     138409 TGTAAGAAAGATCAGAAAGACCT----TGCTTCTGAGGCCTTTCTAGTTTCCTTCAGTTC 138464

Qy        897 AAGTACTTCTTAGCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAACA 956
                | | |  | || || |||| |  | |||||| ||  |||  | ||||||||  | ||||
Db     138465 --GAAGTATTCAGTATGCCAAGGTGCCATACTTCGGGGTATCATGTTCTGAGCCCCAACA 138522

Qy        957 CTGCAATCA 965
                |   || |
Db     138523 ATAGCATAA 138531


RESULT 56
AC024411
LOCUS       AC024411               170800 bp    DNA     linear   HTG 25-MAY-2000
DEFINITION  Homo sapiens chromosome 21 clone RP11-383H8 map 21, WORKING DRAFT
            SEQUENCE, 14 unordered pieces.
ACCESSION   AC024411
VERSION     AC024411.3  GI:8076906
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 170800)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 21, clone RP11-383H8
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 170800)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M.,
            Fenestor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., Landers,T., Largocque,K., Lehoczky,J., Levine,R.,
            Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
            McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivar,T.M.,
            Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
            Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
            Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
            Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
            Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 25, 2000 this sequence version replaced gi:7272333.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L6873
                Center clone name: 383_H_8
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
```

```
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 161079 bases at least Q40
            Consensus quality: 165699 bases at least Q30
            Consensus quality: 167817 bases at least Q20
            Insert size: 170000; agarose-fp
            Insert size: 169500; sum-of-contigs
            Quality coverage: 4.1 in Q20 bases; agarose-fp
            Quality coverage: 4.2 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 14 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1034: contig of 1034 bp in length
            *     1035     1134: gap of 100 bp
            *     1135     2818: contig of 1684 bp in length
            *     2819     2918: gap of 100 bp
            *     2919     5025: contig of 2107 bp in length
            *     5026     5125: gap of 100 bp
            *     5126     6457: contig of 1332 bp in length
            *     6458     6557: gap of 100 bp
            *     6558     9772: contig of 3215 bp in length
            *     9773     9872: gap of 100 bp
            *     9873    14119: contig of 4247 bp in length
            *    14120    14219: gap of 100 bp
            *    14220    18249: contig of 4030 bp in length
            *    18250    18349: gap of 100 bp
            *    18350    23124: contig of 4775 bp in length
            *    23125    23224: gap of 100 bp
            *    23225    26842: contig of 3618 bp in length
            *    26843    26942: gap of 100 bp
            *    26943    43018: contig of 16076 bp in length
            *    43019    43118: gap of 100 bp
            *    43119    63582: contig of 20464 bp in length
            *    63583    63682: gap of 100 bp
            *    63683    88966: contig of 25284 bp in length
            *    88967    89066: gap of 100 bp
            *    89067   126431: contig of 37365 bp in length
            *   126432   126531: gap of 100 bp
            *   126532   170800: contig of 44269 bp in length.
FEATURES             Location/Qualifiers
     source          1. .170800
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21"
                     /clone="RP11-383H8"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1034
                     /note="assembly_fragment"
     misc_feature    1135. .2818
                     /note="assembly_fragment"
     misc_feature    2919. .5025
                     /note="assembly_fragment"
     misc_feature    5126. .6457
                     /note="assembly_fragment"
     misc_feature    6558. .9772
                     /note="assembly_fragment"
     misc_feature    9873. .14119
                     /note="assembly_fragment"
     misc_feature    14220. .18249
                     /note="assembly_fragment"
     misc_feature    18350. .23124
                     /note="assembly_fragment"
     misc_feature    23225. .26842
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    26943. .43018
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    43119. .63582
                     /note="assembly_fragment"
     misc_feature    63683. .88966
                     /note="assembly_fragment"
     misc_feature    89067. .126431
                     /note="assembly_fragment"
     misc_feature    126532. .170800
                     /note="assembly_fragment"
ORIGIN

  Query Match             4.3%;  Score 261.8;  DB 2;  Length 170800;
  Best Local Similarity   71.6%;  Pred. No. 3.4e-47;
  Matches  462;  Conservative    0;  Mismatches  157;  Indels   26;  Gaps    8;

Qy        308 TGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAA 367
              ||  ||||  |   ||| || ||||||||||||||  | || || |||       | |
Db     164063 TGTTGTAGAAGCAGAAAGGTATGATACATTTTCTCACCAATTATCAGG-------CTGAC 164115

Qy        368 ACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAG 427
              | | |||||  |||||||||    || |||    ||  | |||||   |  | ||||||||||
Db     164116 ATTTCTATAACAAAAGACAGAAAAACAAGGAAAAGGCATAACAGGTGCA-TTAATCAAAG 164174

Qy        428 TTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTG 487
              |||||  |||||||| |||||||||||||||||| |||||||||||||||| ||||||| |
Db     164175 TTTTATATGACATGAGAGTCTTCAGAAATGAAAACCCAAAGACCCAGGGAAAACTGT-TT 164233

Qy        488 TTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAA 547
              |||||  |||| || || |||||||  |||| ||||     |  |    ||||  ||||||
Db     164234 CTTTTTATGCTTAGTTTTGATGAAGTGTGGACAGCAGTGTGGAAATGTGATTGCACAAAA 164293

Qy        548 GGATATGATCTAGTGGTAAAGGACTCAGG-GGGAAACACAGCAAGGCCTGTCTATTCAGA 606
              || |||||| || |||||   ||    |||  ||||||  ||||  |||||||||  | ||||
Db     164294 GGGTATGATTTAATGGTAGTAGATGGAGGAAGGAAACCCAGCCCGGCCTGTCTGTCCAGA 164353

Qy        607 TTCTTCTTGATCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGA 666
              ||||||||||  ||||          |||||  ||  ||||||||  ||||||||||||||
Db     164354 TTCTTCTTGACTTCTC--------CACGTAGCAATCCATCCTCCTGGGTATGGGGCAGGA 164405

Qy        667 CTCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATG 726
              |  ||||  |||||||||| ||| | || |||||| |||||| ||||  ||| |||||||
Db     164406 CCTTTCTGGAATGAGGGTC-TCAGGAGAAAAGGGATAAAGTGACCTTCCTAGGTTTTATG 164464

Qy        727 GCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCT 786
              ||||| || ||   ||||| ||||||||||||||||| ||| |||||| |||| ||||||
Db     164465 GCTTGATTTGGAAGAGAGGGGTTCTAGTTTCTATGATCCACCTTGGGCAAGAGAAATTCT 164524

Qy        787 GGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGG 846
              |||| || | |||   |||  | || ||||||  ||   || ||||||  |||| ||| ||
Db     164525 GGTTACTATAACTCAGTTTGGGGGAGGAAGAGGGGTGGAAGACAGGAGAGCAGCAGAGGG 164584

Qy        847 TCAGAAAGAGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAGTACTTCT 906
              ||||| |||   |    | |      |  ||||  | |||||  ||||  |||||    |
Db     164585 TCAGAGAGACCTT----GGTTCTGAGGCCTTTCCAATGTCCTTCAGTTTAAAGTA---TT 164637

Qy        907 TAGCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTC 951
              |||||  |||||  ||| |||||||||  ||| || || ||| |||
Db     164638 GAGCATGCCAAAACACCATACTTTGGGGTATTGTGTTTTGAACTC 164682


RESULT 57
AC021279/c
LOCUS       AC021279              154413 bp    DNA     linear   HTG 01-APR-2000
DEFINITION  Homo sapiens clone RP11-21C8, WORKING DRAFT SEQUENCE, 11 unordered
            pieces.
ACCESSION   AC021279
VERSION     AC021279.3  GI:7382150
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
```

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 154413)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-21C8
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 154413)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Apr 1, 2000 this sequence version replaced gi:6910881.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L3980
                Center clone name: 21_C_8
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 146591 bases at least Q40
                Consensus quality: 150615 bases at least Q30
                Consensus quality: 152040 bases at least Q20
                Insert size: 153000; agarose-fp
                Insert size: 153413; sum-of-contigs
                Quality coverage: 4.7 in Q20 bases; agarose-fp
                Quality coverage: 4.7 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2433: contig of 2433 bp in length
            *     2434     2533: gap of 100 bp
            *     2534     6724: contig of 4191 bp in length
            *     6725     6824: gap of 100 bp
            *     6825    10238: contig of 3414 bp in length
            *    10239    10338: gap of 100 bp
            *    10339    17995: contig of 7657 bp in length
            *    17996    18095: gap of 100 bp
            *    18096    23868: contig of 5773 bp in length
            *    23869    23968: gap of 100 bp
            *    23969    36673: contig of 12705 bp in length
            *    36674    36773: gap of 100 bp
            *    36774    54829: contig of 18056 bp in length
            *    54830    54929: gap of 100 bp
            *    54930    73991: contig of 19062 bp in length
            *    73992    74091: gap of 100 bp
            *    74092    91636: contig of 17545 bp in length
            *    91637    91736: gap of 100 bp
            *    91737   117471: contig of 25735 bp in length
            *   117472   117571: gap of 100 bp
            *   117572   154413: contig of 36842 bp in length.
FEATURES             Location/Qualifiers
     source          1. .154413
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="RP11-21C8"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .2433
                     /note="assembly_fragment"
     misc_feature    2534. .6724
                     /note="assembly_fragment"
     misc_feature    6825. .10238
                     /note="assembly_fragment"
     misc_feature    10339. .17995
                     /note="assembly_fragment"
     misc_feature    18096. .23868
                     /note="assembly_fragment"
     misc_feature    23969. .36673
                     /note="assembly_fragment"
     misc_feature    36774. .54829
                     /note="assembly_fragment"
     misc_feature    54930. .73991
                     /note="assembly_fragment"
     misc_feature    74092. .91636
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     misc_feature    91737. .117471
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    117572. .154413
                     /note="assembly_fragment"
ORIGIN

  Query Match             4.3%;  Score 261.6;  DB 2;  Length 154413;
  Best Local Similarity   69.7%;  Pred. No. 3.8e-47;
  Matches  449;  Conservative    0;  Mismatches  169;  Indels   26;  Gaps    6;

Qy        314 AGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTCCT 373
              |||| || ||| |||||||||  || ||  |||||||||| |||| ||||||| ||||||
Db      33655 AGGGATGGAAATGTGTGATACCATTCCTTGTCCATCATAATGGTCACAACCAACACTCCT 33596

Qy        374 ATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTAC 433
              |||  |||||||||| ||| ||||| ||| | |||  |||||| ||||| |||||||||
Db      33595 ATAACAAAAGACAGGCTAACAAGAGAAAAGCATAAACAATTTACTTAATTAAAGTTTTAT 33536

Qy        434 ATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTTTT 493
              |||     ||||||||||||||||||  ||  |    |  ||  ||    | |  |||||
Db      33535 GTGATGCAGGAGTCTTCAGAAATGAAACCCAGAGACTCAGGGAAAACTATTTTTATTTTT 33476

Qy        494 TTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGATAT 553
              ||||  ||||| |||| || || | |||      |     | | ||||||||| | |
Db      33475 ATGCTTGGGTTCAATGAGGAGTGAACAGCTGTGTAGAAATGTAAGTGGACAAAAGGGTGT 33416

Qy        554 GATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTTCT 613
              |||||| ||||||  |||| |||| | ||   | | ||    |||| |||||||| ||||
Db      33415 GATCTAATGGTAATAGACTGAGGGAGGAA---ACCCAGCAGGGTCTGTTCAGATTTTTCT 33359

Qy        614 TGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTTCT 673
              ||   ||||         | | |||||||  ||| |||   ||||| || |||     | |
Db      33358 TGGCTTCTC--------TTTGTAGCATTGCTTCTTCCAAGGTATGAGGTAGGGTCTCTTT 33307

Qy        674 TCAATGAGGGTCTTC-------AAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATG 726
                |||||| |||  |        ||  |||||||||| |||| | ||| | | |||||||
```

```
Db      33306 GGAATGAGAGTCCCCTTGGAATGAGACAGAAGGGAGAGAGTGACTTTTCTGGGTTTTATG 33247

Qy        727 GCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCT 786
              ||| ||||   ||  |||||||||||||| ||||||| |||||||||||||| ||||| |
Db      33246 GCTCGCTTTATGGG-GAGGAGTTCTAGTATCTATGAACCATCTTGGGGAAGAAGAATTAT 33188

Qy        787 GGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGG 846
              |||||||||||| ||||| | | | |||||  || || |||| ||||||| |||| || ||
Db      33187 GGTTTCTGTGATTTGCTCTGAGGGAGAAAAGGGGGTGAGAGACAGGAGGACAGGGGAAGG 33128

Qy        847 TCAGAAAGAGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAGTACTTCT 906
              |||||  |     |    |||||||||||| | | |       | ||  || |||  | | ||
Db      33127 TCAGAGA-----TACCTTGCTTCTGAGGTCCTTCCAGTGTCCTTCATTTCAA--AGTACT 33075

Qy        907 TAGCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCT 950
              |||||  |||| |||| |||||||||  ||| || |||||||||
Db      33074 CAGCATGCCAAGGCACCATACTTTGGGGTATTGTGTTCTGAGCT 33031
```

RESULT 58
AC130399/c
LOCUS AC130399 156228 bp DNA linear HTG 10-AUG-2002
DEFINITION Homo sapiens chromosome 12 clone RP11-21C8, *** SEQUENCING IN PROGRESS ***, 8 ordered pieces.
ACCESSION AC130399
VERSION AC130399.1 GI:22203155
KEYWORDS HTG; HTGS_PHASE2.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 156228)
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T., Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G. and Gibbs,R.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 156228)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (10-AUG-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT ---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
Center project name: HCNC
Center clone name: RP11-21C8
---------- Summary Statistics
Chemistry: Dye-primer Bodipy: Infinity% of reads
Chemistry: Dye-terminator Big Dye: Infinity% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 149580 bases at least Q40
Consensus quality: 152343 bases at least Q30
Consensus quality: 153504 bases at least Q20
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 3144: contig of 3144 bp in length
* 3145 3244: gap of unknown length
* 3245 9096: contig of 5852 bp in length
* 9097 9196: gap of unknown length
* 9197 22106: contig of 12910 bp in length
* 22107 22206: gap of unknown length
* 22207 40512: contig of 18306 bp in length
* 40513 40612: gap of unknown length
* 40613 59686: contig of 19074 bp in length
* 59687 59786: gap of unknown length
* 59787 80941: contig of 21155 bp in length
* 80942 81041: gap of unknown length
* 81042 106851: contig of 25810 bp in length
* 106852 106951: gap of unknown length
* 106952 156228: contig of 49277 bp in length.
FEATURES Location/Qualifiers
source 1. .156228
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="12"
/clone="RP11-21C8"
ORIGIN

```
  Query Match             4.3%;  Score 261.6;  DB 2;  Length 156228;
  Best Local Similarity   69.7%;  Pred. No. 3.8e-47;
  Matches  449;  Conservative    0;  Mismatches  169;  Indels   26;  Gaps    6;

Qy        314 AGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTCCT 373
              |||| || ||| |||||||||  || ||  |||||||||| |||| ||||||| ||||||
Db      19042 AGGGATGGAAATGTGTGATACCATTCCTTGTCCATCATAATGGTCACAACCAACACTCCT 18983

Qy        374 ATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTAC 433
              |||  |||||||||| ||| ||||| ||| | |||  |||||| ||||| |||||||||
Db      18982 ATAACAAAAGACAGGCTAACAAGAGAAAAGCATAAACAATTTACTTAATTAAAGTTTTAT 18923

Qy        434 ATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTTTT 493
              |||    |||||||||||||||||  ||  |    |  ||  ||    | |  |||||
Db      18922 GTGATGCAGGAGTCTTCAGAAATGAAACCCAGAGACTCAGGGAAAACTATTTTTATTTTT 18863
```

```
Qy        494 TTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGATAT 553
              |||| ||||| |||| || || | |||         |      | | |||||||||| | |
Db      18862 ATGCTTGGGTTCAATGAGGAGTGAACAGCTGTGTAGAAATGTAAGTGGACAAAAGGGTGT 18803

Qy        554 GATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTTCT 613
              |||||| ||||||  |||| |||| | ||   | | ||    |||| |||||||| ||||
Db      18802 GATCTAATGGTAATAGACTGAGGGAGGAA---ACCCAGCAGGGTCTGTTCAGATTTTTCT 18746

Qy        614 TGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTTCT 673
              ||   ||||         | | ||||||| ||| |||   ||||| || |||      | |
Db      18745 TGGCTTCTC--------TTTGTAGCATTGCTTCTTCCAAGGTATGAGGTAGGGTCTCTTT 18694

Qy        674 TCAATGAGGGTCTTC-------AAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATG 726
              |||||| |||  |         ||  |||||||||| |||| | ||| | | |||||||
Db      18693 GGAATGAGAGTCCCCTTGGAATGAGACAGAAGGGAGAGAGTGACTTTTCTGGGTTTTATG 18634

Qy        727 GCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCT 786
              ||| ||||   ||  |||||||||||||| ||||||| |||||||||||||||| ||||| |
Db      18633 GCTCGCTTTATGGG-GAGGAGTTCTAGTATCTATGAACCATCTTGGGGAAGAAGAATTAT 18575

Qy        787 GGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGG 846
              |||||||||| ||||| | | | |||||  || || |||| ||||||| |||| || ||
Db      18574 GGTTTCTGTGATTTGCTCTGAGGGAGAAAAGGGGGTGAGAGACAGGAGGACAGGGGAAGG 18515

Qy        847 TCAGAAAGAGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAGTACTTCT 906
              ||||| |      |    |||||||||| | | |      | ||  || |||  | | ||
Db      18514 TCAGAGA-----TACCTTGCTTCTGAGGTCCTTCCAGTGTCCTTCATTTCAA--AGTACT 18462

Qy        907 TAGCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCT 950
              ||||| |||| |||| |||||||||  ||| || |||||||||
Db      18461 CAGCATGCCAAGGCACCATACTTTGGGGTATTGTGTTCTGAGCT 18418
```

RESULT 59
AC078922/c
LOCUS AC078922 181000 bp DNA linear PRI 07-MAR-2002
DEFINITION Homo sapiens 12q BAC RP11-611E13 (Roswell Park Cancer Institute Human BAC Library) complete sequence.
ACCESSION AC078922
VERSION AC078922.14 GI:15133248
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 181000)
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Emerling,S., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Han,J., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Ioshikhes,I., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lee,E., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Marondel,I., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Merscher,S., Metzker,M., Miller,A., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Montgomery,K.T., Morgan,M., Morris,S., Moser,M., Neal,D., Nelson,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shim,C., Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Kucherlapati,R., Weinstock,G. and Gibbs,R.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 181000)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (11-AUG-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 181000)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (09-AUG-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 4 (bases 1 to 181000)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (07-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Aug 9, 2001 this sequence version replaced gi:15022643.
INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE:Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL:
http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

```
          QUALSTAT-REPORT.
FEATURES             Location/Qualifiers
     source          1. .181000
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="12"
                     /clone="RP11-611E13"
     repeat_region   470. .499
                     /rpt_family="(TC)n"
     repeat_region   complement(2183. .3097)
                     /rpt_family="L1"
     repeat_region   complement(3018. .3543)
                     /rpt_family="L1M2"
     repeat_region   complement(3544. .3843)
                     /rpt_family="AluSx"
     repeat_region   complement(3844. .3903)
                     /rpt_family="L1M2"
     repeat_region   complement(3908. .4300)
                     /rpt_family="MSTB"
     repeat_region   complement(4312. .6604)
                     /rpt_family="L1M2"
     repeat_region   complement(7341. .7650)
                     /rpt_family="AluSx"
     repeat_region   8209. .8269
                     /rpt_family="L2"
     repeat_region   8985. .9041
                     /rpt_family="(TCC)n"
     repeat_region   complement(11081. .11311)
                     /rpt_family="MIR"
     repeat_region   11913. .12017
                     /rpt_family="L2"
     repeat_region   12068. .12218
                     /rpt_family="MIR"
     repeat_region   12219. .12561
                     /rpt_family="LTR50"
     repeat_region   12630. .12883
                     /rpt_family="LTR50"
     repeat_region   12884. .12984
                     /rpt_family="MIR"
     repeat_region   13232. .13416
                     /rpt_family="MER104"
     repeat_region   13520. .13563
                     /rpt_family="(CA)n"
     repeat_region   complement(14015. .14139)
                     /rpt_family="AluSg/x"
     repeat_region   14261. .14282
                     /rpt_family="(A)n"
     repeat_region   14532. .14564
                     /rpt_family="(TTTTA)n"
     repeat_region   complement(14565. .14712)
                     /rpt_family="AluSg/x"
     repeat_region   complement(15010. .15186)
                     /rpt_family="L2"
     repeat_region   15205. .15272
                     /rpt_family="L2"
     repeat_region   complement(15653. .15855)
                     /rpt_family="MER63A"
     repeat_region   complement(16022. .16351)
                     /rpt_family="MER58B"
     repeat_region   16438. .16743
                     /rpt_family="AluY"
     repeat_region   complement(16753. .16965)
                     /rpt_family="MIR"
     repeat_region   17195. .17223
                     /rpt_family="AT_rich"
     repeat_region   17661. .17967
                     /rpt_family="AluSx"
     repeat_region   18481. .18810
                     /rpt_family="L1ME"
     repeat_region   complement(19288. .19582)
                     /rpt_family="AluSg"
     repeat_region   complement(21033. .21229)
                     /rpt_family="L1ME"
     repeat_region   complement(21424. .21477)
                     /rpt_family="L2"
     repeat_region   21630. .21748
                     /rpt_family="L2"
     repeat_region   22140. .22228
                     /rpt_family="MIR"
     repeat_region   22229. .22249
                     /rpt_family="(A)n"
     repeat_region   22422. .22503
                     /rpt_family="MIR"
     repeat_region   complement(22512. .22623)
                     /rpt_family="MIR"
     repeat_region   22828. .22853
                     /rpt_family="(CA)n"
     repeat_region   24088. .24120
                     /rpt_family="(TG)n"

  Query Match             4.3%;  Score 261.6;  DB 9;  Length 181000;
  Best Local Similarity   69.7%;  Pred. No. 3.8e-47;
  Matches  449;  Conservative    0;  Mismatches  169;  Indels   26;  Gaps    6;

Qy        314 AGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTCCT 373
              |||| || ||| |||||||||  || ||  |||||||||| |||| ||||||| ||||||
Db     147077 AGGGATGGAAATGTGTGATACCATTCCTTGTCCATCATAATGGTCACAACCAACACTCCT 147018

Qy        374 ATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTAC 433
              ||| |||||||||| ||| ||||| ||| | |||  |||||| ||||| |||||||||
Db     147017 ATAACAAAAGACAGGCTAACAAGAGAAAAGCATAAACAATTTACTTAATTAAAGTTTTAT 146958

Qy        434 ATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTTTT 493
              |||    |||||||||||||||||  ||  |    |  ||  ||    | |  |||||
Db     146957 GTGATGCAGGAGTCTTCAGAAATGAAACCCAGAGACTCAGGGAAAACTATTTTTATTTTT 146898

Qy        494 TTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGATAT 553
              ||||  ||||| |||| || || | |||         |      | | |||||||||| | |
Db     146897 ATGCTTGGGTTCAATGAGGAGTGAACAGCTGTGTAGAAATGTAAGTGGACAAAAGGGTGT 146838

Qy        554 GATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTTCT 613
              |||||| |||||  |||| |||| | ||   | | ||    |||| |||||||| ||||
Db     146837 GATCTAATGGTAATAGACTGAGGGAGGAA---ACCCAGCAGGGTCTGTTCAGATTTTTCT 146781

Qy        614 TGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTTCT 673
              ||   ||||       | | |||||||  ||| |||   ||||| || |||     | |
Db     146780 TGGCTTCTC--------TTGTAGCATTGCTTCTTCCAAGGTATGAGGTAGGGTCTCTTT 146729

Qy        674 TCAATGAGGGTCTTC-------AAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATG 726
              ||||||  |||  |         ||  |||||||||| |||| | ||| | | |||||||
Db     146728 GGAATGAGAGTCCCCTTGGAATGAGACAGAAGGGAGAGAGTGACTTTTCTGGGTTTTATG 146669

Qy        727 GCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCT 786
              ||| ||||   ||  |||||||||||| |||||| |||||||||||||||| ||||| |
Db     146668 GCTCGCTTTATGGG-GAGGAGTTCTAGTATCTATGAACCATCTTGGGGAAGAAGAATTAT 146610

Qy        787 GGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGG 846
              |||||||||| ||||| | | | |||||  || || |||| ||||||| |||| || ||
Db     146609 GGTTTCTGTGATTTGCTCTGAGGGAGAAAAGGGGGTGAGAGACAGGAGGACAGGGGAAGG 146550

Qy        847 TCAGAAAGAGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAGTACTTCT 906
              ||||| |     |    |||||||||||| | | |       | ||  || |||  | | ||
Db     146549 TCAGAGA-----TACCTTGCTTCTGAGGTCCTTCCAGTGTCCTTCATTTCAA--AGTACT 146497

Qy        907 TAGCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCT 950
              ||||| |||| |||| |||||||||  ||| || |||||||||
Db     146496 CAGCATGCCAAGGCACCATACTTTGGGGTATTGTGTTCTGAGCT 146453


RESULT 60
AC091772
LOCUS       AC091772              158559 bp    DNA     linear   PRI 17-JUN-2003
DEFINITION  Homo sapiens chromosome 10 clone RP11-257L6, complete sequence.
ACCESSION   AC091772
VERSION     AC091772.5  GI:31795104
```

```
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 158559)
  AUTHORS   Smith,D.R.
  TITLE     Genome Therapeutics Corporation Sequencing Center: Human Genome
            Sequence Data
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 158559)
  AUTHORS   Smith,D.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-2001) Genome Therapeutics Corporation, 100 Beaver
            Street, Waltham, MA 02453, USA
REFERENCE   3  (bases 1 to 158559)
  AUTHORS   Smith,D.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-NOV-2001) Genome Therapeutics Corporation, 100 Beaver
            Street, Waltham, MA 02453, USA
REFERENCE   4  (bases 1 to 158559)
  AUTHORS   Smith,D.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-APR-2003) Genome Therapeutics Corporation, 100 Beaver
            Street, Waltham, MA 02453, USA
  REMARK    Sequence updated to remove bacterial transposon
REFERENCE   5  (bases 1 to 158559)
  AUTHORS   Smith,D.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JUN-2003) Genome Therapeutics Corporation, 100 Beaver
            Street, Waltham, MA 02453, USA
  REMARK    Sequence updated to remove vector contamination
COMMENT     On Jun 17, 2003 this sequence version replaced gi:29570805.
FEATURES             Location/Qualifiers
     source          1. .158559
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP11-257L6"
                     /clone_lib="RPCI-11"
ORIGIN

  Query Match          4.3%;  Score 260.2;  DB 9;  Length 158559;
  Best Local Similarity   70.3%;  Pred. No. 7.8e-47;
  Matches  472;  Conservative    0;  Mismatches  168;  Indels   31;  Gaps    8;

Qy        304 AAGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAAC 363
              ||| |   |||||||    ||  |||||||  ||| |||  |||| |||||||||  | |
Db     126822 AAGATTTTGTAGGGGCAGGAAAATGTGATAACTTTCCTCACCCATTATAAGGGTCATAGC 126881

Qy        364 CAAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAA---CCTAACAAATTTATTTA 420
              | | |||| ||  ||||| ||| |||| || | ||||   | |||||||| |||||
Db     126882 TGACATTCCTTTAACAAAAGGCAGATTAACAACAAAAAAAATGCATAACAAATTCATTTA 126941

Qy        421 ATCAAAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAA 480
              ||||||||||||||  |||   ||| |||| ||||| | |||| ||||||||||| |||
Db     126942 ATCAAAGTTTTACACAACAAAAGAGACTTCTGAAATTATGACCAAAAGACCCAGGGAAAA 127001

Qy        481 CTGTCTGTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTA 540
              |||     ||||| |||| || || ||||||||||||| |||       |    ||||
Db     127002 CTG-----TTTTTATGCTTAGTTTTGATGAAGAATGGACAGCCATGTAGAAATGTGATTG 127056

Qy        541 GACAAAAGGATATGATCTAGTGGTAAAGGACTCAGGGGGA-AACACAGCAAGGCCTGTCT 599
              | ||||||| |||||||| ||||||   ||||||| ||| |||  ||||||||||||
Db     127057 GGCAAAAGGGTATGATCTAATGGTAATAAACTCAGGTGGAGAACCTAGCAAGGCCTGTC- 127115

Qy        600 ATTCAGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGG 659
                  |   ||  ||||   ||||||   ||      ||||||  ||   ||||||
Db     127116 ---------CAGATTCTTCTCAGCCTCTCTGAATAGCATTCTCTTTCTACCCAGGTATGG 127166

Qy        660 GGCAGGACTCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAA---AGTGGCCTTTTT 716
              ||||| || | |||  |||||  |||||||||| |||| ||||||   || | |||||||
```

```
Db     127167 GGCAGAACCCCTCTAGAATGAAAGTCTTCAAGGAAGAAAGGAGAAGAGAGAGACCTTTTT 127226

Qy        717 AGATTTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAA 776
              || |||||||||||||||  ||  | | | |||| ||||| |||||||| ||||| |||
Db     127227 AGGTTTTATGGCTTGCTTTAGGAGAAAAGGGTTCCAGTTTTTATGACCCACCTTGGAGAA 127286

Qy        777 GAGGAATTCTGGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGG 836
              ||| |||||| |||||| ||||||||||     |  |||||||  |     |  ||||||||
Db     127287 GAGAAATTCTAGTTTCTATGACTTGCTTCAGAGGTGAAAGAGAGGCTGTAAACAGGAGGG 127346

Qy        837 CAGGAGATGGTCAGAAAGAGACTTGGCTGCTTCTGAGG-GCTTCCGCTCTCCTTTAGTTC 895
              |||||| ||||||| |||         |||||||||||  |||||  | ||||  ||||
Db     127347 CAGGAGAAGGTCAGAGAGA-----CCTTGCTTCTGAGGCCCTTCCAGTGTCCTAAAGTTT 127401

Qy        896 CAAGTACTTCTTAGCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAAC 955
              |||||   || ||||| |  |||  |||||||| ||    |  | ||||||||||| |||
Db     127402 GAAGTA---CTCAGCATGCTGAAGTGCTATACTTCGGGTCTTCTTGTTCTGAGCTCCAAC 127458

Qy        956 ACTGCAATCAT 966
              |    | |||
Db     127459 AACTTCAACAT 127469
```

```
RESULT 61
AC069541/c
LOCUS       AC069541              158568 bp    DNA     linear   PRI 17-MAY-2004
DEFINITION  Homo sapiens chromosome 10 clone RP11-389O2, complete sequence.
ACCESSION   AC069541
VERSION     AC069541.5  GI:20330781
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 158568)
  AUTHORS   Smith,D.R.
  TITLE     Genome Therapeutics Corporation Sequencing Center: Human Genome
            Sequence Data
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 158568)
  AUTHORS   Smith,D.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-2000) Genome Therapeutics Corporation, 100 Beaver
            Street, Waltham, MA 02453, USA
REFERENCE   3  (bases 1 to 158568)
  AUTHORS   Smith,D.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-APR-2001) Genome Therapeutics Corporation, 100 Beaver
            Street, Waltham, MA 02453, USA
REFERENCE   4  (bases 1 to 158568)
  AUTHORS   Smith,D.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-APR-2002) Genome Therapeutics Corporation, 100 Beaver
            Street, Waltham, MA 02453, USA
COMMENT     On Apr 26, 2002 this sequence version replaced gi:13621215.
            1329 bases of Tn10 (J01829) transposon removed.
FEATURES             Location/Qualifiers
     source          1. .158568
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP11-389O2"
                     /clone_lib="RPCI-11"
ORIGIN

  Query Match          4.3%;  Score 260.2;  DB 9;  Length 158568;
  Best Local Similarity   70.3%;  Pred. No. 7.8e-47;
  Matches  472;  Conservative    0;  Mismatches  168;  Indels   31;  Gaps    8;

Qy        304 AAGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAAC 363
              ||| |   |||||||    ||  |||||||  ||| |||  |||| |||||||||  | |
Db      31747 AAGATTTTGTAGGGGCAGGAAAATGTGATAACTTTCCTCACCCATTATAAGGGTCATAGC 31688
```

```
Qy        364 CAAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAA---CCTAACAAATTTATTTA 420
              | | |||| ||  ||||| ||| |||| || |  ||||   | |||||||||| |||||
Db      31687 TGACATTCCTTTAACAAAAGGCAGATTAACAACAAAAAAAATGCATAACAAATTCATTTA 31628

Qy        421 ATCAAAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAA 480
              ||||||||||||||  |||   ||| |||| ||||| | |||| ||||||||||| |||
Db      31627 ATCAAAGTTTTACACAACAAAAGAGACTTCTGAAATTATGACCAAAAGACCCAGGGAAAA 31568

Qy        481 CTGTCTGTTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTA 540
              |||     ||||| |||| || || |||||||||||||| |||       |    ||||
Db      31567 CTG-----TTTTTATGCTTAGTTTTGATGAAGAATGGACAGCCATGTAGAAATGTGATTG 31513

Qy        541 GACAAAAGGATATGATCTAGTGGTAAAGGACTCAGGGGGA-AACACAGCAAGGCCTGTCT 599
              | ||||||| |||||||| ||||||   ||||||| ||| |||  ||||||||||||
Db      31512 GGCAAAAGGGTATGATCTAATGGTAATAAACTCAGGTGGAGAACCTAGCAAGGCCTGTC- 31454

Qy        600 ATTCAGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGG 659
                  |   ||  ||||   ||||||   ||      ||||||  ||   ||||||
Db      31453 ---------CAGATTCTTCTCAGCCTCTCTGAATAGCATTCTCTTTCTACCCAGGTATGG 31403

Qy        660 GGCAGGACTCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAA---AGTGGCCTTTTT 716
              ||||| || | |||  |||||  |||||||||| |||| ||||||   || | |||||||
Db      31402 GGCAGAACCCCTCTAGAATGAAAGTCTTCAAGGAAGAAAGGAGAAGAGAGAGACCTTTTT 31343

Qy        717 AGATTTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAA 776
              || |||||||||||||||  ||  | | | |||| ||||| |||||||| ||||| |||
Db      31342 AGGTTTTATGGCTTGCTTTAGGAGAAAAGGGTTCCAGTTTTTATGACCCACCTTGGAGAA 31283

Qy        777 GAGGAATTCTGGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGG 836
              ||| |||||| |||||| ||||||||||     |  |||||||  |     |  ||||||||
Db      31282 GAGAAATTCTAGTTTCTATGACTTGCTTCAGAGGTGAAAGAGAGGCTGTAAACAGGAGGG 31223

Qy        837 CAGGAGATGGTCAGAAAGAGACTTGGCTGCTTCTGAGG-GCTTCCGCTCTCCTTTAGTTC 895
              ||||||| ||||||| |||        |||||||||||  |||||  | ||||  ||||
Db      31222 CAGGAGAAGGTCAGAGAGA-----CCTTGCTTCTGAGGCCCTTCCAGTGTCCTAAAGTTT 31168

Qy        896 CAAGTACTTCTTAGCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAAC 955
              |||||   || ||||| |  |||  |||||||| ||    |  | |||||||||| |||
Db      31167 GAAGTA---CTCAGCATGCTGAAGTGCTATACTTCGGGTCTTCTTGTTCTGAGCTCCAAC 31111

Qy        956 ACTGCAATCAT 966
              |     | |||
Db      31110 AACTTCAACAT 31100
```

```
RESULT 62
AC034261
LOCUS       AC034261              155797 bp    DNA     linear   HTG 04-OCT-2000
DEFINITION  Homo sapiens clone RP11-14M21, WORKING DRAFT SEQUENCE, 14 unordered
            pieces.
ACCESSION   AC034261
VERSION     AC034261.2  GI:10567954
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 155797)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-14M21
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 155797)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-APR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Oct 4, 2000 this sequence version replaced gi:7417750.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L3424
                Center clone name: 14_M_21
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 147315 bases at least Q40
                Consensus quality: 151520 bases at least Q30
                Consensus quality: 153258 bases at least Q20
                Insert size: 188000; agarose-fp
                Insert size: 154497; sum-of-contigs
                Quality coverage: 4.2 in Q20 bases; agarose-fp
                Quality coverage: 5.2 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 14 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    18121: contig of 18121 bp in length
            *    18122    18221: gap of 100 bp
            *    18222    19632: contig of 1411 bp in length
            *    19633    19732: gap of 100 bp
            *    19733    23072: contig of 3340 bp in length
            *    23073    23172: gap of 100 bp
            *    23173    26430: contig of 3258 bp in length
            *    26431    26530: gap of 100 bp
            *    26531    30992: contig of 4462 bp in length
            *    30993    31092: gap of 100 bp
            *    31093    36194: contig of 5102 bp in length
            *    36195    36294: gap of 100 bp
            *    36295    42695: contig of 6401 bp in length
            *    42696    42795: gap of 100 bp
            *    42796    52136: contig of 9341 bp in length
            *    52137    52236: gap of 100 bp
            *    52237    60509: contig of 8273 bp in length
            *    60510    60609: gap of 100 bp
            *    60610    70259: contig of 9650 bp in length
            *    70260    70359: gap of 100 bp
            *    70360    85204: contig of 14845 bp in length
            *    85205    85304: gap of 100 bp
            *    85305   114230: contig of 28926 bp in length
            *   114231   114330: gap of 100 bp
            *   114331   136562: contig of 22232 bp in length
            *   136563   136662: gap of 100 bp
            *   136663   155797: contig of 19135 bp in length.
```

```
FEATURES             Location/Qualifiers
     source          1. .155797
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="RP11-14M21"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .18121
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    18222. .19632
                     /note="assembly_fragment"
     misc_feature    19733. .23072
                     /note="assembly_fragment"
     misc_feature    23173. .26430
                     /note="assembly_fragment"
     misc_feature    26531. .30992
                     /note="assembly_fragment"
     misc_feature    31093. .36194
                     /note="assembly_fragment"
     misc_feature    36295. .42695
                     /note="assembly_fragment"
     misc_feature    42796. .52136
                     /note="assembly_fragment"
     misc_feature    52237. .60509
                     /note="assembly_fragment"
     misc_feature    60610. .70259
                     /note="assembly_fragment"
     misc_feature    70360. .85204
                     /note="assembly_fragment"
     misc_feature    85305. .114230
                     /note="assembly_fragment"
     misc_feature    114331. .136562
                     /note="assembly_fragment"
     misc_feature    136663. .155797
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
ORIGIN

  Query Match             4.3%;  Score 259.4;  DB 2;  Length 155797;
  Best Local Similarity   69.9%;  Pred. No. 1.2e-46;
  Matches  462;  Conservative    0;  Mismatches  166;  Indels   33;  Gaps    7;

Qy        303 AAAGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAA 362
              | || |   |||| |||| ||  || |||||| ||| |||  ||||||   ||
Db     122214 AGAGATAGTGTAGAGGTGGAATGGTTTGATACCTTTCCTCGCCCATCACTTGG------- 122266

Qy        363 CCAAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAAT 422
              ||      || |  |||||| ||||| |  ||||| ||| | |  |||||||||||||
Db     122267 TCATGCTGACTCTCTTAAAAGTCAGGTGACCAAGAGAAAAGCATGGCAAATTTATTTAAT 122326

Qy        423 CAAAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACT 482
              | ||| ||||||| || || | | |||||||||||||| |||||| ||||  | |||||
Db     122327 CGAAGATTTACAT-ACGTGAGGGCCTTCAGAAATGAAGATCCAAAGGCCCAAAGAAAACT 122385

Qy        483 GTCTGTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGA 542
              |||| | |||     |   | | |||||||||||||| ||         |    |||| ||
Db     122386 GTCTATCTTTAAGCTT--AGGTTGATGAAGAATGGACAGTTGTGTAAAAATGTGATTGGA 122443

Qy        543 CAAAAGGATATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATT 602
              |||||   | ||||||  ||||||  |||| ||||| ||||  ||||||| |||||| ||
Db     122444 CAAAATAGTTTGATCTCTTGGTAATAGACTGAGGGGAAAACCTAGCAAGGACTGTCTGTT 122503

Qy        603 CAGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGC 662
              |  |||||||||  |||||        ||| ||| |||| |  |||||  |||| ||||
Db     122504 CCAATTCTTCTTGGCCTCTC--------TGTGTAGGATTCCTCTCTCCTAGGTATAGGGC 122555

Qy        663 AGGACTCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAA-------AGTGGCCTTTT 715
              ||||||  |||  |||  | ||||||||||  ||||||||||       |||| |||||
Db     122556 TGGACTCCTCTGGAATAAAGGTCTTCAAGGAAGAAGGGAGAAGGAAGAGAGTGACCTTTC 122615

Qy        716 TAGATTTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGA 775
              |||  |||||||  |||  |||  | ||  |||||    |  || ||||||||||||||  || | |
Db     122616 TAGGTTTTATGTCTTCCTTTGAGGGAGAGGGACTGTACTTTCTATGACCCATAGTGAGAA 122675

Qy        776 AGAGGAATTCTGGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGG 835
              |||||||||  |||||  |||  ||||||||      | |||||||||   |  ||||  ||||| |
Db     122676 GGAGGAATTCAGGTTTTTGTCACTTGCTTCAGGGGAGAAAGAGGGATGAGAGACAGGAAG 122735

Qy        836 GCAGGAGATGGTCAGAAAGAGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTC 895
              | |||||  ||||||||||| |      || |||  | | |||| | | ||||| |||||
Db     122736 ACGGGAGAAAGTCAGAAAGACA-----TTGATTCCAAAGCCTTCTGATGTCCTTCAGTTC 122790

Qy        896 CAAGTACTTCTTAGCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAAC 955
              |||||    | ||||| |||| |  | ||||||  ||  |||  || |||||||| | |||
Db     122791 AAAGTA---TTCAGCATGCCAAGGTGCCATACTTCGGGGTATCGTGTTCTGAGCCCCAAC 122847

Qy        956 A 956
              |
Db     122848 A 122848


RESULT 63
AC009964/c
LOCUS       AC009964              197087 bp    DNA     linear   PRI 07-NOV-2001
DEFINITION  Homo sapiens BAC clone RP11-387G23 from 2, complete sequence.
ACCESSION   AC009964
VERSION     AC009964.11  GI:13435280
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 197087)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
   PUBMED   9847074
REFERENCE   2  (bases 1 to 197087)
  AUTHORS   Du,H., Haakenson,W. and Reitz,L.
  TITLE     The sequence of Homo sapiens BAC clone RP11-387G23
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 197087)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-SEP-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 197087)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-MAR-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 197087)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-AUG-2001) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   6  (bases 1 to 197087)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-NOV-2001) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Mar 23, 2001 this sequence version replaced gi:11323432.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0387G23
            --------------
```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (http://bacpac.med.buffalo.edu)
VECTOR: pBACe3.6
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is AC015581; the clone sequenced to the right is AC040960. Actual start of this clone is at base position 1 of RP11-387G23; actual end is at base position 197087 of RP11-387G23.

FEATURES	Location/Qualifiers
source	1. .197087
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="2"
	/map="2"
	/clone="RP11-387G23"
	/clone_lib="RPCI-11"
repeat_region	1. .150
	/rpt_family="MaLR"
repeat_region	1519. .1738
	/rpt_family="MIR"
repeat_region	1912. .2116
	/rpt_family="L2"
repeat_region	2117. .2185
	/rpt_family="(TCTA)n"
repeat_region	2186. .2231
	/rpt_family="L2"
repeat_region	2290. .2339
	/rpt_family="MIR"
repeat_region	2396. .2572
	/rpt_family="MIR"
repeat_region	3081. .3236
	/rpt_family="MER2_type"
repeat_region	3934. .3965
	/rpt_family="AT_rich"
repeat_region	4054. .4361
	/rpt_family="Alu"
repeat_region	4422. .4510
	/rpt_family="MER1_type"
misc_feature	4559. .5068
	/note="similar to EST AA909716 (NID:g3049121) om47f04.s1"
repeat_region	4718. .4765
	/rpt_family="MIR"
repeat_region	4768. .5210
	/rpt_family="L1"
repeat_region	5211. .6535
	/rpt_family="ERV1"
repeat_region	6540. .7246
	/rpt_family="L1"
repeat_region	7249. .7567
	/rpt_family="Alu"
repeat_region	7646. .7954
	/rpt_family="Alu"
repeat_region	8017. .8301
	/rpt_family="Alu"
repeat_region	9766. .9804
	/rpt_family="ERV1"
repeat_region	9807. .9952
	/rpt_family="MaLR"
repeat_region	11399. .11709
	/rpt_family="Alu"
repeat_region	11938. .12220
	/rpt_family="Alu"
repeat_region	12359. .12852
	/rpt_family="ERVL"
repeat_region	13006. .13286
	/rpt_family="ERVL"
repeat_region	14129. .14717
	/rpt_family="L2"
repeat_region	14745. .15051
	/rpt_family="Alu"
repeat_region	15057. .15429
	/rpt_family="MaLR"
repeat_region	15497. .15809
	/rpt_family="Alu"
repeat_region	16164. .16327
	/rpt_family="MIR"
repeat_region	17372. .18056
	/rpt_family="L1"
repeat_region	18057. .18361
	/rpt_family="Alu"
repeat_region	18375. .18765
	/rpt_family="L1"
repeat_region	18766. .19185
	/rpt_family="L1"
repeat_region	19200. .19323
	/rpt_family="L1"
repeat_region	19356. .19683
	/rpt_family="L1"
repeat_region	19682. .20221
	/rpt_family="L1"
repeat_region	20229. .20795
	/rpt_family="L1"
repeat_region	21817. .22055
	/rpt_family="MER1_type"
repeat_region	22118. .22243
	/rpt_family="MIR"
repeat_region	22247. .22690
	/rpt_family="MaLR"
repeat_region	22885. .23013
	/rpt_family="MIR"
repeat_region	23097. .23124
	/rpt_family="AT_rich"
repeat_region	23198. .23469
	/rpt_family="Alu"
repeat_region	23610. .23908
	/rpt_family="Alu"
repeat_region	24017. .24040
	/rpt_family="AT_rich"
repeat_region	24380. .24690
	/rpt_family="Alu"
repeat_region	25108. .25292
	/rpt_family="MER2_type"
repeat_region	25618. .25939
	/rpt_family="MaLR"
repeat_region	26503. .26804
	/rpt_family="Alu"
repeat_region	26826. .26863
	/rpt_family="AT_rich"

```
     repeat_region   28092. .28234
                     /rpt_family="MIR"
     repeat_region   28666. .29064
                     /rpt_family="L1"

  Query Match             4.2%;  Score 257.8;  DB 9;  Length 197087;
  Best Local Similarity   69.7%;  Pred. No. 2.6e-46;
  Matches  461;  Conservative    0;  Mismatches  167;  Indels   33;  Gaps    7;

Qy        303 AAAGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAA 362
              | || |   |||| |||| ||  || |||||| ||| |||  | ||||   ||
Db     119844 AGAGATAGTGTAGAGGTGGAATGGTTTGATACCTTTCCTCGCCTATCACTTGG------- 119792

Qy        363 CCAAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAAT 422
              ||      || |  |||||| ||||| |  ||||| ||| | |  ||||||||||||||
Db     119791 TCATGCTGACTCTCTTAAAAGTCAGGTGACCAAGAGAAAAGCATGGCAAATTTATTTAAT 119732

Qy        423 CAAAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACT 482
              | ||| ||||||| || || | | ||||||||||||||| |||||| ||||  | |||||
Db     119731 CGAAGATTTACAT-ACGTGAGGGCCTTCAGAAATGAAGATCCAAAGGCCCAAAGAAAACT 119673

Qy        483 GTCTGTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGA 542
              |||| | |||      |   | | |||||||||||||| ||          |    |||| ||
Db     119672 GTCTATCTTTAAGCTT--AGGTTGATGAAGAATGGACAGTTGTGTAAAAATGTGATTGGA 119615

Qy        543 CAAAAGGATATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATT 602
              |||||    | ||||||  ||||||  |||| ||||| ||||  ||||||| |||||| ||
Db     119614 CAAAATAGTTTGATCTCTTGGTAATAGACTGAGGGGAAAACCTAGCAAGGACTGTCTGTT 119555

Qy        603 CAGATTCTTCTTGATCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGC 662
              |  ||||||||||  |||||         ||| ||| |||| |  |||||  |||| ||||
Db     119554 CCAATTCTTCTTGGCCTCTC--------TGTGTAGGATTCCTCTCTCCTAGGTATAGGGC 119503

Qy        663 AGGACTCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAA-------AGTGGCCTTTT 715
              ||||||  |||  ||| | |||||||||||| ||||||||||       |||| |||||
Db     119502 TGGACTCCTCTGGAATAAAGGTCTTCAAGGAAGAAGGGAGAAGGAAGAGAGTGACCTTTC 119443

Qy        716 TAGATTTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGA 775
              |||  ||||||| ||| ||| | || |||||    | || ||||||||||||||  || | |
Db     119442 TAGGTTTTATGTCTTCCTTTGAGGGAGAGGGACTGTACTTTCTATGACCCATAGTGAGAA 119383

Qy        776 AGAGGAATTCTGGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGG 835
              |||||||||  ||||| ||| ||||||||    | |||||||||  | |||| ||||| |
Db     119382 GGAGGAATTCAGGTTTTTGTCACTTGCTTCAGGGGAGAAAGAGGGATGAGAGACAGGAAG 119323

Qy        836 GCAGGAGATGGTCAGAAAGAGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTC 895
              | |||||  ||||||||||  |       || |||  | | |||| | |  ||||| |||||
Db     119322 ACGGGAGAAAGTCAGAAAGACA-----TTGATTCCAAAGCCTTCTGATGTCCTTCAGTTC 119268

Qy        896 CAAGTACTTCTTAGCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAAC 955
              |||||    | ||||| |||| |  | |||||| ||  ||| || |||||||| | |||
Db     119267 AAAGTA---TTCAGCATGCCAAGGTGCCATACTTCGGGGTATCGTGTTCTGAGCCCCAAC 119211

Qy        956 A 956
              |
Db     119210 A 119210


RESULT 64
AC097357
LOCUS       AC097357               162277 bp    DNA     linear   PRI 29-APR-2002
DEFINITION  Homo sapiens chromosome 3 clone RP11-137N22, complete sequence.
ACCESSION   AC097357
VERSION     AC097357.2  GI:20336785
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 162277)
  AUTHORS   Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
            Saenphimmachak,C., Phelps,K.A., Buckley,D., Raymond,C. and
            Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 162277)
  AUTHORS   Kaul,R.K., Olson,M.V., Raymond,C., Clendenning,J. and Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-OCT-2001) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
REFERENCE   3  (bases 1 to 162277)
  AUTHORS   Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
            Saenphimmachak,C., Phelps,K.A., Buckley,D., Raymond,C. and
            Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-APR-2002) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
COMMENT     On Apr 29, 2002 this sequence version replaced gi:16152307.
            -------------- Genome Center
                Center: University of Washington Genome Center
                Center Code: UWGC
                Web site: http://www.genome.washington.edu
                Contact: uwgchtgs@u.washington.edu
            -------------- Project Information
                Center project name: chr-3
                Center clone name: RP11-137N22 (bc0255)
            -------------- Summary Statistics
                Sequencing vector: plasmid; L08752; 100% of reads
                Chemistry: Dye-terminator ET; 92% of reads
                Chemistry: Dye-terminator Big Dye; 8% of reads
                Assembly program: Phrap; version 0.990319
                Consensus quality: 162026 bases at least Q40
                Consensus quality: 162268 bases at least Q30
                Consensus quality: 162277 bases at least Q20
                Insert size: 162277; sum-of-contigs
                Quality coverage: 8.6x in Q20 bases; sum-of-contigs
            --------------.
            Overlapping Sequences:
                5':  RP11-963F9 (UWGC:bc0582) AC013250
                3':  RP11-495E23 (UWGC:bc0449) AC114398
            --------------------
            Sequence Quality Assessment:
             This entry has been annotated with sequence quality
             estimates computed by the Phrap assembly program.
             All manually edited bases have been reduced to quality zero.
             Quality levels above 40 are expected to have less than
             1 error in 10,000 bp.
             Base-by-base quality values are not generally visible from the
             GenBank flat file format but are available as part
             of this entry's ASN.1 file.

            This sequence was finished as follows unless otherwise noted:
            all regions were either double-stranded or sequenced with an
            alternate chemistry or covered by high quality data (i.e., Phred
            quality >= 30); an attempt was made to resolve all sequencing
            problems, such as compressions and repeats; all regions were
            covered by at least one plasmid subclone or more than one M13
            subclone; and the assembly was confirmed by restriction digest.
            --------------------
            Sequence Validation:
             This sequence has been validated by Multiple Complete Digest
             fingerprinting. Comparison of the experimentally derived digest
             fragments with sequence-predicted fragments is given below.
             The electronically-digested sequence consists of both insert and
             vector, in order to accurately represent the entire circular BAC.
             Small fragments below a variable cutoff (approximately 400-800 bp)
             are not resolved in the fingerprint and hence do not appear
             in the table. There are no significant remaining discrepancies
             between the experimental and predicted values. Uniquely ordered
             fragments are separated by dashed lines.
                   EcoRI                HindIII               BglII
            ------------------   ------------------   ------------------
            SeqDerMap  FngrPrnt   SeqDerMap  FngrPrnt   SeqDerMap  FngrPrnt
            ---------  --------   ---------  --------   ---------  --------

            ---------             ---------             ---------
```

8696	8795	447	<800	2321	2302
6	<800	6382	6528	2067	2061
996	996	512	<800	17260	17209
5488	5516	449	<800	528	<800
4379	4358	4640	4605	677	<800
1224	1212	1433	1420	10464	10255
8022	8054	5435	5327	10361	10255
7627	7602	4809	4850	262	<800
3644	3663	170	<800	1058	1065
4964	4928	2479	2488	1267	1254
6423	6468	1020	1023	831	834
2727	2770	4237	4271	834	834
1949	1942	1969	1906	3302	3315
5998	6017	2272	2290	3490	3516
19964	20199	702	<800	2733	2722
3839	3884	457	<800	5156	5176
3833	3884	3572	3627	2610	2722
3307	3365	2367	2488	3053	3090
3204	3210	1828	1906	1850	1839
4339	4358	419	<800	166	<800
5625	5709	2250	2290	6199	6166
1489	1480	4935	4850	173	<800
4278	4235	280	<800	5354	5378
1127	1130	7224	7185	5286	5378
4235	4235	164	<800	1633	1596
489	<800	56	<800	443	<800
1620	1597	2852	2947	945	946
2940	2980	7514	7541	1617	1596
5487	5516	3799	3627	750	737
2685	2650	116	<800	6582	6682
406	<800	285	<800	3239	3315
727	741	1433	1420	796	834
1605	1597	5637	5643	2456	2485
4675	4611	3680	3627	6865	6682
2666	2650	1891	1906	1103	1065
220	<800	3830	3627	6679	6682
12857	12765	1887	1906	3329	3315

13219	12765	7123	7185	884	946
40	<800	1249	1236	1759	1707
1710	1678	429	<800	1528	1505
1342	1329	1988	1906	4783	4747
586	<800	1171	1236	3879	3940
319	<800	6507	6528	779	737
		3437	3627	230	<800
		140	<800	19135	19298
		3038	2947	4406	4420
		5355	5327	3977	3940
		192	<800	2713	2722
		17	<800	2441	2485
		2940	2947	723	<800
		12747	12670		
		563	<800		
		2964	2947		
		1294	1236		
		3546	3627		
		764	776		
		8230	8298		

Query Match 4.2%; Score 257.6; DB 9; Length 162277;
Best Local Similarity 68.3%; Pred. No. 2.9e-46;
Matches 462; Conservative 0; Mismatches 174; Indels 40; Gaps 6;

```
Qy      310 CAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAAC 369
            | |||||| || ||| |||||||||| ||| |||  |  ||| ||| ||| || |||| |
Db    64241 CTGTAGGGATGGAAAGGTGTGATACCTTTCCTCACCTCTCACAAGAGTCACAGCCAACAT 64300

Qy      370 TCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTT 429
            |||||||  |||||||||| |  ||||| ||| | |||||||| ||||||||||||||||
Db    64301 TCCTATAACAAAAGACAGGTCAGCAAGAGAAAAGCATAACAAATGTATTTAATCAAAGTT 64360

Qy      430 TTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTT 489
            |||  |||||| | || |||| ||||||||||||||||||  | || |||    |  |
Db    64361 TTATGTGACATAGCAGCCTTCCGAAATGAAGACCCAAAGACTAAAGGAAAA-----TAGT 64415

Qy      490 TTTTTTGCTGAGGTTCGATGAAGAATGGATAGC-ATGTAGCCATGTAGATTAGACAAAAG 548
            |||| |||| |||||| |||||||||| | |||  |||||  |||| |   |   ||| |
Db    64416 TTTTATGCTTAGGTTCAGTGAAGAATGCACAGCTGTGTAGAAATGTGGTAGAACAAAAGG 64475

Qy      549 GATATGATCTAGTGGTAA----------------------AGGACTCAGGGGGAAACACA 586
            | | || |||| ||| ||                      ||  |   || |     ||
Db    64476 GGTGTGTTCTACTGGAAATAGAAAATAGATTGGCGGTGGTGGGGGTGGTGGTGGGGATCA 64535

Qy      587 GCAAGGCCTGTCTATTCAGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTC 646
            |||||||||| |||||||||||||| |||  |||||         ||| || | |||||||
Db    64536 GCAAGGCCTGACTATTCAGATTCTTATTGGCCTCTC--------TGTGTAACGTTCTTTC 64587

Qy      647 CTCCTGAGTATGGGGCAGGACTCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAG 706
            |||~ ||||||| ||||| |   | |  ||||||| ||||||| |||||||| ||| ||
Db    64588 TTCCC-AGTATGGAGCAGGGCCTCTATGGAATGAGGATCTTCAAAGGAGAAGGAAGAGAG 64646
```

```
Qy        707 TGGCCTTTTTAGATTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCA 766
              || | ||| ||| ||| |||||||||  |||||||||   ||||||||||||||||||
Db      64647 TGACTTTTCTAGGTTTAATGGCTTGCTGTGGGGAAGAGACATTCTAGTTTCTATGACCCA 64706

Qy        767 TCTTGGGGAAGAGGAATTCTGGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGA 826
               ||| ||||||||||||||||||| |  |||||||||    | || ||||| |||| ||
Db      64707 CCTTAGGGAAGAGGAATTCTGGTTTATACGACTTGCTTAAGGGGAGGAAGAGAAGTAGGA 64766

Qy        827 GGCAGGAGGGCAGGAGATGGTCAGAAAGAGACTTGGCTGCTTCTGAGGGCTTCCGCTCTC 886
              | ||| |     |||||  | || |||||   | |              |       | ||
Db      64767 GACAGTATTATGGGAGACTGCCACAAAGACCTTGGTTCTGAGGGCTTTCCAATGTATTTC 64826

Qy        887 CTTTAGTTCCAAGTACTTCTTAGCATACCAAAGCACTATACTTTGGCATATGGTTTTCTG 946
              ||| ||||| |||||   || || || | || | || |||||||||   || |  ||||
Db      64827 CTTCAGTTCAAAGTA---CTCAGGATGCTAAGGTACCATACTTTGGGGGATTGGGTTCTA 64883

Qy        947 AGCTCTAACACTGCAA 962
              ||  |  ||||  | |
Db      64884 AGACCCGACACCACTA 64899
```

```
RESULT 65
AC098480
LOCUS       AC098480              188681 bp    DNA     linear   PRI 29-MAR-2002
DEFINITION  Homo sapiens chromosome 3 clone RP11-149J4, complete sequence.
ACCESSION   AC098480 AC060824
VERSION     AC098480.2  GI:19807858
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 188681)
  AUTHORS   Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
            Saenphimmachak,C., Phelps,K.A., Raymond,C. and Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 188681)
  AUTHORS   Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-OCT-2001) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
REFERENCE   3  (bases 1 to 188681)
  AUTHORS   Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
            Saenphimmachak,C., Phelps,K.A., Raymond,C. and Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAR-2002) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
COMMENT     On Mar 29, 2002 this sequence version replaced gi:16328286.
            -------------- Genome Center
                Center: University of Washington Genome Center
                Center Code: UWGC
                Web site: http://www.genome.washington.edu
                Contact: uwgchtgs@u.washington.edu
                Drafting Center: BCM
            -------------- Project Information
                Center project name: chr-3
                Center clone name: RP11-149J4 (bc0625)
            ------------- Summary Statistics
                Sequencing vector: unknown; 54% of reads
                Sequencing vector: plasmid; L08752; 46% of reads
                Chemistry: Dye-terminator ET; 89% of reads
                Chemistry: Dye-terminator Big Dye; 11% of reads
                Assembly program: Phrap; version 0.990319
                Consensus quality: 188471 bases at least Q40
                Consensus quality: 188646 bases at least Q30
                Consensus quality: 188673 bases at least Q20
                Insert size: 188681; sum-of-contigs
                Quality coverage: 8.8x in Q20 bases; sum-of-contigs
            --------------
            Overlapping Sequences:
                5':  CTD-2007I9 (UWGC:bc0773)
                3':  RP11-963F9 (UWGC:bc0582) AC013250
            --------------------
            Sequence Quality Assessment:
             This entry has been annotated with sequence quality
             estimates computed by the Phrap assembly program.
             All manually edited bases have been reduced to quality zero.
             Quality levels above 40 are expected to have less than
             1 error in 10,000 bp.
             Base-by-base quality values are not generally visible from the
             GenBank flat file format but are available as part
             of this entry's ASN.1 file.

            This sequence was finished as follows unless otherwise noted:
            all regions were either double-stranded or sequenced with an
            alternate chemistry or covered by high quality data (i.e., Phred
            quality >= 30); an attempt was made to resolve all sequencing
            problems, such as compressions and repeats; all regions were
            covered by at least one plasmid subclone or more than one M13
            subclone; and the assembly was confirmed by restriction digest.
            --------------------
            Sequence Validation:
             This sequence has been validated by Multiple Complete Digest
             fingerprinting. Comparison of the experimentally derived digest
             fragments with sequence-predicted fragments is given below.
             The electronically-digested sequence consists of both insert and
             vector, in order to accurately represent the entire circular BAC.
             Small fragments below a variable cutoff (approximately 400-800 bp)
             are not resolved in the fingerprint and hence do not appear
             in the table. There are no significant remaining discrepancies
             between the experimental and predicted values. Uniquely ordered
             fragments are separated by dashed lines.
```

EcoRI		HindIII		BglII	
SeqDerMap	FngrPrnt	SeqDerMap	FngrPrnt	SeqDerMap	FngrPrnt
8696	8900	1385	1355	2717	2710
6	<800	6382	6504	2067	2059
5625	5689	512	<800	6196	6292
1489	1488	449	<800	796	<800
4278	4212	2087	2055	2456	2464
1127	1121	429	<800	6865	6751
4236	4212	1988	1960	1103	1079
489	<800	1171	1158	6680	6751
1620	1593	6507	6504	3329	3321
2940	2960	3437	3275	884	871
5488	5489	140	<800	1759	1708
2685	2665	3039	2993	1529	1481
406	<800	5355	5300	4783	4702
727	741	192	<800	3879	3951
1605	1593	17	<800	779	<800
4675	4653	2940	2993	230	<800
2666	2665	12748	12679	19135	19287
220	<800	563	<800	4406	4170

12857	13005	2964	2993	3977	3951
13219	13005	1294	1279	2713	2710
40	<800	3546	3639	2441	2464
1710	1674	764	796	723	<800
1342	1328	8230	8281	2241	2209
586	<800	584	<800	870	871
316	<800	612	<800	4306	4385
15309	15312	244	<800	829	871
365	<800	6551	6504	9716	9636
9590	9623	262	<800	355	<800
8426	8485	3644	3639	4256	4170
1531	1488	1953	1960	665	<800
2324	2318	947	956	73	<800
6495	6435	1137	1158	2082	2059
4094	4071	3001	2993	320	<800
5116	5062	700	<800	5265	5223
876	917	3237	3275	700	<800
1789	1781	219	<800	108	<800
32	<800	3326	3275	3549	3560
933	917	209	<800	2706	2710
3585	3563	2459	2458	4242	4170
11	<800	775	796	6310	6292
1057	1070	1462	1442	750	<800
58	<800	7846	7812	77	<800
1219	1202	5036	5004	4104	4170
11079	10915	449	<800	5608	5595
195	<800	9	<800	663	<800
2233	2318	7184	7158	11289	10954
6672	6725	3720	3639	2186	2209
3359	3379	1391	1355	1682	1708
2901	2960	781	796	240	<800
6077	6075	9363	9345	1251	1237
2340	2318	2757	2782	3824	3951
329	<800	1316	1279	4082	4170
4710	4653	1715	1681	5926	5931
1287	1253	776	796	3118	3103
1109	1121	1015	1027	5662	5595

2035	1973	3025	2993	212	<800
4062	4071	14	<800	7529	7483

Query Match 4.2%; Score 257.6; DB 9; Length 188681;
Best Local Similarity 68.3%; Pred. No. 2.9e-46;
Matches 462; Conservative 0; Mismatches 174; Indels 40; Gaps 6;

```
Qy        310 CAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAAC 369
              | |||||| || ||| |||||||||| ||| |||  |  ||| ||| ||| || |||| |
Db     182572 CTGTAGGGATGGAAAGGTGTGATACCTTTCCTCACCTCTCACAAGAGTCACAGCCAACAT 182631

Qy        370 TCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTT 429
              |||||||  |||||||||| |  ||||| ||| | |||||||| ||||||||||||||||
Db     182632 TCCTATAACAAAAGACAGGTCAGCAAGAGAAAAGCATAACAAATGTATTTAATCAAAGTT 182691

Qy        430 TTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTT 489
              |||  |||||| | || |||| ||||||||||||||||||  | || |||     |  |
Db     182692 TTATGTGACATAGCAGCCTTCCGAAATGAAGACCCAAAGACTAAAGGAAAA-----TAGT 182746

Qy        490 TTTTTTGCTGAGGTTCGATGAAGAATGGATAGC-ATGTAGCCATGTAGATTAGACAAAAG 548
              |||| |||| ||||||  |||||||||  | |||  |||||  |||| |   |   ||| |
Db     182747 TTTTATGCTTAGGTTCAGTGAAGAATGCACAGCTGTGTAGAAATGTGGTAGAACAAAAGG 182806

Qy        549 GATATGATCTAGTGGTAA----------------------AGGACTCAGGGGGAAACACA 586
              | | || |||| ||| ||                      ||  |   ||| |     ||
Db     182807 GGTGTGTTCTACTGGAAATAGAAAATAGATTGGCGGTGGTGGGGGTGGTGGTGGGGATCA 182866

Qy        587 GCAAGGCCTGTCTATTCAGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTC 646
              |||||||||| |||||||||||||| |||  |||||          .||| || | |||||||
Db     182867 GCAAGGCCTGACTATTCAGATTCTTATTGGCCTCTC--------TGTGTAACGTTCTTTC 182918

Qy        647 CTCCTGAGTATGGGGCAGGACTCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAG 706
              |||  ||||||| ||||| |   | |  ||||||| ||||||| |||||||| ||| ||
Db     182919 TTCCC-AGTATGGAGCAGGGCCTCTATGGAATGAGGATCTTCAAAGGAGAAGGAAGAGAG 182977

Qy        707 TGGCCTTTTTAGATTTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCA 766
              || | ||| ||| ||| |||||||||  |||||||||   |||||||||||||||||||
Db     182978 TGACTTTTCTAGGTTTAATGGCTTGCTGTGGGGAAGAGACATTCTAGTTTCTATGACCCA 183037

Qy        767 TCTTGGGGAAGAGGAATTCTGGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGA 826
              ||| |||||||||||||||||| |  |||||||||    | || ||||| |||| ||
Db     183038 CCTTAGGGAAGAGGAATTCTGGTTTATACGACTTGCTTAAGGGGAGGAAGAGAAGTAGGA 183097

Qy        827 GGCAGGAGGGCAGGAGATGGTCAGAAAGAGACTTGGCTGCTTCTGAGGGCTTCCGCTCTC 886
              | ||| |     |||||  | || |||||   | |                |     | ||
Db     183098 GACAGTATTATGGGAGACTGCCACAAAGACCTTGGTTCTGAGGGCTTTCCAATGTATTTC 183157

Qy        887 CTTTAGTTCCAAGTACTTCTTAGCATACCAAAGCACTATACTTTGGCATATGGTTTTCTG 946
              ||| ||||| |||||    || || || | || | || |||||||||    || |  ||||
Db     183158 CTTCAGTTCAAAGTA---CTCAGGATGCTAAGGTACCATACTTTGGGGGATTGGGTTCTA 183214

Qy        947 AGCTCTAACACTGCAA 962
              ||  |  ||||  | |
Db     183215 AGACCCGACACCACTA 183230
```

RESULT 66
AC013250
LOCUS AC013250 192252 bp DNA linear HTG 14-JAN-2001
DEFINITION Homo sapiens chromosome 3 clone RP11-963F9, WORKING DRAFT SEQUENCE, 8 unordered pieces.
ACCESSION AC013250
VERSION AC013250.24 GI:12225309
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 192252)
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,

```
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.
            and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 192252)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-NOV-1999) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jan 14, 2001 this sequence version replaced gi:12083981.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: HMLZ
                Center clone name: RP11-963F9
            -------------- Summary Statistics
                Sequencing vector: M13; L08821
                Chemistry: Dye-primer Bodipy: 1% of reads
                Chemistry: Dye-terminator Big Dye: 99% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 184999 bases at least Q40
                Consensus quality: 188400 bases at least Q30
                Consensus quality: 189989 bases at least Q20
                Estimated insert size: 189170; sum-of-contigs estimation
                Estimated insert size: 160778; agarose-fp estimation
                Quality coverage: 7.1x in Q20 bases; agarose-fp estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    67400: contig of 67400 bp in length
            *   67401    67500: gap of unknown length
            *   67501   111630: contig of 44130 bp in length
            *  111631   111730: gap of unknown length
            *  111731   135827: contig of 24097 bp in length
            *  135828   135927: gap of unknown length
            *  135928   163154: contig of 27227 bp in length
            *  163155   163254: gap of unknown length
            *  163255   174531: contig of 11277 bp in length
            *  174532   174631: gap of unknown length
            *  174632   183466: contig of 8835 bp in length
            *  183467   183566: gap of unknown length
            *  183567   190769: contig of 7203 bp in length
            *  190770   190869: gap of unknown length
            *  190870   192252: contig of 1383 bp in length.
FEATURES             Location/Qualifiers
     source          1. .192252
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-963F9"
ORIGIN

  Query Match             4.2%;  Score 257.6;  DB 2;  Length 192252;
  Best Local Similarity   68.3%;  Pred. No. 2.9e-46;
  Matches  462;  Conservative    0;  Mismatches  174;  Indels   40;  Gaps    6;

Qy        310 CAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAAC 369
              | |||||| || ||| |||||||| ||| |||  |  ||| ||| ||| || |||| |
Db     177363 CTGTAGGGATGGAAAGGTGTGATACCTTTCCTCACCTCTCACAAGAGTCACAGCCAACAT 177422

Qy        370 TCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTT 429
              |||||||  |||||||||| |  |||| ||| | |||||||| ||||||||||||||||
Db     177423 TCCTATAACAAAAGACAGGTCAGCAAGAGAAAAGCATAACAAATGTATTTAATCAAAGTT 177482

Qy        430 TTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTT 489
              |||  |||||| | || |||| ||||||||||||||||||  | || |||     | |
Db     177483 TTATGTGACATAGCAGCCTTCCGAAATGAAGACCCAAAGACTAAAGGAAAA-----TAGT 177537

Qy        490 TTTTTTGCTGAGGTTCGATGAAGAATGGATAGC-ATGTAGCCATGTAGATTAGACAAAAG 548
              |||| |||| |||||  |||||||| | |||  |||||  |||| |    |  ||| |
Db     177538 TTTTATGCTTAGGTTCAGTGAAGAATGCACAGCTGTGTAGAAATGTGGTAGAACAAAAGG 177597

Qy        549 GATATGATCTAGTGGTAA----------------------AGGACTCAGGGGGAAACACA 586
              | | || |||| ||| ||                      ||  |   || |     ||
Db     177598 GGTGTGTTCTACTGGAAATAGAAAATAGATTGGCGGTGGTGGGGGTGGTGGTGGGGATCA 177657

Qy        587 GCAAGGCCTGTCTATTCAGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTC 646
              |||||||||| |||||||||||||| |||  |||||        ||| || | |||||||
Db     177658 GCAAGGCCTGACTATTCAGATTCTTATTGGCCTCTC--------TGTGTAACGTTCTTTC 177709

Qy        647 CTCCTGAGTATGGGGCAGGACTCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAG 706
              |||  ||||||| ||||| |   | |  |||||||  ||||||| ||||||| ||| ||
Db     177710 TTCCC-AGTATGGAGCAGGGCCTCTATGGAATGAGGATCTTCAAAGGAGAAGGAAGAGAG 177768

Qy        707 TGGCCTTTTTAGATTTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCA 766
              || | ||| ||| ||| |||||||||  |||||||||   ||||||||||||||||||
Db     177769 TGACTTTTCTAGGTTTAATGGCTTGCTGTGGGGAAGAGACATTCTAGTTTCTATGACCCA 177828

Qy        767 TCTTGGGGAAGAGGAATTCTGGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGA 826
              |||  ||||||||||||||||||| |  |||||||||    | || ||||| |||| ||
Db     177829 CCTTAGGGAAGAGGAATTCTGGTTTATACGACTTGCTTAAGGGGAGGAAGAGAAGTAGGA 177888

Qy        827 GGCAGGAGGGCAGGAGATGGTCAGAAAGAGACTTGGCTGCTTCTGAGGGCTTCCGCTCTC 886
              | ||| |     |||||  | || |||||   | |                |      | ||
Db     177889 GACAGTATTATGGGAGACTGCCACAAAGACCTTGGTTCTGAGGGCTTTCCAATGTATTTC 177948
```

```
Qy        887 CTTTAGTTCCAAGTACTTCTTAGCATACCAAAGCACTATACTTTGGCATATGGTTTTCTG 946
              ||| ||||| |||||   || || || | || | || |||||||||   || |  ||||
Db     177949 CTTCAGTTCAAAGTA---CTCAGGATGCTAAGGTACCATACTTTGGGGGATTGGGTTCTA 178005

Qy        947 AGCTCTAACACTGCAA 962
              ||  |  ||||  | |
Db     178006 AGACCCGACACCACTA 178021


RESULT 67
AP000825/c
LOCUS       AP000825              181254 bp    DNA     linear   HTG 30-MAY-2000
DEFINITION  Homo sapiens chromosome 11 clone RP11-688N14 map 11q24, WORKING
            DRAFT SEQUENCE, 32 unordered pieces.
ACCESSION   AP000825
VERSION     AP000825.2  GI:8118977
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 181254)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens 181,254 genomic DNA of 11q24
  JOURNAL   Published Only in DataBase (1999)
REFERENCE   2  (bases 1 to 181254)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-DEC-1999) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
            Japan (E-mail:hattori@gsc.riken.go.jp,
            URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
            Fax:81-42-778-9924)
COMMENT     On May 31, 2000 this sequence version replaced gi:6997660.
            -------------- Genome Center
                Center: RIKEN Genomic Sciences Center(GSC)
                Center code: RIKEN
                Web site: http://hgp.gsc.riken.go.jp/
                Contact: hattori@gsc.riken.go.jp
            -------------- Project Information
                Center project name: HumDraft11
                Center clone name: RP11-688N14
            -------------- Summary Statistics
                Sequencing vector: PCR products; 100% of reads
                Chemistry: Dye-terminator ET-amersham; 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 159896 bases at least Q40
                Consensus quality: 169710 bases at least Q30
                Consensus quality: 175115 bases at least Q20
                Insert size: 178154; sum-of-contigs
                Quality coverage: 4.11x in Q20 bases; sum-of-contigs
            --------------.
            NOTE: This is a 'working draft' sequence. It currently consists of
            32 contigs. The true order of the pieces is not known and their
            order in this sequence record is arbitrary. Gaps between the
            contigs are represented as runs N, but the exact sizes of the gaps
            are unknown. This record will be updated with the finished sequence
            as soon as it is available and the accession number will be
            preserved
                   1     23073 contig of  23073 bp in length
               23174     40103 contig of  16930 bp in length
               40204     56985 contig of  16782 bp in length
               57086     65370 contig of   8285 bp in length
               65471     73486 contig of   8016 bp in length
               73587     80589 contig of   7003 bp in length
               80690     87043 contig of   6354 bp in length
               87144     93246 contig of   6103 bp in length
               93347    100970 contig of   7624 bp in length
              101071    107338 contig of   6268 bp in length
              107439    113794 contig of   6356 bp in length
              113895    118957 contig of   5063 bp in length
              119058    125207 contig of   6150 bp in length
              125308    129354 contig of   4047 bp in length
              129455    134539 contig of   5085 bp in length
              134640    139757 contig of   5118 bp in length
              139858    145263 contig of   5406 bp in length
              145364    150109 contig of   4746 bp in length
              150210    152377 contig of   2168 bp in length
              152478    154773 contig of   2296 bp in length
              154874    157366 contig of   2493 bp in length
              157467    160511 contig of   3045 bp in length
              160612    163246 contig of   2635 bp in length
              163347    165600 contig of   2254 bp in length
              165701    168617 contig of   2917 bp in length
              168718    170830 contig of   2113 bp in length
              170931    172271 contig of   1341 bp in length
              172372    174333 contig of   1962 bp in length
              174434    176518 contig of   2085 bp in length
              176619    178758 contig of   2140 bp in length
              178859    179921 contig of   1063 bp in length
              180022    181254 contig of   1233 bp in length
            Sequence updated (26-May-2000).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 32 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     23073: contig of 23073 bp in length
            *    23074     23173: gap of 100 bp
            *    23174     40103: contig of 16930 bp in length
            *    40104     40203: gap of 100 bp
            *    40204     56985: contig of 16782 bp in length
            *    56986     57085: gap of 100 bp
            *    57086     65370: contig of 8285 bp in length
            *    65371     65470: gap of 100 bp
            *    65471     73486: contig of 8016 bp in length
            *    73487     73586: gap of 100 bp
            *    73587     80589: contig of 7003 bp in length
            *    80590     80689: gap of 100 bp
            *    80690     87043: contig of 6354 bp in length
            *    87044     87143: gap of 100 bp
            *    87144     93246: contig of 6103 bp in length
            *    93247     93346: gap of 100 bp
            *    93347    100970: contig of 7624 bp in length
            *   100971    101070: gap of 100 bp
            *   101071    107338: contig of 6268 bp in length
            *   107339    107438: gap of 100 bp
            *   107439    113794: contig of 6356 bp in length
            *   113795    113894: gap of 100 bp
            *   113895    118957: contig of 5063 bp in length
            *   118958    119057: gap of 100 bp
            *   119058    125207: contig of 6150 bp in length
            *   125208    125307: gap of 100 bp
            *   125308    129354: contig of 4047 bp in length
            *   129355    129454: gap of 100 bp
            *   129455    134539: contig of 5085 bp in length
            *   134540    134639: gap of 100 bp
            *   134640    139757: contig of 5118 bp in length
            *   139758    139857: gap of 100 bp
            *   139858    145263: contig of 5406 bp in length
            *   145264    145363: gap of 100 bp
            *   145364    150109: contig of 4746 bp in length
            *   150110    150209: gap of 100 bp
            *   150210    152377: contig of 2168 bp in length
            *   152378    152477: gap of 100 bp
            *   152478    154773: contig of 2296 bp in length
            *   154774    154873: gap of 100 bp
            *   154874    157366: contig of 2493 bp in length
            *   157367    157466: gap of 100 bp
            *   157467    160511: contig of 3045 bp in length
```

```
            *   160512   160611: gap of 100 bp
            *   160612   163246: contig of 2635 bp in length
            *   163247   163346: gap of 100 bp
            *   163347   165600: contig of 2254 bp in length
            *   165601   165700: gap of 100 bp
            *   165701   168617: contig of 2917 bp in length
            *   168618   168717: gap of 100 bp
            *   168718   170830: contig of 2113 bp in length
            *   170831   170930: gap of 100 bp
            *   170931   172271: contig of 1341 bp in length
            *   172272   172371: gap of 100 bp
            *   172372   174333: contig of 1962 bp in length
            *   174334   174433: gap of 100 bp
            *   174434   176518: contig of 2085 bp in length
            *   176519   176618: gap of 100 bp
            *   176619   178758: contig of 2140 bp in length
            *   178759   178858: gap of 100 bp
            *   178859   179921: contig of 1063 bp in length
            *   179922   180021: gap of 100 bp
            *   180022   181254: contig of 1233 bp in length.
FEATURES             Location/Qualifiers
     source          1. .181254
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q24"
                     /clone="RP11-688N14"
     misc_feature    1. .23073
                     /note="assembly_fragment"
     misc_feature    23174. .40103
                     /note="assembly_fragment"
     misc_feature    40204. .56985
                     /note="assembly_fragment"
     misc_feature    57086. .65370
                     /note="assembly_fragment"
     misc_feature    65471. .73486
                     /note="assembly_fragment"
     misc_feature    73587. .80589
                     /note="assembly_fragment"
     misc_feature    80690. .87043
                     /note="assembly_fragment clone_end:T7 vector_side:left"
     misc_feature    87144. .93246
                     /note="assembly_fragment"
     misc_feature    93347. .100970
                     /note="assembly_fragment"
     misc_feature    101071. .107338
                     /note="assembly_fragment"
     misc_feature    107439. .113794
                     /note="assembly_fragment"
     misc_feature    113895. .118957
                     /note="assembly_fragment"
     misc_feature    119058. .125207
                     /note="assembly_fragment"
     misc_feature    125308. .129354
                     /note="assembly_fragment clone_end:SP6 vector_side:left"
     misc_feature    129455. .134539
                     /note="assembly_fragment"
     misc_feature    134640. .139757
                     /note="assembly_fragment"
     misc_feature    139858. .145263
                     /note="assembly_fragment"
     misc_feature    145364. .150109
                     /note="assembly_fragment"
     misc_feature    150210. .152377

  Query Match             4.2%;  Score 257.4;  DB 2;  Length 181254;
  Best Local Similarity   69.5%;  Pred. No. 3.2e-46;
  Matches  444;  Conservative    0;  Mismatches  171;  Indels   24;  Gaps    6;

Qy        329 TGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTCCTATAGTAAAAGACAGG 388
              |||||| ||| ||| ||||||| || ||  |  | |  |     |  |  || |||||| |
Db      36873 TGATACCTTTCCTCATCCATCAAAAAGGGTGACAGCTGA----CACTCCTATAAGACAAG 36818

Qy        389 TTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTACATGACATGGGAGTCT 448
              |||| ||||| ||| | ||||||||||||| | |||||||||||| || |  |||  | ||
Db      36817 TTAACAAGAGAAAAGCATAACAAATTTATTCAGTCAAAGTTTTATATTATGTGGATGCCT 36758

Qy        449 TCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTTTTTTGCTGAGGTTCGAT 508
              |||||||  |||| |||||||  ||||| || ||||||  ||||| |||| ||||| |||
Db      36757 TCAGAAACAAAGATCCAAAGATTCAGGGAAAGCTGTCT-ATTTTTATGCTTAGGTTTGAT 36699

Qy        509 GAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGATATGATCTAGTGGTAAAG 568
              | ||||| || |||          |    |  | |||||| |||||  ||| | ||||||
Db      36698 GGAGAATAGACAGCCATGCAGAAATGTGGCTGGACAAAATGATATAATCAAATGGTAATA 36639

Qy        569 GACTCAG-GGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTTCTTGATCTCTCTCTCT 627
              |||  || | | |||  |||||| |||||||| |||||| |||| |||  |||||
Db      36638 GACCGAGAGTGAAAATCCAGCAAAGCCTGTCTGTTCAGACTCTTTTTGGCCTCTC----- 36584

Qy        628 CTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTTCTTCAATGAGGGTCTT 687
                | | |||||||| |||||||| ||||  | ||| |  | | |  || |||||||||
Db      36583 ---TATTTAGCATTCCTTCCTCCTGGGTATTAGACAGAATCCCCTTTGGAAGGAGGGTCTT 36527

Qy        688 CAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATGGCTTGCTTCGGGGAAGAGGAG 747
              |||| |||||||||| |||| ||||| ||| |||||||  || ||| | || | ||  |
Db      36526 CAAGAGAGAAGGGAGAGAGTGACCTTTCTAGGTTTTATGAATTTCTTTGAGGGAAAGAGG 36467

Qy        748 TTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCTGGTTTCTGTGACTTGC----- 802
              |||||||||||  ||    |||||||||||||||||||| ||| | |||||  |
Db      36466 CTCTAGTTTCTACAACTTGACTTGGGGAAGAGGAATTCTGATTTTTATGACTCTCTCCCC 36407

Qy        803 --TTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGGTCAGAAAGAGACTT 860
                ||    | ||||||||  || | ||  ||||||||||  | ||||||||||||||||
Db      36406 CTTTCAGGGTAGAAAGAGAGGTGAAAGATAGGAGGGCAGGCAAAGGTCAGAAAGAGACTT 36347

Qy        861 GGCTGCT---TCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAGTACTTCTTAGCATACCAA 917
              ||| ||    |||           ||  || |   |       | | |||| ||| ||||
Db      36346 TGCTTCTGAGGCTGCATCTGAGGCCTTCCCATCTCTAGTTCAAAGTACTTATCATGCCAA 36287

Qy        918 AGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAACA 956
              ||||  |||||||||  |||  || |||||||| ||||
Db      36286 AGCATCATACTTTGGGGTATCATTCTCTGAGCTCCAACA 36248


RESULT 68
AP000868/c
LOCUS       AP000868               181589 bp    DNA     linear   PRI 15-MAR-2003
DEFINITION  Homo sapiens genomic DNA, chromosome 11q, clone:RP11-688B18,
            complete sequence.
ACCESSION   AP000868
VERSION     AP000868.4  GI:15320463
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens genomic DNA
  JOURNAL   Published Only in Database (1999)
REFERENCE   2  (bases 1 to 181589)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-DEC-1999) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     On Aug 27, 2001 this sequence version replaced gi:11320565.
FEATURES             Location/Qualifiers
     source          1. .181589
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
```

```
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q"
                     /clone="RP11-688B18"
ORIGIN

  Query Match             4.2%;  Score 257.4;  DB 9;  Length 181589;
  Best Local Similarity   69.5%;  Pred. No. 3.2e-46;
  Matches  444;  Conservative    0;  Mismatches  171;  Indels   24;  Gaps    6;

Qy        329 TGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTCCTATAGTAAAAGACAGG 388
              |||||| ||| ||| ||||||| || ||  |  | |  |    |  |  || |||||| |
Db      78866 TGATACCTTTCCTCATCCATCAAAAAGGGTGACAGCTGA----CACTCCTATAAGACAAG 78811

Qy        389 TTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTACATGACATGGGAGTCT 448
              |||| ||||| ||| | |||||||||||||| | |||||||||| || |  |||  | ||
Db      78810 TTAACAAGAGAAAAGCATAACAAATTTATTCAGTCAAAGTTTTATATTATGTGGATGCCT 78751

Qy        449 TCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTTTTTTTGCTGAGGTTCGAT 508
              |||||||  |||| |||||||  ||||| || ||||||  |||||| |||| ||||| |||
Db      78750 TCAGAAACAAAGATCCAAAGATTCAGGGAAAGCTGTCT-ATTTTTATGCTTAGGTTTGAT 78692

Qy        509 GAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGATATGATCTAGTGGTAAAG 568
              | ||||| || |||       |    |  | |||||||| ||||| ||| | ||||||
Db      78691 GGAGAATAGACAGCCATGCAGAAATGTGGCTGGACAAAATGATATAATCAAATGGTAATA 78632

Qy        569 GACTCAG-GGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTTCTTGATCTCTCTCTCT 627
              ||| || | | |||  |||||| |||||||| |||||| |||| |||  |||||
Db      78631 GACCGAGAGTGAAAATCCAGCAAAGCCTGTCTGTTCAGACTCTTTTTGGCCTCTC----- 78577

Qy        628 CTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTTCTTCAATGAGGGTCTT 687
                | | |||||||| ||||||||| ||||  | |||  |  | |  || |||||||||
Db      78576 ---TATTTAGCATTCCTTCCTCCTGGGTATTAGACAGAATCCCTTTGGAAGGAGGGTCTT 78520

Qy        688 CAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATGGCTTGCTTCGGGGAAGAGGAG 747
              |||| |||||||||||| |||| ||||| ||| |||||||  || ||| | || | ||  |
Db      78519 CAAGAGAGAAGGGAGAGAGTGACCTTTCTAGGTTTTATGAATTTCTTTGAGGGAAAGAGG 78460

Qy        748 TTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCTGGTTTCTGTGACTTGC----- 802
              ||||||||||  ||    |||||||||||||||||||||||| ||| | |||||  |
Db      78459 CTCTAGTTTCTACAACTTGACTTGGGGAAGAGGAATTCTGATTTTTATGACTCTCTCCCC 78400

Qy        803 --TTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGGTCAGAAAGAGACTT 860
                ||    | ||||||||  || | ||  ||||||||||  | ||||||||||||||||
Db      78399 CTTTCAGGGTAGAAAGAGAGGTGAAAGATAGGAGGGCAGGCAAAGGTCAGAAAGAGACTT 78340

Qy        861 GGCTGCT---TCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAGTACTTCTTAGCATACCAA 917
              |||  ||    |||         ||  || |   |       | | |||| ||| ||||
Db      78339 TGCTTCTGAGGCTGCATCTGAGGCCTTCCCATCTCTAGTTCAAAGTACTTATCATGCCAA 78280

Qy        918 AGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAACA 956
              ||||  |||||||||  |||  || |||||||| ||||
Db      78279 AGCATCATACTTTGGGGTATCATTCTCTGAGCTCCAACA 78241


RESULT 69
AC022539/c
LOCUS       AC022539               96512 bp    DNA     linear   PRI 09-APR-2002
DEFINITION  Homo sapiens chromosome 10 clone RP11-456A18, complete sequence.
ACCESSION   AC022539
VERSION     AC022539.13  GI:20087101
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 96512)
  AUTHORS   Smith,D.R.
  TITLE     Genome Therapeutics Corporation Sequencing Center: Human Genome
            Sequence Data
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 96512)
  AUTHORS   Smith,D.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-FEB-2000) Genome Therapeutics Corporation, 100 Beaver
            Street, Waltham, MA 02453, USA
REFERENCE   3  (bases 1 to 96512)
  AUTHORS   Smith,D.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUL-2001) Genome Therapeutics Corporation, 100 Beaver
            Street, Waltham, MA 02453, USA
REFERENCE   4  (bases 1 to 96512)
  AUTHORS   Smith,D.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUL-2001) Genome Therapeutics Corporation, 100 Beaver
            Street, Waltham, MA 02453, USA
REFERENCE   5  (bases 1 to 96512)
  AUTHORS   Smith,D.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAR-2002) Genome Therapeutics Corporation, 100 Beaver
            Street, Waltham, MA 02453, USA
REFERENCE   6  (bases 1 to 96512)
  AUTHORS   Smith,D.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-APR-2002) Genome Therapeutics Corporation, 100 Beaver
            Street, Waltham, MA 02453, USA
COMMENT     On Apr 9, 2002 this sequence version replaced gi:19774371.
            -------------- Genome Center
                Center: Genome Therapeutics Corporation
                Center code: GTC
                Web site: http://www.genomecorp.com/
                Contact: gtc-seqcenter@genomecorp.com
            -------------- Project Information
                Center project name: hg142

                IMPORTANT: This sequence is not the entire
                insert of clone RP11-456A18.  It may be
                shorter because we sequence overlapping
                sections only once, except for a short overlap.
                The true left end of clone RP11-561G1 is at
                2000 in this sequence.
            -------------- Summary Statistics
                Sequencing vector: N/A
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 990315
            --------------
FEATURES             Location/Qualifiers
     source          1. .96512
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP11-456A18"
                     /clone_lib="RPCI-11"
ORIGIN

  Query Match             4.2%;  Score 257.2;  DB 9;  Length 96512;
  Best Local Similarity   69.1%;  Pred. No. 3.7e-46;
  Matches  463;  Conservative    0;  Mismatches  178;  Indels   29;  Gaps    7;

Qy        304 AAGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAAC 363
              ||| |   |||||||    ||  ||||||| ||| |||  |||| |||||||||  | |
Db      68185 AAGATTTTGTAGGGGCAGGAAAATGTGATAACTTTCCTCACCCATTATAAGGGTCATAGC 68126

Qy        364 CAAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAA---CCTAACAAATTTATTTA 420
                | | |||| ||  ||||| ||| |||| || |  ||||    | |||||||| |||||
Db      68125 TGACATTCCTTTAACAAAAGGCAGATTAACAACAAAAAAAATGCATAACAAATTCATTTA 68066

Qy        421 ATCAAAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAA 480
              |||||||||||||  |||  ||| |||| |||||  | |||| ||||||||||| |||
Db      68065 ATCAAAGTTTTACACAACAAAAGAGACTTCTGAAATTATGACCAAAAGACCCAGGGAAAA 68006

Qy        481 CTGTCTGTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTA 540
              |||     ||||| |||| || || |||||||||||||| |||         |    ||||
Db      68005 CTG-----TTTTTATGCTTAGTTTTGATGAAGAATGGACAGCCATGTAGAAATGTGATTG 67951
```

```
Qy        541 GACAAAAGGATATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTA 600
              | ||||||| |||||||||| ||||||   ||| ||| ||| |           |||  | |
Db      67950 GGCAAAAGGGTATGATCTAATGGTAATAAACTGAGGTGGAGA---------ACCTAGCAA 67900

Qy        601 TTCAGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGG 660
                |  | |   ||  ||||   ||||||   ||      ||||||  ||   |||||||
Db      67899 GGCCCGTCCAGATTCTTCTCAGCCTCTCTGAATAGCATTCTCTTTCTACCCAGGTATGGG 67840

Qy        661 GCAGGACTCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAA---AGTGGCCTTTTTA 717
              |||| || | |||  |||||  |||||||||| |||| ||||||   || | ||||||||
Db      67839 GCAGAACCCCTCTAGAATGAAAGTCTTCAAGGAAGAAAGGAGAAGAGAGAGACCTTTTTA 67780

Qy        718 GATTTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAG 777
              | ||||||||||||||   ||  | | | |||| ||||| |||||||||| ||||| ||||
Db      67779 GGTTTTATGGCTTGCTTTAGGAGAAAAGGGTTCCAGTTTTTATGACCCACCTTGGAGAAG 67720

Qy        778 AGGAATTCTGGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGC 837
              || |||||| |||||| ||||||||||     |  |||||||  |     |  |||||||||
Db      67719 AGAAATTCTAGTTTCTATGACTTGCTTCAGAGGTGAAAGAGAGGCTGTAAACAGGAGGGC 67660

Qy        838 AGGAGATGGTCAGAAAGAGACTTGGCTGCTTCTGAGG-GCTTCCGCTCTCCTTTAGTTCC 896
              |||||| ||||||| |||        |||||||||||  |||||  | ||||  ||||
Db      67659 AGGAGAAGGTCAGAGAGA-----CCTTGCTTCTGAGGCCCTTCCAGTGTCCTAAAGTTTG 67605

Qy        897 AAGTACTTCTTAGCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAACA 956
              |||||   || ||||| |  |||  |||||||| ||     |  | ||||||||||| ||||
Db      67604 AAGTA---CTCAGCATGCTGAAGTGCTATACTTCGGGTCTTCTTGTTCTGAGCTCCAACA 67548

Qy        957 CTGCAATCAT 966
                  | |||
Db      67547 ACTTCAACAT 67538
```

```
RESULT 70
AC021718
LOCUS       AC021718               154359 bp    DNA     linear   HTG 12-MAR-2000
DEFINITION  Homo sapiens chromosome 11 clone RP11-207E21 map 11, WORKING DRAFT
            SEQUENCE, 15 unordered pieces.
ACCESSION   AC021718
VERSION     AC021718.3  GI:7229845
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 154359)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 11, clone RP11-207E21
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 154359)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 12, 2000 this sequence version replaced gi:6970561.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
               Center: Whitehead Institute/ MIT Center for Genome Research
               Center code: WIBR
               Web site: http://www-seq.wi.mit.edu
               Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
               Center project name: L5944
               Center clone name: 207_E_21
            -------------- Summary Statistics
               Sequencing vector: M13; M77815; 100% of reads
               Chemistry: Dye-terminator Big Dye; 100% of reads
               Assembly program: Phrap; version 0.960731
               Consensus quality: 147423 bases at least Q40
               Consensus quality: 150978 bases at least Q30
               Consensus quality: 152145 bases at least Q20
               Insert size: 158000; agarose-fp
               Insert size: 152959; sum-of-contigs
               Quality coverage: 4.3 in Q20 bases; agarose-fp
               Quality coverage: 4.4 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 15 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1155: contig of 1155 bp in length
            *     1156     1255: gap of 100 bp
            *     1256     3249: contig of 1994 bp in length
            *     3250     3349: gap of 100 bp
            *     3350     5935: contig of 2586 bp in length
            *     5936     6035: gap of 100 bp
            *     6036     8676: contig of 2641 bp in length
            *     8677     8776: gap of 100 bp
            *     8777    11533: contig of 2757 bp in length
            *    11534    11633: gap of 100 bp
            *    11634    15331: contig of 3698 bp in length
            *    15332    15431: gap of 100 bp
            *    15432    19618: contig of 4187 bp in length
            *    19619    19718: gap of 100 bp
            *    19719    26030: contig of 6312 bp in length
            *    26031    26130: gap of 100 bp
            *    26131    34060: contig of 7930 bp in length
            *    34061    34160: gap of 100 bp
            *    34161    43748: contig of 9588 bp in length
            *    43749    43848: gap of 100 bp
            *    43849    55829: contig of 11981 bp in length
            *    55830    55929: gap of 100 bp
            *    55930    68307: contig of 12378 bp in length
            *    68308    68407: gap of 100 bp
            *    68408    82534: contig of 14127 bp in length
            *    82535    82634: gap of 100 bp
            *    82635   102988: contig of 20354 bp in length
            *   102989   103088: gap of 100 bp
            *   103089   154359: contig of 51271 bp in length.
FEATURES             Location/Qualifiers
     source          1. .154359
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11"
                     /clone="RP11-207E21"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1155
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     misc_feature    1256. .3249
                     /note="assembly_fragment"
```

```
     misc_feature    3350. .5935
                     /note="assembly_fragment"
     misc_feature    6036. .8676
                     /note="assembly_fragment"
     misc_feature    8777. .11533
                     /note="assembly_fragment"
     misc_feature    11634. .15331
                     /note="assembly_fragment"
     misc_feature    15432. .19618
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    19719. .26030
                     /note="assembly_fragment"
     misc_feature    26131. .34060
                     /note="assembly_fragment"
     misc_feature    34161. .43748
                     /note="assembly_fragment"
     misc_feature    43849. .55829
                     /note="assembly_fragment"
     misc_feature    55930. .68307
                     /note="assembly_fragment"
     misc_feature    68408. .82534
                     /note="assembly_fragment"
     misc_feature    82635. .102988
                     /note="assembly_fragment"
     misc_feature    103089. .154359
                     /note="assembly_fragment"
ORIGIN

  Query Match             4.2%;  Score 257.2;  DB 2;  Length 154359;
  Best Local Similarity   69.1%;  Pred. No. 3.6e-46;
  Matches  463;  Conservative    0;  Mismatches  178;  Indels   29;  Gaps    7;

Qy        304 AAGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAAC 363
              |||| |   |||||||    ||  |||||||  ||| |||  |||| |||||||||  | |
Db     145025 AAGATTTTGTAGGGGCAGGAAAATGTGATAACTTTCCTCACCCATTATAAGGGTCATAGC 145084

Qy        364 CAAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAA---CCTAACAAATTTATTTA 420
              | | |||| ||  ||||| ||| |||| || |  ||||    | |||||||| |||||
Db     145085 TGACATTCCTTTAACAAAAGGCAGATTAACAACAAAAAAAATGCATAACAAATTCATTTA 145144

Qy        421 ATCAAAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAA 480
              |||||||||||||  |||   ||| |||| ||||| | |||| ||||||||||| |||
Db     145145 ATCAAAGTTTTACACAACAAAAGAGACTTCTGAAATTATGACCAAAAGACCCAGGGAAAA 145204

Qy        481 CTGTCTGTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTA 540
              |||     ||||| |||| || || |||||||||||||| |||       |    ||||
Db     145205 CTG-----TTTTTATGCTTAGTTTTGATGAAGAATGGACAGCCATGTAGAAATGTGATTG 145259

Qy        541 GACAAAAGGATATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTA 600
              | ||||||| |||||||| ||||||   ||| ||| ||| |          |||  | |
Db     145260 GGCAAAAGGGTATGATCTAATGGTAATAAACTGAGGTGGAGA---------ACCTAGCAA 145310

Qy        601 TTCAGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGG 660
                |  | |   ||  ||||   ||||||   ||      ||||||  ||   |||||||
Db     145311 GGCCCGTCCAGATTCTTCTCAGCCTCTCTGAATAGCATTCTCTTTCTACCCAGGTATGGG 145370

Qy        661 GCAGGACTCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAA---AGTGGCCTTTTTA 717
              |||| || | |||  |||||  |||||||||| |||| ||||||   || | ||||||||
Db     145371 GCAGAACCCCTCTAGAATGAAAGTCTTCAAGGAAGAAAGGAGAAGAGAGAGACCTTTTTA 145430

Qy        718 GATTTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAG 777
              | |||||||||||||||   || | | | |||| ||||| |||||||| ||||| ||||
Db     145431 GGTTTTATGGCTTGCTTTAGGAGAAAAGGGTTCCAGTTTTTATGACCCACCTTGGAGAAG 145490

Qy        778 AGGAATTCTGGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGC 837
              || |||||| |||||| |||||||||    |  |||||||  |    |  |||||||||
Db     145491 AGAAATTCTAGTTTCTATGACTTGCTTCAGAGGTGAAAGAGAGGCTGTAAACAGGAGGGC 145550

Qy        838 AGGAGATGGTCAGAAAGAGACTTGGCTGCTTCTGAGG-GCTTCCGCTCTCCTTTAGTTCC 896
              |||||| ||||||| |||        ||||||||||  |||||  | ||||  ||||
Db     145551 AGGAGAAGGTCAGAGAGA-----CCTTGCTTCTGAGGCCCTTCCAGTGTCCTAAAGTTTG 145605

Qy        897 AAGTACTTCTTAGCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAACA 956
              |||||   || ||||| |  |||  |||||||| ||    |  | |||||||||| ||||
Db     145606 AAGTA---CTCAGCATGCTGAAGTGCTATACTTCGGGTCTTCTTGTTCTGAGCTCCAACA 145662

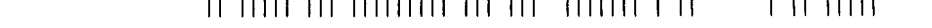
Qy        957 CTGCAATCAT 966
                    | |||
Db     145663 ACTTCAACAT 145672


RESULT 71
AL161743/c
LOCUS       AL161743              100170 bp    DNA     linear   PRI 09-APR-2001
DEFINITION  Human DNA sequence from clone RP4-730B13 on chromosome 1q42.11-42.3
            Contains part of the DISC1 (disrupted in schizophrenia 1) gene,
            STSs and GSSs, complete sequence.
ACCESSION   AL161743
VERSION     AL161743.20  GI:11121016
KEYWORDS    HTG; DISC1; schizophrenia.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 100170)
  AUTHORS   Coville,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Nov 8, 2000 this sequence version replaced gi:10715861.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            This sequence is the entire insert of clone RP4-730B13 This
            sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. RP4-730B13 is from
            the library RPCI-4 constructed by the group of Pieter de Jong. For
            further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pCYPAC2.
FEATURES             Location/Qualifiers
     source          1. .100170
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /map="q42.11-42.3"
                     /clone="RP4-730B13"
                     /clone_lib="RPCI-4"
     gene            complement(join(137. .255,17641. .17905,67802. .67862))
                     /gene="DISC1"
     CDS             complement(join(<137. .255,17641. .17905,67802. .>67862))
                     /gene="DISC1"
                     /note="match: cDNAs: Em:AF222980 Em:AB007926
                     match: proteins: Tr:O75045"
                     /codon_start=3
```

```
                     /evidence=not_experimental
                     /product="dJ730B13.1 (disrupted in schizophrenia 1)"
                     /protein_id="CAC36349.1"
                     /db_xref="GI:13591447"
                     /db_xref="GOA:Q9NRI5"
                     /db_xref="UniProt/Swiss-Prot:Q9NRI5"
                     /translation="TSVKENTMKYMETLKNKLCSCKCPLLGKVWEADLEACRLLIQSL
                     QLQEARGSLSVEDERQMDDLEGAAPPIPPRLHSEDKRKTPLKVLEEWKTHLIPSLHCA
                     GGEQKEESYILSAELGEKCEDIGKKLLYLEDQLHTAIHSHDEDLI"
repeat_region        826. .1092
                     /note="AluY repeat: matches 32. .298 of consensus"
misc_feature         complement(959. .1362)
                     /gene="DISC1"
                     /note="match: GSS: Em:AQ738086"
misc_feature         complement(981. .1392)
                     /gene="DISC1"
                     /note="match: GSS: Em:B88351"
repeat_region        1093. .1191
                     /note="L2 repeat: matches 2606. .2705 of consensus"
misc_feature         complement(1109. .1392)
                     /gene="DISC1"
                     /note="match: GSS: Em:B71142"
repeat_region        1801. .2092
                     /note="AluJo repeat: matches 1. .295 of consensus"
repeat_region        2603. .2680
                     /note="39 copies 2 mer ta 79% conserved"
misc_feature         3674. .4066
                     /note="match: STS: Em:G60191"
misc_feature         3693. .3781
                     /note="match: STS: Em:G43347"
repeat_region        4085. .4381
                     /note="AluSc repeat: matches 1. .298 of consensus"
repeat_region        4534. .4992
                     /note="MLT1G repeat: matches 39. .512 of consensus"
repeat_region        5316. .5355
                     /note="20 copies 2 mer tg 100% conserved"
repeat_region        6320. .6630
                     /note="AluSp repeat: matches 1. .313 of consensus"
repeat_region        7439. .7654
                     /note="L2 repeat: matches 2428. .2710 of consensus"
repeat_region        12707. .13261
                     /note="MER39 repeat: matches 3. .549 of consensus"
misc_feature         complement(13490. .14103)
                     /gene="DISC1"
                     /note="match: GSS: Em:AQ041978"
repeat_region        14437. .14548
                     /note="LTR34 repeat: matches 1. .116 of consensus"
misc_feature         complement(14853. .15183)
                     /gene="DISC1"
                     /note="match: GSS: Em:AQ542245"
repeat_region        14925. .15148
                     /note="L1ME1 repeat: matches 5931. .6150 of consensus"
misc_feature         15195. .15594
                     /note="match: GSS: Em:AQ168058"
repeat_region        15638. .16158
                     /note="MER39 repeat: matches 6. .543 of consensus"
repeat_region        17283. .17372
                     /note="10 copies 9 mer tatataata 68% conserved"
repeat_region        18189. .18382
                     /note="L2 repeat: matches 2298. .2496 of consensus"
repeat_region        18477. .18525
                     /note="L2 repeat: matches 2601. .2649 of consensus"
repeat_region        18533. .18908
                     /note="THE1C repeat: matches 1. .371 of consensus"
repeat_region        18909. .20484
                     /note="THE1C-internal repeat: matches 35. .1580 of
                     consensus"
repeat_region        20804. .21002
                     /note="L1MEc repeat: matches 1217. .1427 of consensus"
repeat_region        21033. .21399
                     /note="L1MEc repeat: matches 1497. .1881 of consensus"
repeat_region        21549. .21786
                     /note="119 copies 2 mer ag 56% conserved"
repeat_region        21828. .21951
                     /note="L2 repeat: matches 966. .1088 of consensus"
repeat_region        21952. .22071
                     /note="L1MC/D repeat: matches 5371. .5496 of consensus"
repeat_region        22107. .22677
                     /note="L1M4 repeat: matches 2922. .3512 of consensus"
repeat_region        22825. .23558
                     /note="L1M4 repeat: matches 2131. .2925 of consensus"
repeat_region        23635. .25145
                     /note="L1MD2 repeat: matches 4794. .6307 of consensus"
repeat_region        25313. .25611
                     /note="AluSx repeat: matches 2. .302 of consensus"
repeat_region        25800. .26697
                     /note="L1MD repeat: matches 979. .1932 of consensus"
repeat_region        26792. .26881
                     /note="10 copies 9 mer tgtttttgt 71% conserved"
repeat_region        26799. .26878
                     /note="40 copies 2 mer tt 63% conserved"
repeat_region        27364. .27547
                     /note="L1MD repeat: matches -17. .167 of consensus"
misc_feature         complement(27754. .28002)
                     /gene="DISC1"
                     /note="match: GSS: Em:AQ019129"
repeat_region        28092. .28203
                     /note="HAL1 repeat: matches 626. .730 of consensus"
repeat_region        30215. .31339
                     /note="L1PA2 repeat: matches 5020. .6144 of consensus"
repeat_region        31342. .31377
                     /note="18 copies 2 mer aa 83% conserved"
repeat_region        31979. .32041
                     /note="7 copies 9 mer tcatcatca 73% conserved"
repeat_region        32052. .32121
                     /note="MIR repeat: matches 158. .227 of consensus"
repeat_region        32137. .32407
                     /note="MLT1G repeat: matches 1. .261 of consensus"
repeat_region        32979. .33095
                     /note="L2 repeat: matches 2593. .2705 of consensus"
repeat_region        34127. .34397
                     /note="L1ME repeat: matches 5108. .5386 of consensus"
repeat_region        34495. .34664
                     /note="L1ME repeat: matches 5567. .5747 of consensus"
repeat_region        36753. .36960
                     /note="MIR repeat: matches 35. .262 of consensus"
repeat_region        37937. .38298
                     /note="MLT1A1 repeat: matches 1. .365 of consensus"
repeat_region        38332. .38471
                     /note="MER5A repeat: matches 9. .173 of consensus"
repeat_region        38492. .38783
                     /note="AluSx repeat: matches 1. .295 of consensus"
misc_feature         39472. .39939
                     /note="match: GSS: Em:AQ683363"
misc_feature         39501. .39917
                     /note="match: GSS: Em:AQ088543"
misc_feature         complement(39885. .40383)
                     /gene="DISC1"
                     /note="match: GSS: Em:AQ121406"
repeat_region        40529. .40699
                     /note="MIR repeat: matches 59. .233 of consensus"
repeat_region        41234. .41325
                     /note="MIR repeat: matches 166. .252 of consensus"
repeat_region        41326. .41651
                     /note="AluSq repeat: matches 1. .311 of consensus"
repeat_region        41652. .41792
                     /note="MIR repeat: matches 7. .166 of consensus"
repeat_region        41864. .42129
                     /note="L2 repeat: matches 2408. .2688 of consensus"

 Query Match             4.2%;  Score 256.4;  DB 9;  Length 100170;
 Best Local Similarity   72.2%;  Pred. No. 5.5e-46;
 Matches  468;  Conservative    0;  Mismatches  146;  Indels   34;  Gaps    9;

Qy       313 TAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTCC 372
```



```
Db      13249 TACGGGTGGAAAGGTGTGATACCTTTCCTCAGCCATCATTAAGG-----GCTAACACTCC 13195

Qy        373 TATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTA 432
              |||| |||| || |  | ||| ||||| ||||| |||||||||||||| |||||||
Db      13194 TATAACAAAAACAAGTT---AACGAGAAAAACATAACACATTTATTTAATCAGAGTTTTA 13138

Qy        433 CATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTTT 492
              |||| ||  |||| |||||||||||||||||||||||||||||  |||||      ||||
Db      13137 CATGGCACAGGAGGCTTCAGAAATGAAGACCCAAAGACCCAGGAAAAACT-----ATTTT 13083

Qy        493 TTTGCTGAGGTTCGATGAAGAATGGATAGC-ATGTAGCCATGTAGATTAGACAAAAGGAT 551
              | |||| || || ||| ||||| |  ||| ||| ||  ||||   |     ||| || |
Db      13082 TATGCTCAGATTTAATGCAGAATAGTCAGCTATGAAGGAATGTGATTGGACAAAATGGGT 13023

Qy        552 ATGATCTAGTGGTAAAGGACTCAGGGGGAAAC---ACAGCAAGGCCTGTCTATTCAGATT 608
              |||||||| |||| |  |||| |||  | |     ||||||| ||||||| ||||||||
Db      13022 ATGATCTAATGGTTATAGACTGGGGGGTGAGCGGACCAGCAAGCCCTGTCTGTTCAGATT 12963

Qy        609 CTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACT 668
              |||||||  | |||         || |||| |||  |||||| || ||||||||| |||||
Db      12962 CTTCTTGGACGCTC--------TGCATAGTATTTATTCCTCTTGGGTATGGGGCCGGACT 12911

Qy        669 CTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTATGGC 728
              | |||  |  ||||| |||||||||||||| ||| |||| ||||| ||| |||||| |
Db      12910 CCTCTAAATGGAGGGCCTTCAAGGGAGAAGGAAGAGAGTGACCTTTCTAGGTTTTAT-GT 12852

Qy        729 TTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCTGG 788
              |||||| | || |||||||| |||||||||| |||     ||| ||||||||||||||||
Db      12851 TTGCTTTGAGGGAGAGGAGTCCTAGTTTCTAAGACTGGCTTTGAGGAAGAGGAATTCTGG 12792

Qy        789 TTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGGTC 848
              ||||| || ||||| | || | ||||| ||||||  |||   |||||||| |||| ||
Db      12791 TTTCTATGGCTTGCATCCAGGGAGAAAAAGGAGTGTGAGAGGGGAGGGCAAGAGAAGGCA 12732

Qy        849 AGAAAGAGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAGTACTTCTTA 908
                |          |||||||||| |||||  |||||||| |||| ||||||
Db      12731 GAGA------GATCTTGCTTCTGAGGCCTTCCAATCTCCTTTGGTTCAAAGTACT--CAG 12680

Qy        909 GCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAACA 956
              ||||||||||    | |||| ||  |||  |  | |||||   ||||
Db      12679 GCATACCAAAGTGTCACACTTAGGGGTATCATACTGTGAGCCACAACA 12632


RESULT 72
AL512600/c
LOCUS       AL512600              141172 bp    DNA     linear   PRI 01-MAR-2001
DEFINITION  Human DNA sequence from clone RP11-452A14 on chromosome 6, complete
            sequence.
ACCESSION   AL512600
VERSION     AL512600.5  GI:13235022
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 141172)
  AUTHORS   Williams,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Mar 5, 2001 this sequence version replaced gi:13160332.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            RP11-452A14 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-452A14 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true left end of clone RP11-337M11 is at 141073 in this
            sequence. The true right end of clone RP1-177I10 is at 100 in this
            sequence.
FEATURES              Location/Qualifiers
     source           1. .141172
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:9606"
                      /chromosome="6"
                      /clone="RP11-452A14"
                      /clone_lib="RPCI-11.2"
     repeat_region    3413. .3448
                      /note="L2 repeat: matches 2708. .2737 of consensus"
     repeat_region    3449. .3751
                      /note="MER33 repeat: matches 2. .324 of consensus"
     repeat_region    3752. .3914
                      /note="L2 repeat: matches 2574. .2708 of consensus"
     repeat_region    3915. .4224
                      /note="AluSg repeat: matches 1. .310 of consensus"
     repeat_region    4225. .4499
                      /note="L2 repeat: matches 2197. .2574 of consensus"
     repeat_region    4535. .4791
                      /note="MLT1H repeat: matches 41. .320 of consensus"
     repeat_region    4985. .6590
                      /note="L2 repeat: matches 246. .2015 of consensus"
     repeat_region    8444. .8597
                      /note="FRAM repeat: matches 4. .166 of consensus"
     repeat_region    9535. .9581
                      /note="MADE1 repeat: matches 1. .47 of consensus"
     repeat_region    10554. .10712
                      /note="MER5B repeat: matches 16. .172 of consensus"
     repeat_region    12407. .12512
                      /note="2 copies 53 mer 94% conserved"
     repeat_region    13008. .13192
                      /note="MER39b repeat: matches 356. .549 of consensus"
     repeat_region    13194. .13554
                      /note="MER39 repeat: matches 13. .380 of consensus"
     repeat_region    14862. .14899
                      /note="L2 repeat: matches 2607. .2641 of consensus"
     repeat_region    14900. .15203
                      /note="AluJo repeat: matches 2. .301 of consensus"
     repeat_region    15204. .15302
                      /note="L2 repeat: matches 2641. .2750 of consensus"
     repeat_region    15569. .15652
                      /note="21 copies 4 mer atac 72% conserved"
     repeat_region    15667. .15960
                      /note="AluSx repeat: matches 1. .294 of consensus"
     repeat_region    18313. .18562
                      /note="MIR repeat: matches 2. .262 of consensus"
     repeat_region    18905. .19289
                      /note="MER31A repeat: matches 7. .466 of consensus"
     repeat_region    19640. .20026
                      /note="L2 repeat: matches 2010. .2460 of consensus"
     repeat_region    20027. .20396
                      /note="THE1C repeat: matches 1. .371 of consensus"
```

```
repeat_region   20397. .20684
                /note="L2 repeat: matches 2460. .2750 of consensus"
repeat_region   20981. .21023
                /note="MIR repeat: matches 86. .129 of consensus"
repeat_region   22062. .22128
                /note="THE1C repeat: matches 3. .54 of consensus"
repeat_region   22894. .23033
                /note="70 copies 2 mer at 75% conserved"
repeat_region   22904. .23023
                /note="4 copies 30 mer 80% conserved"
repeat_region   23060. .23296
                /note="MIR repeat: matches 5. .244 of consensus"
repeat_region   24053. .24253
                /note="MIR repeat: matches 7. .249 of consensus"
repeat_region   24933. .25057
                /note="MIR repeat: matches 78. .196 of consensus"
repeat_region   28904. .29463
                /note="L1ME3 repeat: matches 5291. .5912 of consensus"
repeat_region   29542. .29693
                /note="L1ME3 repeat: matches 5015. .5162 of consensus"
repeat_region   30264. .30466
                /note="MER20 repeat: matches 2. .201 of consensus"
repeat_region   30598. .33346
                /note="L1P repeat: matches 817. .3559 of consensus"
repeat_region   33978. .34161
                /note="L1M4 repeat: matches 4027. .4215 of consensus"
repeat_region   34183. .34630
                /note="112 copies 4 mer tata 54% conserved"
repeat_region   34195. .34695
                /note="167 copies 3 mer tat 58% conserved"
repeat_region   34919. .35248
                /note="L1M4 repeat: matches 3662. .4001 of consensus"
repeat_region   35298. .35586
                /note="AluSx repeat: matches 1. .282 of consensus"
repeat_region   35833. .35960
                /note="L1 repeat: matches 3256. .3389 of consensus"
repeat_region   36175. .36410
                /note="MER8 repeat: matches 1. .239 of consensus"
repeat_region   38030. .38077
                /note="24 copies 2 mer ct 75% conserved"
repeat_region   38036. .38079
                /note="11 copies 4 mer ctct 79% conserved"
repeat_region   38527. .38633
                /note="AluY repeat: matches 185. .291 of consensus"
repeat_region   38675. .38882
                /note="AluJo repeat: matches 1. .203 of consensus"
repeat_region   40007. .40357
                /note="L1M4 repeat: matches 3661. .4043 of consensus"
repeat_region   40722. .41112
                /note="L2 repeat: matches 47. .497 of consensus"
repeat_region   41218. .41246
                /note="L1PBa repeat: matches -1304. .-1274 of consensus"
repeat_region   45343. .45529
                /note="MIR repeat: matches 35. .210 of consensus"
repeat_region   46891. .47016
                /note="MIR repeat: matches 25. .146 of consensus"
repeat_region   47533. .47838
                /note="AluY repeat: matches 1. .306 of consensus"
repeat_region   49370. .49548
                /note="MIR repeat: matches 1. .191 of consensus"
repeat_region   52427. .52501
                /note="L1PA5 repeat: matches 6071. .6145 of consensus"
repeat_region   52906. .53453
                /note="L1PB3 repeat: matches 5539. .6073 of consensus"
repeat_region   55607. .55695
                /note="MLT1I repeat: matches 69. .164 of consensus"
repeat_region   55995. .56267
                /note="AluJb repeat: matches 13. .286 of consensus"
repeat_region   56268. .56375
                /note="27 copies 4 mer gaag 83% conserved"
repeat_region   56318. .56377
                /note="30 copies 2 mer ag 75% conserved"
repeat_region   58596. .58976
                /note="MSTD repeat: matches 1. .394 of consensus"
repeat_region   59106. .59325
                /note="L2 repeat: matches 2517. .2749 of consensus"
repeat_region   60792. .60896
                /note="3 copies 35 mer 76% conserved"
repeat_region   60807. .60848
                /note="14 copies 3 mer tcc 81% conserved"
repeat_region   61373. .61475
                /note="MLT1E repeat: matches 23. .122 of consensus"
repeat_region   61501. .61532
                /note="16 copies 2 mer ac 96% conserved"
repeat_region   61793. .62096
                /note="MLT1E repeat: matches 265. .566 of consensus"
repeat_region   62440. .62804
                /note="MLT1A2 repeat: matches 1. .374 of consensus"
repeat_region   62808. .62867
                /note="30 copies 2 mer tt 68% conserved"
repeat_region   63180. .63241
                /note="L2 repeat: matches 2433. .2495 of consensus"
repeat_region   64021. .64180
                /note="40 copies 4 mer cttc 71% conserved"
repeat_region   64035. .64139
                /note="3 copies 35 mer 77% conserved"
repeat_region   64069. .64170
                /note="51 copies 2 mer ct 63% conserved"
repeat_region   65800. .65889

Query Match             4.2%;  Score 256;  DB 9;  Length 141172;
Best Local Similarity   70.0%;  Pred. No. 6.6e-46;
Matches  452;  Conservative    0;  Mismatches  170;  Indels   24;  Gaps    7;

Qy        312 GTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTC 371
              | ||||||  ||| |  ||| || |||||||  ||||||||||||||  | |||| ||||
Db      13553 GAAGGGGTAGAAAGGCATGACACTTTTTCTCACCCATCATAAGGGTCATAGCCAACACTC 13494

Qy        372 CTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTT 431
              || |   |||||||||||| || || |  ||| | |||||||| ||||||||||
Db      13493 CTGTTACAAAAGACAGGTCAAAAACAAAAAAGCGTAACAAATGTATTTAATCAA------ 13440

Qy        432 ACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTT 491
              |||||||||||||  |||||||||  ||||||||||||  |||| ||| ||||||||||
Db      13439 ACATGACATGGGAACCTTCAGAAA-CAAGACCCAAAGACATAGGGAAAATTGTCTGTTTG 13381

Qy        492 TTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGAT 551
              ||   ||| |   |   ||  |   |   ||| ||   |||| ||||     ||||| |
Db      13380 CTTGAATGAAGAATGGACAATCACATAGGTCATATAAAAATGT-GATTGAGAGAAAGGGT 13322

Qy        552 ATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTT 611
              || ||||| |||| |  | || | |||| ||  ||||||||||||||  ||||||||||
Db      13321 ATAATCTAATGGTTAGAGGCTGACGGGGGAAGCCAGCAAGGCCTGTCAGTTCAGATTCTT 13262

Qy        612 CTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTT 671
              ||||           ||   ||| || ||| | ||||||||||| || |||||  ||   |
Db      13261 CTTG--------GCCTTGTTGTGTAACATACCTTCCTCCTGAGGATAGGGCAAAACCTCT 13210

Qy        672 CTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATGGCTTG 731
              ||  |||||||||||||||||| |||  |   |||   |||| ||| ||||||| ||||
Db      13209 CTGAAATGAGGGTCTTCAAGGGA-AAGAAAAGGAGTTATCTTTCTAGGTTTTATGCCTTG 13151

Qy        732 CTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCTGGTTT 791
              ||| ||||||||||||||||||||||||||||   |||| ||||||  ||||||||||||
Db      13150 CTTTGGGGAAGAGGAGTTCTAGTTTCTATGACCAGCCTTGTGGAAGAAGAATTCTGGTTT 13091

Qy        792 CTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGGTCAGA 851
              |||||||  ||||   | ||||  ||||||   || ||||||    | ||| ||||||||
Db      13090 CTGTGACTCACTTTTGGGGAGAATAAGGAGTTGAAGACAGGAGAATGGCAGAAGGTCAGA 13031

Qy        852 AAGAGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAGTACTTCTTAGCA 911
              ||||         ||  ||||||| | |||    | | ||   | |||  | |  | ||||
Db      13030 AAGA-----CCTTGGATCTGAGGCCCTCCCGAGTTCCTTCTGTGCAA--AGTATTCAGCA 12978

Qy        912 TACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAACAC 957
                || |   || | ||||  | |||  || || ||||  |   |||
```

```
Db      12977 CGCCCAGATACCACACTTCAGGATACTGTGTTTTGAGTGCCCCCAC 12932

RESULT 73
AC109928
LOCUS       AC109928               112353 bp    DNA     linear   PRI 16-APR-2002
DEFINITION  Homo sapiens BAC clone RP11-755F21 from 4, complete sequence.
ACCESSION   AC109928
VERSION     AC109928.4  GI:19920305
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 112353)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
   PUBMED   9847074
REFERENCE   2  (bases 1 to 112353)
  AUTHORS   Trani,L., Haakenson,W., Dignan,G. and Nguyen,C.
  TITLE     The sequence of Homo sapiens BAC clone RP11-755F21
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 112353)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-FEB-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 112353)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAR-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 112353)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-APR-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   6  (bases 1 to 112353)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-APR-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Apr 4, 2002 this sequence version replaced gi:19807975.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0755F21
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. John D.
            McPherson, Department of Genetics, Washington University, St. Louis
            MO. For additional information about the map position of this
            sequence, see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries. Genomics 51:1-8. The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at http://www.chori.org
            VECTOR:  pBACe3.6

            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the right is AP002075, 2000 bp overlap.
            Actual start of this clone is at base position 1 of RP11-755F21.
FEATURES             Location/Qualifiers
     source          1. .112353
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /map="4"
                     /clone="RP11-755F21"
                     /clone_lib="RPCI-11"
     repeat_region   1. .281
                     /rpt_family="MaLR"
     repeat_region   1124. .1155
                     /rpt_family="AT_rich"
     repeat_region   1197. .1223
                     /rpt_family="(TG)n"
     repeat_region   1466. .1648
                     /rpt_family="MIR"
     repeat_region   1794. .1814
                     /rpt_family="AT_rich"
     repeat_region   2030. .2140
                     /rpt_family="MIR"
     repeat_region   3147. .3216
                     /rpt_family="(TCTA)n"
     repeat_region   3447. .3857
                     /rpt_family="L1"
     misc_feature    3512. .3733
                     /note="match to EST BE247533 (NID:g9099290)"
     misc_feature    3523. .3733
                     /note="match to EST BI906977 (NID:g16169756)"
     repeat_region   4397. .4424
                     /rpt_family="AT_rich"
     repeat_region   4692. .4712
                     /rpt_family="AT_rich"
     repeat_region   5225. .5253
                     /rpt_family="(TG)n"
     repeat_region   5306. .5397
                     /rpt_family="MIR"
     repeat_region   5771. .5791
                     /rpt_family="AT_rich"
     repeat_region   5852. .5879
                     /rpt_family="AT_rich"
     repeat_region   6382. .6628
                     /rpt_family="MIR"
     repeat_region   7012. .7041
                     /rpt_family="AT_rich"
     repeat_region   10157. .11032
                     /rpt_family="L2"
     repeat_region   11277. .12232
                     /rpt_family="L1"
     repeat_region   12233. .12375
                     /rpt_family="L1"
     repeat_region   12376. .12681
                     /rpt_family="Alu"
     repeat_region   12682. .13076
                     /rpt_family="L1"
     repeat_region   13104. .13192
                     /rpt_family="CT-rich"
```

```
   repeat_region   13552. .13807
                   /rpt_family="Alu"
   repeat_region   13880. .14170
                   /rpt_family="MER1_type"
   repeat_region   14390. .14530
                   /rpt_family="L1"
   repeat_region   14722. .14956
                   /rpt_family="ERV1"
   repeat_region   15437. .15559
                   /rpt_family="MIR"
   repeat_region   15787. .15885
                   /rpt_family="MER103"
   repeat_region   16351. .16498
                   /rpt_family="MER2_type"
   repeat_region   16499. .16808
                   /rpt_family="Alu"
   repeat_region   16809. .17394
                   /rpt_family="MER2_type"
   repeat_region   17463. .17706
                   /rpt_family="MER2_type"
   repeat_region   18340. .18457
                   /rpt_family="(TA)n"
   misc_feature    18709. .18960
                   /note="match to EST AA430999 (NID:g2113719)"
   repeat_region   18848. .18891
                   /rpt_family="AT_rich"
   repeat_region   19127. .19149
                   /rpt_family="AT_rich"
   repeat_region   19922. .20308
                   /rpt_family="MER1_type?"
   repeat_region   20408. .20581
                   /rpt_family="MIR"
   repeat_region   20582. .20883
                   /rpt_family="Alu"
   repeat_region   20886. .20911
                   /rpt_family="AT_rich"
   repeat_region   20912. .20953
                   /rpt_family="MIR"
   repeat_region   21335. .22272
                   /rpt_family="L1"
   misc_feature    22134. .22137
                   /note="match to EST AA430999 (NID:g2113719)"
   repeat_region   22273. .22294
                   /rpt_family="AT_rich"
   repeat_region   22629. .23142
                   /rpt_family="MaLR"
   repeat_region   23520. .23663
                   /rpt_family="MIR"
   repeat_region   23787. .23885
                   /rpt_family="MIR"
   repeat_region   25035. .25055
                   /rpt_family="AT_rich"
   repeat_region   26251. .26342
                   /rpt_family="MER1_type"
   misc_feature    26649. .26721
                   /note="similar to EST BM918504 (NID:g19368883)"
   misc_feature    26652. .26711
                   /note="match to EST BM151775 (NID:g17175503)"
   repeat_region   27427. .27508

  Query Match             4.2%;  Score 254.8;  DB 9;  Length 112353;
  Best Local Similarity   69.4%;  Pred. No. 1.2e-45;
  Matches  487;  Conservative    0;  Mismatches  182;  Indels   33;  Gaps    9;

Qy        260 CTCAGCATATTCGTGCACCTTCATTTATCTCTCCTTCTCTCTCAAAGCTGCAGTAGGGGT 319
              | |  || | |||  ||| ||  || | | |          || ||      ||  |
Db      77290 CACTACAAAATCGACCACGTTTCTTGAACCCATTAGAGAACTAAAGTTGCAAGGTGTTAG 77349

Qy        320 GAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTCCTATAGTA 379
                | ||||||||| ||| |||  |||||||||||||| ||      |||||| || ||
Db      77350 ATGCAGGTGTGATACCTTTCCTC-ACCATCATAAGGGTCACAGTTGTCACTCCTGTAATA 77408

Qy        380 AAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTACATGACA 439
              |||||  || |||| ||||| ||| |||||||||||||||||||||||| |||       |
Db      77409 AAAGAGAGATTAACAAGAGAAAAGCCTAACAAATTTATTTAATCAAAGATTT-------A 77461

Qy        440 TGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTTTTTTGCTG 499
              |||||| |||||||||||||||||| |||| | ||||||||||    ||||| ||||
Db      77462 TGGGAGCCTTCAGAAATGAAGACCCAAGGACCAATGGGAAACTGTC-CATTTTTATGCTT 77520

Qy        500 AGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGATATGATCTA 559
              |||||| | ||||||| | |||         |    ||||  ||||||| |||||| |||
Db      77521 AGGTTCAACGAAGAATAAACAGCCGTGTAGGACAGTGATTGAACAAAAGAATATGAACTA 77580

Qy        560 GTGGTAAAGGACTCAGGGGGA----AACACAGCAAGGCCTGTCTATTCAGATTCTTCTTG 615
              |||||||  | ||   |||||     || ||| ||| ||||||| | |||||| ||  ||
Db      77581 GTGGTAACAGGCTTGAGGGGAGGAGGACCCAGTAAGTCCTGTCTGTCCAGATTTTTGCTG 77640

Qy        616 ATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTTCTTC 675
                |||||  | ||         |||||||||||||   ||||||  |||||   |  |||
Db      77641 GCTTCTCTATGTC--------ACATTCTTTCCTCCCAGGTATGGAGCAGGCTTTCTCTGA 77692

Qy        676 AATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGG-CCTTTTTAGATTTTATGGCTTGCTT 734
              ||||||| |||| |||| ||||| ||||| |   | ||||||| | |||| ||| || |||
Db      77693 AATGAGGATCTTTAAGGCAGAAGAGAGAAGGGAGACCTTTTTTGGTTTTGTGGTTTTCTT 77752

Qy        735 CGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCTGGTTTCTG 794
               ||| |||||||| ||||||||||||||||   ||||||||||||||||||||| ||||
Db      77753 TGGGAAAGAGGAGCTCTAGTTTCTATGACCTGCCTTGGGGAAGAGGAATTCTGGCTTCTA 77812

Qy        795 TGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGGTCAGAAAG 854
              |||||||  |      | || |||| | ||  ||| |||||||       || ||||||||||
Db      77813 TGACTTGATCAGGGGCAGGAAGAAGGGTTGGAGACAGGAGG---ATGGAAGGTCAGAAAG 77869

Qy        855 AGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAGTACTTCTTAGCATAC 914
              |          |||||||||| | |    ||| | ||||||| |||||      |||| ||
Db      77870 A-----CCTTGCTTCTGAGGCCCTTTAGTCTTCATTAGTTCAAAGTA---TGCAGCACAC 77921

Qy        915 CAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAACA 956
              |||  |   |||||||||  ||| || |||||| |||| |||
Db      77922 CAACACGTCATACTTTGGGGTATTGTCTTCTGAACTCTGACA 77963


RESULT 74
AP002738
LOCUS       AP002738               149700 bp    DNA     linear   HTG 08-JUL-2000
DEFINITION  Homo sapiens chromosome 4 clone 2069K23 map 4q22-q24, ***
            SEQUENCING IN PROGRESS ***, 17 unordered pieces.
ACCESSION   AP002738
VERSION     AP002738.1  GI:8980654
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 149700)
  AUTHORS   Tsai,S.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUL-2000) Shih-Feng Tsai, National Yang-Ming
            University, Institute of Genetics; 155 Li-Rong St. Section 2,
            Peitou, Taipei, Taiwan 11221, Republic of China
            (E-mail:ympetsai@ym.edu.tw, URL:http://genome.ym.edu.tw/,
            Tel:886-2-28267043, Fax:886-2-28264930)
COMMENT     gaps between the contigs are represented as 100 N.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 17 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     3416: contig of 3416 bp in length
            *    3417     3516: gap of 100 bp
            *    3517     4356: contig of 840 bp in length
            *    4357     4456: gap of 100 bp
```

```
            *     4457     8998: contig of 4542 bp in length
            *     8999     9098: gap of 100 bp
            *     9099    10692: contig of 1594 bp in length
            *    10693    10792: gap of 100 bp
            *    10793    16507: contig of 5715 bp in length
            *    16508    16607: gap of 100 bp
            *    16608    22129: contig of 5522 bp in length
            *    22130    22229: gap of 100 bp
            *    22230    28080: contig of 5851 bp in length
            *    28081    28180: gap of 100 bp
            *    28181    35293: contig of 7113 bp in length
            *    35294    35393: gap of 100 bp
            *    35394    40647: contig of 5254 bp in length
            *    40648    40747: gap of 100 bp
            *    40748    50480: contig of 9733 bp in length
            *    50481    50580: gap of 100 bp
            *    50581    61309: contig of 10729 bp in length
            *    61310    61409: gap of 100 bp
            *    61410    70771: contig of 9362 bp in length
            *    70772    70871: gap of 100 bp
            *    70872    80509: contig of 9638 bp in length
            *    80510    80609: gap of 100 bp
            *    80610    94180: contig of 13571 bp in length
            *    94181    94280: gap of 100 bp
            *    94281   103497: contig of 9217 bp in length
            *   103498   103597: gap of 100 bp
            *   103598   118885: contig of 15288 bp in length
            *   118886   118985: gap of 100 bp
            *   118986   149700: contig of 30715 bp in length.
FEATURES             Location/Qualifiers
     source          1. .149700
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /map="4q22-q24"
                     /clone="2069K23"
ORIGIN

  Query Match             4.2%;  Score 254.8;  DB 2;  Length 149700;
  Best Local Similarity   69.4%;  Pred. No. 1.2e-45;
  Matches  487;  Conservative    0;  Mismatches  182;  Indels   33;  Gaps    9;

Qy        260 CTCAGCATATTCGTGCACCTTCATTTATCTCTCCTTCTCTCTCAAAGCTGCAGTAGGGGT 319
              | |  || | |||  ||| ||  || | | |        || ||      ||  |
Db       1287 CACTACAAAATCGACCACGTTTCTTGAACCCATTAGAGAACTAAAGTTGCAAGGTGTTAG 1346

Qy        320 GAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTCCTATAGTA 379
              | |||||||||| ||| |||  |||||||||||||| ||       |||||| || ||
Db       1347 ATGCAGGTGTGATACCTTTCCTC-ACCATCATAAGGGTCACAGTTGTCACTCCTGTAATA 1405

Qy        380 AAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTACATGACA 439
              ||||| || |||| ||||| ||| |||||||||||||||||||||||| |||      |
Db       1406 AAAGAGAGATTAACAAGAGAAAAGCCTAACAAATTTATTTAATCAAAGATTT-------A 1458

Qy        440 TGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTTTTTTGCTG 499
              |||||| |||||||||||||||||||| |||| | ||||||||||   ||||| ||||
Db       1459 TGGGAGCCTTCAGAAATGAAGACCCAAGGACCAATGGGAAACTGTC-CATTTTTATGCTT 1517

Qy        500 AGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGATATGATCTA 559
              |||||| | ||||||| | |||         |    ||||  |||||| |||||| |||
Db       1518 AGGTTCAACGAAGAATAAACAGCCGTGTAGGACAGTGATTGAACAAAAGAATATGAACTA 1577

Qy        560 GTGGTAAAGGACTCAGGGGGA----AACACAGCAAGGCCTGTCTATTCAGATTCTTCTTG 615
              ||||||| | ||   |||||     || ||| ||| ||||||| | |||||| ||  ||
Db       1578 GTGGTAACAGGCTTGAGGGGAGGAGGACCCAGTAAGTCCTGTCTGTCCAGATTTTTGCTG 1637

Qy        616 ATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTTCTTC 675
                |||||  | ||         |||||||||||||   |||||| |||||  |  |||
Db       1638 GCTTCTCTATGTC--------ACATTCTTTCCTCCCAGGTATGGAGCAGGCTTTCTCTGA 1689

Qy        676 AATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGG-CCTTTTTAGATTTTATGGCTTGCTT 734
              ||||||| |||| |||| ||||| ||||| |   | ||||||| | |||| ||| || |||

Db       1690 AATGAGGATCTTTAAGGCAGAAGAGAGAAGGGAGACCTTTTTTGGTTTTGTGGTTTTCTT 1749

Qy        735 CGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCTGGTTTCTG 794
              ||| |||||||| |||||||||||||||    |||||||||||||||||||| ||||
Db       1750 TGGGAAAGAGGAGCTCTAGTTTCTATGACCTGCCTTGGGGAAGAGGAATTCTGGCTTCTA 1809

Qy        795 TGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGGTCAGAAAG 854
              ||||||| |      | || |||| | ||  ||| |||||||       || ||||||||||
Db       1810 TGACTTGATCAGGGGCAGGAAGAAGGGTTGGAGACAGGAGG---ATGGAAGGTCAGAAAG 1866

Qy        855 AGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAGTACTTCTTAGCATAC 914
              |          |||||||||||| | |    ||| | ||||||| |||||      |||| ||
Db       1867 A-----CCTTGCTTCTGAGGCCCTTTAGTCTTCATTAGTTCAAAGTA---TGCAGCACAC 1918

Qy        915 CAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAACA 956
              |||  |   |||||||||  ||| || |||||| |||| |||
Db       1919 CAACACGTCATACTTTGGGGTATTGTCTTCTGAACTCTGACA 1960


RESULT 75
AC093694
LOCUS       AC093694               158566 bp    DNA     linear   PRI 09-MAR-2002
DEFINITION  Homo sapiens BAC clone RP11-769N3 from 4, complete sequence.
ACCESSION   AC093694
VERSION     AC093694.4  GI:19071710
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 158566)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
   PUBMED   9847074
REFERENCE   2  (bases 1 to 158566)
  AUTHORS   Tomlinson,C., Cotton,M. and Creason,K.
  TITLE     The sequence of Homo sapiens BAC clone RP11-769N3
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 158566)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-SEP-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 158566)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-MAR-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 158566)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAR-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Mar 2, 2002 this sequence version replaced gi:17933878.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0769N03
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
```

all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at http://www.chori.org
VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-680I, 2000 bp overlap; the clone sequenced to the right is RP11-755F21. Actual start of this clone is at base position 148706 of RP11-680I; actual end is at base position 158566 of RP11-769N3.

There are single plasmid subclone regions from 82994 to 83124 and 107231 to 107292. Polymorphisms have been identified between AC093694 and AC021940.

FEATURES Location/Qualifiers

source 1. .158566
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone="RP11-769N3"
/clone_lib="RPCI-11"

repeat_region 1. .96
/rpt_family="L1"

repeat_region 97. .408
/rpt_family="Alu"

repeat_region 409. .1187
/rpt_family="L1"

repeat_region 1255. .1504
/rpt_family="Alu"

repeat_region 1538. .2365
/rpt_family="L1"

repeat_region 2394. .3233
/rpt_family="L1"

repeat_region 3235. .3677
/rpt_family="MaLR"

repeat_region 3685. .4638
/rpt_family="L1"

repeat_region 5746. .6221
/rpt_family="MaLR"

repeat_region 6222. .6537
/rpt_family="MER2_type"

repeat_region 6538. .6592
/rpt_family="MaLR"

repeat_region 6747. .6778
/rpt_family="AT_rich"

repeat_region 7388. .7425
/rpt_family="(TG)n"

repeat_region 8761. .8841
/rpt_family="U6"

repeat_region 8846. .8870
/rpt_family="AT_rich"

repeat_region 9538. .9591
/rpt_family="MIR"

repeat_region 9787. .9824
/rpt_family="(TA)n"

repeat_region 9825. .10119
/rpt_family="Alu"

repeat_region 10388. .10580
/rpt_family="MIR"

repeat_region 10685. .10901
/rpt_family="MER1_type"

repeat_region 11613. .11698
/rpt_family="MER103"

repeat_region 11959. .12423
/rpt_family="L1"

repeat_region 12424. .12508
/rpt_family="L1"

repeat_region 12509. .12553
/rpt_family="Mariner"

repeat_region 12556. .12928
/rpt_family="L1"

repeat_region 13142. .13402
/rpt_family="Alu"

repeat_region 13471. .13619
/rpt_family="L1"

repeat_region 13620. .13711
/rpt_family="L1"

repeat_region 13712. .14019
/rpt_family="Alu"

repeat_region 14020. .14312
/rpt_family="L1"

repeat_region 14366. .14487
/rpt_family="L1"

repeat_region 14779. .14799
/rpt_family="AT_rich"

repeat_region 14887. .15073
/rpt_family="MaLR"

repeat_region 15429. .15700
/rpt_family="MaLR"

repeat_region 16283. .16409
/rpt_family="L1"

repeat_region 16740. .16829
/rpt_family="MIR"

repeat_region 17004. .17077
/rpt_family="L2"

repeat_region 17349. .17534
/rpt_family="MIR"

repeat_region 17801. .17828
/rpt_family="(TTTA)n"

repeat_region 17836. .18121
/rpt_family="Alu"

repeat_region 19520. .19546
/rpt_family="(TTTG)n"

repeat_region 19847. .20159
/rpt_family="L1"

repeat_region 20158. .20217
/rpt_family="L1"

repeat_region 20218. .20496
/rpt_family="Alu"

repeat_region 20497. .21139
/rpt_family="L1"

repeat_region 21140. .21167
/rpt_family="AT_rich"

repeat_region 21386. .22273
/rpt_family="L1"

repeat_region 22770. .23017
/rpt_family="MIR"

repeat_region 29004. .29319
/rpt_family="L1"

repeat_region 30077. .30301
/rpt_family="Alu"

repeat_region 31379. .31726
/rpt_family="MaLR"

repeat_region 31922. .32070
/rpt_family="L2"

repeat_region 32334. .32449

```
                     /rpt_family="L2"
     repeat_region   32692. .32720

  Query Match             4.2%;  Score 254.8;  DB 9;  Length 158566;
  Best Local Similarity   69.4%;  Pred. No. 1.2e-45;
  Matches  487;  Conservative    0;  Mismatches  182;  Indels   33;  Gaps    9;

Qy        260 CTCAGCATATTCGTGCACCTTCATTTATCTCTCCTTCTCTCTCAAAGCTGCAGTAGGGGT 319
              | |  || | |||  ||| ||  || | | |        || ||      ||  |
Db     149245 CACTACAAAATCGACCACGTTTCTTGAACCCATTAGAGAACTAAAGTTGCAAGGTGTTAG 149304

Qy        320 GAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTCCTATAGTA 379
              | |||||||| ||| |||  |||||||||||||| ||      |||||| || ||
Db     149305 ATGCAGGTGTGATACCTTTCCTC-ACCATCATAAGGGTCACAGTTGTCACTCCTGTAATA 149363

Qy        380 AAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTACATGACA 439
              ||||| || |||| ||||| ||| |||||||||||||||||||||||| |||      |
Db     149364 AAAGAGAGATTAACAAGAGAAAAGCCTAACAAATTTATTTAATCAAAGATTT-------A 149416

Qy        440 TGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTTTTTTGCTG 499
              |||||| |||||||||||||||||||| |||| | ||||||||||   ||||| ||||
Db     149417 TGGGAGCCTTCAGAAATGAAGACCCAAGGACCAATGGGAAACTGTC-CATTTTTATGCTT 149475

Qy        500 AGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGATATGATCTA 559
              |||||| | ||||||| | |||        |    ||||  ||||||| |||||| |||
Db     149476 AGGTTCAACGAAGAATAAACAGCCGTGTAGGACAGTGATTGAACAAAAGAATATGAACTA 149535

Qy        560 GTGGTAAAGGACTCAGGGGGA----AACACAGCAAGGCCTGTCTATTCAGATTCTTCTTG 615
              |||||||  | ||   |||||     || ||| ||| ||||||| | |||||| ||  ||
Db     149536 GTGGTAACAGGCTTGAGGGGAGGAGGACCCAGTAAGTCCTGTCTGTCCAGATTTTTGCTG 149595

Qy        616 ATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTTCTTC 675
                |||||  | ||         ||||||||||||   |||||| |||||  |  |||
Db     149596 GCTTCTCTATGTC--------ACATTCTTTCCTCCCAGGTATGGAGCAGGCTTTCTCTGA 149647

Qy        676 AATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGG-CCTTTTTAGATTTTATGGCTTGCTT 734
              ||||||| |||| |||| ||||| ||||| |   | ||||||| | |||| ||| || |||
Db     149648 AATGAGGATCTTTAAGGCAGAAGAGAGAAGGGAGACCTTTTTTGGTTTTGTGGTTTTCTT 149707

Qy        735 CGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCTGGTTTCTG 794
              ||| |||||||| ||||||||||||||||   |||||||||||||||||||| ||||
Db     149708 TGGGAAAGAGGAGCTCTAGTTTCTATGACCTGCCTTGGGGAAGAGGAATTCTGGCTTCTA 149767

Qy        795 TGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGGTCAGAAAG 854
              ||||||| |     | || |||| | ||  ||| |||||||      || ||||||||||
Db     149768 TGACTTGATCAGGGGCAGGAAGAAGGGTTGGAGACAGGAGG---ATGGAAGGTCAGAAAG 149824

Qy        855 AGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAGTACTTCTTAGCATAC 914
              |        |||||||||| | |     ||| | ||||||| |||||      |||| ||
Db     149825 A-----CCTTGCTTCTGAGGCCCTTTAGTCTTCATTAGTTCAAAGTA---TGCAGCACAC 149876

Qy        915 CAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAACA 956
              |||  |   |||||||||  ||| || |||||| |||| |||
Db     149877 CAACACGTCATACTTTGGGGTATTGTCTTCTGAACTCTGACA 149918


RESULT 76
BS000174/c
LOCUS       BS000174              193756 bp    DNA     linear   PRI 12-JUN-2004
DEFINITION  Pan troglodytes chromosome 22 clone:PTB-137B16, map 22, complete
            sequences.
ACCESSION   BS000174 BA000046
VERSION     BS000174.1  GI:37537441
KEYWORDS    HTG.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   The International Chimpanzee Chromosome 22 Consortium.
  TITLE     DNA sequence and comparative analysis of chimpanzee chromosome 22
  JOURNAL   Nature 429, 382-388 (2004)
REFERENCE   2  (bases 1 to 193756)
  AUTHORS   Wang,S., Cai,Z., Wang,B., Zheng,H., Zhang,Y., Zhang,X., Zhu,G.,
            Lu,G., Fu,G. and Chen,Z.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-MAY-2003) Shengyue Wang, Chinese National Human
            Genome Center at Shanghai, Genomic Sequencing; No.250 BiBo Road,
            Zhang Jiang HI-TECH Park, Shanghai 201203, CHINA
            (E-mail:wangsy@chgc.sh.cn, URL:http://www.chgc.sh.cn,
            Tel:86-21-50801919, Fax:86-21-50801922)
COMMENT     The Chimpanzee Chromosome 22 Sequencing Consortium consists of:
            *Chinese National Human Genome Center at Shanghai, Shanghai, China;
            *GBF, Dept. of Genome Analysis, Braunschweig, Germany; *Institute
            of Molecular Biotechnology, Jena, Germany; *KRIBB Genome Research
            Center, Daejeon, Korea;
            *Max-Planck-Institute for Molecular Genetics, Berlin, Germany;
            *National Institute of Genetics, Mishima, Japan;
            *National Yang Ming University Genome Research Center, Taipei,
            Taiwan;
            *RIKEN Genomic Sciences Center, Yokohama, Japan.
            -------------- Genome Center
            Center: Chinese National Human Genome Center at Shanghai Center
            code: CHGCS
            Web site: http://chgc.sh.cn
            Contact: wangsy@chgc.sh.cn
            -------------- Project Information
            Center project name:The Chimpanzee Chromosome 22 Sequencing Project
            Center clone name: PTB-137B16
            -------------- Summary Statistics
            Sequencing vector: pUC18,100% of reads
            Chemistry: Dye-terminator Big Dye and ET; 100% of reads Assembly
            program: Phrap; version 0.990329
            Consensus quality:   193668 bases at least Q40
            Consensus quality:   193750 bases at least Q30
            Consensus quality:   193756 bases at least Q20
            Quality coverage: 10.1x
            --------------
            This sequence was finished as follows unless otherwise noted: all
            regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30);
            an attempt was made to resolve all sequencing problems, such as
            compressions and repeats; all regions were covered by at one
            plasmid
            subclone or more than one M13 subclone;
            and the assembly was confirmed by restriction digest.
            --------------
            Source information:
            The PTB1 chimpanzee BAC library was prepared from DNA isolated from
            cultured cells established from the blood of a single male
            chimpanzee.
            Clones may be obtained from Asao Fujiyama and co-workers
            (http://www.gsc.riken.go.jp).
            VECTOR:  pKS145
            Sequence Quality Assessment:
            This entry has been annotated with sequence
            estimates computed by the Phrap assembly program.
            All manually edited bases have been reduced to quality zero.
            Quality levels above 40 are expected to have less than 1 error in
            10,000 bp.
            ----------
            Neighboring clones: PTB-258K22(left) and PTB-091H17(right).
FEATURES              Location/Qualifiers
     source           1. .193756
                      /organism="Pan troglodytes"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:9598"
                      /chromosome="22"
                      /clone="PTB-137B16"
                      /clone_lib="PTB1 chimpanzee BAC"
ORIGIN

  Query Match             4.2%;  Score 254.8;  DB 9;  Length 193756;
  Best Local Similarity   71.7%;  Pred. No. 1.2e-45;
  Matches  452;  Conservative    0;  Mismatches  152;  Indels   26;  Gaps    8;
```

```
Qy        330 GATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTCCTATAGTAAAAGACAGGT 389
              ||||| ||| |||  |||||||||   || | ||  | | |||||||  ||||||||||
Db     156107 GATACCTTTCCTCACCCATCATAACAATCACCACTGACATTCCTATAACAAAAGACAGGT 156048

Qy        390 TAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTACATGACATGGGAGTCTT 449
              ||| ||||| ||| | || ||||||| ||||| ||||||||||||||||   |||| |||
Db     156047 TAACAAGAGAAAATCATAGCAAATTTGTTTAACCAAAGTTTTACATGACTCAGGAGGCTT 155988

Qy        450 CAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTTTTTTGCTGAGGTTCGATG 509
              |||||||||| |||||||| |||| |||||||||  ||| | |||| || ||| |||
Db     155987 CAGAAATGAAGA-CCAAAGACCTAGGGAAAACTGTCT-ATTTCTATGCTTAGATTCCATG 155930

Qy        510 AAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGATATGATCTAGTGGTAAAGG 569
              ||||||| | | ||||| |  |||| |||| ||||||||| ||  |||     |  | |||
Db     155929 AAGAATGAACAACATGTGGAAATGT-GATTGGACAAAAGGGTAGAATCAGTGGTAATAGG 155871

Qy        570 ACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTTCTTGATCTCTCTCTCTCT 629
                |||| || | |||||||||||||| | ||| ||||||||||   |||||
Db     155870 CTGAGGGGGAAAGCCCAGCAAGGCCTGTGTGTTCGGATTCTTCTTAGCCTCTC------- 155818

Qy        630 ATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTTCTTCAATGAGGGTCTTCA 689
               |||  ||  ||| |||||||||| ||||| |||| | | |  | |||| |||||||
Db     155817 -TGTGCAGTGTTCCTTCCTCCTGAGCATGGGTCAGGGCCCCTATGAAGTGAGAGTCTTCA 155759

Qy        690 AGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATGGCTTGCTTCGGGGAAGAGGAGTT 749
              |||||  ||| ||          |||| ||||||||||  |||||||  ||| |||||||||
Db     155758 AGGGAAAAGAGAAT-------CTTTCTAGATTTTATATCTTGCTTTTGGGGAGAGGAGTT 155706

Qy        750 CTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCTGGTTTCTGTGACTTGCTTTCATG 809
              ||||||||||||  ||| |||||||||| | ||||||  |||||  |||||  | |   ||
Db     155705 CTAGTTTCTATGTTCCA-CTTGGGGAAGTGAAATTCTACTTTCTATGACTCACGTCAGTG 155647

Qy        810 AAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGGTCAGAAAGAGACTTGGCTGCTTC 869
               |  |||||| ||| ||   | || || |||||||  ||||||||||        || |||
Db     155646 GATTAAGAGGGGTAGGATACCGGGGGTCAGGAGAAAGTCAGAAAGA-----CCTTGTTTC 155592

Qy        870 TGAGGGCTTCCGCTCTCCTTTAGTTCCAAGTACTTCTTAGCATACCAAAGCACTATACTT 929
              ||||  ||| |        |||   | |||  | |  | ||||| |||| |  | ||| ||
Db     155591 TGAGAACTTGCCAGTCTTTTTCAGTTCAA--AGTAGTCAGCATGCCAAGGTGCCATATTT 155534

Qy        930 TGGCATATGGTTTTCTGAGCTCTAACACTG 959
              |||  | |  | |||||||| | |||| ||
Db     155533 TGGGGTGTCATGTTCTGAGCCCCAACAGTG 155504
```

```
RESULT 77
AC141282/c
LOCUS       AC141282               138007 bp    DNA     linear   HTG 11-MAR-2003
DEFINITION  Homo sapiens chromosome 16 clone RP11-347C20, WORKING DRAFT
            SEQUENCE.
ACCESSION   AC141282
VERSION     AC141282.1  GI:28913062
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 138007)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 138007)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-MAR-2003) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 539681
            Center clone name: RPCI-11_347C20
            -------
            Summary Statistics
            Consensus quality: 138007 bases at least Q40
            Consensus quality: 138007 bases at least Q30
            Consensus quality: 138007 bases at least Q20
            Estimated insert size: 175000; agarose-fp estimation
            Estimated insert size: 138007; sum-of-contigs estimation
            Quality coverage:  14.06 in Q20 bases; agarose-fp estimation
            Quality coverage:  17.82 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   138007: contig of 138007 bp in length.
FEATURES             Location/Qualifiers
     source          1. .138007
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone="RP11-347C20"
                     /clone_lib="RPCI human BAC library 11"
ORIGIN
```

```
  Query Match             4.2%;  Score 254.6;  DB 2;  Length 138007;
  Best Local Similarity   71.3%;  Pred. No. 1.3e-45;
  Matches  471;  Conservative    0;  Mismatches  159;  Indels   31;  Gaps    9;

Qy        304 AAGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAAC 363
              | |  || |||||| ||  || ||||||  | ||| ||  ||||||  |||||||  |  |
Db      30025 ATGGGGCTGTAGGGCTGTGAAGGTGTGAGTCCTTTACTTATCCATCTCAAGGGTCACGGC 29966

Qy        364 CAAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATC 423
              ||| |||||||||  ||||||||  |||| |||||  |||||  | ||||||||||||||
Db      29965 CAACACTCCTATAACAAAAGACATATTAACAAGAGAAAAACAGAGCAAATTTATTTAATC 29906

Qy        424 AAAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTG 483
              ||||||||  |||||||||||| ||||||||||||  ||||||||  |||||||  ||||||
Db      29905 CAAGTTTTATGTGACATGGGAGGCTTCAGAAATGACAACCCAAAGGCCCAGGGAAAACTG 29846

Qy        484 TCTGTTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGAC 543
              | |  ||||| |||| || || ||||||||||||| |||  |      |    |||| |||
Db      29845 T-TAATTTTTATGCTTAGATTAGATGAAGAATGGACAGCCAGATAGAAATGTGATTGGAC 29787

Qy        544 AAAAGGATATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTC 603
              ||| || ||||| ||| ||| ||  |||| |  ||| | |      ||   ||||||  ||
Db      29786 AAACGGGTATGACCTACTGGGAACAGACTGAAAGGGGACCCAGCAAA--ACTGTCTGCTC 29729

Qy        604 AGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCA 663
              |||||||         ||||   |||| ||| |||| | | |||||||| ||||||||||
Db      29728 AGATTCT--------TCTCAGCCTCTCTGTGTAGCCTCCATTCCTCCTGGGTATGGGGCA 29677

Qy        664 GGACTC-TTCTTCAATGAGGGTCTTCAAGGGAGAAGGG------AGAAAGTGGCCTTTTT 716
              |||| | ||||  |||||| || ||||| ||||||||       ||| |||| ||||| |
Db      29676 GGACCCTTTCTGAAATGAGAGTTTTCAAAGGAGAAGGAGAAGATAGAGAGTGACCTTTCT 29617

Qy        717 AGATTTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAA 776
              || |||||||| ||||||   | | ||||  |||||| |||||||||||||||| ||||
Db      29616 AGGTTTTATGGTTTGCTT-TAGCAGGAGGGTTTCTAGGTTCTATGACCCATCTTGAGGAA 29558

Qy        777 GAGGAATTCTGGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGG 836
              ||  |||||||||||||||||  ||||   | |  |||   || |  ||  |||
Db      29557 GATAAATTCTGGTTTCTGTGACTCACTTTGGAGGATGAAGGCAAGCAGAAGAGAGGGTAT 29498

Qy        837 CAGGAGATGGTCAGAAAGAGACTTGGCTGCTTCTGAG-GGCTTCCGCTCTCCTTTAGTTC 895
              ||||||  |||  ||||||         ||||||||||   |||||  | |||||  |||||
Db      29497 GAGGAGAAGGTTGGAAAGA-----CCTTGCTTCTGAGTTCCTTCCAATGTCCTTCAGTTC 29443
```

```
Qy        896 CAAGTACTTCTTAGCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAAC 955
              || ||||      ||  |  | |||| || ||||||  | |  | |||| ||| |  ||
Db      29442 AAAATACT------CAGGCTGAGGCACCATGCTTTGGGGTGTCATGTTCTAAGCCCCGAC 29389

Qy        956 A 956
              |
Db      29388 A 29388
```

RESULT 78
AC006453/c
LOCUS AC006453 155313 bp DNA linear PRI 26-JUN-2002
DEFINITION Homo sapiens PAC clone RP4-614C10 from 2, complete sequence.
ACCESSION AC006453
VERSION AC006453.3 GI:4753285
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 155313)
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
PUBMED 9847074
REFERENCE 2 (bases 1 to 155313)
AUTHORS Ali,J., Wohldmann,P., Ames,M. and Le,T.
TITLE The sequence of Homo sapiens PAC clone RP4-614C10
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 155313)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (28-JAN-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 155313)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (05-MAY-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 5 (bases 1 to 155313)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (26-JUN-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On May 5, 1999 this sequence version replaced gi:4337294.
---------- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
---------- Summary Statistics
Center project name: H_DJ0614C10
----------.

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 2 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see http://www.nhgri.nih.gov/DIR/GTB/CHR2 , send mailto:egreen@nhgri.nih.gov , or see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
This clone was derived from human PAC library RPCI-4, prepared by Pieter de Jong and coworkers at http://www.chori.org using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994).
The library is from one male donor.
The clone may be obtained either from Genome Systems, Inc. (http://www.genomesystems.com) or Research Genetics, Inc. (http://www.resgen.com); or from Pieter de Jong.
VECTOR: pCYPAC2

NEIGHBORING SEQUENCE INFORMATION:
Actual start of this clone is at base position 1 of RP4-614C10; actual end is at base position 155313 of RP4-614C10.

FEATURES	Location/Qualifiers
source	1. .155313
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="2"
	/map="2"
	/clone="RP4-614C10"
	/clone_lib="RPCI-4"
repeat_region	1. .38
	/rpt_family="MaLR"
repeat_region	39. .337
	/rpt_family="Alu"
repeat_region	338. .563
	/rpt_family="MaLR"
repeat_region	1066. .1302
	/rpt_family="MIR"
repeat_region	2094. .2392
	/rpt_family="L1"
repeat_region	3283. .3359
	/rpt_family="MIR"
repeat_region	3591. .3677
	/rpt_family="MIR"
repeat_region	3678. .3740
	/rpt_family="(TA)n"
repeat_region	3746. .4041
	/rpt_family="Alu"
repeat_region	4042. .4169
	/rpt_family="MIR"
repeat_region	4335. .4643
	/rpt_family="Alu"
repeat_region	4698. .4991
	/rpt_family="Alu"
repeat_region	5490. .5788
	/rpt_family="Alu"
repeat_region	5918. .5966
	/rpt_family="(GA)n"
repeat_region	5966. .5993
	/rpt_family="(CA)n"
repeat_region	6006. .6304
	/rpt_family="Alu"
repeat_region	6618. .6644
	/rpt_family="AT_rich"
repeat_region	6807. .6840
	/rpt_family="(TAGA)n"
repeat_region	8099. .8471
	/rpt_family="L1"
repeat_region	8472. .8651
	/rpt_family="Alu"
repeat_region	8652. .8943
	/rpt_family="Alu"
repeat_region	8944. .9084

```
                /rpt_family="Alu"
     repeat_region   9085. .9131
                /rpt_family="L1"
     repeat_region   9225. .9464
                /rpt_family="L2"
     repeat_region   9467. .9791
                /rpt_family="MaLR"
     repeat_region   9880. .10074
                /rpt_family="L2"
     repeat_region   10805. .10871
                /rpt_family="L2"
     repeat_region   11017. .11037
                /rpt_family="AT_rich"
     repeat_region   11038. .11324
                /rpt_family="Alu"
     repeat_region   11325. .11342
                /rpt_family="AT_rich"
     repeat_region   11881. .11939
                /rpt_family="L2"
     repeat_region   12173. .12490
                /rpt_family="L1"
     repeat_region   12821. .13222
                /rpt_family="MaLR"
     repeat_region   13264. .13329
                /rpt_family="purine-rich"
     repeat_region   13330. .14448
                /rpt_family="L2"
     repeat_region   14480. .14723
                /rpt_family="MER2_type"
     repeat_region   14724. .15038
                /rpt_family="Alu"
     repeat_region   15039. .15111
                /rpt_family="MER2_type"
     repeat_region   15158. .15721
                /rpt_family="MaLR"
     repeat_region   15899. .15977
                /rpt_family="MIR"
     repeat_region   17577. .17605
                /rpt_family="AT_rich"
     repeat_region   17801. .17862
                /rpt_family="MER1_type"
     repeat_region   17863. .17895
                /rpt_family="AT_rich"
     repeat_region   17903. .18115
                /rpt_family="L1"
     repeat_region   18141. .18222
                /rpt_family="MER1_type"
     repeat_region   18283. .18594
                /rpt_family="Alu"
     repeat_region   18595. .18610
                /rpt_family="AT_rich"
     repeat_region   18611. .18910
                /rpt_family="Alu"
     repeat_region   18911. .18925
                /rpt_family="AT_rich"
     repeat_region   19036. .19169
                /rpt_family="MER1_type"
     repeat_region   19921. .20412
                /rpt_family="L1"
     repeat_region   20620. .20664
                /rpt_family="AT_rich"
     repeat_region   20981. .21345
                /rpt_family="MaLR"
     repeat_region   21508. .21590
                /rpt_family="MER1_type"
     repeat_region   22790. .23079

Query Match             4.2%;  Score 254.6;  DB 9;  Length 155313;
Best Local Similarity   71.3%;  Pred. No. 1.3e-45;
Matches  471;  Conservative    0;  Mismatches  159;  Indels   31;  Gaps    9;

Qy        304 AAGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAAC 363
              | |  || |||||| ||  || ||||||  | ||| ||  ||||||  ||||||| |  |

Db      28366 ATGGGGCTGTAGGGCTGTGAAGGTGTGAGTCCTTTACTTATCCATCTCAAGGGTCACGGC 28307

Qy        364 CAAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATC 423
              ||| |||||||||  |||||||  |||| ||||| |||||  | ||||||||||||||
Db      28306 CAACACTCCTATAACAAAAGACATATTAACAAGAGAAAAACAGAGCAAATTTATTTAATC 28247

Qy        424 AAAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTG 483
              ||||||||  |||||||||||| |||||||||||  ||||||||  |||||||  |||||
Db      28246 CAAGTTTTATGTGACATGGGAGGCTTCAGAAATGACAACCCAAAGGCCCAGGGAAAACTG 28187

Qy        484 TCTGTTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGAC 543
              | |  ||||| |||| || || |||||||||||||| |||  |      |    |||| |||
Db      28186 T-TAATTTTTATGCTTAGATTAGATGAAGAATGGACAGCCAGATAGAAATGTGATTGGAC 28128

Qy        544 AAAAGGATATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTC 603
              ||| || ||||| ||| ||| ||  |||| |  ||| | |      ||   ||||||  ||
Db      28127 AAACGGGTATGACCTACTGGGAACAGACTGAAAGGGGACCCAGCAAA--ACTGTCTGCTC 28070

Qy        604 AGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCA 663
              |||||||        ||||   |||| ||| |||| | | |||||||||| ||||||||||
Db      28069 AGATTCT--------TCTCAGCCTCTCTGTGTAGCCTCCATTCCTCCTGGGTATGGGGCA 28018

Qy        664 GGACTC-TTCTTCAATGAGGGTCTTCAAGGGAGAAGGG------AGAAAGTGGCCTTTTT 716
              |||| | ||||  |||||| || ||||| ||||||||        ||| |||| ||||| |
Db      28017 GGACCCTTTCTGAAATGAGAGTTTTCAAAGGAGAAGGAGAAGATAGAGAGTGACCTTTCT 27958

Qy        717 AGATTTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAA 776
              || |||||||| ||||||   | | ||||  |||||| ||||||||||||||||| ||||
Db      27957 AGGTTTTATGGTTTGCTT-TAGCAGGAGGGTTTCTAGGTTCTATGACCCATCTTGAGGAA 27899

Qy        777 GAGGAATTCTGGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGG 836
              ||  |||||||||||||||||  ||||   | |  |||   || |  ||  |||
Db      27898 GATAAATTCTGGTTTCTGTGACTCACTTTGGAGGATGAAGGCAAGCAGAAGAGAGGGTAT 27839

Qy        837 CAGGAGATGGTCAGAAAGAGACTTGGCTGCTTCTGAG-GGCTTCCGCTCTCCTTTAGTTC 895
              |||||| |||  ||||||         ||||||||||   |||||  | ||||| |||||
Db      27838 GAGGAGAAGGTTGGAAAGA-----CCTTGCTTCTGAGTTCCTTCCAATGTCCTTCAGTTC 27784

Qy        896 CAAGTACTTCTTAGCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAAC 955
              || ||||      ||  |  | |||| || ||||||  | |  | |||| ||| |  ||
Db      27783 AAAATACT------CAGGCTGAGGCACCATGCTTTGGGGTGTCATGTTCTAAGCCCCGAC 27730

Qy        956 A 956
              |
Db      27729 A 27729
```

```
RESULT 79
AC116050/c
LOCUS       AC116050              165225 bp    DNA     linear   PRI 03-JUL-2002
DEFINITION  Homo sapiens BAC clone RP11-427F2 from 2, complete sequence.
ACCESSION   AC116050
VERSION     AC116050.3  GI:21686938
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 165225)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
   PUBMED   9847074
REFERENCE   2  (bases 1 to 165225)
  AUTHORS   Belter,E., Kozlowicz,A. and Doebber,A.
  TITLE     The sequence of Homo sapiens BAC clone RP11-427F2
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 165225)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-MAR-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
```

```
            MO 63108, USA
REFERENCE   4  (bases 1 to 165225)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-APR-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 165225)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUL-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Jul 3, 2002 this sequence version replaced gi:20153316.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0427F02
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. John D.
            McPherson, Department of Genetics, Washington University, St. Louis
            MO. For additional information about the map position of this
            sequence, see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries. Genomics 51:1-8. The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at http://www.chori.org
            VECTOR:  pBACe3.6

            NEIGHBORING SEQUENCE INFORMATION:
            Actual start of this clone is at base position 1 of RP11-427F2;
            actual end is at base position 165225 of RP11-427F2.

            Single plasmid regions exist between 83459 and 83487; 88788 and
            88803; 88857 and 88870.
FEATURES             Location/Qualifiers
     source          1. .165225
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2"
                     /clone="RP11-427F2"
                     /clone_lib="RPCI-11"
     repeat_region   507. .654
                     /rpt_family="L1"
     repeat_region   789. .1092
                     /rpt_family="Alu"
     repeat_region   1202. .1638
                     /rpt_family="L1"
     repeat_region   1633. .1792
                     /rpt_family="L1"
     repeat_region   1793. .1951
                     /rpt_family="L1"
     repeat_region   1953. .2108
                     /rpt_family="L1"
     repeat_region   2112. .2348
                     /rpt_family="L1"
     repeat_region   2349. .2625
                     /rpt_family="Alu"
     repeat_region   2626. .2716
                     /rpt_family="L1"
     repeat_region   2722. .2993
                     /rpt_family="L1"
     repeat_region   2994. .3437
                     /rpt_family="L1"
     repeat_region   3438. .3496
                     /rpt_family="L1"
     repeat_region   3497. .3521
                     /rpt_family="(A)n"
     repeat_region   3522. .3595
                     /rpt_family="L1"
     repeat_region   3596. .3631
                     /rpt_family="L1"
     repeat_region   3636. .3923
                     /rpt_family="Alu"
     repeat_region   3985. .4008
                     /rpt_family="(CAAAA)n"
     repeat_region   4431. .4572
                     /rpt_family="Alu"
     repeat_region   4573. .4883
                     /rpt_family="Alu"
     repeat_region   4884. .5059
                     /rpt_family="Alu"
     repeat_region   5144. .5247
                     /rpt_family="Alu"
     repeat_region   5249. .6067
                     /rpt_family="ERVL"
     repeat_region   6071. .6175
                     /rpt_family="(CCAA)n"
     repeat_region   6178. .6364
                     /rpt_family="ERVL"
     repeat_region   6365. .6393
                     /rpt_family="(TG)n"
     repeat_region   6406. .7330
                     /rpt_family="ERV1"
     repeat_region   7343. .7441
                     /rpt_family="ERVL"
     repeat_region   7476. .7577
                     /rpt_family="ERVL"
     repeat_region   7894. .8016
                     /rpt_family="MIR"
     repeat_region   8805. .8934
                     /rpt_family="MIR"
     repeat_region   9228. .9469
                     /rpt_family="MaLR"
     repeat_region   9711. .9765
                     /rpt_family="(TA)n"
     repeat_region   11206. .11365
                     /rpt_family="MIR"
     repeat_region   12145. .12440
                     /rpt_family="Alu"
     repeat_region   13235. .13356
                     /rpt_family="ERV1"
     repeat_region   13357. .13682
                     /rpt_family="ERV1"
     repeat_region   14428. .15177
                     /rpt_family="L1"
     repeat_region   15937. .16031
                     /rpt_family="L2"
     repeat_region   17538. .17589
                     /rpt_family="AT_rich"
     repeat_region   18059. .18364
                     /rpt_family="Alu"
```

```
    repeat_region   18737. .19031
                    /rpt_family="Alu"
    repeat_region   19199. .19268
                    /rpt_family="MIR"
    repeat_region   19583. .20241
                    /rpt_family="L1"
    repeat_region   20244. .20467
                    /rpt_family="L1"
    repeat_region   20468. .20777
                    /rpt_family="Alu"
    repeat_region   20778. .21285
                    /rpt_family="L1"
    repeat_region   21928. .21965
                    /rpt_family="AT_rich"
    repeat_region   22020. .22184
                    /rpt_family="(CATTT)n"
    repeat_region   22189. .22858
                    /rpt_family="L1"
    repeat_region   22874. .23110
                    /rpt_family="(CATCT)n"
    repeat_region   23111. .24484
                    /rpt_family="L1"
    repeat_region   24485. .24532
                    /rpt_family="(CATCT)n"
    repeat_region   24536. .24700
                    /rpt_family="(CATTT)n"
    repeat_region   24732. .24864
                    /rpt_family="(CATCT)n"
    repeat_region   24955. .24995
                    /rpt_family="AT_rich"

  Query Match             4.2%;  Score 254.6;  DB 9;  Length 165225;
  Best Local Similarity   71.3%;  Pred. No. 1.3e-45;
  Matches  471;  Conservative    0;  Mismatches  159;  Indels   31;  Gaps    9;

Qy        304 AAGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAAC 363
              | |  || |||||| ||  || ||||||  | ||| ||  ||||||  ||||||| |  |
Db     105905 ATGGGGCTGTAGGGCTGTGAAGGTGTGAGTCCTTTACTTATCCATCTCAAGGGTCACGGC 105846

Qy        364 CAAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATC 423
              ||| |||||||||  ||||||||  |||| |||||  |||||   | ||||||||||||||
Db     105845 CAACACTCCTATAACAAAAGACATATTAACAAGAGAAAAACAGAGCAAATTTATTTAATC 105786

Qy        424 AAAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTG 483
              ||||||||  |||||||||||  ||||||||||||  |||||||| |||||| ||||||
Db     105785 CAAGTTTTATGTGACATGGGAGGCTTCAGAAATGACAACCCAAAGGCCCAGGGAAAACTG 105726

Qy        484 TCTGTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGAC 543
              | |  ||||| |||| || || ||||||||||||| |||  |     |    |||| |||
Db     105725 T-TAATTTTTATGCTTAGATTAGATGAAGAATGGACAGCTAGATAGAAATGTGATTGGAC 105667

Qy        544 AAAAGGATATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTC 603
              ||| || ||||| ||| ||| ||  |||| |  ||| | |      ||   ||||||  ||
Db     105666 AAACGGGTATGACCTACTGGGAACAGACTGAAAGGGGACCCAGCAAA--ACTGTCTGCTC 105609

Qy        604 AGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCA 663
              |||||||        ||||   |||| ||| |||| | | |||||||| ||||||||||
Db     105608 AGATTCT--------TCTCAGCCTCTCTGTGTAGCCTCCATTCCTCCTGGGTATGGGGCA 105557

Qy        664 GGACTC-TTCTTCAATGAGGGTCTTCAAGGGAGAAGGG------AGAAAGTGGCCTTTTT 716
              |||| | ||||  |||||| || ||||| ||||||||       ||| |||| ||||| |
Db     105556 GGACCCTTTCTGAAATGAGAGTTTTCAAAGGAGAAGGAGAAGATAGAGAGTGACCTTTCT 105497

Qy        717 AGATTTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAA 776
              || |||||||| ||||||   | | ||||  |||||| ||||||||||||||||| ||||
Db     105496 AGGTTTTATGGTTTGCTT-TAGCAGGAGGGTTTCTAGGTTCTATGACCCATCTTGAGGAA 105438

Qy        777 GAGGAATTCTGGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGG 836
              ||  |||||||||||||||||  ||||   | |  |||   || |  ||  |||
Db     105437 GATAAATTCTGGTTTCTGTGACTCACTTTGGAGGATGAAGGCAAGCAGAAGAGAGGGTAT 105378

Qy        837 CAGGAGATGGTCAGAAAGAGACTTGGCTGCTTCTGAG-GGCTTCCGCTCTCCTTTAGTTC 895
              |||||| |||  ||||||       ||||||||||   |||||  | ||||| |||||
```

```
Db     105377 GAGGAGAAGGTTGGAAAGA-----CCTTGCTTCTGAGTTCCTTCCAATGTCCTTCAGTTC 105323

Qy        896 CAAGTACTTCTTAGCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAAC 955
              || ||||      ||  |  | |||| || ||||||  | |  | |||| ||| |  ||
Db     105322 AAAATACT------CAGGCTGAGGCACCATGCTTTGGGGTGTCATGTTCTAAGCCCCGAC 105269

Qy        956 A 956
              |
Db     105268 A 105268


RESULT 80
AC141297/c
LOCUS       AC141297              173505 bp    DNA     linear   HTG 11-MAR-2003
DEFINITION  Homo sapiens chromosome 16 clone RP11-677C7, WORKING DRAFT
            SEQUENCE, 9 unordered pieces.
ACCESSION   AC141297
VERSION     AC141297.1  GI:28913077
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 173505)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 173505)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-MAR-2003) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 1554527
            Center clone name: RPCI-11_677C7
            -------
            Summary Statistics
            Consensus quality: 159377 bases at least Q40
            Consensus quality: 165628 bases at least Q30
            Consensus quality: 169187 bases at least Q20
            Estimated insert size: 175000; agarose-fp estimation
            Estimated insert size: 172705; sum-of-contigs estimation
            Quality coverage:  5.32 in Q20 bases; agarose-fp estimation
            Quality coverage:  5.39 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     2517: contig of 2517 bp in length
            *    2518     2617: gap of unknown length
            *    2618     6174: contig of 3557 bp in length
            *    6175     6274: gap of unknown length
            *    6275    17421: contig of 11147 bp in length
            *   17422    17521: gap of unknown length
            *   17522    41363: contig of 23842 bp in length
            *   41364    41463: gap of unknown length
            *   41464    62902: contig of 21439 bp in length
            *   62903    63002: gap of unknown length
            *   63003    93375: contig of 30373 bp in length
            *   93376    93475: gap of unknown length
            *   93476   119278: contig of 25803 bp in length
            *  119279   119378: gap of unknown length
            *  119379   145421: contig of 26043 bp in length
            *  145422   145521: gap of unknown length
```

```
          *   145522   173505: contig of 27984 bp in length.
FEATURES             Location/Qualifiers
     source          1. .173505
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone="RP11-677C7"
                     /clone_lib="RPCI human BAC library 11"
ORIGIN

  Query Match             4.2%;  Score 254.6;  DB 2;  Length 173505;
  Best Local Similarity   71.3%;  Pred. No. 1.3e-45;
  Matches  471;  Conservative    0;  Mismatches  159;  Indels   31;  Gaps    9;

Qy        304 AAGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAAC 363
              | |  || |||||| ||  || ||||||  | ||| ||  ||||||  ||||||| |  |
Db     154972 ATGGGGCTGTAGGGCTGTGAAGGTGTGAGTCCTTTACTTATCCATCTCAAGGGTCACGGC 154913

Qy        364 CAAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATC 423
              ||| |||||||||  ||||||||  |||| ||||| |||||  | ||||||||||||||
Db     154912 CAACACTCCTATAACAAAAGACATATTAACAAGAGAAAAACAGAGCAAATTTATTTAATC 154853

Qy        424 AAAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTG 483
              ||||||||  ||||||||||| |||||||||||  ||||||||  |||||||  ||||||
Db     154852 CAAGTTTTATGTGACATGGGAGGCTTCAGAAATGACAACCCAAAGGCCCAGGGAAAACTG 154793

Qy        484 TCTGTTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGAC 543
              | |  ||||| |||| || || ||||||||||||| |||  |     |    |||| |||
Db     154792 T-TAATTTTTATGCTTAGATTAGATGAAGAATGGACAGCCAGATAGAAATGTGATTGGAC 154734

Qy        544 AAAAGGATATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTC 603
              ||| || ||||| ||| ||| ||  |||| |  ||| | |     ||   ||||||  ||
Db     154733 AAACGGGTATGACCTACTGGGAACAGACTGAAAGGGGACCCAGCAAA--ACTGTCTGCTC 154676

Qy        604 AGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCA 663
              |||||||         ||||   |||| ||| |||| | | |||||||||  ||||||||||
Db     154675 AGATTCT--------TCTCAGCCTCTCTGTGTAGCCTCCATTCCTCCTGGGTATGGGGCA 154624

Qy        664 GGACTC-TTCTTCAATGAGGGTCTTCAAGGGAGAAGGG------AGAAAGTGGCCTTTTT 716
              ||||  | ||||  |||||| || ||||| ||||||||       ||| |||| ||||| |
Db     154623 GGACCCTTTCTGAAATGAGAGTTTTCAAAGGAGAAGGAGAAGATAGAGAGTGACCTTTCT 154564

Qy        717 AGATTTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAA 776
              || |||||||| ||||||   | | ||||  |||||| ||||||||||||||||| ||||
Db     154563 AGGTTTTATGGTTTGCTT-TAGCAGGAGGGTTTCTAGGTTCTATGACCCATCTTGAGGAA 154505

Qy        777 GAGGAATTCTGGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGG 836
              ||  ||||||||||||||||||  ||||   | |  |||   || |  ||  |||
Db     154504 GATAAATTCTGGTTTCTGTGACTCACTTTGGAGGATGAAGGCAAGCAGAAGAGAGGGTAT 154445

Qy        837 CAGGAGATGGTCAGAAAGAGACTTGGCTGCTTCTGAG-GGCTTCCGCTCTCCTTTAGTTC 895
              ||||||  |||  ||||||        |||||||||   |||||  | |||||  |||||
Db     154444 GAGGAGAAGGTTGGAAAGA-----CCTTGCTTCTGAGTTCCTTCCAATGTCCTTCAGTTC 154390

Qy        896 CAAGTACTTCTTAGCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAAC 955
              || ||||      ||  |  | |||| || ||||||  | |  | |||| ||| |  ||
Db     154389 AAAATACT------CAGGCTGAGGCACCATGCTTTGGGGTGTCATGTTCTAAGCCCCGAC 154336

Qy        956 A 956
              |
Db     154335 A 154335


RESULT 81
AC147647/c
LOCUS       AC147647              184788 bp    DNA     linear   HTG 21-DEC-2003
DEFINITION  Homo sapiens chromosome UNK clone RP11-1347F12, WORKING DRAFT
            SEQUENCE, 15 unordered pieces.
ACCESSION   AC147647
VERSION     AC147647.1  GI:40286567
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 184788)
  AUTHORS   Wilson,R.K.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 184788)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-DEC-2003) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT
            ------------------- Genome Center --------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu
            Contact: submissions@watson.wustl.edu
            ---------------- Project Information -----------------
            Center project name: H_NH1347F12
            ------------------------------------------------------
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 0%
            Sequencing vector: plasmid; 100%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 182112 bases at least Q40
            Consensus quality: 185176 bases at least Q30
            Consensus quality: 187499 bases at least Q20
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 15 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      1963: contig of 1963 bp in length
            *     1964      2063: gap of unknown length
            *     2064      3765: contig of 1702 bp in length
            *     3766      3865: gap of unknown length
            *     3866      5598: contig of 1733 bp in length
            *     5599      5698: gap of unknown length
            *     5699      7035: contig of 1337 bp in length
            *     7036      7135: gap of unknown length
            *     7136     11260: contig of 4125 bp in length
            *    11261     11360: gap of unknown length
            *    11361     13954: contig of 2594 bp in length
            *    13955     14054: gap of unknown length
            *    14055     18946: contig of 4892 bp in length
            *    18947     19046: gap of unknown length
            *    19047     20917: contig of 1871 bp in length
            *    20918     21017: gap of unknown length
            *    21018     26019: contig of 5002 bp in length
            *    26020     26119: gap of unknown length
            *    26120     40113: contig of 13994 bp in length
            *    40114     40213: gap of unknown length
            *    40214     60429: contig of 20216 bp in length
            *    60430     60529: gap of unknown length
            *    60530     86354: contig of 25825 bp in length
            *    86355     86454: gap of unknown length
            *    86455    117100: contig of 30646 bp in length
            *   117101    117200: gap of unknown length
            *   117201    139607: contig of 22407 bp in length
            *   139608    139707: gap of unknown length
            *   139708    184788: contig of 45081 bp in length.
FEATURES             Location/Qualifiers
     source          1. .184788
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
```

```
                     /db_xref="taxon:9606"
                     /chromosome="UNK"
                     /clone="RP11-1347F12"
     misc_feature    1. .1963
                     /note="assembly_name:Contig68"
     misc_feature    2064. .3765
                     /note="assembly_name:Contig70"
     misc_feature    3866. .5598
                     /note="assembly_name:Contig72"
     misc_feature    5699. .7035
                     /note="assembly_name:Contig73"
     misc_feature    7136. .11260
                     /note="assembly_name:Contig74"
     misc_feature    11361. .13954
                     /note="assembly_name:Contig75"
     misc_feature    14055. .18946
                     /note="assembly_name:Contig76"
     misc_feature    19047. .20917
                     /note="assembly_name:Contig77"
     misc_feature    21018. .26019
                     /note="assembly_name:Contig78"
     misc_feature    26120. .40113
                     /note="assembly_name:Contig79"
     misc_feature    40214. .60429
                     /note="assembly_name:Contig80"
     misc_feature    60530. .86354
                     /note="assembly_name:Contig81"
     misc_feature    86455. .117100
                     /note="assembly_name:Contig82"
     misc_feature    117201. .139607
                     /note="assembly_name:Contig83"
     misc_feature    139708. .184788
                     /note="assembly_name:Contig84"
ORIGIN

  Query Match             4.2%;  Score 254.6;  DB 2;  Length 184788;
  Best Local Similarity   71.3%;  Pred. No. 1.3e-45;
  Matches  471;  Conservative    0;  Mismatches  159;  Indels   31;  Gaps    9;

Qy        304 AAGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAAC 363
              | |  || |||||| ||  || |||||| | ||| ||  ||||||  |||||||  |   |
Db      92450 ATGGGGCTGTAGGGCTGTGAAGGTGTGAGTCCTTTACTTATCCATCTCAAGGGTCACGGC 92391

Qy        364 CAAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATC 423
              ||| |||||||||  ||||||||  |||| ||||| |||||  | ||||||||||||||
Db      92390 CAACACTCCTATAACAAAAGACATATTAACAAGAGAAAAACAGAGCAAATTTATTTAATC 92331

Qy        424 AAAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTG 483
              ||||||||  |||||||||| ||||||||||||  |||||||| ||||||| ||||||
Db      92330 CAAGTTTTATGTGACATGGGAGGCTTCAGAAATGACAACCCAAAGGCCCAGGGAAAACTG 92271

Qy        484 TCTGTTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGAC 543
              | |  ||||| |||| || || |||||||||||||| |||  |      |    |||| |||
Db      92270 T-TAATTTTTATGCTTAGATTAGATGAAGAATGGACAGCTAGATAGAAATGTGATTGGAC 92212

Qy        544 AAAAGGATATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTC 603
              ||| || ||||| ||| ||| ||  |||| |  ||| | |      ||   ||||||  ||
Db      92211 AAACGGGTATGACCTACTGGGAACAGACTGAAAGGGGACCCAGCAAA--ACTGTCTGCTC 92154

Qy        604 AGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCA 663
              |||||||        ||||   |||| ||| |||| | | |||||||| ||||||||||
Db      92153 AGATTCT--------TCTCAGCCTCTCTGTGTAGCCTCCATTCCTCCTGGGTATGGGGCA 92102

Qy        664 GGACTC-TTCTTCAATGAGGGTCTTCAAGGGAGAAGGG------AGAAAGTGGCCTTTTT 716
              |||| | ||||  |||||| || ||||| ||||||||       ||| |||| ||||| |
Db      92101 GGACCCTTTCTGAAATGAGAGTTTTCAAAGGAGAAGGAGAAGATAGAGAGTGACCTTTCT 92042

Qy        717 AGATTTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAA 776
              || |||||||| ||||||   | | ||||  |||||| ||||||||||||||||| ||||
Db      92041 AGGTTTTATGGTTTGCTT-TAGCAGGAGGGTTTCTAGGTTCTATGACCCATCTTGAGGAA 91983

Qy        777 GAGGAATTCTGGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGG 836
              ||  |||||||||||||||||  ||||   | |  |||   || |  ||  |||
Db      91982 GATAAATTCTGGTTTCTGTGACTCACTTTGGAGGATGAAGGCAAGCAGAAGAGAGGGTAT 91923

Qy        837 CAGGAGATGGTCAGAAAGAGACTTGGCTGCTTCTGAG-GGCTTCCGCTCTCCTTTAGTTC 895
              ||||||  |||  ||||||        ||||||||||   |||||  | ||||| |||||
Db      91922 GAGGAGAAGGTTGGAAAGA-----CCTTGCTTCTGAGTTCCTTCCAATGTCCTTCAGTTC 91868

Qy        896 CAAGTACTTCTTAGCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAAC 955
              || ||||      ||  |  | |||| || ||||||  | |  | |||| ||| |  ||
Db      91867 AAAATACT------CAGGCTGAGGCACCATGCTTTGGGGTGTCATGTTCTAAGCCCCGAC 91814

Qy        956 A 956
              |
Db      91813 A 91813


RESULT 82
AC141251
LOCUS       AC141251               195985 bp    DNA     linear   HTG 11-MAR-2003
DEFINITION  Homo sapiens chromosome 16 clone RP11-1259H4, WORKING DRAFT
            SEQUENCE, 2 unordered pieces.
ACCESSION   AC141251
VERSION     AC141251.1  GI:28913031
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 195985)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 195985)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-MAR-2003) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 2703262
            Center clone name: RPCI-11_1259H4
            -------
            Summary Statistics
            Consensus quality: 194077 bases at least Q40
            Consensus quality: 195169 bases at least Q30
            Consensus quality: 195528 bases at least Q20
            Estimated insert size: 175000; agarose-fp estimation
            Estimated insert size: 195885; sum-of-contigs estimation
            Quality coverage:  10.2 in Q20 bases; agarose-fp estimation
            Quality coverage:  9.11 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     6828: contig of 6828 bp in length
            *    6829     6928: gap of unknown length
            *    6929   195985: contig of 189057 bp in length.
FEATURES                Location/Qualifiers
     source             1. .195985
                        /organism="Homo sapiens"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9606"
                        /chromosome="16"
                        /clone="RP11-1259H4"
                        /clone_lib="RPCI human BAC library 11"
ORIGIN
```

```
  Query Match             4.2%;  Score 254.6;  DB 2;  Length 195985;
  Best Local Similarity   71.3%;  Pred. No. 1.3e-45;
  Matches  471;  Conservative    0;  Mismatches  159;  Indels   31;  Gaps    9;

Qy        304 AAGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAAC 363
              | |  || |||||| ||  || ||||||  | ||| ||  ||||||  ||||||| | |
Db     182645 ATGGGGCTGTAGGGCTGTGAAGGTGTGAGTCCTTTACTTATCCATCTCAAGGGTCACGGC 182704

Qy        364 CAAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATC 423
              ||| |||||||||  ||||||||  |||| ||||| |||||  | ||||||||||||||
Db     182705 CAACACTCCTATAACAAAAGACATATTAACAAGAGAAAAACAGAGCAAATTTATTTAATC 182764

Qy        424 AAAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTG 483
              ||||||||  |||||||||||  ||||||||||||  ||||||| ||||||| ||||||
Db     182765 CAAGTTTTATGTGACATGGGAGGCTTCAGAAATGACAACCCAAAGGCCCAGGGAAAACTG 182824

Qy        484 TCTGTTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGAC 543
              | |  ||||| |||| || || ||||||||||||| ||| |      |    |||| |||
Db     182825 T-TAATTTTTATGCTTAGATTAGATGAAGAATGGACAGCTAGATAGAAATGTGATTGGAC 182883

Qy        544 AAAAGGATATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTC 603
              ||| || ||||| ||| ||| ||  |||| |  ||| | |      ||   ||||||  ||
Db     182884 AAACGGGTATGACCTACTGGGAACAGACTGAAAGGGGACCCAGCAAA--ACTGTCTGCTC 182941

Qy        604 AGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCA 663
              |||||||         ||||   |||| ||| |||| | | |||||||||| ||||||||||
Db     182942 AGATTCT--------TCTCAGCCTCTCTGTGTAGCCTCCATTCCTCCTGGGTATGGGGCA 182993

Qy        664 GGACTC-TTCTTCAATGAGGGTCTTCAAGGGAGAAGGG------AGAAAGTGGCCTTTTT 716
              ||||  |  ||||  |||||| || |||||  ||||||||       ||| |||| ||||| |
Db     182994 GGACCCTTTCTGAAATGAGAGTTTTCAAAGGAGAAGGAGAAGATAGAGAGTGACCTTTCT 183053

Qy        717 AGATTTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAA 776
              || |||||||| ||||||   | | ||||  |||||| ||||||||||||||||| ||||
Db     183054 AGGTTTTATGGTTTGCTT-TAGCAGGAGGGTTTCTAGGTTCTATGACCCATCTTGAGGAA 183112

Qy        777 GAGGAATTCTGGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGG 836
              ||  ||||||||||||||||||  ||||   | |  |||   || |  ||  |||
Db     183113 GATAAATTCTGGTTTCTGTGACTCACTTTGGAGGATGAAGGCAAGCAGAAGAGAGGGTAT 183172

Qy        837 CAGGAGATGGTCAGAAAGAGACTTGGCTGCTTCTGAG-GGCTTCCGCTCTCCTTTAGTTC 895
              ||||||  |||  ||||||         ||||||||||    |||||  | |||||  |||||
Db     183173 GAGGAGAAGGTTGGAAAGA-----CCTTGCTTCTGAGTTCCTTCCAATGTCCTTCAGTTC 183227

Qy        896 CAAGTACTTCTTAGCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAAC 955
              || ||||       ||  |  | |||| || ||||||  | |  | |||| ||| |  ||
Db     183228 AAAATACT------CAGGCTGAGGCACCATGCTTTGGGGTGTCATGTTCTAAGCCCCGAC 183281

Qy        956 A 956
              |
Db     183282 A 183282
```

```
RESULT 83
AC141252/c
LOCUS       AC141252              197109 bp    DNA     linear   HTG 11-MAR-2003
DEFINITION  Homo sapiens chromosome 16 clone RP11-1261B23, WORKING DRAFT
            SEQUENCE, 5 unordered pieces.
ACCESSION   AC141252
VERSION     AC141252.1  GI:28913032
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 197109)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 197109)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-MAR-2003) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 2703905
            Center clone name: RPCI-11_1261B23
            -------
            Summary Statistics
            Consensus quality: 193827 bases at least Q40
            Consensus quality: 195005 bases at least Q30
            Consensus quality: 195504 bases at least Q20
            Estimated insert size: 175000; agarose-fp estimation
            Estimated insert size: 196709; sum-of-contigs estimation
            Quality coverage:  11.47 in Q20 bases; agarose-fp estimation
            Quality coverage:  10.2 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     6482: contig of 6482 bp in length
            *     6483     6582: gap of unknown length
            *     6583    11868: contig of 5286 bp in length
            *    11869    11968: gap of unknown length
            *    11969    18767: contig of 6799 bp in length
            *    18768    18867: gap of unknown length
            *    18868    78894: contig of 60027 bp in length
            *    78895    78994: gap of unknown length
            *    78995   197109: contig of 118115 bp in length.
FEATURES             Location/Qualifiers
     source          1. .197109
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone="RP11-1261B23"
                     /clone_lib="RPCI human BAC library 11"
ORIGIN
```

```
  Query Match             4.2%;  Score 254.6;  DB 2;  Length 197109;
  Best Local Similarity   71.3%;  Pred. No. 1.3e-45;
  Matches  471;  Conservative    0;  Mismatches  159;  Indels   31;  Gaps    9;

Qy        304 AAGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAAC 363
              | |  || |||||| ||  || ||||||  | ||| ||  ||||||  ||||||| | |
Db      25601 ATGGGGCTGTAGGGCTGTGAAGGTGTGAGTCCTTTACTTATCCATCTCAAGGGTCACGGC 25542

Qy        364 CAAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATC 423
              ||| |||||||||  ||||||||  |||| ||||| |||||  | ||||||||||||||
Db      25541 CAACACTCCTATAACAAAAGACATATTAACAAGAGAAAAACAGAGCAAATTTATTTAATC 25482

Qy        424 AAAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTG 483
              ||||||||  |||||||||||  ||||||||||||  ||||||| ||||||| ||||||
Db      25481 CAAGTTTTATGTGACATGGGAGGCTTCAGAAATGACAACCCAAAGGCCCAGGGAAAACTG 25422

Qy        484 TCTGTTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGAC 543
              | |  ||||| |||| || || ||||||||||||| ||| |      |    |||| |||
Db      25421 T-TAATTTTTATGCTTAGATTAGATGAAGAATGGACAGCTAGATAGAAATGTGATTGGAC 25363

Qy        544 AAAAGGATATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTC 603
              ||| || ||||| ||| ||| ||  |||| |  ||| | |      ||   ||||||  ||
Db      25362 AAACGGGTATGACCTACTGGGAACAGACTGAAAGGGGACCCAGCAAA--ACTGTCTGCTC 25305

Qy        604 AGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCA 663
              |||||||         ||||   |||| ||| |||| | | |||||||||| ||||||||||
Db      25304 AGATTCT--------TCTCAGCCTCTCTGTGTAGCCTCCATTCCTCCTGGGTATGGGGCA 25253
```

```
Qy        664 GGACTC-TTCTTCAATGAGGGTCTTCAAGGGAGAAGGG------AGAAAGTGGCCTTTTT 716
              |||| | ||||  |||||| || ||||| ||||||||       ||| |||| ||||| |
Db      25252 GGACCCTTTCTGAAATGAGAGTTTTCAAAGGAGAAGGAGAAGATAGAGAGTGACCTTTCT 25193

Qy        717 AGATTTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAA 776
              || |||||||| ||||||    | | ||||  |||||| ||||||||||||||||||| ||||
Db      25192 AGGTTTTATGGTTTGCTT-TAGCAGGAGGGTTTCTAGGTTCTATGACCCATCTTGAGGAA 25134

Qy        777 GAGGAATTCTGGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGG 836
              ||  ||||||||||||||||||  ||||    | |  |||    || |  ||  |||
Db      25133 GATAAATTCTGGTTTCTGTGACTCACTTTGGAGGATGAAGGCAAGCAGAAGAGAGGGTAT 25074

Qy        837 CAGGAGATGGTCAGAAAGAGACTTGGCTGCTTCTGAG-GGCTTCCGCTCTCCTTTAGTTC 895
              |||||| |||  ||||||        ||||||||||   |||||  | |||||  |||||
Db      25073 GAGGAGAAGGTTGGAAAGA-----CCTTGCTTCTGAGTTCCTTCCAATGTCCTTCAGTTC 25019

Qy        896 CAAGTACTTCTTAGCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAAC 955
              || ||||      ||  |  | |||| || ||||||  | |  | |||| ||| |  ||
Db      25018 AAAATACT------CAGGCTGAGGCACCATGCTTTGGGGTGTCATGTTCTAAGCCCCGAC 24965

Qy        956 A 956
              |
Db      24964 A 24964
```

```
RESULT 84
AC139261
LOCUS       AC139261              215386 bp    DNA     linear   HTG 29-JAN-2003
DEFINITION  Homo sapiens chromosome 16 clone RP11-1291J19, WORKING DRAFT
            SEQUENCE, 3 unordered pieces.
ACCESSION   AC139261
VERSION     AC139261.1  GI:28009253
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 215386)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 215386)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JAN-2003) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 2715613
            Center clone name: RPCI-11_1291J19
            -------
            Summary Statistics
            Consensus quality: 213901 bases at least Q40
            Consensus quality: 214007 bases at least Q30
            Consensus quality: 214205 bases at least Q20
            Estimated insert size: 175000; agarose-fp estimation
            Estimated insert size: 215186; sum-of-contigs estimation
            Quality coverage:  15.49 in Q20 bases; agarose-fp estimation
            Quality coverage:  12.59 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     1205: contig of 1205 bp in length
            *    1206     1305: gap of unknown length
            *    1306     2518: contig of 1213 bp in length
            *    2519     2618: gap of unknown length
            *    2619   215386: contig of 212768 bp in length.
FEATURES             Location/Qualifiers
     source          1. .215386
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone="RP11-1291J19"
                     /clone_lib="RPCI human BAC library 11"
ORIGIN
```

```
  Query Match             4.2%;  Score 254.6;  DB 2;  Length 215386;
  Best Local Similarity   71.3%;  Pred. No. 1.3e-45;
  Matches  471;  Conservative    0;  Mismatches  159;  Indels   31;  Gaps    9;

Qy        304 AAGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAAC 363
              | |  || |||||| ||  || ||||||  | ||| ||  ||||||  |||||||  | |
Db     197684 ATGGGGCTGTAGGGCTGTGAAGGTGTGAGTCCTTTACTTATCCATCTCAAGGGTCACGGC 197743

Qy        364 CAAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATC 423
              ||| |||||||||  ||||||||  |||| |||||  |||||  | ||||||||||||||
Db     197744 CAACACTCCTATAACAAAAGACATATTAACAAGAGAAAAACAGAGCAAATTTATTTAATC 197803

Qy        424 AAAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTG 483
              ||||||||  ||||||||||| ||||||||||  |||||||  |||||||  ||||||
Db     197804 CAAGTTTTATGTGACATGGGAGGCTTCAGAAATGACAACCCAAAGGCCCAGGGAAAACTG 197863

Qy        484 TCTGTTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGAC 543
              | |  ||||| |||| || || ||||||||||||| |||  |      |    |||| |||
Db     197864 T-TAATTTTTATGCTTAGATTAGATGAAGAATGGACAGCCAGATAGAAATGTGATTGGAC 197922

Qy        544 AAAAGGATATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTC 603
              ||| || ||||| ||| ||| ||  |||| |  ||| | |      ||   ||||||  ||
Db     197923 AAACGGGTATGACCTACTGGGAACAGACTGAAAGGGGACCCAGCAAA--ACTGTCTGCTC 197980

Qy        604 AGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCA 663
              |||||||          |||| |||| |||| | | |||||||||  ||||||||||
Db     197981 AGATTCT---------TCTCAGCCTCTCTGTGTAGCCTCCATTCCTCCTGGGTATGGGGCA 198032

Qy        664 GGACTC-TTCTTCAATGAGGGTCTTCAAGGGAGAAGGG------AGAAAGTGGCCTTTTT 716
              |||| | ||||  |||||| || ||||| ||||||||       ||| |||| ||||| |
Db     198033 GGACCCTTTCTGAAATGAGAGTTTTCAAAGGAGAAGGAGAAGATAGAGAGTGACCTTTCT 198092

Qy        717 AGATTTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAA 776
              || |||||||| ||||||    | | ||||  |||||| ||||||||||||||||||| ||||
Db     198093 AGGTTTTATGGTTTGCTT-TAGCAGGAGGGTTTCTAGGTTCTATGACCCATCTTGAGGAA 198151

Qy        777 GAGGAATTCTGGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGG 836
              ||  ||||||||||||||||||  ||||    | |  |||    || |  ||  |||
Db     198152 GATAAATTCTGGTTTCTGTGACTCACTTTGGAGGATGAAGGCAAGCAGAAGAGAGGGTAT 198211

Qy        837 CAGGAGATGGTCAGAAAGAGACTTGGCTGCTTCTGAG-GGCTTCCGCTCTCCTTTAGTTC 895
              |||||| |||  ||||||        ||||||||||   |||||  | |||||  |||||
Db     198212 GAGGAGAAGGTTGGAAAGA-----CCTTGCTTCTGAGTTCCTTCCAATGTCCTTCAGTTC 198266

Qy        896 CAAGTACTTCTTAGCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAAC 955
              || ||||      ||  |  | |||| || ||||||  | |  | |||| ||| |  ||
Db     198267 AAAATACT------CAGGCTGAGGCACCATGCTTTGGGGTGTCATGTTCTAAGCCCCGAC 198320

Qy        956 A 956
              |
Db     198321 A 198321
```

```
RESULT 85
AP004195/c
LOCUS       AP004195              151834 bp    DNA     linear   PRI 26-SEP-2001
DEFINITION  Homo sapiens genomic DNA, chromosome 8q23, clone: KB1434A12.
ACCESSION   AP004195
VERSION     AP004195.1  GI:15778159
```

```
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Shimizu,N. and Asakawa,S.
  TITLE     Homo sapiens DNA chromosome 8 SEQUENCE
  JOURNAL   Published Only in Database (2001)
REFERENCE   2  (bases 1 to 151834)
  AUTHORS   Shimizu,N. and Asakawa,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-SEP-2001) Nobuyoshi Shimizu, Keio University, School
            of Medicine, Molecular Biology; 35 Shinanomachi, Shinjuku-ku, Tokyo
            160-8582, Japan (E-mail:nshimizu@dmb.med.keio.ac.jp,
            Tel:81-3-3351-2370, Fax:81-3-3351-2370)
FEATURES             Location/Qualifiers
     source          1. .151834
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8q23"
                     /clone="KB1434A12"
                     /cell_line="FLEB 14-14"
                     /clone_lib="Keio BAC library"
     repeat_region   388. .528
                     /evidence=not_experimental
                     /rpt_family="MIR"
     repeat_region   complement(1592. .1887)
                     /evidence=not_experimental
                     /rpt_family="AluSx"
     repeat_region   complement(2056. .2325)
                     /evidence=not_experimental
                     /rpt_family="AluJo"
     repeat_region   complement(2942. .3581)
                     /evidence=not_experimental
                     /rpt_family="LTR13"
     repeat_region   complement(3581. .3948)
                     /evidence=not_experimental
                     /rpt_family="LTR13"
     repeat_region   3996. .4204
                     /evidence=not_experimental
                     /rpt_family="AluJb"
     repeat_region   4333. .4617
                     /evidence=not_experimental
                     /rpt_family="AluSx"
     repeat_region   4620. .4657
                     /evidence=not_experimental
                     /rpt_family="(TAAA)n"
     repeat_region   6448. .6700
                     /evidence=not_experimental
                     /rpt_family="AluSq"
     repeat_region   6753. .7039
                     /evidence=not_experimental
                     /rpt_family="AluSg"
     repeat_region   7048. .7129
                     /evidence=not_experimental
                     /rpt_family="MER30"
     repeat_region   7130. .7235
                     /evidence=not_experimental
                     /rpt_family="AluJo/FRAM"
     repeat_region   8542. .8565
                     /evidence=not_experimental
                     /rpt_family="(A)n"
     repeat_region   complement(9235. .9525)
                     /evidence=not_experimental
                     /rpt_family="AluJo"
     repeat_region   10195. .10478
                     /evidence=not_experimental
                     /rpt_family="AluSq"
     repeat_region   10869. .10909
                     /evidence=not_experimental
                     /rpt_family="(CA)n"
     repeat_region   11869. .12162
                     /evidence=not_experimental
                     /rpt_family="AluSx"
     repeat_region   complement(12441. .12572)
                     /evidence=not_experimental
                     /rpt_family="FAM"
     repeat_region   complement(12761. .13054)
                     /evidence=not_experimental
                     /rpt_family="AluSp"
     repeat_region   17032. .17122
                     /evidence=not_experimental
                     /rpt_family="(CCCG)n"
     repeat_region   17604. .17624
                     /evidence=not_experimental
                     /rpt_family="GC_rich"
     repeat_region   17726. .17774
                     /evidence=not_experimental
                     /rpt_family="AT_rich"
     repeat_region   17933. .18024
                     /evidence=not_experimental
                     /rpt_family="GA-rich"
     repeat_region   19421. .19553
                     /evidence=not_experimental
                     /rpt_family="AluSq/x"
     repeat_region   19556. .19846
                     /evidence=not_experimental
                     /rpt_family="AluSc"
     repeat_region   20386. .20579
                     /evidence=not_experimental
                     /rpt_family="L2"
     repeat_region   20575. .20795
                     /evidence=not_experimental
                     /rpt_family="L2"
     repeat_region   22525. .22831
                     /evidence=not_experimental
                     /rpt_family="AluSg"
     repeat_region   complement(22904. .23201)
                     /evidence=not_experimental
                     /rpt_family="AluSq"
     repeat_region   23611. .23878
                     /evidence=not_experimental
                     /rpt_family="AluSx"
     repeat_region   23879. .23955
                     /evidence=not_experimental
                     /rpt_family="AluS"
     repeat_region   23977. .24102
                     /evidence=not_experimental
                     /rpt_family="AluJb"
     repeat_region   24103. .24125
                     /evidence=not_experimental
                     /rpt_family="(CAAAA)n"
     repeat_region   24126. .24287
                     /evidence=not_experimental
                     /rpt_family="AluJb"
     repeat_region   complement(24392. .24706)
                     /evidence=not_experimental
                     /rpt_family="AluSq"
     repeat_region   complement(25360. .25473)
                     /evidence=not_experimental
                     /rpt_family="L2"
     repeat_region   25507. .25543
                     /evidence=not_experimental
                     /rpt_family="AT_rich"
     repeat_region   25605. .25665
                     /evidence=not_experimental
                     /rpt_family="CT-rich"
     repeat_region   complement(25666. .25951)
                     /evidence=not_experimental
                     /rpt_family="AluJo"
     repeat_region   25983. .26194
                     /evidence=not_experimental
                     /rpt_family="AluSg/x"
```

```
     repeat_region   26347. .26642
                     /evidence=not_experimental
                     /rpt_family="AluSg"
     repeat_region   26643. .26925
                     /evidence=not_experimental
                     /rpt_family="AluSq"
     repeat_region   26926. .26954
                     /evidence=not_experimental
                     /rpt_family="(TAAAA)n"
     repeat_region   28147. .28378
                     /evidence=not_experimental
                     /rpt_family="L1MC4"
     repeat_region   28385. .28694
                     /evidence=not_experimental
                     /rpt_family="AluSx"
     repeat_region   28695. .28719
                     /evidence=not_experimental
                     /rpt_family="AT_rich"
     repeat_region   complement(29257. .29552)
                     /evidence=not_experimental
                     /rpt_family="AluSx"
     repeat_region   29616. .29827
                     /evidence=not_experimental
                     /rpt_family="MER3"
     repeat_region   complement(29852. .30126)
                     /evidence=not_experimental
                     /rpt_family="AluJo"
     repeat_region   30183. .30224
                     /evidence=not_experimental
                     /rpt_family="(A)n"
     repeat_region   30707. .30831
                     /evidence=not_experimental
                     /rpt_family="FLAM_C"
     repeat_region   30982. .31229
                     /evidence=not_experimental
                     /rpt_family="AluSg1"
     repeat_region   31375. .31564
                     /evidence=not_experimental
                     /rpt_family="L2"
     repeat_region   31605. .31736
                     /evidence=not_experimental
                     /rpt_family="AluJ"
     repeat_region   31737. .32009
                     /evidence=not_experimental
                     /rpt_family="AluSx"
     repeat_region   32025. .32197
                     /evidence=not_experimental
                     /rpt_family="AluJ"
     repeat_region   32265. .32437
                     /evidence=not_experimental
```

```
  Query Match             4.2%;  Score 253.8;  DB 9;  Length 151834;
  Best Local Similarity   72.6%;  Pred. No. 2e-45;
  Matches  468;  Conservative    0;  Mismatches  132;  Indels   45;  Gaps    9;

Qy        307 CTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAA 366
              || | ||||||| | ||| ||||||| | ||| |||   ||||||||||||| || || |
Db     109281 CTACTGTAGGGGCGGAAAGGTGTGATCCCTTTCCTCGCGCATCATAAGGGTCACAGCCGA 109222

Qy        367 AACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAA 426
               | | ||   |||| |||||||||| ||| | ||||| ||||| ||||  |||||||||
Db     109221 CATTTCT---GTAACAGACAGGTTAACAAGGGAAAAACGTAACAGATTTGCTTAATCAAA 109165

Qy        427 GTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCT 486
              ||||||  |||||  ||| | ||||||    ||||||||||       || ||| ||
Db     109164 GTTTTATGTGACACAGGACTATTCAGACCCAAAGACCCAAA-------GGAAAAATGGTC 109112

Qy        487 GTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAA 546
              |||||  |||| |||||| |  ||| | ||| |||| ||     | |     | | ||||
Db     109111 CATTTTTATGCTTAGGTTCAACAAAGCAGGGACAGCA-GTGCAGAAATGTGATGGGCAAA 109053

Qy        547 AGGATATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGA 606
              ||| | ||||||| |||||| ||||| |||||| | ||            |||  |||||
```

```
Db     109052 AGGGTGTGATCTAATGGTAATGGACTGAGGGGGGACCA-----------TCTGCTCAGA 109005

Qy        607 TTCTTCTTGATCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGA 666
              |||||||||  |||          ||| |||||||  ||| ||| | || ||||||||||
Db     109004 TTCTTCTTGGGCTC----------TGTGTAGCATTTCTTCTTCCCGGGTGTGGGGCAGGA 108955

Qy        667 CTCTTCTT-CAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTAT 725
              | | |||| |||||||||||||||||||| ||||||||  |||||||||  ||||| |||
Db     108954 CCCCTCTTGGAATGAGGGTCTTCAAGGGAAAAGGGAGAGTGTGGCCTTTCTAGATGTTAC 108895

Qy        726 GGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTC 785
              |||||||   ||| ||| | |||||||||||||  |||| ||||||||||||||||||
Db     108894 TGCTTGCTGTAGGGGAGAAGGGTTCTAGTTTCTATTGCCCACCTTGGGGAAGAGGAATTC 108835

Qy        786 TGGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATG 845
              |||||||| ||||||||||    | | |||||||   ||||| || ||||||||||||
Db     108834 TGGTTTCTATGACTTGCTTCAGGGGAAAAAGAGGGACAAGAGACA-GAGGGCAGGAGAAG 108776

Qy        846 GTCAGAAAGAGACT-------TGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAA 898
              |||||| |||    |         |||||||||||||| |||||   | ||||| ||||| ||
Db     108775 GTCAGAGAGAACTTGCACCTGAGGCTGCTTCTGAGGACTTCCAATTTCCTTCAGTTCAAA 108716

Qy        899 GTACTTCTTAGCATACCAAAGCACTATACTTTGGCATATGGTTTT 943
              |    ||| ||||| ||||||| |  ||||||||  |||  ||||
Db     108715 G---CTCTCAGCATGCCAAAGCCCCTTACTTTGGGGTATTTTTTT 108674
```

```
RESULT 86
HS24M15
LOCUS       HS24M15               134539 bp    DNA     linear   PRI 04-MAR-2003
DEFINITION  Human DNA sequence from clone RP1-24M15 on chromosome 1q24-25,
            complete sequence.
ACCESSION   Z94055
VERSION     Z94055.1  GI:1944546
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 134539)
  AUTHORS   Wilkinson,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     ---------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            ----------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
            Group.  Further information can be found at
```

```
           http://www.sanger.ac.uk/HGP/Chr1
           RP1-24M15 is from the library RPCI-1 constructed by the group of
           Pieter de Jong. For further details see
           http://www.chori.org/bacpac/home.htm
           VECTOR: pCYPAC2.
FEATURES             Location/Qualifiers
     source          1. .134539
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="RZPD:RPCIP704M1524"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /map="q24-25"
                     /clone="RP1-24M15"
                     /clone_lib="RPCI-1"
ORIGIN

  Query Match             4.2%;  Score 253.6;  DB 9;  Length 134539;
  Best Local Similarity   73.7%;  Pred. No. 2.2e-45;
  Matches  407;  Conservative    0;  Mismatches  129;  Indels   16;  Gaps    6;

Qy        313 TAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCA-ACCAAAACTC 371
              | ||||   ||| |||||||  |  || |||  ||||||||   |    || | ||| |||
Db     122035 TGGGGGCAGAAAGGTGTGATGCCCTTGCTCACCCATCATATAAGCATCACAACAACACTG 122094

Qy        372 CTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTT 431
              ||||   |||||||||||||| ||||| ||| | ||||||||||||||||||||||||
Db     122095 CTATCACAAAAGACAGGTTAACAAGAG-AAAGCGTAACAAATTTATTTAATCAAAGTTTC 122153

Qy        432 ACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTT 491
              ||||||||  |||| |||||||||||||||||||||||||  |||||||  ||||||  |||
Db     122154 ACATGACAAAGGAGCCTTCAGAAATGAAGACCCAAAGACCTGGGGGAAATTGTCTG-TTT 122212

Qy        492 TTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGAT 551
              ||  ||| ||||| || ||| ||||||  |||         |      ||| ||||||  || |
Db     122213 TTACGCTTAGGTTTGAGGAATAATGGACAGCCATGTAGAAACGTAATTGGACAAATGGCT 122272

Qy        552 ATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTT 611
              |||||||| ||||||  | || ||||||| || ||||||||||||||   |||||||||
Db     122273 ATGATCTAATGGTAATAGTCTAAGGGGGAGACCCAGCAAGGCCTGTCTGCCCAGATTCTT 122332

Qy        612 CTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTT 671
              ||||  |||||| |       |    |       |||||      | |||| || |||
Db     122333 CTTGGCCTCTCTGTGAAGCATTTCTTCTCCACCTCCTC-----TGTGGGACAAGACCTCA 122387

Qy        672 CTTCAATGAGGGTCTTCAAGGGAGAAGGGAG-------AAAGTGGCCTTTTTAGATTTTA 724
              ||  |||||||||| ||||||||||||||        |||||  ||||| ||| |||||
Db     122388 CTGGAATGAGGGTCCTCAAGGGAGAAGGGAGGAGGGAAAAAGTTACCTTTCTAGGTTTTA 122447

Qy        725 TGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATT 784
              |||||||||| | |||| |||||||||||||||||| |||    | || ||||||||||||
Db     122448 TGGCTTGCTTTGAGGAAAAGGAGTTCTAGTTTCTATTACCTGCCCTGAGGAAGAGGAATT 122507

Qy        785 CTGGTTTCTGTGACTTGCTT-TCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGA 843
              ||| |||||||||  |  |||   | | ||||  |||    | |||    |||||  | | |||
Db     122508 CTGATTTCTGTGAATCACTTCAAAGGGAGAAGAAGGGTCAGGAGATGGGAGGACTGCAGA 122567

Qy        844 TGGTCAGAAAGA 855
              ||||||||||||
Db     122568 TGGTCAGAAAGA 122579


RESULT 87
AC027373
LOCUS       AC027373               140783 bp    DNA     linear   PRI 06-JUN-2002
DEFINITION  Homo sapiens chromosome 8, clone RP11-521I2, complete sequence.
ACCESSION   AC027373
VERSION     AC027373.11  GI:21327537
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE   1  (bases 1 to 140783)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 8, clone RP11-521I2
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 140783)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 140783)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,
            Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N.,
            Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C.,
            Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S.,
            Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
            Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JUN-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 140783)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,
            Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N.,
            Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C.,
            Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S.,
```

```
            Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
            Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUN-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jun 6, 2002 this sequence version replaced gi:21326334.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L5256
                Center clone name: 521_I_2
            --------------

            Only the last 140.8 kilobases of this clone are being submitted.
            The remainder overlaps accession number AC018781 [WICGR project
            L5210].
FEATURES             Location/Qualifiers
     source          1. .140783
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="RP11-521I2"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   complement(38. .328)
                     /rpt_family="AluY"
     unsure          301. .503
                     /note="PCR product sequence only"
     repeat_region   complement(796. .1084)
                     /rpt_family="AluJo"
     unsure          921. .926
                     /note="<30 qual SNGL region"
     unsure          1349. .1389
                     /note="single clone coverage"
     repeat_region   complement(1684. .1963)
                     /rpt_family="AluSx"
     unsure          2015. .2019
                     /note="<30 qual SNGL region"
     repeat_region   2108. .2127
                     /rpt_family="(A)n"
     unsure          2126. .2130
                     /note="<30 qual SNGL region"
     unsure          2137. .2142
                     /note="<30 qual SNGL region"
     unsure          2193. .2202
                     /note="<30 qual SNGL region"
     unsure          2212. .2217
                     /note="<30 qual SNGL region"
     repeat_region   2224. .2244
                     /rpt_family="AT_rich"
     unsure          2262. .2267
                     /note="<30 qual SNGL region"
     repeat_region   3648. .3780
                     /rpt_family="AluJb"
     repeat_region   4435. .4473
                     /rpt_family="AT_rich"
     repeat_region   5643. .5948
                     /rpt_family="AluSg1"
     repeat_region   complement(6174. .6277)
                     /rpt_family="MIR3"
     repeat_region   complement(6633. .6711)
                     /rpt_family="MADE1"
     repeat_region   6997. .7282
                     /rpt_family="AluSc"
     repeat_region   8342. .8614
                     /rpt_family="MER31-int"
     repeat_region   9490. .9589
                     /rpt_family="L2"
     repeat_region   9556. .9595
                     /rpt_family="MIR"
     repeat_region   9882. .9917
                     /rpt_family="(TA)n"
     repeat_region   complement(9957. .10263)
                     /rpt_family="AluSx"
     repeat_region   complement(10588. .10884)
                     /rpt_family="AluY"
     repeat_region   11846. .12139
                     /rpt_family="AluSq"
     repeat_region   12312. .12429
                     /rpt_family="MER5A"
     repeat_region   complement(12522. .12648)
                     /rpt_family="FLAM_A"
     repeat_region   13122. .13561
                     /rpt_family="MER1A"
     repeat_region   complement(15218. .15506)
                     /rpt_family="AluSq"
     repeat_region   complement(15550. .15711)
                     /rpt_family="L1MC4"
     unsure          15700. .15745
                     /note="single clone coverage"
     repeat_region   15712. .16017
                     /rpt_family="AluSp"
     repeat_region   complement(16018. .16532)
                     /rpt_family="L1MC4"
     repeat_region   16819. .16931
                     /rpt_family="L2"
     repeat_region   17173. .17478
                     /rpt_family="AluJo"
     repeat_region   18049. .18194
                     /rpt_family="MER5B"
     repeat_region   18231. .18540
                     /rpt_family="MER33"
     repeat_region   complement(18599. .18768)
                     /rpt_family="MLT1J"
     repeat_region   complement(18817. .18927)
                     /rpt_family="MER5A"
     repeat_region   18963. .19158
                     /rpt_family="MIR"

  Query Match             4.1%;  Score 252.2;  DB 9;  Length 140783;
  Best Local Similarity   72.4%;  Pred. No. 4.6e-45;
  Matches  467;  Conservative    0;  Mismatches  133;  Indels   45;  Gaps    9;

Qy        307 CTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAA 366
              || | ||||||| | ||| ||||||| | ||| |||   ||||||||||||| || || |
Db     114789 CTACTGTAGGGGCGGAAAGGTGTGATCCCTTTCCTCGCGCATCATAAGGGTCACAGCCGA 114848

Qy        367 AACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAA 426
              | | |||   ||| |||||||||||| ||| | ||||| ||||| ||||  |||||||||
Db     114849 CATTTCTA---TAACAGACAGGTTAACAAGGGAAAAACGTAACAGATTTGCTTAATCAAA 114905

Qy        427 GTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCT 486
              ||||||  |||||  ||| | ||||||    ||||||||||       || ||| ||
Db     114906 GTTTTATGTGACACAGGACTATTCAGACCCAAAGACCCAAA-------GGAAAAATGGTC 114958

Qy        487 GTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAA 546
              |||||  |||| |||||| |  ||| | ||| |||| ||    | |    | | ||||
Db     114959 CATTTTTATGCTTAGGTTCAACAAAGCAGGGACAGCA-GTGCAGAAATGTGATGGGCAAA 115017

Qy        547 AGGATATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGA 606
              ||| | ||||||| |||||| ||||| |||||| | ||            |||  |||||
Db     115018 AGGGTGTGATCTAATGGTAATGGACTGAGGGGGGACCA-----------TCTGCTCAGA 115065

Qy        607 TTCTTCTTGATCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGA 666
              |||||||||  |||          ||| |||||||  ||| ||| | || ||||||||||
Db     115066 TTCTTCTTGGGCTC----------TGTGTAGCATTTCTTCTTCCCGGGTGTGGGGCAGGA 115115
```

```
Qy        667 CTCTTCTT-CAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTAT 725
              | | ||||  |||||||||||||||||||| ||||||||  |||||||| ||||  |||
Db     115116 CCCCTCTTGGAATGAGGGTCTTCAAGGGAAAAGGGAGAGTGTGGCCTTTCTAGACGTTAC 115175

Qy        726 GGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTC 785
              |||||||   ||| ||| | ||||||||||||||  |||| ||||||||||||||||||
Db     115176 TGCTTGCTGTAGGGGAGAAGGGTTCTAGTTTCTATTGCCCACCTTGGGGAAGAGGAATTC 115235

Qy        786 TGGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATG 845
              |||||||| ||||||||||    | | |||||||   ||||| || |||||||||||| |
Db     115236 TGGTTTCTATGACTTGCTTCAGGGGAAAAAGAGGGACAAGAGACA-GAGGGCAGGAGAAG 115294

Qy        846 GTCAGAAAGAGACT-------TGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAA 898
              |||||| |||   |       |||||||||||||| |||||  | ||||| ||||| ||
Db     115295 GTCAGAGAGAACTTGCACCTGAGGCTGCTTCTGAGGACTTCCAATTTCCTTCAGTTCAAA 115354

Qy        899 GTACTTCTTAGCATACCAAAGCACTATACTTTGGCATATGGTTTT 943
              |   ||| ||||| ||||||| |  ||||||||  |||  ||||
Db     115355 G---CTCTCAGCATGCCAAAGCCCCTTACTTTGGGGTATTTTTTT 115396
```

```
RESULT 88
AC079093/c
LOCUS       AC079093              141612 bp    DNA     linear   PRI 01-NOV-2002
DEFINITION  Homo sapiens chromosome 8, clone RP11-265H5, complete sequence.
ACCESSION   AC079093
VERSION     AC079093.16  GI:24462426
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 141612)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 8, clone RP11-265H5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 141612)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Beda,F., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,
            FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,
            Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L.,
            Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G.,
            Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,
            O'Donnell,P., O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Rieback,M., Riley,R.,
            Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-AUG-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 141612)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J.,
            Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
            Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
            Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-OCT-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 141612)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J.,
            Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
            Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
            Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-NOV-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Nov 1, 2002 this sequence version replaced gi:24414568.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L10659
                Center clone name: 265_H_5
            --------------.
FEATURES             Location/Qualifiers
     source          1. .141612
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="RP11-265H5"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   2774. .2864
                     /rpt_family="(CCCG)n"
     repeat_region   3346. .3366
                     /rpt_family="GC_rich"
     repeat_region   3468. .3516
                     /rpt_family="AT_rich"
     repeat_region   3675. .3766
                     /rpt_family="GA-rich"
     repeat_region   complement(4873. .5008)
                     /rpt_family="MIR"
     repeat_region   5163. .5295
                     /rpt_family="AluSq/x"
     repeat_region   5298. .5588
                     /rpt_family="AluSc"
     repeat_region   6005. .6537
                     /rpt_family="L2"
     repeat_region   8265. .8573
                     /rpt_family="AluSg"
     repeat_region   8574. .8638
```

```
                     /rpt_family="MIR"
     repeat_region   complement(8640. .8937)
                     /rpt_family="AluSx"
     repeat_region   9341. .9608
                     /rpt_family="AluSx"
     repeat_region   9609. .9685
                     /rpt_family="AluS"
     repeat_region   9707. .9832
                     /rpt_family="AluJb"
     repeat_region   9833. .9855
                     /rpt_family="(CAAAA)n"
     repeat_region   9856. .10028
                     /rpt_family="AluJb"
     repeat_region   complement(10123. .10437)
                     /rpt_family="AluSq"
     repeat_region   complement(10460. .10555)
                     /rpt_family="MIR"
     repeat_region   complement(11091. .11208)
                     /rpt_family="L2"
     repeat_region   11238. .11274
                     /rpt_family="AT_rich"
     repeat_region   11336. .11396
                     /rpt_family="CT-rich"
     repeat_region   complement(11397. .11682)
                     /rpt_family="AluJb"
     repeat_region   11714. .11925
                     /rpt_family="AluSg/x"
     repeat_region   12078. .12374
                     /rpt_family="AluJb"
     repeat_region   12375. .12695
                     /rpt_family="AluSq"
     repeat_region   13871. .14109
                     /rpt_family="L1MC4a"
     repeat_region   14116. .14425
                     /rpt_family="AluSx"
     repeat_region   14438. .14612
                     /rpt_family="L1MC4"
     repeat_region   complement(14988. .15281)
                     /rpt_family="AluSx"
     repeat_region   15345. .15556
                     /rpt_family="MER3"
     repeat_region   complement(15581. .15855)
                     /rpt_family="AluJo"
     repeat_region   15912. .15953
                     /rpt_family="(A)n"
     repeat_region   16434. .16560
                     /rpt_family="FLAM_C"
     repeat_region   16711. .16958
                     /rpt_family="AluSx"
     repeat_region   16961. .16991
                     /rpt_family="(CAAA)n"
     repeat_region   17066. .17331
                     /rpt_family="L2"
     repeat_region   17338. .17469
                     /rpt_family="AluJ"
     repeat_region   17470. .17742
                     /rpt_family="AluSx"
     repeat_region   17758. .17930
                     /rpt_family="AluJ"
     repeat_region   17947. .18170
                     /rpt_family="L2"
     repeat_region   18563. .18869
                     /rpt_family="AluJb"
     repeat_region   complement(18896. .18927)
                     /rpt_family="MIR"
     repeat_region   complement(18928. .19048)
                     /rpt_family="FLAM_C"
     repeat_region   complement(19049. .19161)
                     /rpt_family="MIR"
     repeat_region   complement(19204. .19479)
                     /rpt_family="AluSx"
     repeat_region   complement(20007. .20405)
                     /rpt_family="L2"
     unsure          20277. .20281
                     /note="single clone coverage"
     repeat_region   20651. .20940
                     /rpt_family="AluSx"
     repeat_region   20950. .21243

  Query Match             4.1%;  Score 252.2;  DB 9;  Length 141612;
  Best Local Similarity   72.4%;  Pred. No. 4.5e-45;
  Matches  467;  Conservative    0;  Mismatches  133;  Indels   45;  Gaps    9;

Qy        307 CTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAA 366
              || | ||||||| | ||| ||||||| | ||| |||   |||||||||||| || || |
Db      93213 CTACTGTAGGGGCGGAAAGGTGTGATCCCTTTCCTCGCGCATCATAAGGGTCACAGCCGA 93154

Qy        367 AACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAA 426
               | | |||   ||| |||||||||| ||| | ||||| ||||| ||||  |||||||||
Db      93153 CATTTCTA---TAACAGACAGGTTAACAAGGGAAAAACGTAACAGATTTGCTTAATCAAA 93097

Qy        427 GTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCT 486
              ||||||  |||||  ||| | ||||||    ||||||||||       || ||| ||
Db      93096 GTTTTATGTGACACAGGACTATTCAGACCCAAAGACCCAAA-------GGAAAAATGGTC 93044

Qy        487 GTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAA 546
              |||||  |||| |||||| |  ||| | ||| |||| ||     |  |     | | ||||
Db      93043 CATTTTTATGCTTAGGTTCAACAAAGCAGGGACAGCA-GTGCAGAAATGTGATGGGCAAA 92985

Qy        547 AGGATATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGA 606
              ||| | ||||||| |||||| ||||| |||||| | ||           |||  |||||
Db      92984 AGGGTGTGATCTAATGGTAATGGACTGAGGGGGGACCA-----------TCTGCTCAGA 92937

Qy        607 TTCTTCTTGATCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGA 666
              |||||||||  |||          ||| |||||||  ||| ||| | || ||||||||||
Db      92936 TTCTTCTTGGGCTC----------TGTGTAGCATTTCTTCTTCCCGGGTGTGGGGCAGGA 92887

Qy        667 CTCTTCTT-CAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTAT 725
              | | ||||  |||||||||||||||||| |||||||  |||||||| |||| |||
Db      92886 CCCCTCTTGGAATGAGGGTCTTCAAGGGAAAAGGGAGAGTGTGGCCTTTCTAGACGTTAC 92827

Qy        726 GGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTC 785
              |||||||   ||| ||| | ||||||||||||||  |||| ||||||||||||||||||
Db      92826 TGCTTGCTGTAGGGGAGAAGGGTTCTAGTTTCTATTGCCCACCTTGGGGAAGAGGAATTC 92767

Qy        786 TGGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATG 845
              |||||||| ||||||||||     | | |||||||   ||||| || |||||||||||| |
Db      92766 TGGTTTCTATGACTTGCTTCAGGGGAAAAAGAGGGACAAGAGACA-GAGGGCAGGAGAAG 92708

Qy        846 GTCAGAAAGAGACT-------TGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAA 898
              |||||| |||    |       |||||||||||||| |||||  | ||||| ||||| ||
Db      92707 GTCAGAGAGAACTTGCACCTGAGGCTGCTTCTGAGGACTTCCAATTTCCTTCAGTTCAAA 92648

Qy        899 GTACTTCTTAGCATACCAAAGCACTATACTTTGGCATATGGTTTT 943
              |    ||| ||||| ||||||| |  ||||||||  |||  ||||
Db      92647 G---CTCTCAGCATGCCAAAGCCCCTTACTTTGGGGTATTTTTTT 92606


RESULT 89
AC025931/c
LOCUS       AC025931               168042 bp    DNA     linear   HTG 24-AUG-2002
DEFINITION  Homo sapiens chromosome 8 clone RP11-304H10 map 8, WORKING DRAFT
            SEQUENCE, 33 unordered pieces.
ACCESSION   AC025931
VERSION     AC025931.3  GI:8077106
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 168042)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 8, clone RP11-304H10
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 168042)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
```

```
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-MAR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 168042)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 25, 2000 this sequence version replaced gi:7408047.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L8125
                Center clone name: 304_H_10
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 151058 bases at least Q40
                Consensus quality: 159522 bases at least Q30
                Consensus quality: 162718 bases at least Q20
                Insert size: 170000; agarose-fp
                Insert size: 164842; sum-of-contigs
                Quality coverage: 3.9 in Q20 bases; agarose-fp
                Quality coverage: 4.0 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 33 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1050: contig of 1050 bp in length
            *     1051     1150: gap of 100 bp
            *     1151     2842: contig of 1692 bp in length
            *     2843     2942: gap of 100 bp
            *     2943     4216: contig of 1274 bp in length
            *     4217     4316: gap of 100 bp
            *     4317     6658: contig of 2342 bp in length
            *     6659     6758: gap of 100 bp
            *     6759     7880: contig of 1122 bp in length
            *     7881     7980: gap of 100 bp
            *     7981     9768: contig of 1788 bp in length
            *     9769     9868: gap of 100 bp
            *     9869    11201: contig of 1333 bp in length
            *    11202    11301: gap of 100 bp
            *    11302    13028: contig of 1727 bp in length
            *    13029    13128: gap of 100 bp
            *    13129    15537: contig of 2409 bp in length
            *    15538    15637: gap of 100 bp
            *    15638    16349: contig of 712 bp in length
            *    16350    16449: gap of 100 bp
            *    16450    18401: contig of 1952 bp in length
            *    18402    18501: gap of 100 bp
            *    18502    20890: contig of 2389 bp in length
            *    20891    20990: gap of 100 bp
            *    20991    23341: contig of 2351 bp in length
            *    23342    23441: gap of 100 bp
            *    23442    25914: contig of 2473 bp in length
            *    25915    26014: gap of 100 bp
            *    26015    28004: contig of 1990 bp in length
            *    28005    28104: gap of 100 bp
            *    28105    30871: contig of 2767 bp in length
            *    30872    30971: gap of 100 bp
            *    30972    34015: contig of 3044 bp in length
            *    34016    34115: gap of 100 bp
            *    34116    38117: contig of 4002 bp in length
            *    38118    38217: gap of 100 bp
            *    38218    41978: contig of 3761 bp in length
            *    41979    42078: gap of 100 bp
            *    42079    46550: contig of 4472 bp in length
            *    46551    46650: gap of 100 bp
            *    46651    51745: contig of 5095 bp in length
            *    51746    51845: gap of 100 bp
            *    51846    57322: contig of 5477 bp in length
            *    57323    57422: gap of 100 bp
            *    57423    61546: contig of 4124 bp in length
            *    61547    61646: gap of 100 bp
            *    61647    66565: contig of 4919 bp in length
            *    66566    66665: gap of 100 bp
            *    66666    72365: contig of 5700 bp in length
            *    72366    72465: gap of 100 bp
            *    72466    80205: contig of 7740 bp in length
            *    80206    80305: gap of 100 bp
            *    80306    89094: contig of 8789 bp in length
            *    89095    89194: gap of 100 bp
            *    89195    99286: contig of 10092 bp in length
            *    99287    99386: gap of 100 bp
            *    99387   108845: contig of 9459 bp in length
            *   108846   108945: gap of 100 bp
            *   108946   120518: contig of 11573 bp in length
            *   120519   120618: gap of 100 bp
            *   120619   131381: contig of 10763 bp in length
            *   131382   131481: gap of 100 bp
            *   131482   142718: contig of 11237 bp in length
            *   142719   142818: gap of 100 bp
            *   142819   168042: contig of 25224 bp in length.
FEATURES             Location/Qualifiers
```

```
     source          1. .168042
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="RP11-304H10"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1050
                     /note="assembly_fragment"
     misc_feature    1151. .2842
                     /note="assembly_fragment"
     misc_feature    2943. .4216
                     /note="assembly_fragment"
     misc_feature    4317. .6658
                     /note="assembly_fragment"
     misc_feature    6759. .7880
                     /note="assembly_fragment"
     misc_feature    7981. .9768
                     /note="assembly_fragment"
     misc_feature    9869. .11201
                     /note="assembly_fragment"
     misc_feature    11302. .13028
                     /note="assembly_fragment"
     misc_feature    13129. .15537
                     /note="assembly_fragment"
     misc_feature    15638. .16349
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    16450. .18401
                     /note="assembly_fragment"
     misc_feature    18502. .20890
                     /note="assembly_fragment"
     misc_feature    20991. .23341
                     /note="assembly_fragment"
     misc_feature    23442. .25914
                     /note="assembly_fragment"

  Query Match             4.1%;  Score 252.2;  DB 2;  Length 168042;
  Best Local Similarity   72.4%;  Pred. No. 4.5e-45;
  Matches  467;  Conservative    0;  Mismatches  133;  Indels   45;  Gaps    9;

Qy        307 CTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAA 366
              || | ||||||| | ||| ||||||| | ||| |||   ||||||||||||| || || |
Db     100977 CTACTGTAGGGGCGGAAAGGTGTGATCCCTTTCCTCGCGCATCATAAGGGTCACAGCCGA 100918

Qy        367 AACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAA 426
               | | |||   ||| |||||||||||| ||| | ||||| ||||| ||||  |||||||||
Db     100917 CATTTCTA---TAACAGACAGGTTAACAAGGGAAAAACGTAACAGATTTGCTTAATCAAA 100861

Qy        427 GTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCT 486
              ||||||  |||||  ||| | ||||||     ||||||||||        || ||| ||
Db     100860 GTTTTATGTGACACAGGACTATTCAGACCCAAAGACCCAAA-------GGAAAAATGGTC 100808

Qy        487 GTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAA 546
               |||||  |||| |||||| |  ||| | ||| |||| ||    |  |    | | ||||
Db     100807 CATTTTTATGCTTAGGTTCAACAAAGCAGGGACAGCA-GTGCAGAAATGTGATGGGCAAA 100749

Qy        547 AGGATATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGA 606
              ||| | ||||||| |||||| ||||| |||||| | ||            |||  |||||
Db     100748 AGGGTGTGATCTAATGGTAATGGACTGAGGGGGGACCA------------TCTGCTCAGA 100701

Qy        607 TTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGA 666
              |||||||||  |||          ||| |||||||  ||| ||| | || ||||||||||
Db     100700 TTCTTCTTGGGCTC----------TGTGTAGCATTTCTTCTTCCCGGGTGTGGGGCAGGA 100651

Qy        667 CTCTTCTT-CAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTAT 725
              | | ||||  |||||||||||||||||||| ||||||||  |||||||| ||||  |||
Db     100650 CCCCTCTTGGAATGAGGGTCTTCAAGGGAAAAGGGAGAGTGTGGCCTTTCTAGACGTTAC 100591

Qy        726 GGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTC 785
              |||||||   ||| ||| | ||||||||||||||  |||| ||||||||||||||||||
```

```
Db     100590 TGCTTGCTGTAGGGGAGAAGGGTTCTAGTTTCTATTGCCCACCTTGGGGAAGAGGAATTC 100531

Qy        786 TGGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATG 845
              |||||||| ||||||||||    | | |||||||   ||||| || |||||||||||| |
Db     100530 TGGTTTCTATGACTTGCTTCAGGGGAAAAAGAGGGACAAGAGACA-GAGGGCAGGAGAAG 100472

Qy        846 GTCAGAAAGAGACT-------TGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAA 898
              |||||| |||    |       ||||||||||||||| |||||  | |||||| ||||| ||
Db     100471 GTCAGAGAGAACTTGCACCTGAGGCTGCTTCTGAGGACTTCCAATTTCCTTCAGTTCAAA 100412

Qy        899 GTACTTCTTAGCATACCAAAGCACTATACTTTGGCATATGGTTTT 943
              |    ||| ||||| ||||||| |  |||||||  |||  ||||
Db     100411 G---CTCTCAGCATGCCAAAGCCCCTTACTTTGGGGTATTTTTTT 100370


RESULT 90
AC112641
LOCUS       AC112641              165619 bp    DNA     linear   PRI 29-AUG-2002
DEFINITION  Homo sapiens 3 BAC RP11-431I8 (Roswell Park Cancer Institute Human
            BAC Library) complete sequence.
ACCESSION   AC112641
VERSION     AC112641.3  GI:19774278
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 165619)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., He,X.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M., Neal,D.,
            Nelson,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N.,
            Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N.,
            Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L.,
            Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M.,
            Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S.,
            Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E.,
            Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A.,
            Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C.,
            Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L.,
            Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S.,
            Warren,R., Washington,C., Watlington,S., Williams,G.,
            Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y.,
            Wu,Y.F., Zhou,J., Zorrilla,S., Naylor,S.L., Weinstock,G. and
            Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 165619)
  AUTHORS   Worley,K.C.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (22-FEB-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 165619)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   4  (bases 1 to 165619)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   5  (bases 1 to 165619)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-AUG-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Mar 28, 2002 this sequence version replaced gi:19747074.
            INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
            gc-help@bcm.tmc.edu

            CLONE LENGTH: This sequence does not necessarily represent the
            entire insert of this clone. Overlapping regions of clones are only
            sequenced and submitted once, so the sequence for the remainder of
            the insert may be found in the record for the adjacent clones.
            Overlapping clones are noted at the beginning and end of the
            Features listing.

            ANNOTATION OF FEATURES:
               STSs are identified using ePCR (Genome Res. 7:541-550) searches
            of a local database that includes entries from dbSTS, GDB, and
            local mapping efforts.
               Repeats are identified using RepeatMasker (A. Smit and P. Green,
            unpublished.) for Human and Mouse sequences.
               Genes and Region of sequence similarity are identified by BLAST
            (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
            EST and cDNA sequences. Genes demonstrate at least two exons
            flanked by consensus splice sites that maintained sequence
            continuity across the splice junctions. Sequences that are not
            identical matches are annotated as similar.

            SEQUENCING READ COVERAGE:Sequencing is completed to a minimum
            standard of double strand coverage with a minimum of 2 clones and 2
            reads with no ambiguities or 2 chemistries with a minimum of 2
            clones and 3 reads with no ambiguities. If the sequence quality for
            a region does not meet this standard, it will be indicated in the
            annotation as Low Coverage.

            QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality
            standards - estimated error rate less than 1 per 10,000 bases.
            Reports of lowest quality individual bases and measures of base
            quality are listed below. Description of the metrics can be found
            at URL:
            http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

            QUALSTAT-REPORT.
FEATURES             Location/Qualifiers
     source          1. .165619
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-431I8"
     repeat_region   complement(1. .421)
                     /rpt_family="LTR42"
     repeat_region   complement(932. .1013)
                     /rpt_family="L1ME2"
     repeat_region   complement(1427. .1593)
                     /rpt_family="L1ME"
     repeat_region   1594. .1711
                     /rpt_family="FLAM_C"
     repeat_region   complement(1712. .1872)
                     /rpt_family="L1ME"
     repeat_region   complement(2006. .2309)
                     /rpt_family="AluSx"
     repeat_region   complement(2320. .2372)
                     /rpt_family="FLAM"
     repeat_region   2408. .2915
                     /rpt_family="MER6"
     repeat_region   2916. .2946
                     /rpt_family="T-rich"
     repeat_region   2947. .3132
                     /rpt_family="(TTTTA)n"
     repeat_region   3133. .3191
                     /rpt_family="MER6"
     repeat_region   complement(3192. .3487)
                     /rpt_family="AluSx"
     repeat_region   3488. .3778
                     /rpt_family="MER6"
     repeat_region   complement(4022. .4324)
                     /rpt_family="AluSx"
     repeat_region   5123. .5346
                     /rpt_family="MER58A"
     repeat_region   complement(5426. .5556)
                     /rpt_family="L2"
     repeat_region   6094. .6402
                     /rpt_family="AluSq"
     repeat_region   7220. .7279
                     /rpt_family="MER2"
     repeat_region   7952. .8168
                     /rpt_family="MER44"
     repeat_region   complement(8169. .8329)
                     /rpt_family="AluJb"
     repeat_region   complement(8330. .8645)
                     /rpt_family="AluSg"
     repeat_region   complement(8646. .8793)
                     /rpt_family="AluJb"
     repeat_region   8794. .8898
                     /rpt_family="MER44"
     repeat_region   8896. .8940
                     /rpt_family="MER44D"
     repeat_region   complement(11856. .12921)
                     /rpt_family="MER11C"
     repeat_region   complement(14733. .15018)
                     /rpt_family="AluSq"
     repeat_region   complement(15070. .15292)
                     /rpt_family="AluSg/x"
     repeat_region   16955. .16978
                     /rpt_family="(CCCCAG)n"
     repeat_region   complement(18734. .18751)
                     /rpt_family="AluSc"
     repeat_region   18752. .18771
                     /rpt_family="(TTA)n"
     repeat_region   complement(18772. .19051)
                     /rpt_family="AluSc"
     repeat_region   complement(21281. .21594)
                     /rpt_family="AluSq"
     repeat_region   complement(21646. .21867)
                     /rpt_family="AluSg/x"
     repeat_region   23390. .23431
                     /rpt_family="(CCCCAG)n"
     repeat_region   complement(25396. .25696)
                     /rpt_family="AluSx"
     repeat_region   complement(28569. .28817)
                     /rpt_family="AluSx"
     repeat_region   complement(28832. .28878)
                     /rpt_family="AluSp/q"
     repeat_region   complement(28901. .28966)
                     /rpt_family="MER5A"

Query Match              4.1%;  Score 252;  DB 9;  Length 165619;
```

```
  Best Local Similarity   69.2%;  Pred. No. 5e-45;
  Matches  393;  Conservative    0;  Mismatches  160;  Indels   15;  Gaps    3;

Qy        312 GTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTC 371
              |||   | | | |     ||   | || ||  |||||    | ||||     || | || ||||
Db      58133 GTATAAGGGTAGAAAATTGTGATATCTTTCCTTCCCATAGAAGGCGGACAGCTAACACTC 58192

Qy        372 CTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTT 431
              || || | |||||||||||| || || ||| | |||||||| |||||||||||||||  |
Db      58193 TTAAAGCACAAGACAGGTTAACAAAAGAAAAGCATAACAAATCTATTTAATCAAAGTTGT 58252

Qy        432 ACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTT 491
                |||||   ||| |||||||| |||||  | ||||| ||||||||||||||     |||
Db      58253 GTGTGACACAAGAGCCTTCAGAAGTGAAGGTCTAAAGATCCAGGGGAAACTG-----TTT 58307

Qy        492 TTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGAT 551
              || || | ||||| ||| ||||| |||||||          |    |||| |||   ||| |
Db      58308 TTATGTTTAGGTTGGATAAAGAACGGATAGCCATGTAGAACTGTGATTGGAC---AGGTT 58364

Qy        552 ATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTT 611
              |||||||| ||||||  |||| | |||||||| | ||||| ||||||| ||||| |||||
Db      58365 ATGATCTAATGGTAATAGACTGAAGGGGAAACCCCGCAAGACCTGTCTGTTCAGGTTCTT 58424

Qy        612 CTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTT 671
              |||| |||||||   |          | |||||||    ||||  ||  ||||| | |
Db      58425 CTTGGTCTCTCTGCAT-------AGCATTCCTTTCCTGTCCAGTAGAGGTTAGGACACCT 58477

Qy        672 CTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATGGCTTG 731
              ||  |||| |||| ||||| ||||||||| |||||  ||||| ||| |||||||| ||||
Db      58478 GTTACATGAAGGTCCTCAAGAGAGAAGGGACAAAGTTACCTTTCTAGGTTTTATGACTTG 58537

Qy        732 CTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCTGGTTT 791
              ||| | |||||| || || ||||||||||||||| ||||| |||| ||||||||||||
Db      58538 CTTTGCGGAAGAAGACTTTTAGTTTCTATGACCCAACTTGGTGAAGAAGAATTCTGGTTT 58597

Qy        792 CTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGGTCAGA 851
              || |||||  ||||  || |  || || | |    | || | || |||||||    |
Db      58598 CTATGACTCACTTTGCAGAGGGTAGTGGGGAAGACAGGAGAATGGGAGGAGATCAAGATG 58657

Qy        852 AAGAGACTTGGCTGCTTCTGAGGGCTTC 879
                ||        || ||||||||  | |
Db      58658 TTGAAGAGATCTTGTTTCTGAGGCATGC 58685
```

```
RESULT 91
AC023505
LOCUS       AC023505               168177 bp    DNA     linear   HTG 08-JAN-2003
DEFINITION  Homo sapiens chromosome 3 clone RP11-145N22, WORKING DRAFT
            SEQUENCE, 15 unordered pieces.
ACCESSION   AC023505
VERSION     AC023505.15  GI:20335475
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 168177)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 168177)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-FEB-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 168177)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JAN-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Apr 28, 2002 this sequence version replaced gi:9255951.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: HAFU
                Center clone name: RP11-145N22
            -------------- Summary Statistics
                Sequencing vector: M13;
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 155068 bases at least Q40
                Consensus quality: 159325 bases at least Q30
                Consensus quality: 161802 bases at least Q20
                Estimated insert size: 166654; sum-of-contigs estimation
                Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 15 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2092: contig of 2092 bp in length
            *     2093     2192: gap of unknown length
            *     2193     4938: contig of 2746 bp in length
            *     4939     5038: gap of unknown length
            *     5039     9171: contig of 4133 bp in length
            *     9172     9271: gap of unknown length
```

```
            *     9272    14631: contig of 5360 bp in length
            *    14632    14731: gap of unknown length
            *    14732    20935: contig of 6204 bp in length
            *    20936    21035: gap of unknown length
            *    21036    28250: contig of 7215 bp in length
            *    28251    28350: gap of unknown length
            *    28351    38101: contig of 9751 bp in length
            *    38102    38201: gap of unknown length
            *    38202    51839: contig of 13638 bp in length
            *    51840    51939: gap of unknown length
            *    51940    66672: contig of 14733 bp in length
            *    66673    66772: gap of unknown length
            *    66773    79483: contig of 12711 bp in length
            *    79484    79583: gap of unknown length
            *    79584    94902: contig of 15319 bp in length
            *    94903    95002: gap of unknown length
            *    95003   111622: contig of 16620 bp in length
            *   111623   111722: gap of unknown length
            *   111723   128057: contig of 16335 bp in length
            *   128058   128157: gap of unknown length
            *   128158   147895: contig of 19738 bp in length
            *   147896   147995: gap of unknown length
            *   147996   168177: contig of 20182 bp in length.
FEATURES             Location/Qualifiers
     source          1. .168177
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-145N22"
ORIGIN

  Query Match             4.1%;  Score 252;  DB 2;  Length 168177;
  Best Local Similarity   73.6%;  Pred. No. 5e-45;
  Matches  406;  Conservative    0;  Mismatches  130;  Indels   16;  Gaps    6;

Qy        313 TAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCA-ACCAAAACTC 371
              | ||||   ||| |||||||  |  || |||  ||||||||   |   || | ||| |||
Db     131189 TGGGGGCAGAAAGGTGTGATGCCCTTGCTCACCCATCATATAAGCATCACAACAACACTG 131248

Qy        372 CTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTT 431
              ||||   ||||||||||||||| ||||| ||| | ||||||||||||||||||||||||
Db     131249 CTATCACAAAAGACAGGTTAACAAGAG-AAAGCGTAACAAATTTATTTAATCAAAGTTTC 131307

Qy        432 ACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTT 491
              ||||||||  ||||  |||||||||||||||||||||||||  |||||||  |||||| |||
Db     131308 ACATGACAAAGGAGCCTTCAGAAATGAAGACCCAAAGACCTGGGGGAAATTGTCTG-TTT 131366

Qy        492 TTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGAT 551
              ||  ||| ||||| || ||| ||||||  |||        |     ||| ||||||  || |
Db     131367 TTACGCTTAGGTTTGAGGAATAATGGACAGCCATGTAGAAACGTAATTGGACAAATGGCT 131426

Qy        552 ATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTT 611
              |||||||| ||||||  | || |||||||  || ||||||||||||||   |||||||||
Db     131427 ATGATCTAATGGTAATAGTCTAAGGGGGAGACCCAGCAAGGCCTGTCTGCCCAGATTCTT 131486

Qy        612 CTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTT 671
              ||||  ||||||  |      |     |       |||||      | |||| || |||
Db     131487 CTTGGCCTCTCTGTGAAGCATTTCTTCTCCACCTCCTC-----TGTGGGACAAGACCTCA 131541

Qy        672 CTTCAATGAGGGTCTTCAAGGGAGAAGGGAG-------AAAGTGGCCTTTTTAGATTTTA 724
              ||  |||||||||| ||||||||||||||||       |||||  ||||| ||| |||||
Db     131542 CTGGAATGAGGGTCCTCAAGGGAGAAGGGAGGAGGGAAAAAGTTACCTTTCTAGGTTTTA 131601

Qy        725 TGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATT 784
              ||||| |||| | |||| ||||||||||||||||||| |||   | || ||||||||||||
Db     131602 TGGCTGGCTTTGAGGAAAAGGAGTTCTAGTTTCTATTACCTGCCCTGAGGAAGAGGAATT 131661

Qy        785 CTGGTTTCTGTGACTTGCTT-TCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGA 843
              ||| |||||||| |  |||   | | ||||  |||   | |||   ||||| | | |||
Db     131662 CTGATTTCTGTGAATCACTTCAAAGGGAGAAGAAGGGTCAGGAGATGGGAGGACTGCAGA 131721

Qy        844 TGGTCAGAAAGA 855
              ||||||||||||
Db     131722 TGGTCAGAAAGA 131733


RESULT 92
AP000238/c
LOCUS       AP000238               62873 bp    DNA     linear   PRI 15-MAR-2003
DEFINITION  Homo sapiens genomic DNA, chromosome 21q, clone:RP11-313P22,
            complete sequence.
ACCESSION   AP000238
VERSION     AP000238.2  GI:14861140
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens genomic DNA
  JOURNAL   Published Only in Database (1999)
REFERENCE   2  (bases 1 to 62873)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-1999) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     On Jul 17, 2001 this sequence version replaced gi:4835607.
FEATURES             Location/Qualifiers
     source          1. .62873
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q"
                     /clone="RP11-313P22"
ORIGIN

  Query Match             4.1%;  Score 251.6;  DB 9;  Length 62873;
  Best Local Similarity   71.4%;  Pred. No. 6.6e-45;
  Matches  450;  Conservative    0;  Mismatches  154;  Indels   26;  Gaps    8;

Qy        330 GATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTCCTATAGTAAAAGACAGGT 389
              ||||| ||| |||  |||||||||   || ||||  | | |||||||  ||||||||||
Db      59157 GATACCTTTCCTCACCCATCATAACAATCACAACTGACATTCCTATAACAAAAGACAGGT 59098

Qy        390 TAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTACATGACATGGGAGTCTT 449
              ||| ||||| ||| | || |||||||| ||||| ||||||||||||||||    |||| |||
Db      59097 TAACAAGAGAAAATCATAGCAAATTTGTTTAACCAAAGTTTTACATGACTCAGGAGGCTT 59038

Qy        450 CAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTTTTTTGCTGAGGTTCGATG 509
              || |||||||| |||||||| |||| |||||||||  ||| | |||| || ||| |||
Db      59037 CAAAAATGAAGA-CCAAAGACCTAGGGAAAACTGTCT-ATTTCTATGCTTAGATTCCATG 58980

Qy        510 AAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGATATGATCTAGTGGTAAAGG 569
              ||||||| | | ||||| |  |||| |||| |||||||||  ||  |||    |  | |||
Db      58979 AAGAATGAACAACATGTGGAAATGT-GATTGGACAAAAGGGTAGAATCAGTGGTAATAGG 58921

Qy        570 ACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTTCTTGATCTCTCTCTCTCT 629
                |||| || | |||||||||||||| | ||| ||||||||||   |||||
Db      58920 CTGAGGGGGAAAGCCCAGCAAGGCCTGTGTGTTCGGATTCTTCTTAGCCTCTC------- 58868

Qy        630 ATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTTCTTCAATGAGGGTCTTCA 689
              |||  ||  ||| |||||||||| ||||| |||| | | | |  ||| || ||||||||
Db      58867 -TGTGCAGTGTTCCTTCCTCCTGAGCATGGGTCAGGGCCCCTATGAAATAAGAGTCTTCA 58809

Qy        690 AGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATGGCTTGCTTCGGGGAAGAGGAGTT 749
              ||||| ||| ||          |||| |||||||||  |||||||  ||| |||||||||
Db      58808 AGGGAAAAGAGAAT-------CTTTCTAGATTTTATATCTTGCTTTTGGGGAGAGGAGTT 58756
```

```
Qy        750 CTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCTGGTTTCTGTGACTTGCTTTCATG 809
              |||| |||||||  ||| ||| |||||| | ||||||  ||||| |||||  | |   ||
Db      58755 CTAGCTTCTATGCTCCA-CTTAGGGAAGTGAAATTCTACTTTCTATGACTCACGTCAGTG 58697

Qy        810 AAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGGTCAGAAAGAGACTTGGCTGCTTC 869
              |  |||||| ||| ||  | || || ||||||| |||||||||          || |||
Db      58696 GATTAAGAGGGGTAGGATACCGGGGGTCAGGAGAAAGTCAGAAAGA-----CCTTGTTTC 58642

Qy        870 TGAGGGCTTCCGCTCTCCTTTAGTTCCAAGTACTTCTTAGCATACCAAAGCACTATACTT 929
              ||||  ||| |       |||   | |||  | |  | ||||| |||| |  | ||| ||
Db      58641 TGAGAACTTTCCAGTCTTTTTCAGTTCAA--AGTAGTCAGCATGCCAAGGTGCCATATTT 58584

Qy        930 TGGCATATGGTTTTCTGAGCTCTAACACTG 959
              |||  | |  | |||||||| | |||| ||
Db      58583 TGGGGTGTCATGTTCTGAGCCCCAACAGTG 58554


RESULT 93
AP000094
LOCUS       AP000094              100000 bp    DNA     linear   PRI 25-MAY-2002
DEFINITION  Homo sapiens genomic DNA of 21q22.1, GART and AML related,
            B335D16-P10G11 region, segment 4/7.
ACCESSION   AP000094
VERSION     AP000094.1  GI:4730855
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Shiba,T. and Sakaki,Y.
  TITLE     Homo sapiens 726,344bp genomic DNA of 21q22.1 GART and AML region
  JOURNAL   Published Only in Database (1999)
REFERENCE   2  (bases 1 to 100000)
  AUTHORS   Hirakawa,M., Yamaguchi,H., Imai,K. and Shimada,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-APR-1999) Mika Hirakawa, Japan Science and Technology
            Corporation (JST), Advanced Databases Department; 5-3, Yonbancho,
            Chiyoda-ku, Tokyo 102-0081, Japan (E-mail:mika@tokyo.jst.go.jp,
            URL:http://www-alis.tokyo.jst.go.jp/, Tel:81-3-5214-8491,
            Fax:81-3-5214-8470)
COMMENT     This sequence is conducted by Kitasato University JST sequencing
            Laboratory as a JST sequencing team.
            Principal Investigator:Yoshiyuki Sakaki Ph.D.
            Phone: +81-3-5449-5622, Fax : +81-3-5449-5445,
            sakaki@hgc.ims.u-tokyo.ac.jp
            Sub-leader: Tadayoshi Shiba Ph.D., Masahira Hattori Ph.D. The
            sequence is submitted by Human Genome Sequencing in ALIS project of
            JST.
            Japan Science and Technology Corporation (JST)
            5-3, Yonbancho, Chiyoda-ku, Tokyo 102-0081 Japan
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www-alis.tokyo.jst.go.jp/HGS/) or send
            email to webmaster@www-alis.tokyo.jst.go.jp.
FEATURES             Location/Qualifiers
     source          1. .100000
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.1"
     STS             16824. .16941
                     /note="OLI-1/OLI-3F;The location is between each flanking
                     site of PCR primers."
                     /db_xref="GDB:201821"
     STS             18639. .18734
                     /standard_name="D21S1699"
                     /note="528E12haR-F/528E12haR-R;The location is between
                     each flanking site of PCR primers."
                     /db_xref="GDB:438564"
ORIGIN
```

```
  Query Match             4.1%;  Score 251.6;  DB 9;  Length 100000;
  Best Local Similarity   71.4%;  Pred. No. 6.3e-45;
  Matches  450;  Conservative    0;  Mismatches  154;  Indels   26;  Gaps    8;

Qy        330 GATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTCCTATAGTAAAAGACAGGT 389
              ||||| ||| |||  |||||||||    || ||||  | | |||||||  ||||||||||
Db      14869 GATACCTTTCCTCACCCATCATAACAATCACAACTGACATTCCTATAACAAAAGACAGGT 14928

Qy        390 TAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTACATGACATGGGAGTCTT 449
              ||| ||||| ||| | || |||||||| ||||| ||||||||||||||||   |||| |||
Db      14929 TAACAAGAGAAAATCATAGCAAATTTGTTTAACCAAAGTTTTACATGACTCAGGAGGCTT 14988

Qy        450 CAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTTTTTTGCTGAGGTTCGATG 509
              || |||||||| |||||||||| |||| |||||||||   ||| | |||| || ||| |||
Db      14989 CAAAAATGAAGA-CCAAAGACCTAGGGAAAACTGTCT-ATTTCTATGCTTAGATTCCATG 15046

Qy        510 AAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGATATGATCTAGTGGTAAAGG 569
              ||||||| | | ||||| |  |||| |||| ||||||||| ||  |||    |  | |||
Db      15047 AAGAATGAACAACATGTGGAAATGT-GATTGGACAAAAGGGTAGAATCAGTGGTAATAGG 15105

Qy        570 ACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTTCTTGATCTCTCTCTCTCT 629
                  |||| || | |||||||||||||| | ||| ||||||||||   |||||
Db      15106 CTGAGGGGGAAAGCCCAGCAAGGCCTGTGTGTTCGGATTCTTCTTAGCCTCTC------- 15158

Qy        630 ATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTTCTTCAATGAGGGTCTTCA 689
              |||  ||  ||| |||||||||| |||| |||| | | | |  ||| || |||||||
Db      15159 -TGTGCAGTGTTCCTTCCTCCTGAGCATGGGTCAGGGCCCCTATGAAATAAGAGTCTTCA 15217

Qy        690 AGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATGGCTTGCTTCGGGGAAGAGGAGTT 749
              ||||| ||| ||          |||| ||||||||||  |||||||  ||| |||||||||
Db      15218 AGGGAAAAGAGAAT-------CTTTCTAGATTTTATATCTTGCTTTTGGGGAGAGGAGTT 15270

Qy        750 CTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCTGGTTTCTGTGACTTGCTTTCATG 809
              |||| |||||||  ||| ||| |||||| | ||||||  ||||| |||||  | |   ||
Db      15271 CTAGCTTCTATGCTCCA-CTTAGGGAAGTGAAATTCTACTTTCTATGACTCACGTCAGTG 15329

Qy        810 AAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGGTCAGAAAGAGACTTGGCTGCTTC 869
              |  |||||| ||| ||  | || || ||||||| |||||||||          || |||
Db      15330 GATTAAGAGGGGTAGGATACCGGGGGTCAGGAGAAAGTCAGAAAGA-----CCTTGTTTC 15384

Qy        870 TGAGGGCTTCCGCTCTCCTTTAGTTCCAAGTACTTCTTAGCATACCAAAGCACTATACTT 929
              ||||  ||| |       |||   | |||  | |  | ||||| |||| |  | ||| ||
Db      15385 TGAGAACTTTCCAGTCTTTTTCAGTTCAA--AGTAGTCAGCATGCCAAGGTGCCATATTT 15442

Qy        930 TGGCATATGGTTTTCTGAGCTCTAACACTG 959
              |||  | |  | |||||||| | |||| ||
Db      15443 TGGGGTGTCATGTTCTGAGCCCCAACAGTG 15472


RESULT 94
AP000198
LOCUS       AP000198              100000 bp    DNA     linear   PRI 08-JAN-2000
DEFINITION  Homo sapiens genomic DNA, chromosome 21q22.1, D21S226-AML region,
            clone B355D16-T1073, segment 4/8, complete sequence.
ACCESSION   AP000198
VERSION     AP000198.1  GI:4827096
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 100000)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y. and Sakaki,Y.
  TITLE     Homo sapiens 726,344bp genomic DNA of 21q22.1 (REGION: D21S226-AML
            CLONE RANGE: B355D16-T1073)
  JOURNAL   Published Only in DataBase (1999)
REFERENCE   2  (bases 1 to 100000)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-1999) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
```

```
            Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
            Japan (E-mail:hattori@gsc.riken.go.jp,
            URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
            Fax:81-42-778-9924)
COMMENT     E. coli transposon insertion:The present data does not contain E.
            coli transposon sequences which integrated in the
            original/previous sequences.  We determined the boundary between
            the insertion and genomic sequences experimentally, removed the
            insertion sequences, reconstituted the present data. The sequencing
            project is supported by Japan Science Technology Corporation (JST)
            and The Institute of Physical and Chemical Research (RIKEN).
FEATURES             Location/Qualifiers
     source          1. .100000
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.1"
ORIGIN

  Query Match             4.1%;  Score 251.6;  DB 9;  Length 100000;
  Best Local Similarity   71.4%;  Pred. No. 6.3e-45;
  Matches  450;  Conservative    0;  Mismatches  154;  Indels   26;  Gaps    8;

Qy        330 GATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTCCTATAGTAAAAGACAGGT 389
              ||||| ||| |||  |||||||||   || ||||  | | |||||||  ||||||||||
Db      14869 GATACCTTTCCTCACCCATCATAACAATCACAACTGACATTCCTATAACAAAAGACAGGT 14928

Qy        390 TAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTACATGACATGGGAGTCTT 449
              ||| ||||| ||| | || ||||||| ||||| ||||||||||||||||    |||| |||
Db      14929 TAACAAGAGAAAATCATAGCAAATTTGTTTAACCAAAGTTTTACATGACTCAGGAGGCTT 14988

Qy        450 CAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTTTTTTGCTGAGGTTCGATG 509
              || ||||||||| |||||||| |||| |||||||||  ||| | |||| || ||| |||
Db      14989 CAAAAATGAAGA-CCAAAGACCTAGGGAAAACTGTCT-ATTTCTATGCTTAGATTCCATG 15046

Qy        510 AAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGATATGATCTAGTGGTAAAGG 569
              ||||||| | | ||||| |  |||| |||| |||||||||  ||  |||    |  | |||
Db      15047 AAGAATGAACAACATGTGGAAATGT-GATTGGACAAAAGGGTAGAATCAGTGGTAATAGG 15105

Qy        570 ACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTTCTTGATCTCTCTCTCTCT 629
                  |||| || | |||||||||||||| | ||| ||||||||||   |||||
Db      15106 CTGAGGGGGAAAGCCCAGCAAGGCCTGTGTGTTCGGATTCTTCTTAGCCTCTC------- 15158

Qy        630 ATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTTCTTCAATGAGGGTCTTCA 689
              |||  ||  ||| |||||||||||| ||||| |||| | | |  | ||| || |||||||
Db      15159 -TGTGCAGTGTTCCTTCCTCCTGAGCATGGGTCAGGGCCCCTATGAAATAAGAGTCTTCA 15217

Qy        690 AGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATGGCTTGCTTCGGGGAAGAGGAGTT 749
              ||||| ||| ||         |||| |||||||||  |||||||  ||| |||||||||
Db      15218 AGGGAAAAGAGAAT-------CTTTCTAGATTTTATATCTTGCTTTTGGGGAGAGGAGTT 15270

Qy        750 CTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCTGGTTTCTGTGACTTGCTTTCATG 809
              |||| |||||||  ||| ||| |||||| | ||||||  ||||| |||||  | |   ||
Db      15271 CTAGCTTCTATGCTCCA-CTTAGGGAAGTGAAATTCTACTTTCTATGACTCACGTCAGTG 15329

Qy        810 AAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGGTCAGAAAGAGACTTGGCTGCTTC 869
              |  |||||| ||| ||  | || || |||||||  ||||||||||         || |||
Db      15330 GATTAAGAGGGGTAGGATACCGGGGGTCAGGAGAAAGTCAGAAAGA-----CCTTGTTTC 15384

Qy        870 TGAGGGCTTCCGCTCTCCTTTAGTTCCAAGTACTTCTTAGCATACCAAAGCACTATACTT 929
              ||||  ||| |       |||   | ||| | |  | ||||| |||| |  | ||| ||
Db      15385 TGAGAACTTTCCAGTCTTTTTCAGTTCAA--AGTAGTCAGCATGCCAAGGTGCCATATTT 15442

Qy        930 TGGCATATGGTTTTCTGAGCTCTAACACTG 959
              |||  | |  | |||||||| | |||| ||
Db      15443 TGGGGTGTCATGTTCTGAGCCCCAACAGTG 15472


RESULT 95
AP001706/c
LOCUS       AP001706               340000 bp    DNA     linear   PRI 21-MAY-2003
DEFINITION  Homo sapiens genomic DNA, chromosome 21q, section 50/105.
ACCESSION   AP001706 AL163251 BA000005
VERSION     AP001706.1  GI:7768780
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
            Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E.,
            Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
            Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
            Patterson,D., Reichwald,K., Rump,A., Schillhabel,M., Schudy,A.,
            Zimmermann,W., Rosenthal,A., Kudoh,J., Shibuya,K., Kawasaki,K.,
            Asakawa,S., Shintani,A., Sasaki,K., Nagamine,K., Mitsuyama,S.,
            Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordsiek,G.,
            Hornischer,K., Barandt,P., Scharfe,M., Schoen,O., Desario,A.,
            Reichelt,J., Kauer,G., Bloecker,H., Ramser,J., Beck,A., Klages,S.,
            Hennig,S., Riesselmann,L., Dagand,E., Wehrmeyer,S., Borzym,K.,
            Gardiner,K., Nizetic,D., Francis,F., Lehrach,H., Reinhardt,R. and
            Yaspo,M.L.
  TITLE     The DNA sequence of human chromosome 21
  JOURNAL   Nature 405 (6784), 311-319 (2000)
  MEDLINE   20289799
   PUBMED   10830953
REFERENCE   2  (bases 1 to 340000)
  AUTHORS   Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
            Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E.,
            Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
            Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
            Patterson,D., Reichwald,K., Rump,A., Schillhabel,M., Schudy,A.,
            Zimmermann,W., Rosenthal,A., Kudoh,J., Shibuya,K., Kawasaki,K.,
            Asakawa,S., Shintani,A., Sasaki,T., Nagamine,K., Mitsuyama,S.,
            Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordsiek,G.,
            Hornischer,K., Barandt,P., Scharfe,M., Schoen,O., Desario,A.,
            Reichelt,J., Kauer,G., Bloecker,H., Ramser,J., Beck,A., Klages,S.,
            Hennig,S., Riesselmann,L., Dagand,E., Wehrmaeyer,S., Borzym,K.,
            Gardiner,K., Nizetic,D., Francis,F., Lehrach,H., Reinhardt,R. and
            Yaspo,M.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-APR-2000) The Chromosome 21 Mapping and Sequencing
            Consortium: * RIKEN Genomic Sciences Center, Human Genome Research
            Group * Institute of Molecular Biotechnology, Genome Analysis *
            Keio University School of Medicine, Dept. of Molecular Biology *
            GBF, Dept. of Genome Analysis * Max-Planck Institute for Molecular
            Genetics (addresses see below)
COMMENT     On May 30, 2000 this sequence version replaced gi:7717316.
            The chromosome 21 mapping and sequencing consortium consisting of
            * RIKEN Genomic Sciences Center, Human Genome Research Group, *
            Sagamihara 228-8555, Japan,
            * e.mail: hattori@gsc.riken.go.jp
            * URL: http://hgp.gsc.riken.go.jp/
            and
            * Institute of Molecular Biotechnology, Genome Analysis, *
            Beutenbergstrasse 11, D-07745 Jena, Germany,
            * e.mail: gscj-submit@genome.imb-jena.de
            * URL: http://genome.imb-jena.de/
            and
            * Keio University School of Medicine, Molecular Biology, * Tokyo
            160-8582, Japan,
            * e.mail: nshimizu@dmb-med.keio.ac.jp
            * URL: http://www.dmb.med.keio.ac.jp/
            and
            * GBF, Dept. of Genome Analysis,
            * Mascheroder Weg 1, D-38124 Braunschweig, Germany, * e.mail:
            info.genome@gbf.de
            * URL: http://genome.gbf.de/
            and
            * Max-Planck Institute for Molecular Genetics,
            * Ihnestrasse 73, D-14195 Berlin, Germany,
            * e.mail: info-chr21@molgen.mpg.de
            * URL: http://chr21.rz-berlin.mpg.de/
            AL163251: Submitted (10-Apr-2000).
```

```
FEATURES             Location/Qualifiers
     source          1. .340000
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
     source          <1. .27840
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.1"
                     /clone="R313P22, 5' partial"
                     /clone_lib="RPCI-11 BAC library"
                     /note="Accession No. AP000238"
     source          27220. .179949
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.1"
                     /clone="R269P7"
                     /clone_lib="RPCI-11 BAC library"
                     /note="Accession No. AP000237"
     source          61223. .215637
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.1"
                     /clone="295E05-A"
                     /clone_lib="CEPH BAC library"
                     /note="Accession No. AF107258"
     source          179900. .339148
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.1"
                     /clone="R355D16"
                     /clone_lib="RPCI-11 BAC library"
                     /note="Accession No. AP000236"
     source          189485. .329415
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.1"
                     /clone="P209L12"
                     /clone_lib="RPCI1,3-5 PAC library"
                     /note="Accession No. AF107257"
     source          307039. .>340000
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.1"
                     /clone="P681K20, 3' partial"
                     /clone_lib="RPCI1,3-5 PAC library"
                     /note="Accession No. AF191069"
     source          337401. .>340000
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.1"
                     /clone="R347N6, 3' partial"
                     /clone_lib="RPCI-11 BAC library"
                     /note="Accession No. AP000390"
     repeat_region   complement(1545. .1974)
                     /note="L2"
                     /rpt_family="LINE/L2"
                     /rpt_type=DISPERSED
     repeat_region   3877. .3948
                     /note="L1PA13"
                     /rpt_family="LINE/L1"
                     /rpt_type=DISPERSED
     repeat_region   3960. .4264
                     /note="AluSp"
                     /rpt_family="SINE/Alu"
                     /rpt_type=DISPERSED
     repeat_region   4375. .4467
                     /note="L2"
                     /rpt_family="LINE/L2"
                     /rpt_type=DISPERSED
     repeat_region   4733. .4897
                     /note="MIR"
                     /rpt_family="SINE/MIR"
                     /rpt_type=DISPERSED
     repeat_region   complement(5948. .6378)
                     /note="L2"
                     /rpt_family="LINE/L2"
                     /rpt_type=DISPERSED
     gene            complement(join(6976. .10821,27566. .27823,31733. .31900,
                     277367. .277896))
                     /gene="GRIK1"
     mRNA            complement(join(<6976. .10821,27566. .27823,31733. .31900,
                     277367. .277896))
                     /gene="GRIK1"
                     /product="glutamate receptor 5 subunit GLUR5"
                     /note="Accession No. L19058"
     exon            complement(6976. .10821)
                     /gene="GRIK1"
                     /number=4
     repeat_region   complement(7279. .7324)
                     /note="L2"
                     /rpt_family="LINE/L2"
                     /rpt_type=DISPERSED
     repeat_region   7376. .7604
                     /note="MIR"
                     /rpt_family="SINE/MIR"
                     /rpt_type=DISPERSED
     repeat_region   8950. .8970
                     /note="(TTTTA)n"
                     /rpt_family="Simple_repeat"
                     /rpt_type=TANDEM
     repeat_region   complement(8972. .9224)
                     /note="AluSx"
                     /rpt_family="SINE/Alu"
                     /rpt_type=DISPERSED
     repeat_region   complement(9728. .9861)
                     /note="MER20"
                     /rpt_family="DNA/MER1_type"
                     /rpt_type=DISPERSED
     repeat_region   complement(9895. .10157)

  Query Match             4.1%;  Score 251.6;  DB 9;  Length 340000;
  Best Local Similarity   71.4%;  Pred. No. 5.8e-45;
  Matches  450;  Conservative    0;  Mismatches  154;  Indels   26;  Gaps    8;

Qy        330 GATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTCCTATAGTAAAAGACAGGT 389
              ||||| ||| |||  |||||||||    || ||||  | | |||||||  ||||||||||
Db      24124 GATACCTTTCCTCACCCATCATAACAATCACAACTGACATTCCTATAACAAAAGACAGGT 24065

Qy        390 TAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTACATGACATGGGAGTCTT 449
              ||| ||||| ||| | || |||||||  ||||| ||||||||||||||||    |||| |||
Db      24064 TAACAAGAGAAAATCATAGCAAATTTGTTTAACCAAAGTTTTACATGACTCAGGAGGCTT 24005

Qy        450 CAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTTTTTTTGCTGAGGTTCGATG 509
              || ||||||||| |||||||||  |||| |||||||||   ||| | |||| || ||| |||
Db      24004 CAAAAATGAAGA-CCAAAGACCTAGGGAAAACTGTCT-ATTTCTATGCTTAGATTCCATG 23947

Qy        510 AAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGATATGATCTAGTGGTAAAGG 569
              ||||||| | | ||||| |  |||| |||| |||||||||  ||  |||    |  | |||
Db      23946 AAGAATGAACAACATGTGGAAATGT-GATTGGACAAAAGGGTAGAATCAGTGGTAATAGG 23888
```

```
Qy        570 ACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTTCTTGATCTCTCTCTCTCT 629
              |||| || | |||||||||||||| | ||| ||||||||||   |||||
Db      23887 CTGAGGGGGAAAGCCCAGCAAGGCCTGTGTGTTCGGATTCTTCTTAGCCTCTC------- 23835

Qy        630 ATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTTCTTCAATGAGGGTCTTCA 689
               |||  ||  ||| |||||||||| ||||| |||| | | | |  ||| || |||||||
Db      23834 -TGTGCAGTGTTCCTTCCTCCTGAGCATGGGTCAGGGCCCCTATGAAATAAGAGTCTTCA 23776

Qy        690 AGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATGGCTTGCTTCGGGGAAGAGGAGTT 749
              |||||| ||| ||          |||| ||||||||||  |||||||  ||| ||||||||||
Db      23775 AGGGAAAAGAGAAT-------CTTTCTAGATTTTATATCTTGCTTTTGGGGAGAGGAGTT 23723

Qy        750 CTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCTGGTTTCTGTGACTTGCTTTCATG 809
              |||| |||||||  ||| ||| |||||| | ||||||  ||||| |||||  | |   ||
Db      23722 CTAGCTTCTATGCTCCA-CTTAGGGAAGTGAAATTCTACTTTCTATGACTCACGTCAGTG 23664

Qy        810 AAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGGTCAGAAAGAGACTTGGCTGCTTC 869
               |  |||||| ||| ||  | || || |||||||  ||||||||||        || |||
Db      23663 GATTAAGAGGGGTAGGATACCGGGGGTCAGGAGAAAGTCAGAAAGA-----CCTTGTTTC 23609

Qy        870 TGAGGGCTTCCGCTCTCCTTTAGTTCCAAGTACTTCTTAGCATACCAAAGCACTATACTT 929
              ||||  ||| |        |||   | |||  | |  | ||||| |||| |  | ||| ||
Db      23608 TGAGAACTTTCCAGTCTTTTTCAGTTCAA--AGTAGTCAGCATGCCAAGGTGCCATATTT 23551

Qy        930 TGGCATATGGTTTTCTGAGCTCTAACACTG 959
              |||  | |  | ||||||| | |||| ||
Db      23550 TGGGGTGTCATGTTCTGAGCCCCAACAGTG 23521


RESULT 96
CNS05TCT
LOCUS       CNS05TCT              167875 bp    DNA     linear   PRI 30-APR-2001
DEFINITION  Human chromosome 14 DNA sequence BAC R-705B14 of library RPCI-11
            from chromosome 14 of Homo sapiens (Human), complete sequence.
ACCESSION   AL355838
VERSION     AL355838.3  GI:13928042
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 167875)
  AUTHORS   Heilig,R., Petit,J.L., Vico,V., Dasilva,C., Robert,C., Wincker,P.,
            Brottier,P., Cattolico,L., Barbe,V., Pelletier,E., Artiguenave,F.,
            Levy,M., Eckenberg,R., Bruls,T., deBerardinis,V., Cruaud,C.,
            Gyapay,G., Saurin,W. and Weissenbach,J.
  TITLE     Sequencing of the human chromosome 14
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 167875)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-APR-2001) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     On May 2, 2001 this sequence version replaced gi:8388628.
            --------------- Genome Center
            Center: Genoscope / Centre National de Sequencage
            Center code: GS
            Web site: http://www.genoscope.cns.fr/
            Contact: SeqRef@genoscope.cns.fr
            ---------------
            The following BAC sequence is oriented from the T7 to the SP6 end.
            Upstream BAC (overlapping the T7 end) : R-666E17 (AC=AL358275)
            Downstream BAC (overlapping the SP6 end) : R-226P1 ----------
            Summary Statistics
            Assembly program: Phrap; version 2.0
            Quality coverage: 7.52x in Q20 bases; sum-of-contigs
            --------------
            ----------------------
            Overall quality chart :
            Range    :   bases
            0        :
            1  -  9  :       1
            10 - 19  :      10
            20 - 29  :      23
            30 - 39  :     115
            40 - 49  :    1167
            50 - 59  :    3022
            60 - 69  :    6100
            70 - 79  :   17860
            80 - 89  :   48925
            90 - 99  :   90652
            ----------------------
            Percentage of bases with a quality value >= 40 : 99 %.
FEATURES             Location/Qualifiers
     source          1. .167875
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="14"
                     /clone="R-705B14"
                     /clone_lib="RPCI-11"
ORIGIN

  Query Match             4.1%;  Score 249.8;  DB 9;  Length 167875;
  Best Local Similarity   68.6%;  Pred. No. 1.5e-44;
  Matches  441;  Conservative    0;  Mismatches  177;  Indels   25;  Gaps    6;

Qy        320 GAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTCCTATAGTA 379
              ||||| |||| |||| ||||||   ||||||| || |      |  | | | || ||  |
Db      18786 GAAAAGGTGTAATACCTTTTCTTACCCATCATGAGAGGTATGGCTGACATTTCTGTAACA 18845

Qy        380 AAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTACATGACA 439
              |||| ||||  |||||  | ||| | |||| ||| |||||||||||||||||||||||||
Db      18846 AAAGGCAGGCAAATAATGGAAAAGCATAACTAATATATTTAATCAAAGTTTTACATGACA 18905

Qy        440 TGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTTTTTTGCTG 499
               |||| |||||||||||||||| ||||| ||||| ||||||      |||| ||||
Db      18906 CAGGAGACTTCAGAAATGAAGACCC-AAGACGCAGGGAAAACTG-----GTTTTATGCTT 18959

Qy        500 AGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGATATGATCTA 559
              |||||  |||||||||||| |||        |    |||| ||||||||| ||| ||| |
Db      18960 AGGTTTTATGAAGAATGGACAGCCATATAAAAATGTGATTTGACAAAAGGGTATAATCCA 19019

Qy        560 GTGGTAAAGGACTCAG-GGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTTCTTGATC 618
              ||||||  || | || |||||| |    | ||||||| | | || ||||||||||| |
Db      19020 ATGGTAACAGATTAAGAGGGGAATCCTGGAAAGGCCTATTTGTTTAGATTCTTCTTGGCC 19079

Qy        619 TCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTTCTTCAAT 678
              | ||          || |      | | ||   ||||  | ||| || |  ||  |||
Db      19080 TTTC----------TACACAGCATACTTCCCCCAGGTATAAGACAGAACACCCCTGGAAT 19129

Qy        679 GAGGGTCTTCAAGGGAGAA-------GGGAGAAAGTGGCCTTTTTAGATTTTATGGCTTG 731
              |||| ||||||||||||||       || | |||||   |||| ||| |||||| ||||
Db      19130 GAGGATCTTCAAGGGAGAAGGATGAGGGAAAAAAGTTATCTTTCTAGGTTTTATTGCTTC 19189

Qy        732 CTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCTGGTTT 791
              |||||||  |   | ||||| |||||| ||| |||  || |  |||| |||||| ||||
Db      19190 CTTCGGGAGATCAGGGTTCTGGTTTCTGTGATCCACTTTAGAAAAGAGAAATTCTAGTTT 19249

Qy        792 CTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGGTCAGA 851
              || ||||||||||    | | |||||||  |  ||| ||||||||||||| | |||||||
Db      19250 CTATGACTTGCTTCAGGGGACAAAGAGGGATGGGAGACAGGAGGGCAGGATAAGGTCAGA 19309

Qy        852 AAGAGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAGTACTTCTTAGCA 911
              ||||| || ||| | |  |    |||| |  | | |  | |  |   |  | || | ||
Db      19310 GAGAGAATTTGCTTC-TGAGGCTACTTCTGTGCCCTTCCAATATCTTTTTGTACTAAACA 19368

Qy        912 TACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAA 954
              |  |||||||| |||||||||   |||  |||||| ||| | ||
Db      19369 TCTCAAAGCACCATACTTTGGGGTATCATTTTCTAAGCCCCAA 19411


RESULT 97
AC068492/c
LOCUS       AC068492              172281 bp    DNA     linear   PRI 29-OCT-2002
```

```
DEFINITION  Homo sapiens BAC clone RP11-809C23 from 2, complete sequence.
ACCESSION   AC068492
VERSION     AC068492.2  GI:14010945
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 172281)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
   PUBMED   9847074
REFERENCE   2  (bases 1 to 172281)
  AUTHORS   Lamar,B., Abbott,A. and Ureta,M.
  TITLE     The sequence of Homo sapiens BAC clone RP11-809C23
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 172281)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-MAY-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 172281)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2001) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   5  (bases 1 to 172281)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JAN-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   6  (bases 1 to 172281)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-OCT-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On May 9, 2001 this sequence version replaced gi:7677994.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0809C23
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. John D.
            McPherson, Department of Genetics, Washington University, St. Louis
            MO. For additional information about the map position of this
            sequence, see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries. Genomics 51:1-8. The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at http://www.chori.org
            VECTOR:  pBACe3.6

            NEIGHBORING SEQUENCE INFORMATION:
            Actual start of this clone is at base position 1 of RP11-809C23;
            actual end is at base position 172281 of RP11-809C23.
FEATURES             Location/Qualifiers
     source          1. .172281
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2"
                     /clone="RP11-809C23"
                     /clone_lib="RPCI-11"
     repeat_region   1161. .1218
                     /rpt_family="MIR"
     repeat_region   1245. .1302
                     /rpt_family="MIR"
     repeat_region   1313. .1407
                     /rpt_family="MIR"
     repeat_region   2612. .2712
                     /rpt_family="MaLR"
     repeat_region   2868. .3080
                     /rpt_family="ERVL"
     repeat_region   3088. .3175
                     /rpt_family="MER1_type"
     repeat_region   3872. .3997
                     /rpt_family="MIR"
     repeat_region   4265. .4436
                     /rpt_family="MIR"
     repeat_region   5913. .6082
                     /rpt_family="MER1_type"
     repeat_region   6124. .6274
                     /rpt_family="L2"
     repeat_region   6855. .7214
                     /rpt_family="MaLR"
     repeat_region   7808. .7956
                     /rpt_family="MIR"
     repeat_region   7990. .8112
                     /rpt_family="L2"
     repeat_region   8188. .8540
                     /rpt_family="MaLR"
     repeat_region   8659. .8760
                     /rpt_family="MER53"
     repeat_region   9485. .9630
                     /rpt_family="MaLR"
     repeat_region   9785. .9844
                     /rpt_family="MaLR"
     repeat_region   11223. .11294
                     /rpt_family="MIR"
     repeat_region   12678. .12781
                     /rpt_family="MIR"
     repeat_region   13515. .13713
                     /rpt_family="MIR"
     repeat_region   14305. .14830
                     /rpt_family="L1"
     repeat_region   14831. .15039
                     /rpt_family="MER1_type"
     repeat_region   15040. .15071
                     /rpt_family="L1"
     repeat_region   15072. .15405
                     /rpt_family="MaLR"
     repeat_region   15424. .15498
                     /rpt_family="L1"
     repeat_region   15499. .15814
                     /rpt_family="Alu"
     repeat_region   15815. .16113
                     /rpt_family="L1"
     repeat_region   16173. .16243
```

```
                     /rpt_family="L1"
     repeat_region   16244. .16557
                     /rpt_family="Alu"
     repeat_region   16558. .16645
                     /rpt_family="L1"
     repeat_region   16646. .16830
                     /rpt_family="ERV1"
     repeat_region   16817. .17794
                     /rpt_family="ERV1"
     repeat_region   17795. .18118
                     /rpt_family="ERV1"
     repeat_region   18119. .19275
                     /rpt_family="L1"
     repeat_region   19258. .19362
                     /rpt_family="L1"
     misc_feature    19363. .20255
                     /note="similar to  EST BE900547 (NID:g10388822)"
     misc_feature    19363. .20203
                     /note="similar to  EST AU132006 (NID:g10992360)"
     misc_feature    19363. .20194
                     /note="similar to  EST BE900431 (NID:g10388587)"
     misc_feature    19363. .20184
                     /note="similar to  EST BE733455 (NID:g10147447)"
     misc_feature    19363. .20146
                     /note="similar to  EST AU137312 (NID:g10997851)"
     misc_feature    19363. .20106
                     /note="similar to  EST AU135322 (NID:g10995861)"
     misc_feature    19363. .20093
                     /note="similar to  EST AU123565 (NID:g10948281)"
     misc_feature    19363. .19632
                     /note="similar to  EST N83917 (NID:g1259542)"
     misc_feature    19366. .20209
                     /note="similar to  EST BE300406 (NID:g9184154)"
     misc_feature    19374. .20261
                     /note="similar to  EST BE741444 (NID:g10155436)"
     misc_feature    19374. .20051
                     /note="similar to  EST BE275889 (NID:g9150850)"
     misc_feature    19374. .20027
                     /note="similar to  EST BE410450 (NID:g9346900)"
     misc_feature    19374. .19487
                     /note="similar to  EST BE736692 (NID:g10150684)"
     misc_feature    19375. .19772
                     /note="similar to  EST BF043529 (NID:g10760584)"
     misc_feature    19384. .20095
                     /note="similar to  EST BE903326 (NID:g10394419)"
     misc_feature    19384. .20086
                     /note="similar to  EST BE901941 (NID:g10391598)"
     misc_feature    19384. .20059
                     /note="similar to  EST AU136548 (NID:g10997087)"
     misc_feature    19384. .20010
                     /note="similar to  EST AU136949 (NID:g10997488)"
     misc_feature    19384. .19962
                     /note="similar to  EST BF204034 (NID:g11097620)"

  Query Match             4.1%;  Score 248;  DB 9;  Length 172281;
  Best Local Similarity   70.1%;  Pred. No. 3.8e-44;
  Matches  485;  Conservative    0;  Mismatches  175;  Indels   32;  Gaps   10;

Qy        302 CAAAGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCA 361
              || | |  | |||||||||  |  |||| |  | ||| ||   ||||||||||||||  ||
Db      82589 CATAACCTCTGTAGGGGTGGTAGGGTGTAACCCCTTTCCT-ACCCATCATAAGGGTTACA 82531

Qy        362 ACCAAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAA 421
              |||| ||| |||||  |||||||||||||| ||||| ||||| || ||||| ||||  |
Db      82530 GCCAATACTGCTATAACAAAAGACAGGTTAACAAGAGAAAAACATAGCAAATGTATT--A 82473

Qy        422 TCAAAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAAC 481
                ||  |||| |||||||   || ||||||||||||||| |||||||||| | || |
Db      82472 ATCAAAGTTTATATGACATAAAAGCCTTCAGAAATGAAGACTCAAAGACCCAGAGAAAGC 82413

Qy        482 TGTCTGTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAG 541
              ||     ||||| |||| |||||  |||||||||||| |||  ||||  ||||    | |
Db      82412 TG-----TTTTTATGCTTAGGTTTAATGAAGAATGGACAGCGGGTAGAAATGT--GATTG 82360
```

```
Qy        542 ACAAAAGGATATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTAT 601
              |||||  |||| | || |||| |  |||| ||||||| |  |||  ||| ||||||  |
Db      82359 GAGAAAGGGTATGGTGTAATGGTGATAGACTGAGGGGGATATCCAGGGAGG-CTGTCTGT 82301

Qy        602 TCAGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGG 661
              ||| ||||||||||  |||||          |||  |||||||  ||||||   ||||||||
Db      82300 TCAAATTCTTCTTGGCCTCTC--------TGTGCAGCATTCCCTCCTCCCAGGTATGGGG 82249

Qy        662 CAGGACTCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATT 721
              |   | ||   ||||| |||||||| |||||||||||| |||| ||||| | | ||
Db      82248 GCAGGACCCTCCGGAATGAAGGTCTTCAAAGGAGAAGGGAGAGAGTGACCTTTCTGG-TT 82190

Qy        722 TTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGA 781
              |||||||||||||   || ||||| |||||| ||||||||| |   | ||||||||  ||
Db      82189 TTATGGCTTGCTTTAAGGGAGAGGGGTTCTATTTTCTATGATCTGCCGTGGGGAAGGAGA 82130

Qy        782 ATTCTGGTTTCTGTGACTTGCTTTCATGAAGA----AAGAGGAGTAAGAGGCAGGAGGGC 837
              ||||| |||| | || ||  ||||   | |||    |||||  ||    || ||||||||
Db      82129 ATTCTAGTTTGTATGGCTCACTTTGGGGGAGAAATGAAGAGAGGTGGAAGACAGGAGGGG 82070

Qy        838 AGGAGATGGTCAGAAAGAGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCA 897
              || ||| ||||||||||       |||||||||| |||||  |||||||||||||  |
Db      82069 AGAAGAAGGTCAGAAAGA-----CTTTGCTTCTGAGGTCTTCCAATCTCCTTTAGTTCAA 82015

Qy        898 AGTACTTCTTAGCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAACAC 957
              ||||   || ||| | || |||  | || || |||   |   |  |||| ||  | | |||
Db      82014 AGTA---CTCAGCGTGCCCAAGTGCCATGCTCTGGGTTGCTGCATTCTAAGTCCCAGCAC 81958

Qy        958 TGCAATCATGCTAAACTCCTCTATGACCTTCA 989
                |   | || | |||    |   | | |||
Db      81957 CACTCACCAGCAAGACTAGGGTGCCATCCTCA 81926
```

```
RESULT 98
AP002754/c
LOCUS       AP002754              185766 bp    DNA     linear   PRI 26-MAR-2003
DEFINITION  Homo sapiens genomic DNA, chromosome 11 clone:RP11-855O10, complete
            sequence.
ACCESSION   AP002754
VERSION     AP002754.5  GI:29243385
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Hattori,M., Toyoda,A., Taylor,T.D., Fujiyama,A., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens genomic DNA
  JOURNAL   Published Only in Database (2000)
REFERENCE   2  (bases 1 to 185766)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     On Mar 25, 2003 this sequence version replaced gi:28189495.
FEATURES             Location/Qualifiers
     source          1. .185766
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q"
                     /clone="RP11-855O10"
ORIGIN

  Query Match             4.1%;  Score 248;  DB 9;  Length 185766;
  Best Local Similarity   69.6%;  Pred. No. 3.8e-44;
```

```
  Matches  458;  Conservative    0;  Mismatches  175;  Indels   25;  Gaps    8;

Qy        286 ATCTCTCCTTCTCTCTCAAAGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTC 345
              ||  ||  |  | ||   ||      ||||||||  ||| |||||  ||  ||| ||| |
Db      77999 ATAACTTTTATTATCAGCAAAGCTGTGTAGGGGTAGAAAGGTGTGCTAACTTTCCTCATT 77940

Qy        346 CATCATAAGGGTCGCAACCAAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACC 405
              |||||||| ||||    |  | |||||||||| |||||||| |||||| ||||| ||| |
Db      77939 CATCATAAAGGTCATGGCTGACACTCCTATAATAAAAGACTGGTTAACAAGAGGAAAGCA 77880

Qy        406 TAACAAATTTATTTAATCAAAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCA 465
              |||||||||||||||||||||||||||| ||||  |||| |||||| ||||||||||||
Db      77879 AAACAAATTTATTTAATCAAAGTTTTACAAGACACAGGAGCCTTCAG-AATGAAGACCCA 77821

Qy        466 AAGACCCAGGGGAAACTGTCTGTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATG 525
              |||| ||||||  || || |||  ||||| |||| || ||| ||| | |   || |||  |
Db      77820 AAGATCCAGGGAAACCTATCT-ATTTTTATGCTTAGATTCTATGCAAAGCAGACAGC-CG 77763

Qy        526 TAGCCATGTAGATTAGACAAAAGGATATGATCTAGTGGTAAAGGACTCAGGGGGAAACAC 585
              |    |   |||  ||||||||  |||||  |  || | ||        | ||| || |
Db      77762 TGTAGAAAGTGATCGGACAAAAGAGCATGATTCATGGGAATAGACTGAGCGCGGATACCC 77703

Qy        586 AGCAAGGCCTGTCTATTCAGATTCTTCTTGATCTCTCTCTCTATGTATAGCATTCTTT 645
              |||||| |||||||||| |||| |||||||  |||||       ||||  ||| || ||
Db      77702 AGCAAGTCCTGTCTATTGAGATGCTTCTTGGCCTCTC--------TGTACGGCACTCCTT 77651

Qy        646 CCTCCTGAGTATGGGGCAGGACTCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAA 705
              ||  | |     |||   |  |  ||||  |||  ||||||||| ||||||||||||| |
Db      77650 CCAGCAGGTATCGGGCAGGACCCTTTCTGGAATAGGGGTCTTCAGGGGAGAAGGGAGAGA 77591

Qy        706 GTG-GCCTTTTTAGATTTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACC 764
              ||| ||||| ||| |||||||||||||||  ||| ||||| |||||||||||||||||
Db      77590 GTGACCCTTTCTAGGTTTTATGGCTTGCTTTAGGGGAGAGGGGTTCTAGTTTCTATGAC- 77532

Qy        765 CATCTTGGGGAAGAGGAATTCTGGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAA 824
              |||||||  | |||||||||| ||| ||  ||  || |   |  |||||||| | |
Db      77531 -----TGGGGAAAAAGAATTCTGGTGTCTATGGATTCCTCTGGGGGCGAAAGAGGGGCAG 77477

Qy        825 GAGGCAGGAGGGCAGGAGATGGTCAGAAAGAGACTTGGCTGCTTCTGAGGGCTTCCGCTC 884
              ||| |||||||   |||||  |||||| ||||| ||| |    | |   | | | |
Db      77476 GAGACAGGAGGATGGGAGAACGTCAGAGAGAGAGTTGAC----TTTTGAGACCTTCCAGT 77421

Qy        885 TCCTTTAGTTCCAAGTACTTCTTAGCATACCAAAGCACTATACTTTGGCATATGGTTT 942
              | |||||||| |||||   || ||| | ||| ||  | |||||||||  ||  ||||
Db      77420 TTCTTTAGTTCAAAGTA---CTCAGCCTGCCATAGTGCCATACTTTGGAGTACCGTTT 77366


RESULT 99
AC025689/c
LOCUS       AC025689              206860 bp    DNA     linear   HTG 07-JUL-2000
DEFINITION  Homo sapiens chromosome 11 clone RP11-5F17, WORKING DRAFT SEQUENCE,
            34 unordered pieces.
ACCESSION   AC025689
VERSION     AC025689.3  GI:7582725
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 206860)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 206860)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Apr 18, 2000 this sequence version replaced gi:7331641.

            ----------------- Genome Center -----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0005F17
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 100%
            Sequencing vector: plasmid; 0%
            Chemistry: Dye-primer ET; 100% of reads
            Chemistry: Dye-terminator Big Dye; 0% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 183324 bases at least Q40
            Consensus quality: 191583 bases at least Q30
            Consensus quality: 196031 bases at least Q20
            Insert size: 188000; agarose-fp
            Insert size: 203560; sum-of-contigs
            Quality coverage: 3.58 in Q20 bases; agarose-fp
            Quality coverage: 3.36 in Q20 bases; sum-of-contigs
            ------------------------------------------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 34 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1624: contig of 1624 bp in length
            *     1625     1724: gap of unknown length
            *     1725     2911: contig of 1187 bp in length
            *     2912     3011: gap of unknown length
            *     3012     4387: contig of 1376 bp in length
            *     4388     4487: gap of unknown length
            *     4488     6397: contig of 1910 bp in length
            *     6398     6497: gap of unknown length
            *     6498     7539: contig of 1042 bp in length
            *     7540     7639: gap of unknown length
            *     7640     9556: contig of 1917 bp in length
            *     9557     9656: gap of unknown length
            *     9657    11324: contig of 1668 bp in length
            *    11325    11424: gap of unknown length
            *    11425    13643: contig of 2219 bp in length
            *    13644    13743: gap of unknown length
            *    13744    16115: contig of 2372 bp in length
            *    16116    16215: gap of unknown length
            *    16216    18834: contig of 2619 bp in length
            *    18835    18934: gap of unknown length
            *    18935    21566: contig of 2632 bp in length
            *    21567    21666: gap of unknown length
            *    21667    24821: contig of 3155 bp in length
            *    24822    24921: gap of unknown length
            *    24922    28240: contig of 3319 bp in length
            *    28241    28340: gap of unknown length
            *    28341    31745: contig of 3405 bp in length
            *    31746    31845: gap of unknown length
            *    31846    36638: contig of 4793 bp in length
            *    36639    36738: gap of unknown length
            *    36739    41768: contig of 5030 bp in length
            *    41769    41868: gap of unknown length
            *    41869    47101: contig of 5233 bp in length
            *    47102    47201: gap of unknown length
            *    47202    53462: contig of 6261 bp in length
            *    53463    53562: gap of unknown length
            *    53563    59492: contig of 5930 bp in length
            *    59493    59592: gap of unknown length
            *    59593    64654: contig of 5062 bp in length
            *    64655    64754: gap of unknown length
            *    64755    71460: contig of 6706 bp in length
            *    71461    71560: gap of unknown length
            *    71561    77423: contig of 5863 bp in length
            *    77424    77523: gap of unknown length
            *    77524    83876: contig of 6353 bp in length
            *    83877    83976: gap of unknown length
            *    83977    91292: contig of 7316 bp in length
```

```
            *    91293    91392: gap of unknown length
            *    91393    99736: contig of 8344 bp in length
            *    99737    99836: gap of unknown length
            *    99837   108225: contig of 8389 bp in length
            *   108226   108325: gap of unknown length
            *   108326   115325: contig of 7000 bp in length
            *   115326   115425: gap of unknown length
            *   115426   123536: contig of 8111 bp in length
            *   123537   123636: gap of unknown length
            *   123637   135165: contig of 11529 bp in length
            *   135166   135265: gap of unknown length
            *   135266   146017: contig of 10752 bp in length
            *   146018   146117: gap of unknown length
            *   146118   159245: contig of 13128 bp in length
            *   159246   159345: gap of unknown length
            *   159346   171827: contig of 12482 bp in length
            *   171828   171927: gap of unknown length
            *   171928   187764: contig of 15837 bp in length
            *   187765   187864: gap of unknown length
            *   187865   206860: contig of 18996 bp in length.
FEATURES             Location/Qualifiers
     source          1. .206860
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /clone="RP11-5F17"
     misc_feature    1. .1624
                     /note="assembly_name:Contig6"
     misc_feature    1725. .2911
                     /note="assembly_name:Contig7"
     misc_feature    3012. .4387
                     /note="assembly_name:Contig8"
     misc_feature    4488. .6397
                     /note="assembly_name:Contig9"
     misc_feature    6498. .7539
                     /note="assembly_name:Contig10"
     misc_feature    7640. .9556
                     /note="assembly_name:Contig11"
     misc_feature    9657. .11324
                     /note="assembly_name:Contig12"
     misc_feature    11425. .13643
                     /note="assembly_name:Contig13"
     misc_feature    13744. .16115
                     /note="assembly_name:Contig14"
     misc_feature    16216. .18834
                     /note="assembly_name:Contig15"
     misc_feature    18935. .21566
                     /note="assembly_name:Contig16"
     misc_feature    21667. .24821
                     /note="assembly_name:Contig17"
     misc_feature    24922. .28240
                     /note="assembly_name:Contig18"
     misc_feature    28341. .31745
                     /note="assembly_name:Contig19"
     misc_feature    31846. .36638
                     /note="assembly_name:Contig20"
     misc_feature    36739. .41768
                     /note="assembly_name:Contig21"
     misc_feature    41869. .47101
                     /note="assembly_name:Contig22"
     misc_feature    47202. .53462
                     /note="assembly_name:Contig23"
     misc_feature    53563. .59492
                     /note="assembly_name:Contig24"
     misc_feature    59593. .64654
                     /note="assembly_name:Contig25
                     clone_end:SP6
                     vector_side:right"
     misc_feature    64755. .71460
                     /note="assembly_name:Contig26"
     misc_feature    71561. .77423
                     /note="assembly_name:Contig27"
     misc_feature    77524. .83876
                     /note="assembly_name:Contig28"
     misc_feature    83977. .91292
                     /note="assembly_name:Contig29"
     misc_feature    91393. .99736
                     /note="assembly_name:Contig30
                     clone_end:T7
                     vector_side:right"
     misc_feature    99837. .108225
                     /note="assembly_name:Contig31"
     misc_feature    108326. .115325
                     /note="assembly_name:Contig32"
     misc_feature    115426. .123536
                     /note="assembly_name:Contig33"
     misc_feature    123637. .135165
                     /note="assembly_name:Contig34"
     misc_feature    135266. .146017
                     /note="assembly_name:Contig35"
     misc_feature    146118. .159245
                     /note="assembly_name:Contig36"
     misc_feature    159346. .171827
                     /note="assembly_name:Contig37"
     misc_feature    171928. .187764
                     /note="assembly_name:Contig38"
     misc_feature    187865. .206860
                     /note="assembly_name:Contig39"
ORIGIN

  Query Match             4.1%;  Score 248;  DB 2;  Length 206860;
  Best Local Similarity   69.6%;  Pred. No. 3.7e-44;
  Matches  458;  Conservative    0;  Mismatches  175;  Indels   25;  Gaps    8;

Qy        286 ATCTCTCCTTCTCTCTCAAAGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTC 345
              ||  ||  |  | ||   ||       ||||||||  ||| |||||  ||  ||| ||| |
Db     180352 ATAACTTTTATTATCAGCAAAGCTGTGTAGGGGTAGAAAGGTGTGCTAACTTTCCTCATT 180293

Qy        346 CATCATAAGGGTCGCAACCAAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACC 405
              |||||||| ||||    |  | |||||||| |||||||| |||||| ||||| ||| |
Db     180292 CATCATAAAGGTCATGGCTGACACTCCTATAATAAAAGACTGGTTAACAAGAGGAAAGCA 180233

Qy        406 TAACAAATTTATTTAATCAAAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCA 465
              |||||||||||||||||||||||||||| ||||  |||| |||||| ||||||||||||
Db     180232 AAACAAATTTATTTAATCAAAGTTTTACAAGACACAGGAGCCTTCAG-AATGAAGACCCA 180174

Qy        466 AAGACCCAGGGGAAACTGTCTGTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATG 525
              |||| |||||| || || |||  ||||| |||| || ||| ||| | |   || |||  |
Db     180173 AAGATCCAGGGAAACCTATCT-ATTTTTATGCTTAGATTCTATGCAAAGCAGACAGC-CG 180116

Qy        526 TAGCCATGTAGATTAGACAAAAGGATATGATCTAGTGGTAAAGGACTCAGGGGGAAACAC 585
              |    |   |||  ||||||||   |||||  |  || | ||       | ||| || |
Db     180115 TGTAGAAAGTGATCGGACAAAAGAGCATGATTCATGGGAATAGACTGAGCGCGGATACCC 180056

Qy        586 AGCAAGGCCTGTCTATTCAGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTT 645
              ||||||  |||||||||| |||| |||||||  |||||         ||||  ||| || ||
Db     180055 AGCAAGTCCTGTCTATTGAGATGCTTCTTGGCCTCTC--------TGTACGGCACTCCTT 180004

Qy        646 CCTCCTGAGTATGGGGCAGGACTCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAA 705
              ||  | |      |||   |  |||| |||  |||||||| ||||||||||||| | |
Db     180003 CCAGCAGGTATCGGGCAGGACCCTTTCTGGAATAGGGGTCTTCAGGGGAGAAGGGAGAGA 179944

Qy        706 GTG-GCCTTTTTAGATTTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACC 764
              |||  ||||| ||| ||||||||||||||  ||| ||||| ||||||||||||||||||
Db     179943 GTGACCCTTTCTAGGTTTTATGGCTTGCTTTAGGGGAGAGGGGTTCTAGTTTCTATGAC- 179885

Qy        765 CATCTTGGGGAAGAGGAATTCTGGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAA 824
              |||||||  | |||||||||| ||| ||  || || |    |  |||||||| | |
Db     179884 -----TGGGGAAAAAGAATTCTGGTGTCTATGGATTCCTCTGGGGGCGAAAGAGGGGCAG 179830

Qy        825 GAGGCAGGAGGGCAGGAGATGGTCAGAAAGAGACTTGGCTGCTTCTGAGGGCTTCCGCTC 884
              ||| |||||||   |||||  |||||| ||||| ||| |    | |   | | | |
Db     179829 GAGACAGGAGGATGGGAGAACGTCAGAGAGAGAGTTGAC----TTTTGAGACCTTCCAGT 179774

Qy        885 TCCTTTAGTTCCAAGTACTTCTTAGCATACCAAAGCACTATACTTTGGCATATGGTTT 942
```

```
                 | ||||||||| |||||   || ||| | ||| ||  | |||||||||  ||  ||||
Db     179773 TTCTTTAGTTCAAAGTA---CTCAGCCTGCCATAGTGCCATACTTTGGAGTACCGTTT 179719


RESULT 100
AC087848/c
LOCUS       AC087848               209545 bp    DNA     linear   HTG 13-JUN-2001
DEFINITION  Homo sapiens chromosome 11 clone RP11-5F17 map 11, WORKING DRAFT
            SEQUENCE, 8 unordered pieces.
ACCESSION   AC087848
VERSION     AC087848.3  GI:14389343
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 209545)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 11, clone RP11-5F17
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 209545)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
            Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
            Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
            Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
            Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
            Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-JAN-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jun 13, 2001 this sequence version replaced gi:13487976.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L11772
                Center clone name: 5_F_17
            -------------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 206008 bases at least Q40
                Consensus quality: 208005 bases at least Q30
                Consensus quality: 208534 bases at least Q20
                Insert size: 200000; agarose-fp
                Insert size: 208845; sum-of-contigs
                Quality coverage: 8.7 in Q20 bases; agarose-fp
                Quality coverage: 8.3 in Q20 bases; sum-of-contigs
            --------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    17945: contig of 17945 bp in length
            *    17946    18045: gap of 100 bp
            *    18046    70195: contig of 52150 bp in length
            *    70196    70295: gap of 100 bp
            *    70296    81731: contig of 11436 bp in length
            *    81732    81831: gap of 100 bp
            *    81832   102026: contig of 20195 bp in length
            *   102027   102126: gap of 100 bp
            *   102127   122861: contig of 20735 bp in length
            *   122862   122961: gap of 100 bp
            *   122962   157365: contig of 34404 bp in length
            *   157366   157465: gap of 100 bp
            *   157466   207248: contig of 49783 bp in length
            *   207249   207348: gap of 100 bp
            *   207349   209545: contig of 2197 bp in length.
FEATURES             Location/Qualifiers
     source          1. .209545
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11"
                     /clone="RP11-5F17"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .17945
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    18046. .70195
                     /note="assembly_fragment"
     misc_feature    70296. .81731
                     /note="assembly_fragment"
     misc_feature    81832. .102026
                     /note="assembly_fragment"
     misc_feature    102127. .122861
                     /note="assembly_fragment"
     misc_feature    122962. .157365
                     /note="assembly_fragment"
     misc_feature    157466. .207248
                     /note="assembly_fragment"
     misc_feature    207349. .209545
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
ORIGIN

  Query Match             4.1%;  Score 248;  DB 2;  Length 209545;
  Best Local Similarity   69.6%;  Pred. No. 3.7e-44;
  Matches  458;  Conservative    0;  Mismatches  175;  Indels   25;  Gaps    8;

Qy        286 ATCTCTCCTTCTCTCTCAAAGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTC 345
              ||  ||  |  | ||   ||      ||||||||  ||| |||||  ||  ||| ||| |
Db     146087 ATAACTTTTATTATCAGCAAAGCTGTGTAGGGGTAGAAAGGTGTGCTAACTTTCCTCATT 146028

Qy        346 CATCATAAGGGTCGCAACCAAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACC 405
              |||||||| ||||    |  | |||||||| |||||||| |||||| ||||| ||| |
Db     146027 CATCATAAAGGTCATGGCTGACACTCCTATAATAAAAGACTGGTTAACAAGAGGAAAGCA 145968

Qy        406 TAACAAATTTATTTAATCAAAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCA 465
              |||||||||||||||||||||||||||| ||||  |||| |||||| ||||||||||||
Db     145967 AAACAAATTTATTTAATCAAAGTTTTACAAGACACAGGAGCCTTCAG-AATGAAGACCCA 145909

Qy        466 AAGACCCAGGGGAAACTGTCTGTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATG 525
              |||| |||||| || || |||  ||||| |||| || ||| ||| | |   || ||| |
Db     145908 AAGATCCAGGGAAACCTATCT-ATTTTTATGCTTAGATTCTATGCAAAGCAGACAGC-CG 145851

Qy        526 TAGCCATGTAGATTAGACAAAAGGATATGATCTAGTGGTAAAGGACTCAGGGGGAAACAC 585
              |    |   ||| ||||||||  |||||  |  || | ||        | ||| || |
Db     145850 TGTAGAAAGTGATCGGACAAAAGAGCATGATTCATGGGAATAGACTGAGCGCGGATACCC 145791
```

```
Qy        586 AGCAAGGCCTGTCTATTCAGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTT 645
              |||||| |||||||||| |||| |||||||  |||||         ||||  ||| || ||
Db     145790 AGCAAGTCCTGTCTATTGAGATGCTTCTTGGCCTCTC--------TGTACGGCACTCCTT 145739

Qy        646 CCTCCTGAGTATGGGGCAGGACTCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAA 705
              ||  | |     |||   |  |  ||||  |||  ||||||||| ||||||||||||| |
Db     145738 CCAGCAGGTATCGGGCAGGACCCTTTCTGGAATAGGGGTCTTCAGGGGAGAAGGGAGAGA 145679

Qy        706 GTG-GCCTTTTTAGATTTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACC 764
              |||  ||||| ||| ||||||||||||||||  ||| ||||| ||||||||||||||||
Db     145678 GTGACCCTTTCTAGGTTTTATGGCTTGCTTTAGGGGAGAGGGGTTCTAGTTTCTATGAC- 145620

Qy        765 CATCTTGGGGAAGAGGAATTCTGGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAA 824
                |||||| | |||||||||| ||| ||  || || |   |  |||||||| | |
Db     145619 -----TGGGGAAAAAGAATTCTGGTGTCTATGGATTCCTCTGGGGGCGAAAGAGGGGCAG 145565

Qy        825 GAGGCAGGAGGGCAGGAGATGGTCAGAAAGAGACTTGGCTGCTTCTGAGGGCTTCCGCTC 884
              ||| |||||||   |||||  |||||| ||||| ||| |    | |   | | | |
Db     145564 GAGACAGGAGGATGGGAGAACGTCAGAGAGAGAGTTGAC----TTTTGAGACCTTCCAGT 145509

Qy        885 TCCTTTAGTTCCAAGTACTTCTTAGCATACCAAAGCACTATACTTTGGCATATGGTTT 942
              | ||||||||| |||||   || ||| | ||| ||  | |||||||||  ||  ||||
Db     145508 TTCTTTAGTTCAAAGTA---CTCAGCCTGCCATAGTGCCATACTTTGGAGTACCGTTT 145454
```

Search completed: April 21, 2005, 17:04:48
Job time : 17116.6 secs

This Page Blank (uspto)

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 20, 2005, 22:52:20 ; Search time 9950.71 Seconds
(without alignments)
19011.657 Million cell updates/sec

Title: US-09-939-209A-3_COPY_1_4970
Perfect score: 4970
Sequence: 1 agttcaagaccagcctgagc.....gtggctgaatgtctgccccc 4970

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_htc:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	347.8	7.0	976	9	CL830166	CL830166 OR_CBa005
	2	345.4	6.9	863	9	CL840224	CL840224 OR_CBa006
	3	342	6.9	835	9	CL800294	CL800294 OR_CBa001
	4	341.4	6.9	828	8	CC107667	CC107667 NDL.37F6.
c	5	340.4	6.8	976	9	CL830166	CL830166 OR_CBa005
c	6	338.8	6.8	835	9	CL800294	CL800294 OR_CBa001
c	7	338.4	6.8	863	9	CL840224	CL840224 OR_CBa006
c	8	338.2	6.8	828	8	CC107667	CC107667 NDL.37F6.
c	9	337.4	6.8	620	7	CF626074	CF626074 zmrws05_0
c	10	337.4	6.8	632	7	CF634146	CF634146 zmrww00_0
c	11	337.4	6.8	883	9	CW016544	CW016544 ZMMBLc000
	12	337.2	6.8	883	9	CW016544	CW016544 ZMMBLc000
	13	336.6	6.8	725	8	AQ917102	AQ917102 T233131b
	14	336.4	6.8	827	8	CC067623	CC067623 CSU-K33r.
	15	336.2	6.8	590	9	CR294134	CR294134 Medicago
c	16	335.8	6.8	827	8	CC067623	CC067623 CSU-K33r.
	17	335.2	6.7	620	7	CF626074	CF626074 zmrws05_0
	18	335.2	6.7	632	7	CF634146	CF634146 zmrww00_0
	19	335.2	6.7	717	4	BI505681	BI505681 BB170032B
c	20	334.8	6.7	590	9	CR294134	CR294134 Medicago
c	21	334.2	6.7	717	4	BI505681	BI505681 BB170032B
c	22	333.8	6.7	725	8	AQ917102	AQ917102 T233131b
c	23	333	6.7	744	9	CL623140	CL623140 OR_BBa001
c	24	332.6	6.7	512	8	BZ796807	BZ796807 PUGDF25TD
c	25	331.6	6.7	707	9	CL755178	CL755178 OR_BBa012
	26	331.2	6.7	784	9	CL934691	CL934691 OA_ABa004
	27	331	6.7	512	8	BZ796807	BZ796807 PUGDF25TD
c	28	330.4	6.6	668	9	BX145650	BX145650 Danio rer
	29	330.4	6.6	707	9	CL755178	CL755178 OR_BBa012
c	30	330.2	6.6	496	6	CA916667	CA916667 PCSC21302
c	31	329.8	6.6	751	7	CF924719	CF924719 DD005F11.
	32	329	6.6	668	9	BX145650	BX145650 Danio rer
c	33	328.4	6.6	784	9	CL934691	CL934691 OA_ABa004
c	34	328.2	6.6	587	9	CL776144	CL776144 OR_BBa009
	35	327.2	6.6	751	7	CF924719	CF924719 DD005F11.
	36	326.2	6.6	496	6	CA916667	CA916667 PCSC21302
	37	324.2	6.5	744	9	CL623140	CL623140 OR_BBa001
	38	323.8	6.5	587	9	CL776144	CL776144 OR_BBa009
c	39	322.6	6.5	491	9	CL830207	CL830207 OR_CBa005
	40	318.2	6.4	845	8	CC099198	CC099198 CSU-K34.1
c	41	317	6.4	845	8	CC099198	CC099198 CSU-K34.1
	42	315	6.3	816	9	CL855754	CL855754 OR_CBa008
	43	314.6	6.3	570	4	BI506858	BI506858 BB170024B
c	44	313.2	6.3	570	4	BI506858	BI506858 BB170024B
c	45	311.4	6.3	816	9	CL855754	CL855754 OR_CBa008

ALIGNMENTS

RESULT 1
CL830166
LOCUS CL830166 976 bp DNA linear GSS 09-AUG-2004
DEFINITION OR_CBa0052B14.f OR_CBa Oryza rufipogon genomic clone OR_CBa0052B14 5', genomic survey sequence.
ACCESSION CL830166
VERSION CL830166.1 GI:51075776
KEYWORDS GSS.
SOURCE Oryza rufipogon
ORGANISM Oryza rufipogon
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 976)
AUTHORS Kim,H., Yu,Y., Wissotski,M., Yost,D., Stum,D., Rao,K., Luo,M., Jetty,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
TITLE OMAP project
JOURNAL Unpublished (2004)
COMMENT Contact: Rod A. Wing
Arizona Genomics Institute
University of Arizona
Forbes Building Room 303, Tucson, AZ 85721-0036, USA
Tel: 520 626 9595
Fax: 520 621 1259
Email: http://genome.arizona.edu
PCR PRimers
FORWARD: TAA TAC GAC TCA CTA TAG GG
BACKWARD: CAC TCA TTA GGC ACC CCA
Plate: 0052 row: B column: 14
Seq primer: TAA TAC GAC TCA CTA TAG GG
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .976
/organism="Oryza rufipogon"
/mol_type="genomic DNA"
/db_xref="taxon:4529"
/clone="OR_CBa0052B14"
/tissue_type="young leaves"
/dev_stage="2 week old seedlings"
/lab_host="DH10B T1 phage resistant"
/clone_lib="OR_CBa"
/note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII; drk treated 36 hrs before harvest"
ORIGIN

Query Match 7.0%; Score 347.8; DB 9; Length 976;

```
  Best Local Similarity   69.5%;  Pred. No. 1.4e-49;
  Matches  486;  Conservative    0;  Mismatches  212;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db        188 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 247

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db        248 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 307

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        308 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 367

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||| | ||||||| | ||||| | |||||
Db        368 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 427

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db        428 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 487

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db        488 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 547

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db        548 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 607

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db        608 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 667

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| |||||||||||          | |||          | |   ||    |    |||
Db        668 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 727

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
              |   | | | |  | | |    | |||    | |   | || |      |       | |
Db        728 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 787

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885
              ||        | ||| ||| || | ||   | |  ||| | |||  ||  ||   |  ||
Db        788 TATATATATATATATATATATATATATATATATATATATATAATATATATATATATATAT 847

Qy        886 CACTATATATATATAGTGAAAGCAGTGTAAACCTAAA 924
              | ||||||||||||||   | |    | || |  || |
Db        848 ATATATATATATATATATATATATATATATATATATA 886


RESULT 2
CL840224
LOCUS       CL840224                863 bp    DNA     linear   GSS 09-AUG-2004
DEFINITION  OR_CBa0067C10.f OR_CBa Oryza rufipogon genomic clone OR_CBa0067C10
            5', genomic survey sequence.
ACCESSION   CL840224
VERSION     CL840224.1  GI:51085834
KEYWORDS    GSS.
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 863)
  AUTHORS   Kim,H., Yu,Y., Wissotski,M., Yost,D., Stum,D., Rao,K., Luo,M.,
            Jetty,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and
            Wing,R.
  TITLE     OMAP project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0067  row: C  column: 10
            Seq primer: TAA TAC GAC TCA CTA TAG GG
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .863
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
                     /clone="OR_CBa0067C10"
                     /tissue_type="young leaves"
                     /dev_stage="2 week old seedlings"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OR_CBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII;
                     drk treated 36 hrs before harvest"
ORIGIN

  Query Match             6.9%;  Score 345.4;  DB 9;  Length 863;
  Best Local Similarity   70.3%;  Pred. No. 3.6e-49;
  Matches  477;  Conservative    0;  Mismatches  201;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db        152 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 211

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db        212 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 271

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        272 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 331

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db        332 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 391

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db        392 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 451

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db        452 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 511

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db        512 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 571

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db        572 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 631

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| |||||||||||          | |||          | |   ||    |    |||
Db        632 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 691

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
              |   | | | |  | | |    | |||    | |   | || |      |   |   | |
Db        692 TATATATATATATATATATATATATATATATATATATATATATATATATATATATTTATATA 751

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885
              ||        | ||| ||| || | ||   | |  ||| | ||       |  | ||  ||
Db        752 TATATATATATATATATATATATATATATATATATATATATATATATATATATAATATAT 811
```

```
Qy        886 CACTATATATATATATAGT 904
              ||||||||||||||| |
Db        812 ATATATATATATATATAAT 830


RESULT 3
CL800294
LOCUS       CL800294                835 bp    DNA     linear   GSS 06-AUG-2004
DEFINITION  OR_CBa0011N09.r OR_CBa Oryza rufipogon genomic clone OR_CBa0011N09
            3', genomic survey sequence.
ACCESSION   CL800294
VERSION     CL800294.1  GI:51025055
KEYWORDS    GSS.
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 835)
  AUTHORS   Kim,H., Yu,Y., Wissotski,M., Yost,D., Stum,D., Rao,K., Luo,M.,
            Jetty,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and
            Wing,R.
  TITLE     OMAP project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0011  row: N  column: 09
            Seq primer: CAC TCA TTA GGC ACC CCA
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .835
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
                     /clone="OR_CBa0011N09"
                     /tissue_type="young leaves"
                     /dev_stage="2 week old seedlings"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OR_CBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII;
                     drk treated 36 hrs before harvest"
ORIGIN

  Query Match             6.9%;  Score 342;  DB 9;  Length 835;
  Best Local Similarity   67.4%;  Pred. No. 1.4e-48;
  Matches  511;  Conservative    0;  Mismatches  245;  Indels    2;  Gaps    2;

Qy        212 AAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACATATAT 271
              || |  |  |  || | |||||||||||||||||||||||| | | | | | | ||||||
Db         67 AACAGTATTAAACAGTATATATATATATATATATATATATATATATATATATATATATAT 126

Qy        272 ATATACACATATATATACACATATATACATATATACACATATATACACATATATATACAC 331
              ||||| | |||||||| | ||||||| ||||||| | ||||||| | |||||||| |
Db        127 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 186

Qy        332 ATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACACAC 391
              ||| |||| || | |||||||| | |||| |||| | ||||||| | ||||||| | |
Db        187 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 246

Qy        392 ATATATACACATATATACACACATATATACACATATATACACATATATACACATATACAC 451
              ||||||| | ||||||| | | |||||||| | ||||||| | ||||||| | ||||| |
Db        247 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 306

Qy        452 ATATATACACATATATACATATATACACATATATATA-ATATACACACATATATATACAC 510
              ||||||| | ||||||| ||||||| | ||||||||| ||||| | | |||||||||| |

Db        307 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 366

Qy        511 ATATATACACACATATATACACATATATACACATATATATACACATATATACACATATAT 570
              ||||||| | | ||||||| | ||||||| | |||||||||| | ||||||| | ||||||
Db        367 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 426

Qy        571 ACATATATACACATATATATACATATATACACATATATACATATATACACATATATACAT 630
              | ||||||| | ||||||||| ||||||| | ||||||| ||||||| | ||||||| ||
Db        427 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 486

Qy        631 ATATACACACATATATACACATACATATACACACACATAGATATACATATATATACACAT 690
              ||||| | | ||||||| | ||| ||||| | | | ||| ||||| |||||||||| | ||
Db        487 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 546

Qy        691 ATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGA 750
              |||||||  |||||||| ||||||||||          |  |||          |  |
Db        547 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 606

Qy        751 TTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTT 810
              ||    |  ||| |  | | | |  | | |   | |||   | |  | || |
Db        607 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 666

Qy        811 CAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA-AG 869
              |   | |||       | ||| ||| || | ||  |  | . ||| | || |
Db        667 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 726

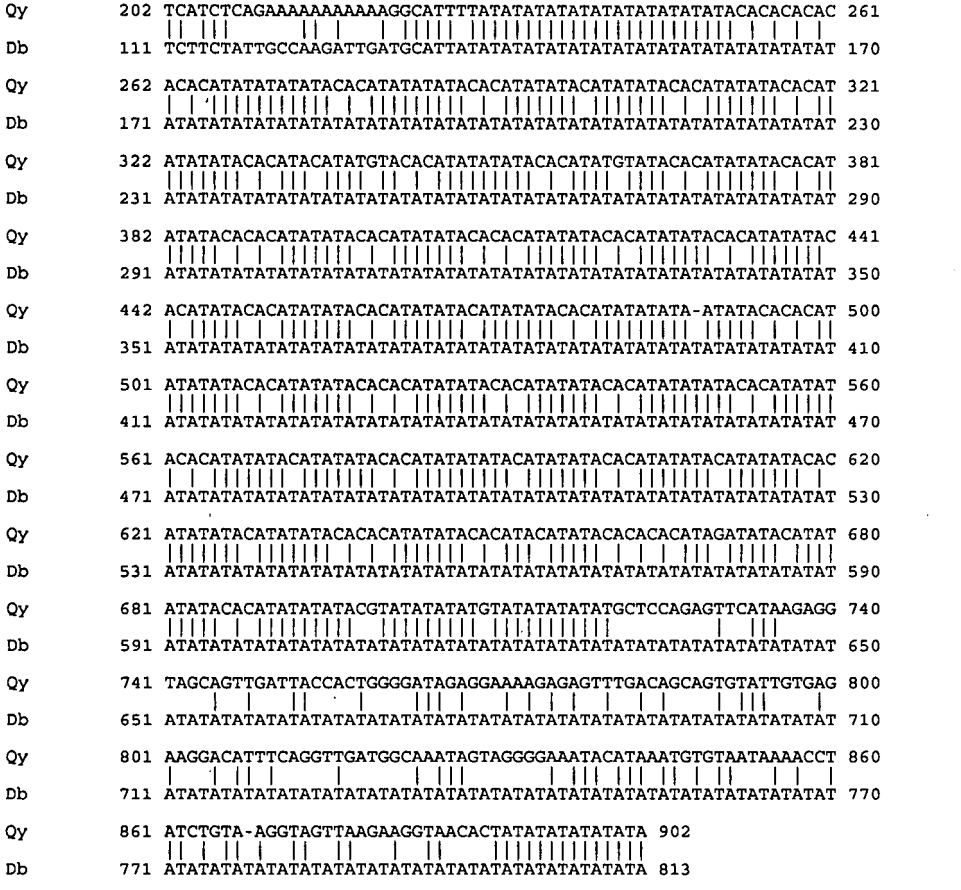
Qy        870 GTAGTTAAGAAGGTAACACTATATATATATATAGTGAAAGCAGTGTAAACCTAAAGGATG 929
              ||  ||   |  ||    |||||||||||||   | |    | || |  || |
Db        727 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 786

Qy        930 GGCCAAGGATTTAAATGTTATAGAAGAATGGCTAAGAT 967
                |  || || || | ||| |   ||   || ||
Db        787 ATATATATATATATATATAATATATATATAAAAAATAT 824


RESULT 4
CC107667
LOCUS       CC107667                828 bp    DNA     linear   GSS 16-APR-2003
DEFINITION  NDL.37F6.T7 Notre Dame Liverpool Aedes aegypti genomic clone
            NDL.37F6, genomic survey sequence.
ACCESSION   CC107667
VERSION     CC107667.1  GI:29976722
KEYWORDS    GSS.
SOURCE      Aedes aegypti (yellow fever mosquito)
  ORGANISM  Aedes aegypti
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Aedes;
            Stegomyia.
REFERENCE   1  (bases 1 to 828)
  AUTHORS   Loftus,B., Shetty,J., Knudson,D. and Severson,D.
  TITLE     BAC end sequencing of Aedes aegypti
  JOURNAL   Unpublished (2003)
COMMENT     Other_GSSs: NDL.37F6.SP6
            Contact: Brendan Loftus
            Department of Eukaryotic Genomics
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-3543
            Fax: 301-838-0208
            Email: enta@tigr.org
            Library was provided by David Severson
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .828
                     /organism="Aedes aegypti"
                     /mol_type="genomic DNA"
                     /strain="Liverpool"
                     /db_xref="taxon:7159"
                     /clone="NDL.37F6"
                     /clone_lib="Notre Dame Liverpool"
                     /note="Vector: pECBAC1; Site_1: Hind III; The library was
                     prepared from whole body tissue of newly hatched L1 larvae
```

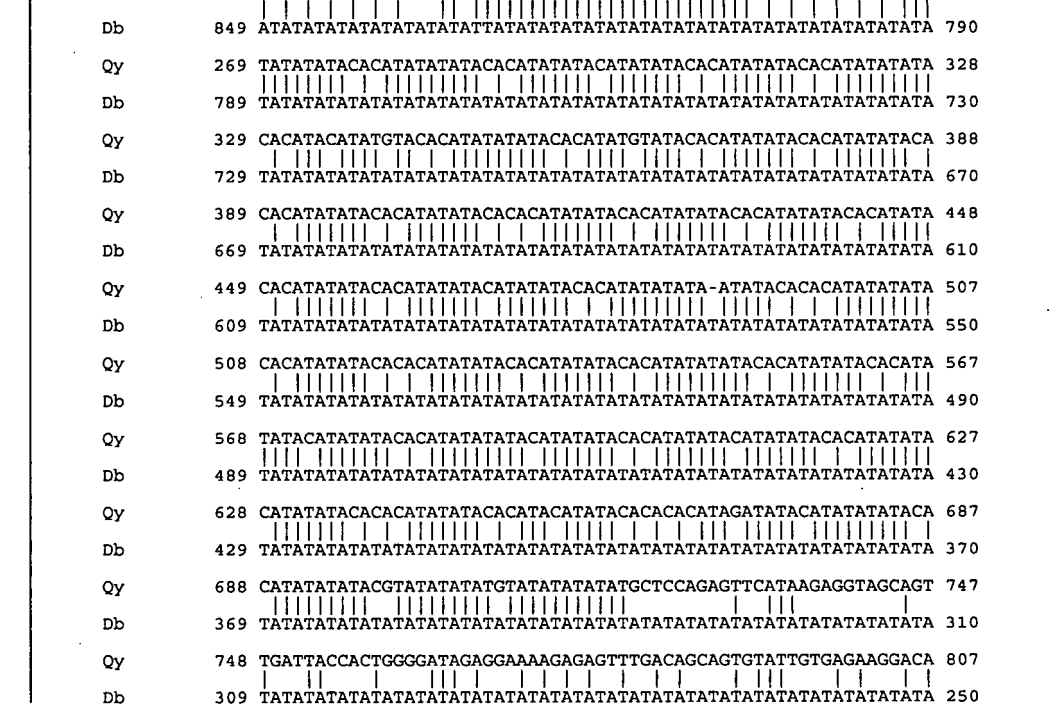
by David Severson at the University of Notre Dame and
Hongbin Zhang"
ORIGIN

Query Match 6.9%; Score 341.4; DB 8; Length 828;
Best Local Similarity 69.7%; Pred. No. 1.8e-48;
Matches 490; Conservative 0; Mismatches 211; Indels 2; Gaps 2;



RESULT 5
CL830166/c
LOCUS CL830166 976 bp DNA linear GSS 09-AUG-2004
DEFINITION OR_CBa0052B14.f OR_CBa Oryza rufipogon genomic clone OR_CBa0052B14 5', genomic survey sequence.
ACCESSION CL830166
VERSION CL830166.1 GI:51075776
KEYWORDS GSS.
SOURCE Oryza rufipogon
ORGANISM Oryza rufipogon
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 976)
AUTHORS Kim,H., Yu,Y., Wissotski,M., Yost,D., Stum,D., Rao,K., Luo,M., Jetty,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
TITLE OMAP project
JOURNAL Unpublished (2004)
COMMENT Contact: Rod A. Wing
Arizona Genomics Institute
University of Arizona
Forbes Building Room 303, Tucson, AZ 85721-0036, USA
Tel: 520 626 9595
Fax: 520 621 1259
Email: http://genome.arizona.edu
PCR PRimers
FORWARD: TAA TAC GAC TCA CTA TAG GG
BACKWARD: CAC TCA TTA GGC ACC CCA
Plate: 0052 row: B column: 14
Seq primer: TAA TAC GAC TCA CTA TAG GG
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .976
/organism="Oryza rufipogon"
/mol_type="genomic DNA"
/db_xref="taxon:4529"
/clone="OR_CBa0052B14"
/tissue_type="young leaves"
/dev_stage="2 week old seedlings"
/lab_host="DH10B T1 phage resistant"
/clone_lib="OR_CBa"
/note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII; drk treated 36 hrs before harvest"
ORIGIN

Query Match 6.8%; Score 340.4; DB 9; Length 976;
Best Local Similarity 68.9%; Pred. No. 2.6e-48;
Matches 495; Conservative 0; Mismatches 221; Indels 2; Gaps 2;



```
Qy        808 TTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA 867
              | |      |       | |||       | ||| ||| || | ||   | |  ||| | ||
Db        249 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 190

Qy        868 -AGGTAGTTAAGAAGGTAACACTATATATATATATAGTGAAAGCAGTGTAAACCTAAA 924
               |  ||  ||   | ||    ||||||||||||||   | |    | || |  || |
Db        189 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 132


RESULT 6
CL800294/c
LOCUS       CL800294                 835 bp    DNA     linear   GSS 06-AUG-2004
DEFINITION  OR_CBa0011N09.r OR_CBa Oryza rufipogon genomic clone OR_CBa0011N09
            3', genomic survey sequence.
ACCESSION   CL800294
VERSION     CL800294.1  GI:51025055
KEYWORDS    GSS.
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 835)
  AUTHORS   Kim,H., Yu,Y., Wissotski,M., Yost,D., Stum,D., Rao,K., Luo,M.,
            Jetty,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and
            Wing,R.
  TITLE     OMAP project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0011  row: N  column: 09
            Seq primer: CAC TCA TTA GGC ACC CCA
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .835
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
                     /clone="OR_CBa0011N09"
                     /tissue_type="young leaves"
                     /dev_stage="2 week old seedlings"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OR_CBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII;
                     drk treated 36 hrs before harvest"
ORIGIN

  Query Match             6.8%;  Score 338.8;  DB 9;  Length 835;
  Best Local Similarity   69.5%;  Pred. No. 4.9e-48;
  Matches  488;  Conservative    0;  Mismatches  212;  Indels    2;  Gaps    2;

Qy        225 ATTTTATATATATATATATATATATATATACACACACACACACATATATATATACACATATA 284
              || |||||||||||||||||||||||| | | | | | |||||||||||| | |||||
Db        809 ATATTATATATATATATATATATATATATATATATATATATATATATATATATATATATA 750

Qy        285 TATACACATATATACATATATACACATATATACACATATATATACACATACATATGTACA 344
              |||| | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || |
Db        749 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 690

Qy        345 CATATATATACACATATGTATACACATATATACACATATATACACACATATATACACATA 404
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||
Db        689 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 630

Qy        405 TATACACACATATATACACATATATACACATATATACACATATACACATATATACACATA 464
              |||| | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||
Db        629 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 570

Qy        465 TATACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACA 523
              |||| ||||||| | |||||||||| ||||| | | ||||||||| | ||||||| | | |
Db        569 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 510

Qy        524 TATATACACATATATACACATATATATACACATATATACACATATATACATATATACACA 583
              |||||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |
Db        509 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 450

Qy        584 TATATATACATATATACACATATATACATATATACACATATATACATATATACACACATA 643
              |||||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||
Db        449 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 390

Qy        644 TATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATA 703
              |||| | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||
Db        389 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 330

Qy        704 TATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGA 763
              ||||| ||||||||||||          | |||         | |   ||    |   |
Db        329 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 270

Qy        764 TAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCA 823
              || |   | | | |   | | |   | |||    | |   | || |     |      |
Db        269 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 210

Qy        824 AATAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA-AGGTAGTTAAGAAGG 882
              |||     | ||| ||| || | ||   | |  ||| | || |  ||  ||   |
Db        209 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 150

Qy        883 TAACACTATATATATATAGTGAAAGCAGTGTAAACCTAAA 924
              ||    |||||||||||||   | |    | || |  || |
Db        149 TATATATATATATATATATATATATATATATATATATATA 108


RESULT 7
CL840224/c
LOCUS       CL840224                 863 bp    DNA     linear   GSS 09-AUG-2004
DEFINITION  OR_CBa0067C10.f OR_CBa Oryza rufipogon genomic clone OR_CBa0067C10
            5', genomic survey sequence.
ACCESSION   CL840224
VERSION     CL840224.1  GI:51085834
KEYWORDS    GSS.
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 863)
  AUTHORS   Kim,H., Yu,Y., Wissotski,M., Yost,D., Stum,D., Rao,K., Luo,M.,
            Jetty,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and
            Wing,R.
  TITLE     OMAP project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0067  row: C  column: 10
            Seq primer: TAA TAC GAC TCA CTA TAG GG
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .863
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
```

```
                     /clone="OR_CBa0067C10"
                     /tissue_type="young leaves"
                     /dev_stage="2 week old seedlings"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OR_CBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII;
                     drk treated 36 hrs before harvest"
ORIGIN

  Query Match             6.8%;  Score 338.4;  DB 9;  Length 863;
  Best Local Similarity   69.1%;  Pred. No. 5.8e-48;
  Matches  492;  Conservative    0;  Mismatches  216;  Indels    4;  Gaps    2;

Qy        204 ATCTCAGAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACAC 263
              ||   | | | | | | |   || ||||||||||||||||||||||||||| | | | | |
Db        830 ATTATATATATATATATATATATATTATATATATATATATATATATATATATATATAT 771

Qy        264 ACATATATATATACACATATATATACACATATATACATATATACACATATATACACATAT 323
              | |||||||||||| | ||||||||| | ||||||| ||||||| | ||||||| | ||||
Db        770 ATATATATATATATATATATATATATAAATATATATATATATATATATATATATATAT 711

Qy        324 ATATACACATACATATGTACACATATATATACACATATGTATACACATATATACACATAT 383
              ||||| | ||| |||| || | ||||||||| | |||| |||| | ||||||| | ||||
Db        710 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 651

Qy        384 ATACACACATATATACACATATATACACACATATATACACATATATACACATATATACAC 443
              ||| | | ||||||| | ||||||| | | ||||||| | ||||||| | ||||||| |
Db        650 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 591

Qy        444 ATATACACATATATACACATATATACATATATACACATATATATA-ATATACACACATAT 502
              ||||| | ||||||| | ||||||| ||||||| | ||||||||| ||||| | | ||||
Db        590 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 531

Qy        503 ATATACACATATATACACACATATATACACATATATACACATATATATACACATATATAC 562
              ||||| | ||||||| | | ||||||| | ||||||||| | ||||||||| | |||||||
Db        530 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 471

Qy        563 ACATATATACATATATACACATATATATACATATATACACATATATACATATATACACAT 622
              | ||||||| ||||||| | ||||||||| ||||||| | ||||||| ||||||| | ||
Db        470 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 411

Qy        623 ATATACATATATACACACATATATACACATACATATACACACACATAGATATACATATAT 682
              ||||| ||||||| | | ||||||| | ||| ||||| | | | ||| ||||| ||||||
Db        410 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 351

Qy        683 ATACACATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTA 742
              ||| | |||||||||  |||||||||  ||||||||||          |  |||
Db        350 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 291

Qy        743 GCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAA 802
              |  |   ||    |   ||| |  | | | |  | | |    | |||    | |
Db        290 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 231

Qy        803 GGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTAT 862
              | || |    |   | |||     | ||| ||| || | ||   | |  |||
Db        230 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 171

Qy        863 CTGTAAGGTAGTTAAGAAGGTAACACTATATATATATATAGTGAAAGCAGTG 914
              | ||   ||  ||   |  ||    |||||||||| | ||   | | |||
Db        170 ATATA---TATATATATATATATATATATATATGTTGTTTCACCTGTG 122


RESULT 8
CC107667/c
LOCUS       CC107667                828 bp    DNA     linear   GSS 16-APR-2003
DEFINITION  NDL.37F6.T7 Notre Dame Liverpool Aedes aegypti genomic clone
            NDL.37F6, genomic survey sequence.
ACCESSION   CC107667
VERSION     CC107667.1  GI:29976722
KEYWORDS    GSS.
SOURCE      Aedes aegypti (yellow fever mosquito)
  ORGANISM  Aedes aegypti
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Aedes;
            Stegomyia.
REFERENCE   1  (bases 1 to 828)
  AUTHORS   Loftus,B., Shetty,J., Knudson,D. and Severson,D.
  TITLE     BAC end sequencing of Aedes aegypti
  JOURNAL   Unpublished (2003)
COMMENT     Other_GSSs: NDL.37F6.SP6
            Contact: Brendan Loftus
            Department of Eukaryotic Genomics
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-3543
            Fax: 301-838-0208
            Email: enta@tigr.org
            Library was provided by David Severson
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .828
                     /organism="Aedes aegypti"
                     /mol_type="genomic DNA"
                     /strain="Liverpool"
                     /db_xref="taxon:7159"
                     /clone="NDL.37F6"
                     /clone_lib="Notre Dame Liverpool"
                     /note="Vector: pECBAC1; Site_1: Hind III; The library was
                     prepared from whole body tissue of newly hatched L1 larvae
                     by David Severson at the University of Notre Dame and
                     Hongbin Zhang"
ORIGIN

  Query Match             6.8%;  Score 338.2;  DB 8;  Length 828;
  Best Local Similarity   70.2%;  Pred. No. 6.3e-48;
  Matches  482;  Conservative    0;  Mismatches  203;  Indels    2;  Gaps    2;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | |||||||||||||||||||||||| | | | | | | ||||||||||| | |||||||
Db        817 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 758

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db        757 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 698

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        697 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 638

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db        637 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 578

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db        577 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 518

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db        517 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 458

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db        457 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 398

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| |||||||||| | |||||||||  ||||||
Db        397 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 338

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||          |  |||          |  |   ||   |   |||
Db        337 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 278

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
                |  | | | |  | | |   | |||   | |   | || |     |     | |
```

```
Db        277 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 218

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTA 884
              ||       | ||| ||| || | ||   | |  ||| | || |  ||  ||   |  ||
Db        217 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 158

Qy        885 ACACTATATATATATATAGTGAAAGCA 911
                  ||||||||||||||   || |||
Db        157 TATATATATATATATATATATAATGCA 131
```

RESULT 9
CF626074/c

LOCUS	CF626074 620 bp mRNA linear EST 02-OCT-2003
DEFINITION	zmrws05_0B10-009-d11.s3 zmrws05 Zea mays cDNA 3', mRNA sequence.
ACCESSION	CF626074
VERSION	CF626074.1 GI:37378736
KEYWORDS	EST.
SOURCE	Zea mays
ORGANISM	Zea mays Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE	1 (bases 1 to 620)
AUTHORS	Bohnert,H., Sharp,R.E., Springer,G.K., Poroyko,V., Fredricksen,M., Sharp,L.G., Spollen,W.G., Ries,J., Guillen,A., Khambati,A., Topinka,C., Davis,G.E., Schachtman,D., Wu,Y. and Nguyen,H.T.
TITLE	NSF Grant DBI-0211842: Functional Genomics of Root Growth and Root Signaling Under Drought
JOURNAL	Unpublished (2003)
COMMENT	Contact: Hans Bohnert University of Illinois, Urbana-Champaign 1201 West Gregory Drive, Urbana, IL 61801, USA Tel: 217-265-5475 Fax: 217-333-5574 Email: bohnerth@life.uiuc.edu POLYA=Yes.

FEATURES Location/Qualifiers

source 1. .620
/organism="Zea mays"
/mol_type="mRNA"
/db_xref="taxon:4577"
/clone_lib="zmrws05"
/note="Samples were collected in Robert E. Sharp's lab (University of Missouri-Columbia) to construct three normalized cDNA libraries. Dark-grown maize seedlings with primary roots 12-20 mm in length were transplanted to high (-0.03 MPa) or low water potential (-1.6 MPa) vermiculite, and harvested at 5 h and 48 h after transplanting. About 1,000 roots were used for each of the low water potential libraries (zmrws05 and zmrws48) while 500 roots were combined from each of the two time points at high water potential (zmrww00). Each root was divided into 4 segments (distances are from the junction of the root apex and root cap): segment 1, 0-3 mm plus the root cap; segment 2, 3-7 mm; segment 3, 7-12 mm; segment 4, 12-20 mm. (For detatils of conditions see (1) with nutrient modifications as in (2)). The three normalized cDNA libraries were constructed in the lab of Hans Bohnert (University of Illinois-UC). Total RNA was extracted by the 'hot Phenol' method (Plant Molecular Biology manual. D5: 1-13, 2nd ed., 1997). This method worked in eliminating carbohydrate material present in the root tips. The integrity of the RNA was verified by denaturing agarose gels and spectrophotometry (ratio A260/280). Poly(A)+mRNA was isolated twice from total RNA using the Oligotex Direct mRNA kit (Qiagen). Poly(A)+ mRNA was converted to double-stranded cDNA and tagged by using modified Oligo(dT) primers. One of 4 sequence tags corresponding to a different segment of the root was added to the 3'-end of the modified Oligo(dT) primers, including a Not1 site and used to reverse transcribe the segment-specific mRNAs into cDNAs. Each library contains all four tags. A suffix (s1, s2, s3, or s4) has been added to each sequence identifier to designate which region of the root (Root_segment_1, 2, 3, or 4) the sequence was found in based on the identification of the tag. A suffix of s0 indicates that the sequence tag, and hence the source segment, could not be identified. The double stranded cDNAs were size-selected (>450 bp). Size selected cDNAs were adaptored with EcoRI adaptors at both ends, and then digested with NotI. The cDNA was directionally cloned into EcoR1-NotI digested pBS II SK(+) phagemid vector (Stratagene) and electroporated into E.coli DH10B. The total number of white colony forming units (cfu) in the primary libraries before amplification was as follows: zmrws05: 3.37 x 106; zmrws48: 4.87 x 106; zmrww00: 3 x 106. The background of empty clones was less than 1%. Inserts ranged from ~0.5kb to >2.5 kb, as determined by PCR. Plasmid DNA from the primary libraries then was converted to single-stranded circles and used as a template for PCR amplification using the T7 and T3 priming sites that flank the cloned cDNA inserts. The purified PCR products, representing the entire cDNA population cloned in each library, were used as a driver for normalization. Hybridization between the single-stranded library and the PCR products was carried out for 44 hours at 30oC. Non-hybridized single-stranded DNA circles were separated from hybridized DNA rendered partially double-stranded and electroporated into DH10B. The total number of clones with insert was: zmrws05: 2.0x107; zmrws48: 4.2x107; zmrww00: 1.1x107. The background of empty clones was less than 2%. Insert size, determined by PCR of the entire library, ranged from 0.5kb to 2.5kb. (1) Sharp R E; Silk W K; Hsiao T C. Growth of the Maize Primary Root at Low Water Potentials I. Spatial Distribution of Expansive Growth. Plant Physiology (Rockville). 87(1). 1988. 50-57. (2) Spollen W G; LeNoble M E; Samuels T D; Bernstein N; Sharp R E. Abscisic acid accumulation maintains maize primary root elongation at low water potentials by restricting ethylene production. Plant Physiology (Rockville). 122(3). March, 2000. 967-976.
TAG_TISSUE=Root_segment_3
TAG_SEQ=TCGCA"

ORIGIN

Query Match 6.8%; Score 337.4; DB 7; Length 620;
Best Local Similarity 78.7%; Pred. No. 8.9e-48;
Matches 415; Conservative 0; Mismatches 111; Indels 1; Gaps 1;

```
Qy        227 TTTATATATATATATATATATATATACACACACACACATATATATATACACATATATA 286
              | |||||||||||||||||||||||| | | | | | | |||||||||||| | |||||||
Db        566 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 507

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | |||||||||| | ||| |||| || | |
Db        506 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 447

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        446 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 387

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db        386 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 327

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db        326 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 267

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db        266 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 207

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db        206 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 147
```

```
Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db        146 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 87

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATT 752
              ||| ||||||||||||       ||  || ||    | ||| |||||
Db         86 TATATATATATATATATATATATATGTAAATTAGCATCACCAGCTGATT 40


RESULT 10
CF634146/c
LOCUS       CF634146                632 bp    mRNA    linear   EST 02-OCT-2003
DEFINITION  zmrww00_0A10-008-g12.s3 zmrww00 Zea mays cDNA 3', mRNA sequence.
ACCESSION   CF634146
VERSION     CF634146.1  GI:37393761
KEYWORDS    EST.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 632)
  AUTHORS   Bohnert,H., Sharp,R.E., Springer,G.K., Poroyko,V., Fredricksen,M.,
            Sharp,L.G., Spollen,W.G., Ries,J., Guillen,A., Khambati,A.,
            Topinka,C., Davis,G.E., Schachtman,D., Wu,Y. and Nguyen,H.T.
  TITLE     NSF Grant DBI-0211842: Functional Genomics of Root Growth and Root
            Signaling Under Drought
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Hans Bohnert
            University of Illinois, Urbana-Champaign
            1201 West Gregory Drive, Urbana, IL 61801, USA
            Tel: 217-265-5475
            Fax: 217-333-5574
            Email: bohnerth@life.uiuc.edu
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .632
                     /organism="Zea mays"
                     /mol_type="mRNA"
                     /db_xref="taxon:4577"
                     /clone_lib="zmrww00"
                     /note="Samples were collected in Robert E. Sharp's lab
                     (University of Missouri-Columbia) to construct three
                     normalized cDNA libraries. Dark-grown maize seedlings with
                     primary roots 12-20 mm in length were transplanted to high
                     (-0.03 MPa) or low water potential (-1.6 MPa) vermiculite,
                     and harvested at 5 h and 48 h after transplanting. About
                     1,000 roots were used for each of the low water potential
                     libraries (zmrws05 and zmrws48) while 500 roots were
                     combined from each of the two time points at high water
                     potential (zmrww00). Each root was divided into 4 segments
                     (distances are from the junction of the root apex and root
                     cap): segment 1, 0-3 mm plus the root cap; segment 2, 3-7
                     mm; segment 3, 7-12 mm; segment 4, 12-20 mm. (For detatils
                     of conditions see (1) with nutrient modifications as in
                     (2)). The three normalized cDNA libraries were constructed
                     in the lab of Hans Bohnert (University of Illinois-UC).
                     Total RNA was extracted by the 'hot Phenol' method (Plant
                     Molecular Biology manual. D5: 1-13, 2nd ed., 1997). This
                     method worked in eliminating carbohydrate material present
                     in the root tips. The integrity of the RNA was verified by
                     denaturing agarose gels and spectrophotometry (ratio
                     A260/280). Poly(A)+mRNA was isolated twice from total RNA
                     using the Oligotex Direct mRNA kit (Qiagen). Poly(A)+ mRNA
                     was converted to double-stranded cDNA and tagged by using
                     modified Oligo(dT) primers. One of 4 sequence tags
                     corresponding to a different segment of the root was added
                     to the 3'-end of the modified Oligo(dT) primers, including
                     a Not1 site and used to reverse transcribe the
                     segment-specific mRNAs into cDNAs. Each library contains
                     all four tags. A suffix (s1, s2, s3, or s4) has been added
                     to each sequence identifier to designate which region of
                     the root (Root_segment_1, 2, 3, or 4) the sequence was
                     found in based on the identification of the tag. A suffix
                     of s0 indicates that the sequence tag, and hence the
                     source segment, could not be identified. The double
                     stranded cDNAs were size-selected (>450 bp). Size selected
                     cDNAs were adaptored with EcoRI adaptors at both ends, and
                     then digested with NotI. The cDNA was directionally cloned
                     into EcoR1-NotI digested pBS II SK(+) phagemid vector
                     (Stratagene) and electroporated into E.coli DH10B. The
                     total number of white colony forming units (cfu) in the
                     primary libraries before amplification was as follows:
                     zmrws05: 3.37 x 106; zmrws48: 4.87 x 106; zmrww00: 3 x
                     106. The background of empty clones was less than 1%.
                     Inserts ranged from ~0.5kb to >2.5 kb, as determined by
                     PCR. Plasmid DNA from the primary libraries then was
                     converted to single-stranded circles and used as a
                     template for PCR amplification using the T7 and T3 priming
                     sites that flank the cloned cDNA inserts. The purified PCR
                     products, representing the entire cDNA population cloned
                     in each library, were used as a driver for normalization.
                     Hybridization between the single-stranded library and the
                     PCR products was carried out for 44 hours at 30oC.
                     Non-hybridized single-stranded DNA circles were separated
                     from hybridized DNA rendered partially double-stranded and
                     electroporated into DH10B. The total number of clones with
                     insert was: zmrws05: 2.0x107; zmrws48: 4.2x107; zmrww00:
                     1.1x107. The background of empty clones was less than 2%.
                     Insert size, determined by PCR of the entire library,
                     ranged from 0.5kb to 2.5kb. (1) Sharp R E; Silk W K; Hsiao
                     T C. Growth of the Maize Primary Root at Low Water
                     Potentials I. Spatial Distribution of Expansive Growth.
                     Plant Physiology (Rockville). 87(1). 1988. 50-57. (2)
                     Spollen W G; LeNoble M E; Samuels T D; Bernstein N; Sharp
                     R E. Abscisic acid accumulation maintains maize primary
                     root elongation at low water potentials by restricting
                     ethylene production. Plant Physiology (Rockville). 122(3).
                     March, 2000. 967-976.
                     TAG_TISSUE=Root_segment_3
                     TAG_SEQ=TCGCA"
ORIGIN

  Query Match             6.8%;  Score 337.4;  DB 7;  Length 632;
  Best Local Similarity   78.7%;  Pred. No. 8.9e-48;
  Matches  415;  Conservative    0;  Mismatches  111;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACATATATATATACACATATATA 286
              |||||||||||||||||||||||||| | | | | | ||||||||||| | |||||||
Db        545 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 486

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db        485 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 426

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        425 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 366

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db        365 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 306

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db        305 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 246

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db        245 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 186

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db        185 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 126
```

```
Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | ||||||||  ||||||
Db        125 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 66

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATT 752
              ||| ||||||||||||        ||  || ||    | ||| |||||
Db         65 TATATATATATATATATATATATGTAAATTAGCATCACCAGCTGATT 19


RESULT 11
CW016544/c
LOCUS       CW016544                883 bp    DNA     linear   GSS 23-SEP-2004
DEFINITION  ZMMBLc0009C11.f ZMMBLc Zea mays genomic clone ZMMBLc0009C11 5',
            genomic survey sequence.
ACCESSION   CW016544
VERSION     CW016544.1  GI:52606273
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 883)
  AUTHORS   Wing,R., Luo,M., Soderlund,C. and Haller,K.
  TITLE     ZMMBL sequences
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            Plate: 0009  row: C  column: 11
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .883
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /cultivar="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBLc0009C11"
                     /tissue_type="immature ears"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="ZMMBLc"
                     /note="Vector: pAGIBAC1; Site_1: Hpa II; Site_2: Hpa II"
ORIGIN

  Query Match             6.8%;  Score 337.4;  DB 9;  Length 883;
  Best Local Similarity   81.4%;  Pred. No. 8.5e-48;
  Matches  403;  Conservative    0;  Mismatches   91;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | |||||||||||||||||||||||||| | | ||| | | ||||||||||| |||||||||
Db        872 TATATATATATATATATATATATATATATATACATATATATATATATATATACACATATATA 813

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              |||| ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db        812 TACATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 753

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        752 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 693

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db        692 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 633

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db        632 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 573

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db        572 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 513

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db        512 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 453

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | ||||||||  ||||||
Db        452 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 393

Qy        706 TATGTATATATATAT 720
              ||| ||||||||||||
Db        392 TATATATATATATAT 378


RESULT 12
CW016544
LOCUS       CW016544                883 bp    DNA     linear   GSS 23-SEP-2004
DEFINITION  ZMMBLc0009C11.f ZMMBLc Zea mays genomic clone ZMMBLc0009C11 5',
            genomic survey sequence.
ACCESSION   CW016544
VERSION     CW016544.1  GI:52606273
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 883)
  AUTHORS   Wing,R., Luo,M., Soderlund,C. and Haller,K.
  TITLE     ZMMBL sequences
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            Plate: 0009  row: C  column: 11
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .883
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /cultivar="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBLc0009C11"
                     /tissue_type="immature ears"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="ZMMBLc"
                     /note="Vector: pAGIBAC1; Site_1: Hpa II; Site_2: Hpa II"
ORIGIN

  Query Match             6.8%;  Score 337.2;  DB 9;  Length 883;
  Best Local Similarity   71.3%;  Pred. No. 9.2e-48;
  Matches  458;  Conservative    0;  Mismatches  183;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | |||||||||||||||||||||||||| | | | | | ||||||||||| | |||||||
Db        233 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 292

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db        293 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 352

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        353 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 412

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
```

```
Db        413 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 472

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || |||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db        473 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 532

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db        533 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 592

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db        593 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 652

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| |||||||| | |||||||||  ||||||
Db        653 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 712

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||||          | |||          | |    ||     |    |||
Db        713 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 772

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
              |   | | | |  | | |    | |||    | | | | || | |  |     | |
Db        773 TATATATATATATATATATATATATATATATATATATGTATATATATGTATATATATATA 832

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA 867
              ||     | | ||| ||| || | ||   | |  ||| | ||
Db        833 TATATATGTATATATATATATATATATATATATATATATA 874


RESULT 13
AQ917102
LOCUS       AQ917102                725 bp    DNA     linear   GSS 07-DEC-1999
DEFINITION  T233131b Medicago truncatula BAC library Medicago truncatula
            genomic clone 24E12, genomic survey sequence.
ACCESSION   AQ917102
VERSION     AQ917102.1  GI:6536747
KEYWORDS    GSS.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 725)
  AUTHORS   Kim,D., Limpens,E., Peng,H., Ellis,L. and Cook,D.R.
  TITLE     BAC end sequencing of Medicago truncatula
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Cook DR
            The Crop Biotechnology Center
            Texas A&M University
            Department of Plant Pathology and Microbiology, Rm 120 L.F.
            Peterson Bldg, College Station, TX 77843-2132, USA
            Tel: 409 845 8743
            Fax: 409 862 4790
            Email: dcook@ppserver.tamu.edu
            Other name: DK277-L; date: 8/12/99; Submitted to the Database of
            Genome Survey Sequences (GSS) on 12/06/99; More information is
            available at 'http://chrysie.tamu.edu/medicago'.
            Seq primer: SQ-BAC-L (AACGCCAGGGTTTTCCCAGTCACGACG)
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .725
                     /organism="Medicago truncatula"
                     /mol_type="genomic DNA"
                     /cultivar="genotype A17"
                     /db_xref="taxon:3880"
                     /clone="24E12"
                     /clone_lib="Medicago truncatula BAC library"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII; Nam. Y-W, Penmetsa, R.V., Endre, G., Kim, D., and
                     Cook, D.R.  1999. Construction of a bacterial artificial
                     chromosome  library of Medicago truncatula and
                     identification of clones containing ethylene response
                     genes. Theor Appl Genet 98:  638-646. "
ORIGIN

  Query Match             6.8%;  Score 336.6;  DB 8;  Length 725;
  Best Local Similarity   77.7%;  Pred. No. 1.2e-47;
  Matches  419;  Conservative    0;  Mismatches  119;  Indels    1;  Gaps    1;

Qy        183 AGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATA 242
              ||  || ||   | |  |||  |   |   ||||   | | |||  |  ||||||||||
Db        158 AGCAAAGAAACCCAATTTGTGGTGGAAATGAAAAGGGATGCCATGATGGATATATATATA 217

Qy        243 TATATATATACACACACACACATATATATATACACATATATATACACATATATACATA 302
              |||||||||| | | | | | ||||||||||| | |||||||| | ||||||| |||
Db        218 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 277

Qy        303 TATACACATATATACACATATATATACACATACATATGTACACATATATATACACATATG 362
              |||| | ||||||| | ||||||||| | ||| |||| || | ||||||||| | ||||
Db        278 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 337

Qy        363 TATACACATATATACACATATATACACACATATATACACATATATACACACATATATACA 422
              |||| | ||||||| | ||||||| | | ||||||| | ||||||| | | ||||||| |
Db        338 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 397

Qy        423 CATATATACACATATATACACATATACACATATATACACATATATACATATATACACATA 482
              ||||||| | ||||||| | ||||| | ||||||| | ||||||| ||||||| | |||
Db        398 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 457

Qy        483 TATATA-ATATACACACATATATATACACATATATACACACATATATACACATATATACA 541
              |||||| ||||| | | ||||||||| | ||||||| | | ||||||| | ||||||| |
Db        458 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 517

Qy        542 CATATATATACACATATATACACATATATACATATATACACATATATATACATATATACA 601
              ||||||||| | ||||||| | ||||||| ||||||| | ||||||||| ||||||| |
Db        518 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 577

Qy        602 CATATATACATATATACACATATATACATATATACACACATATATACACATACATATACA 661
              ||||||| ||||||| | ||||||| ||||||| | | ||||||| | ||| ||||| |
Db        578 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 637

Qy        662 CACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATAT 720
              | | ||| ||||| ||||||||| | |||||||||  ||||||||| ||||||||||
Db        638 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 696


RESULT 14
CC067623
LOCUS       CC067623                827 bp    DNA     linear   GSS 16-APR-2003
DEFINITION  CSU-K33r.18M15.SP6 CSU-K33r Aedes aegypti genomic clone
            CSU-K33r.18M15, genomic survey sequence.
ACCESSION   CC067623
VERSION     CC067623.1  GI:29906129
KEYWORDS    GSS.
SOURCE      Aedes aegypti (yellow fever mosquito)
  ORGANISM  Aedes aegypti
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Aedes;
            Stegomyia.
REFERENCE   1  (bases 1 to 827)
  AUTHORS   Loftus,B., Shetty,J., Severson,D., Brown,S. and Knudson,D.
  TITLE     End sequencing of Aedes aegypti BACs
  JOURNAL   Unpublished (2003)
COMMENT     Other_GSSs: CSU-K33r.18M15.T7
            Contact: Brendan Loftus
            Department of Eukaryotic Genomics
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-3543
            Fax: 301-838-0208
            Email: enta@tigr.org
            Library was provided by Susan Brown and Dennis Knudson at Colorado
            State University.
```

```
           Seq primer: SP6
           Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .827
                     /organism="Aedes aegypti"
                     /mol_type="genomic DNA"
                     /strain="Rexville"
                     /db_xref="taxon:7159"
                     /clone="CSU-K33r.18M15"
                     /clone_lib="CSU-K33r"
                     /note="Vector: pBeloBAC11; Site_1: HindIII"
ORIGIN

  Query Match             6.8%;  Score 336.4;  DB 8;  Length 827;
  Best Local Similarity   78.7%;  Pred. No. 1.3e-47;
  Matches  414;  Conservative    0;  Mismatches  111;  Indels    1;  Gaps    1;

Qy        196 GAAATGTCATCTCAGAAAAAAAAAAAGGCATTTTATATATATATATATATATATACAC 255
              ||| |  | ||  |  ||     |||| |     |||||||||||||||||||||| |
Db        238 GAATTTCCTTCAAACGAATGCTAAAACGTGCGATATATATATATATATATATATATAT 297

Qy        256 ACACACACACATATATATATACACATATATATACACATATATACATATATACACATATAT 315
              | | | | | |||||||||||| | |||||||| | ||||||| ||||||| | ||||||
Db        298 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 357

Qy        316 ACACATATATATACACATACATATGTACACATATATATACACATATGTATACACATATAT 375
              | | |||||||||| | ||| |||| || | |||||||||| | |||| |||| | |||||||
Db        358 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 417

Qy        376 ACACATATATACACACATATATACACATATATACACACATATATACACATATATACACAT 435
              | | ||||||| | | |||||||| | ||||||| | ||||||| | ||||||| | ||
Db        418 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 477

Qy        436 ATATACACATATACACATATATACACATATATACATATATACACATATATATA-ATATAC 494
              |||||  |  |||||  |  |||||||  |  |||||||  |||||||  |  |||||||||  |||||
Db        478 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 537

Qy        495 ACACATATATATACACATATATACACACATATATACACATATATACACATATATATACAC 554
              | | |||||||||  | |||||||  | |  |||||||  |  |||||||  |  |||||||||  |
Db        538 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 597

Qy        555 ATATATACACATATATACATATATACACATATATATACATATATACACATATATACATAT 614
              |||||||  |  |||||||  |||||||  |  |||||||||  |||||||  |  |||||||  ||||
Db        598 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 657

Qy        615 ATACACATATATACATATATACACACATATATACACATACATATACACACACATAGATAT 674
              |||  |  |||||||  |||||||  |  |  |||||||  |  |||  |||||  |  |  |||  ||||
Db        658 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 717

Qy        675 ACATATATATACACATATATATACGTATATATATGTATATATATAT 720
              |  |||||||||  |  |||||||||  |||||||||  ||||||||||||
Db        718 ATATATATATATATATATATATATATATATATATATATATATATAT 763


RESULT 15
CR294134
LOCUS       CR294134                590 bp    DNA     linear   GSS 28-FEB-2004
DEFINITION  Medicago truncatula BAC ends cultivar Jemalong A17 of Medicago
            truncatula, genomic survey sequence.
ACCESSION   CR294134
VERSION     CR294134.1  GI:44684558
KEYWORDS    GSS.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 590)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-FEB-2004) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
FEATURES             Location/Qualifiers
     source          1. .590
                     /organism="Medicago truncatula"
                     /mol_type="genomic DNA"
                     /cultivar="Jemalong A17"
                     /db_xref="taxon:3880"
                     /clone_lib="MTE1"
                     /note="Vector: pIndigoBAC ; Site_1: EcoRI ; Site_2: EcoRI
                     ; Debelle F. and Chalhoub B.~Genoscope sequence ID :
                     mte1-1J18FM1"
ORIGIN

  Query Match             6.8%;  Score 336.2;  DB 9;  Length 590;
  Best Local Similarity   79.7%;  Pred. No. 1.4e-47;
  Matches  409;  Conservative    0;  Mismatches  103;  Indels    1;  Gaps    1;

Qy        209 AGAAAAAAAAAAAGGCATTTTATATATATATATATATATATATACACACACACACATA 268
              || | | | | |  | ||  |||||||||||||||||||||||| | | | | | |||
Db         15 AGGAGATATATATAGTATAGTATATATATATATATATATATATATATATATATATATA 74

Qy        269 TATATATACACATATATATACACATATATACATATATACACATATATACACATATATATA 328
              |||||||| | ||||||||| | ||||||| ||||||| | ||||||| | |||||||||
Db         75 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 134

Qy        329 CACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACA 388
              | ||| |||| || | ||||||||| | |||| |||| | ||||||| | ||||||| |
Db        135 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 194

Qy        389 CACATATATACACATATATACACACATATATACACATATATACACATATATACACATATA 448
              | ||||||| | |||||||| | | ||||||| | ||||||| | ||||||| | |||||
Db        195 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 254

Qy        449 CACATATATACACATATATACATATATACACATATATATA-ATATACACACATATATATA 507
              | ||||||| | ||||||| ||||||| | ||||||||| |||||  |  |  |||||||||
Db        255 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 314

Qy        508 CACATATATACACACATATATACACATATATACACATATATATACACATATATACACATA 567
              | ||||||| | | ||||||| | ||||||| | ||||||||| | ||||||| | |||
Db        315 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 374

Qy        568 TATACATATATACACATATATATACATATATACACATATATACATATATACACATATATA 627
              ||||  |||||||  |||||||||  |||||||  |||||||  |||||||  |||||||
Db        375 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 434

Qy        628 CATATATACACACATATATACACATACATATACACACACATAGATATACATATATATACA 687
              |||||||  |  |  |||||||  |  |||  |||||  |  |  |  |||  |||||  |||||||||  |
Db        435 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 494

Qy        688 CATATATATACGTATATATATGTATATATATAT 720
              |||||||||  |||||||||  ||||||||||||
Db        495 TATATATATATATATATATATATATATATATAT 527


RESULT 16
CC067623/c
LOCUS       CC067623                827 bp    DNA     linear   GSS 16-APR-2003
DEFINITION  CSU-K33r.18M15.SP6 CSU-K33r Aedes aegypti genomic clone
            CSU-K33r.18M15, genomic survey sequence.
ACCESSION   CC067623
VERSION     CC067623.1  GI:29906129
KEYWORDS    GSS.
SOURCE      Aedes aegypti (yellow fever mosquito)
  ORGANISM  Aedes aegypti
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Aedes;
            Stegomyia.
REFERENCE   1  (bases 1 to 827)
  AUTHORS   Loftus,B., Shetty,J., Severson,D., Brown,S. and Knudson,D.
  TITLE     End sequencing of Aedes aegypti BACs
  JOURNAL   Unpublished (2003)
COMMENT     Other_GSSs: CSU-K33r.18M15.T7
            Contact: Brendan Loftus
```

Department of Eukaryotic Genomics
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-3543
Fax: 301-838-0208
Email: enta@tigr.org
Library was provided by Susan Brown and Dennis Knudson at Colorado State University.
Seq primer: SP6
Class: BAC ends.

FEATURES Location/Qualifiers
source 1. .827
/organism="Aedes aegypti"
/mol_type="genomic DNA"
/strain="Rexville"
/db_xref="taxon:7159"
/clone="CSU-K33r.18M15"
/clone_lib="CSU-K33r"
/note="Vector: pBeloBAC11; Site_1: HindIII"

ORIGIN

Query Match 6.8%; Score 335.8; DB 8; Length 827;
Best Local Similarity 76.2%; Pred. No. 1.6e-47;
Matches 426; Conservative 0; Mismatches 132; Indels 1; Gaps 1;

```
Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||| | | | | | | |||||||||| | |||||||
Db        766 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 707

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db        706 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 647

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        646 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 587

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db        586 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 527

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || |||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db        526 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 467

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db        466 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 407

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db        406 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 347

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | ||||||||| ||||||
Db        346 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 287

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| |||||||||||    ||   || |  |   ||   |   | | |    | |  |
Db        286 TATATATATATATATATCGCACGTTTTAGCATTCGTTTGAAGGAAATTCAGGAGAAGGGA 227

Qy        766 GAGGAAAAGAGAGTTTGAC 784
              ||| |    | | ||| ||
Db        226 GAGTATTGTATAATTTTAC 208
```

RESULT 17
CF626074
LOCUS CF626074 620 bp mRNA linear EST 02-OCT-2003
DEFINITION zmrws05_0B10-009-d11.s3 zmrws05 Zea mays cDNA 3', mRNA sequence.
ACCESSION CF626074
VERSION CF626074.1 GI:37378736
KEYWORDS EST.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 620)
AUTHORS Bohnert,H., Sharp,R.E., Springer,G.K., Poroyko,V., Fredricksen,M., Sharp,L.G., Spollen,W.G., Ries,J., Guillen,A., Khambati,A., Topinka,C., Davis,G.E., Schachtman,D., Wu,Y. and Nguyen,H.T.
TITLE NSF Grant DBI-0211842: Functional Genomics of Root Growth and Root Signaling Under Drought
JOURNAL Unpublished (2003)
COMMENT Contact: Hans Bohnert
University of Illinois, Urbana-Champaign
1201 West Gregory Drive, Urbana, IL 61801, USA
Tel: 217-265-5475
Fax: 217-333-5574
Email: bohnerth@life.uiuc.edu
POLYA=Yes.

FEATURES Location/Qualifiers
source 1. .620
/organism="Zea mays"
/mol_type="mRNA"
/db_xref="taxon:4577"
/clone_lib="zmrws05"
/note="Samples were collected in Robert E. Sharp's lab (University of Missouri-Columbia) to construct three normalized cDNA libraries. Dark-grown maize seedlings with primary roots 12-20 mm in length were transplanted to high (-0.03 MPa) or low water potential (-1.6 MPa) vermiculite, and harvested at 5 h and 48 h after transplanting. About 1,000 roots were used for each of the low water potential libraries (zmrws05 and zmrws48) while 500 roots were combined from each of the two time points at high water potential (zmrww00). Each root was divided into 4 segments (distances are from the junction of the root apex and root cap): segment 1, 0-3 mm plus the root cap; segment 2, 3-7 mm; segment 3, 7-12 mm; segment 4, 12-20 mm. (For detatils of conditions see (1) with nutrient modifications as in (2)). The three normalized cDNA libraries were constructed in the lab of Hans Bohnert (University of Illinois-UC). Total RNA was extracted by the 'hot Phenol' method (Plant Molecular Biology manual. D5: 1-13, 2nd ed., 1997). This method worked in eliminating carbohydrate material present in the root tips. The integrity of the RNA was verified by denaturing agarose gels and spectrophotometry (ratio A260/280). Poly(A)+mRNA was isolated twice from total RNA using the Oligotex Direct mRNA kit (Qiagen). Poly(A)+ mRNA was converted to double-stranded cDNA and tagged by using modified Oligo(dT) primers. One of 4 sequence tags corresponding to a different segment of the root was added to the 3'-end of the modified Oligo(dT) primers, including a Not1 site and used to reverse transcribe the segment-specific mRNAs into cDNAs. Each library contains all four tags. A suffix (s1, s2, s3, or s4) has been added to each sequence identifier to designate which region of the root (Root_segment_1, 2, 3, or 4) the sequence was found in based on the identification of the tag. A suffix of s0 indicates that the sequence tag, and hence the source segment, could not be identified. The double stranded cDNAs were size-selected (>450 bp). Size selected cDNAs were adaptored with EcoRI adaptors at both ends, and then digested with NotI. The cDNA was directionally cloned into EcoR1-NotI digested pBS II SK(+) phagemid vector (Stratagene) and electroporated into E.coli DH10B. The total number of white colony forming units (cfu) in the primary libraries before amplification was as follows: zmrws05: 3.37 x 106; zmrws48: 4.87 x 106; zmrww00: 3 x 106. The background of empty clones was less than 1%. Inserts ranged from ~0.5kb to >2.5 kb, as determined by PCR. Plasmid DNA from the primary libraries then was converted to single-stranded circles and used as a template for PCR amplification using the T7 and T3 priming

sites that flank the cloned cDNA inserts. The purified PCR products, representing the entire cDNA population cloned in each library, were used as a driver for normalization. Hybridization between the single-stranded library and the PCR products was carried out for 44 hours at 30oC. Non-hybridized single-stranded DNA circles were separated from hybridized DNA rendered partially double-stranded and electroporated into DH10B. The total number of clones with insert was: zmrws05: 2.0x107; zmrws48: 4.2x107; zmrww00: 1.1x107. The background of empty clones was less than 2%. Insert size, determined by PCR of the entire library, ranged from 0.5kb to 2.5kb. (1) Sharp R E; Silk W K; Hsiao T C. Growth of the Maize Primary Root at Low Water Potentials I. Spatial Distribution of Expansive Growth. Plant Physiology (Rockville). 87(1). 1988. 50-57. (2) Spollen W G; LeNoble M E; Samuels T D; Bernstein N; Sharp R E. Abscisic acid accumulation maintains maize primary root elongation at low water potentials by restricting ethylene production. Plant Physiology (Rockville). 122(3). March, 2000. 967-976.
TAG_TISSUE=Root_segment_3
TAG_SEQ=TCGCA"

ORIGIN

Query Match 6.7%; Score 335.2; DB 7; Length 620;
Best Local Similarity 78.4%; Pred. No. 2.1e-47;
Matches 414; Conservative 0; Mismatches 113; Indels 1; Gaps 1;

```
Qy        194 GCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATATAC 253
              || ||  |  ||| |    | | | | |      | ||||||||||||||||||||||||
Db         54 GCTAATTTACATATATATATATATATATATATATATATATATATATATATATATATATATAT 113

Qy        254 ACACACACACACATATATATATACACATATATATACACATATATACATATATACACATAT 313
              | | | | | | |||||||||| | ||||||||| | ||||||| ||||||| | ||||
Db        114 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 173

Qy        314 ATACACATATATATACACATACATATGTACACATATATATACACATATGTATACACATAT 373
              ||| | ||||||||| | ||| |||| || | ||||||||| | |||| |||| | ||||
Db        174 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 233

Qy        374 ATACACATATATACACACATATATACACATATATACACACATATATACACATATATACAC 433
              ||| | ||||||| | | ||||||| | ||||||| | | ||||||| | ||||||| |
Db        234 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 293

Qy        434 ATATATACACATATACACATATATACACATATATACATATATACACATATATATA-ATAT 492
              ||||||| | ||||| | ||||||| | ||||||| |||||||.| ||||||||| ||||
Db        294 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 353

Qy        493 ACACACATATATATACACATATATACACACATATATACACATATATACACATATATATAC 552
              | | | ||||||||| | ||||||| | | ||||||| | ||||||| | |||||||||
Db        354 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 413

Qy        553 ACATATATACACATATATACATATATACACATATATATACATATATACACATATATACAT 612
              | ||||||| | ||||||| ||||||| | ||||||||| ||||||| | ||||||| ||
Db        414 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 473

Qy        613 ATATACACATATATACATATATACACACATATATACACATACATATACACACACATAGAT 672
              ||||| | ||||||| ||||||| | | ||||||| | ||| ||||| | | | ||| ||
Db        474 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 533

Qy        673 ATACATATATATACACATATATATACGTATATATATGTATATATATAT 720
              ||| ||||||||| | |||||||||  ||||||||| ||||||||||
Db        534 ATATATATATATATATATATATATATATATATATATATATATATATATATATAT 581
```

RESULT 18
CF634146
LOCUS CF634146 632 bp mRNA linear EST 02-OCT-2003
DEFINITION zmrww00_0A10-008-g12.s3 zmrww00 Zea mays cDNA 3', mRNA sequence.
ACCESSION CF634146
VERSION CF634146.1 GI:37393761
KEYWORDS EST.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 632)
AUTHORS Bohnert,H., Sharp,R.E., Springer,G.K., Poroyko,V., Fredricksen,M., Sharp,L.G., Spollen,W.G., Ries,J., Guillen,A., Khambati,A., Topinka,C., Davis,G.E., Schachtman,D., Wu,Y. and Nguyen,H.T.
TITLE NSF Grant DBI-0211842: Functional Genomics of Root Growth and Root Signaling Under Drought
JOURNAL Unpublished (2003)
COMMENT Contact: Hans Bohnert
University of Illinois, Urbana-Champaign
1201 West Gregory Drive, Urbana, IL 61801, USA
Tel: 217-265-5475
Fax: 217-333-5574
Email: bohnerth@life.uiuc.edu
POLYA=Yes.
FEATURES Location/Qualifiers
source 1. .632
/organism="Zea mays"
/mol_type="mRNA"
/db_xref="taxon:4577"
/clone_lib="zmrww00"
/note="Samples were collected in Robert E. Sharp's lab (University of Missouri-Columbia) to construct three normalized cDNA libraries. Dark-grown maize seedlings with primary roots 12-20 mm in length were transplanted to high (-0.03 MPa) or low water potential (-1.6 MPa) vermiculite, and harvested at 5 h and 48 h after transplanting. About 1,000 roots were used for each of the low water potential libraries (zmrws05 and zmrws48) while 500 roots were combined from each of the two time points at high water potential (zmrww00). Each root was divided into 4 segments (distances are from the junction of the root apex and root cap): segment 1, 0-3 mm plus the root cap; segment 2, 3-7 mm; segment 3, 7-12 mm; segment 4, 12-20 mm. (For detatils of conditions see (1) with nutrient modifications as in (2)). The three normalized cDNA libraries were constructed in the lab of Hans Bohnert (University of Illinois-UC). Total RNA was extracted by the 'hot Phenol' method (Plant Molecular Biology manual. D5: 1-13, 2nd ed., 1997). This method worked in eliminating carbohydrate material present in the root tips. The integrity of the RNA was verified by denaturing agarose gels and spectrophotometry (ratio A260/280). Poly(A)+mRNA was isolated twice from total RNA using the Oligotex Direct mRNA kit (Qiagen). Poly(A)+ mRNA was converted to double-stranded cDNA and tagged by using modified Oligo(dT) primers. One of 4 sequence tags corresponding to a different segment of the root was added to the 3'-end of the modified Oligo(dT) primers, including a Not1 site and used to reverse transcribe the segment-specific mRNAs into cDNAs. Each library contains all four tags. A suffix (s1, s2, s3, or s4) has been added to each sequence identifier to designate which region of the root (Root_segment_1, 2, 3, or 4) the sequence was found in based on the identification of the tag. A suffix of s0 indicates that the sequence tag, and hence the source segment, could not be identified. The double stranded cDNAs were size-selected (>450 bp). Size selected cDNAs were adaptored with EcoRI adaptors at both ends, and then digested with NotI. The cDNA was directionally cloned into EcoR1-NotI digested pBS II SK(+) phagemid vector (Stratagene) and electroporated into E.coli DH10B. The total number of white colony forming units (cfu) in the primary libraries before amplification was as follows: zmrws05: 3.37 x 106; zmrws48: 4.87 x 106; zmrww00: 3 x 106. The background of empty clones was less than 1%. Inserts ranged from ~0.5kb to >2.5 kb, as determined by PCR. Plasmid DNA from the primary libraries then was converted to single-stranded circles and used as a template for PCR amplification using the T7 and T3 priming sites that flank the cloned cDNA inserts. The purified PCR

products, representing the entire cDNA population cloned in each library, were used as a driver for normalization. Hybridization between the single-stranded library and the PCR products was carried out for 44 hours at 30oC. Non-hybridized single-stranded DNA circles were separated from hybridized DNA rendered partially double-stranded and electroporated into DH10B. The total number of clones with insert was: zmrws05: 2.0x107; zmrws48: 4.2x107; zmrww00: 1.1x107. The background of empty clones was less than 2%. Insert size, determined by PCR of the entire library, ranged from 0.5kb to 2.5kb. (1) Sharp R E; Silk W K; Hsiao T C. Growth of the Maize Primary Root at Low Water Potentials I. Spatial Distribution of Expansive Growth. Plant Physiology (Rockville). 87(1). 1988. 50-57. (2) Spollen W G; LeNoble M E; Samuels T D; Bernstein N; Sharp R E. Abscisic acid accumulation maintains maize primary root elongation at low water potentials by restricting ethylene production. Plant Physiology (Rockville). 122(3). March, 2000. 967-976.
TAG_TISSUE=Root_segment_3
TAG_SEQ=TCGCA"

ORIGIN

Query Match 6.7%; Score 335.2; DB 7; Length 632;
Best Local Similarity 78.4%; Pred. No. 2.1e-47;
Matches 414; Conservative 0; Mismatches 113; Indels 1; Gaps 1;

```
Qy      194 GCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATATAC 253
            || ||  |  |||  |     | | | |       | ||||||||||||||||||||||
Db       33 GCTAATTTACATATATATATATATATATATATATATATATATATATATATATATATATATAT 92

Qy      254 ACACACACACACATATATATATACACATATATATACACATATATACATATATACACATAT 313
            | | | | | | |||||||||||| | ||||||||| | ||||||| ||||||| | ||||
Db       93 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 152

Qy      314 ATACACATATATATACACATACATATGTACACATATATATACACATATGTATACACATAT 373
            ||| | |||||||||| | ||| |||| || | ||||||||| | |||| |||| | ||||
Db      153 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 212

Qy      374 ATACACATATATACACACATATATACACATATATACACACATATATACACATATATACAC 433
            ||| | ||||||| | | ||||||| | ||||||| | | ||||||| | ||||||| |
Db      213 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 272

Qy      434 ATATATACACATATACACATATATACACATATATACATATATACACATATATATA-ATAT 492
            ||||||| | ||||| | ||||||| | ||||||| ||||||| | ||||||||| ||||
Db      273 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 332

Qy      493 ACACACATATATATACACATATATACACACATATATACACATATATACACATATATATAC 552
            | | | ||||||||| | ||||||| | | ||||||| | ||||||| | |||||||||
Db      333 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 392

Qy      553 ACATATATACACATATATACATATATACACATATATATACATATATACACATATATACAT 612
            | ||||||| | ||||||| ||||||| | ||||||||| ||||||| | ||||||| ||
Db      393 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 452

Qy      613 ATATACACATATATACATATATACACACATATATACACATACATATACACACACATAGAT 672
            ||||| | ||||||| ||||||| | | ||||||| | ||| ||||| | | | ||| ||
Db      453 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 512

Qy      673 ATACATATATATACACATATATATACGTATATATATGTATATATATAT 720
            ||| ||||||||| | |||||||||  ||||||||| ||||||||||
Db      513 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 560
```

RESULT 19
BI505681
LOCUS BI505681 717 bp mRNA linear EST 08-APR-2002
DEFINITION BB170032B10E06.5 Bee Brain Normalized/Subtracted Library, BB17 Apis mellifera cDNA clone BB170032B10E06 5', mRNA sequence.
ACCESSION BI505681
VERSION BI505681.1 GI:15356055
KEYWORDS EST.
SOURCE Apis mellifera (honey bee)
ORGANISM Apis mellifera
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Apoidea; Apidae; Apis.
REFERENCE 1 (bases 1 to 717)
AUTHORS Whitfield,C.W., Band,M.R., Bonaldo,M.F., Kumar,C.G., Liu,L., Pardinas,J., Robertson,H.M., Soares,B. and Robinson,G.E.
TITLE Annotated expressed sequence tags and cDNA microarrays for studies of brain and behavior in the honey bee
JOURNAL Genome Res. 12 (4), 555-566 (2002)
MEDLINE 21929762
PUBMED 11932240
COMMENT Contact: Gene E. Robinson
Department of Entomology
University of Illinois
505 S. Goodwin Ave., Urbana, IL 61801, USA
Tel: 217 265 0309
Fax: 217 244 3499
Email: generobi@life.uiuc.edu
This research was funded by the University of Illinois Critical Research Initiatives Fund and a Burroughs-Wellcome Trust Innovation Award in Functional Genomics to G.E. Robinson and an NSF Postdoctoral Fellowship in Bioinformatics to C.W. Whitfield.
REPEATS IN THE SEQUENCE
Low_complexity STRAND (+) ELEMENT AT_rich LOCATION [101,132];
Simple_repeat STRAND (+) ELEMENT (TA)n LOCATION [133,717].
PCR PRimers
FORWARD: TAATACGACTCACTATAGGG
BACKWARD: ATTAACCCTCACTAAAG
Plate: BB170032B10 row: E column: 06
Seq primer: AGCGGATAACAATTTCACACAGGA
High quality sequence stop: 717.
FEATURES Location/Qualifiers
source 1. .717
/organism="Apis mellifera"
/mol_type="mRNA"
/strain="mixed strains of European bees, predominantly A.m. ligustica"
/db_xref="taxon:7460"
/clone="BB170032B10E06"
/sex="female"
/tissue_type="brain"
/dev_stage="adult worker honey bee"
/lab_host="DH10B"
/clone_lib="Bee Brain Normalized/Subtracted Library, BB17"
/note="Organ: brain; Vector: pT7T3-Pac; Site_1: EcoR1; Site_2: Not1; This BB17 cDNA library was generated by subtraction of the BB16 library with 4000 previously sequenced clones. The BB16 library was contributed by the Soares laboratory and it was constructed and normalized as described by Bonaldo, M.F., Lennon, G. and Soares, M.B. (1996), Genome Research 6(9): 791-806. RNA was prepared from dissected brains of adult worker bees of various ages and various behavioral groups. "

ORIGIN

Query Match 6.7%; Score 335.2; DB 4; Length 717;
Best Local Similarity 77.2%; Pred. No. 2.1e-47;
Matches 420; Conservative 0; Mismatches 123; Indels 1; Gaps 1;

```
Qy      178 AGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATAT 237
            | | ||  || | ||   | || |      | | || | | |      | |||||||||
Db       96 AACTCAATTAAAATGATTTATATATATATATATATAATATATATATATATATATATATAT 155

Qy      238 ATATATATATATATACACACACACACATATATATATACACATATATATACACATATAT 297
            ||||||||||||||| | | | | | | |||||||||||| | ||||||||| | ||||||
Db      156 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 215

Qy      298 ACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACAC 357
            | ||||||| | ||||||| | ||||||||| | ||| |||| || | ||||||||| |
Db      216 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 275

Qy      358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417
```

```
Db        276 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 335

Qy        418 ATACACATATATACACATATATACACATATACACATATATACACATATATACATATATAC 477

Db        336 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 395

Qy        478 ACATATATATA-ATATACACACATATATATACACATATATACACACATATATACACATAT 536

Db        396 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 455

Qy        537 ATACACATATATATACACATATATACACATATATACATATATACACATATATATACATAT 596

Db        456 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 515

Qy        597 ATACACATATATACATATATACACATATATACATATATACACACATATATACACATACAT 656

Db        516 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 575

Qy        657 ATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATAT 716

Db        576 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 635

Qy        717 ATAT 720

Db        636 ATAT 639
```

RESULT 20
CR294134/c
LOCUS CR294134 590 bp DNA linear GSS 28-FEB-2004
DEFINITION Medicago truncatula BAC ends cultivar Jemalong A17 of Medicago truncatula, genomic survey sequence.
ACCESSION CR294134
VERSION CR294134.1 GI:44684558
KEYWORDS GSS.
SOURCE Medicago truncatula (barrel medic)
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.
REFERENCE 1 (bases 1 to 590)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (25-FEB-2004) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
FEATURES Location/Qualifiers
source 1. .590
/organism="Medicago truncatula"
/mol_type="genomic DNA"
/cultivar="Jemalong A17"
/db_xref="taxon:3880"
/clone_lib="MTE1"
/note="Vector: pIndigoBAC ; Site_1: EcoRI ; Site_2: EcoRI ; Debelle F. and Chalhoub B.~Genoscope sequence ID : mte1-1J18FM1"
ORIGIN

Query Match 6.7%; Score 334.8; DB 9; Length 590;
Best Local Similarity 79.8%; Pred. No. 2.5e-47;
Matches 407; Conservative 0; Mismatches 102; Indels 1; Gaps 1;

```
Qy        227 TTTATATATATATATATATATATATACACACACACACATATATATATACACATATATA 286

Db        532 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 473

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346

Db        472 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 413

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406

Db        412 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 353

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466

Db        352 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 293

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525

Db        292 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 233

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585

Db        232 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 173

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645

Db        172 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 113

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705

Db        112 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 53

Qy        706 TATGTATATATATATGCTCCAGAGTTCATA 735

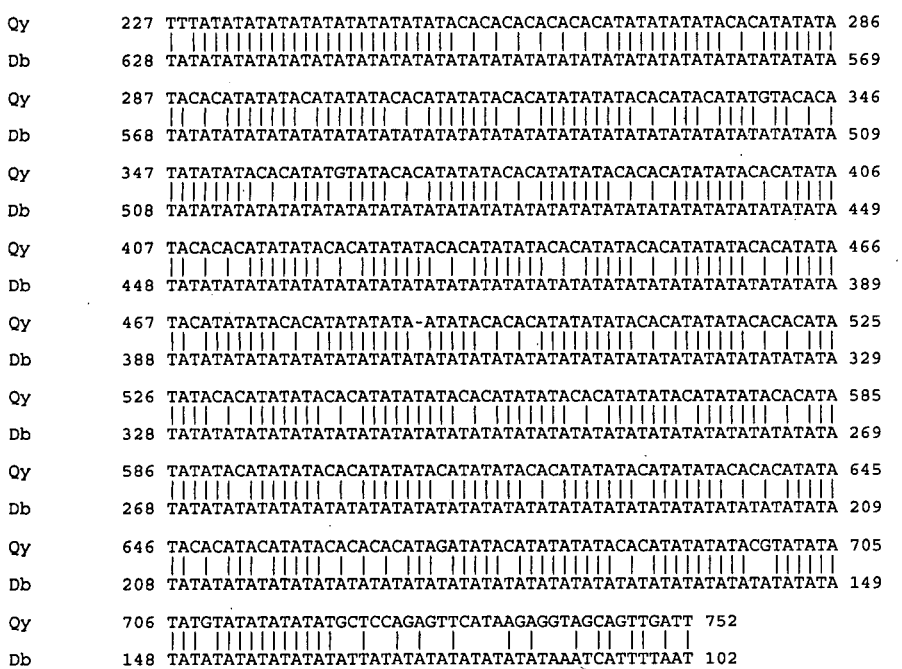
Db         52 TATATATATATATATATACTATACTATATA 23
```

RESULT 21
BI505681/c
LOCUS BI505681 717 bp mRNA linear EST 08-APR-2002
DEFINITION BB170032B10E06.5 Bee Brain Normalized/Subtracted Library, BB17 Apis mellifera cDNA clone BB170032B10E06 5', mRNA sequence.
ACCESSION BI505681
VERSION BI505681.1 GI:15356055
KEYWORDS EST.
SOURCE Apis mellifera (honey bee)
ORGANISM Apis mellifera
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Apoidea; Apidae; Apis.
REFERENCE 1 (bases 1 to 717)
AUTHORS Whitfield,C.W., Band,M.R., Bonaldo,M.F., Kumar,C.G., Liu,L., Pardinas,J., Robertson,H.M., Soares,B. and Robinson,G.E.
TITLE Annotated expressed sequence tags and cDNA microarrays for studies of brain and behavior in the honey bee
JOURNAL Genome Res. 12 (4), 555-566 (2002)
MEDLINE 21929762
PUBMED 11932240
COMMENT Contact: Gene E. Robinson
Department of Entomology
University of Illinois
505 S. Goodwin Ave., Urbana, IL 61801, USA
Tel: 217 265 0309
Fax: 217 244 3499
Email: generobi@life.uiuc.edu
This research was funded by the University of Illinois Critical Research Initiatives Fund and a Burroughs-Wellcome Trust Innovation Award in Functional Genomics to G.E. Robinson and an NSF Postdoctoral Fellowship in Bioinformatics to C.W. Whitfield.
REPEATS IN THE SEQUENCE
Low_complexity STRAND (+) ELEMENT AT_rich LOCATION [101,132];
Simple_repeat STRAND (+) ELEMENT (TA)n LOCATION [133,717].
PCR PRIMERS
FORWARD: TAATACGACTCACTATAGGG
BACKWARD: ATTAACCCTCACTAAAG
Plate: BB170032B10 row: E column: 06
Seq primer: AGCGGATAACAATTTCACACAGGA
High quality sequence stop: 717.
FEATURES Location/Qualifiers
source 1. .717
/organism="Apis mellifera"
/mol_type="mRNA"

/strain="mixed strains of European bees, predominantly A.m. ligustica"
/db_xref="taxon:7460"
/clone="BB170032B10E06"
/sex="female"
/tissue_type="brain"
/dev_stage="adult worker honey bee"
/lab_host="DH10B"
/clone_lib="Bee Brain Normalized/Subtracted Library, BB17"
/note="Organ: brain; Vector: pT7T3-Pac; Site_1: EcoR1; Site_2: Not1; This BB17 cDNA library was generated by subtraction of the BB16 library with 4000 previously sequenced clones. The BB16 library was contributed by the Soares laboratory and it was constructed and normalized as described by Bonaldo, M.F., Lennon, G. and Soares, M.B. (1996), Genome Research 6(9): 791-806. RNA was prepared from dissected brains of adult worker bees of various ages and various behavioral groups. "

ORIGIN

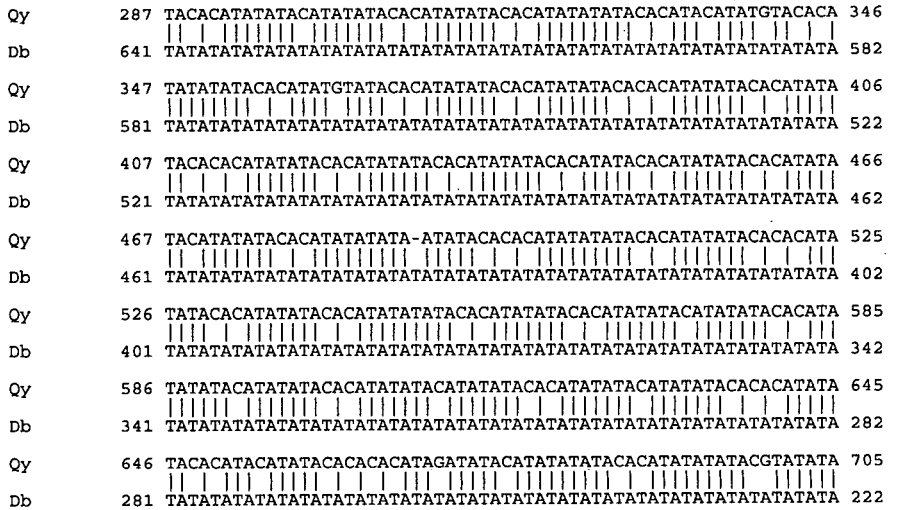
Query Match 6.7%; Score 334.2; DB 4; Length 717;
Best Local Similarity 78.4%; Pred. No. 3.1e-47;
Matches 413; Conservative 0; Mismatches 113; Indels 1; Gaps 1;



RESULT 22
AQ917102/c
LOCUS AQ917102 725 bp DNA linear GSS 07-DEC-1999
DEFINITION T233131b Medicago truncatula BAC library Medicago truncatula genomic clone 24E12, genomic survey sequence.
ACCESSION AQ917102
VERSION AQ917102.1 GI:6536747
KEYWORDS GSS.
SOURCE Medicago truncatula (barrel medic)
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.
REFERENCE 1 (bases 1 to 725)
AUTHORS Kim,D., Limpens,E., Peng,H., Ellis,L. and Cook,D.R.
TITLE BAC end sequencing of Medicago truncatula
JOURNAL Unpublished (1999)
COMMENT Contact: Cook DR
The Crop Biotechnology Center
Texas A&M University
Department of Plant Pathology and Microbiology, Rm 120 L.F. Peterson Bldg, College Station, TX 77843-2132, USA
Tel: 409 845 8743
Fax: 409 862 4790
Email: dcook@ppserver.tamu.edu
Other name: DK277-L; date: 8/12/99; Submitted to the Database of Genome Survey Sequences (GSS) on 12/06/99; More information is available at 'http://chrysie.tamu.edu/medicago'.
Seq primer: SQ-BAC-L (AACGCCAGGGTTTTCCCAGTCACGACG)
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .725
/organism="Medicago truncatula"
/mol_type="genomic DNA"
/cultivar="genotype A17"
/db_xref="taxon:3880"
/clone="24E12"
/clone_lib="Medicago truncatula BAC library"
/note="Vector: pBeloBAC11; Site_1: HindIII; Site_2: HindIII; Nam. Y-W, Penmetsa, R.V., Endre, G., Kim, D., and Cook, D.R. 1999. Construction of a bacterial artificial chromosome library of Medicago truncatula and identification of clones containing ethylene response genes. Theor Appl Genet 98: 638-646. "

ORIGIN

Query Match 6.7%; Score 333.8; DB 8; Length 725;
Best Local Similarity 80.4%; Pred. No. 3.6e-47;
Matches 403; Conservative 0; Mismatches 97; Indels 1; Gaps 1;



```
Db        221 TATATATATATATATCCATCA 201


RESULT 23
CL623140/c
LOCUS       CL623140               744 bp    DNA     linear   GSS 01-JUL-2004
DEFINITION  OR_BBa0018D11.r OR_BBa Oryza rufipogon genomic clone OR_BBa0018D11
            3', genomic survey sequence.
ACCESSION   CL623140
VERSION     CL623140.1  GI:49121028
KEYWORDS    GSS.
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 744)
  AUTHORS   Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R.,
            Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
  TITLE     OMAP Project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Insert Length: 161   Std Error: 0.00
            Plate: 0018  row: D  column: 11
            Seq primer: CAC TCA TTA GGC ACC CCA
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .744
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
                     /clone="OR_BBa0018D11"
                     /tissue_type="young leaves"
                     /lab_host="DH10B-T1 phage resistant"
                     /clone_lib="OR_BBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

  Query Match             6.7%;  Score 333;  DB 9;  Length 744;
  Best Local Similarity   79.3%;  Pred. No. 5e-47;
  Matches  407;  Conservative    0;  Mismatches  105;  Indels    1;  Gaps    1;

Qy        209 AGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACACATA 268
              | | | | | | |   || |||||||| |||||||||||||||| | | | | | |||
Db        702 ATATATATATATATATATATTATATATCTATATATATATATATATATATATATATATA 643

Qy        269 TATATATACACATATATATACACATATATACATATATACACATATATACACATATATATA 328
              |||||||| | ||||||||| | ||||||| ||||||| | ||||||| | |||||||||
Db        642 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 583

Qy        329 CACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACA 388
              | ||| |||| || | ||||||||| | |||| |||| | |||||||| | |||||||| |
Db        582 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 523

Qy        389 CACATATATACACATATATACACACATATATACACATATATACACATATATACACATATA 448
              | ||||||| | ||||||| | | ||||||| | ||||||| | ||||||| | |||||
Db        522 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 463

Qy        449 CACATATATACACATATATACATATATACACATATATATAATATACACACATATATATAC 508
              | |||||||| | ||||||| ||||||| | |||||||| ||||| | | |||||||||
Db        462 TATATATATATATATATATATATATATATATATATATAT-ATATATATATATATATAT 404

Qy        509 ACATATATACACACATATATACACATATATACACATATATATACACATATATACACATAT 568
              | ||||||| | | ||||||| | ||||||| | ||||||||| | ||||||| | ||||

Db        403 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 344

Qy        569 ATACATATATACACATATATATACATATATACACATATATACATATATACACATATATAC 628
              ||| ||||||| | ||||||||| ||||||| | ||||||| ||||||| | |||||||
Db        343 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 284

Qy        629 ATATATACACACATATATACACATACATATACACACACATAGATATACATATATATACAC 688
              ||||||| | | ||||||| | ||| ||||| | | | ||| ||||| |||||||||
Db        283 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 224

Qy        689 ATATATATACGTATATATATGTATATATATATG 721
              |||||||||  |||||||||||||| ||| ||||
Db        223 ATATATATATATATATATATGTATGTATGTATG 191


RESULT 24
BZ796807/c
LOCUS       BZ796807               512 bp    DNA     linear   GSS 17-MAR-2003
DEFINITION  PUGDF25TD ZM_0.6_1.0_KB Zea mays genomic clone ZMMBTa341F02,
            genomic survey sequence.
ACCESSION   BZ796807
VERSION     BZ796807.1  GI:28998547
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 512)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
            Bennetzen,J.
  TITLE     Maize Genomics Consortium
  JOURNAL   Unpublished (2003)
COMMENT     Other_GSSs: PUGDF25TB
            Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .512
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBTa341F02"
                     /clone_lib="ZM_0.6_1.0_KB"
                     /note="Vector: pCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high
                     CoT selected genomic DNA library"
ORIGIN

  Query Match             6.7%;  Score 332.6;  DB 8;  Length 512;
  Best Local Similarity   80.8%;  Pred. No. 6.1e-47;
  Matches  400;  Conservative    0;  Mismatches   94;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATACACATATATA 286
              | ||||||||||||||||||||||||||| | | | | | ||||||||||| | |||||||
Db        508 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 449

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | |||||||| | ||| |||| || | |
Db        448 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 389

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              ||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        388 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 329

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
```

```
Db        328 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 269

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db        268 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 209

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db        208 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 149

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db        148 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 89

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db         88 TATATATATATATATATATATATATATATATATATATATATATATATATATATAAATATATA 29

Qy        706 TATGTATATATATAT 720
              ||| ||||||||||||
Db         28 TATATATATATATAT 14


RESULT 25
CL755178/c
LOCUS       CL755178                 707 bp    DNA     linear   GSS 27-JUL-2004
DEFINITION  OR_BBa0122N12.r OR_BBa Oryza rufipogon genomic clone OR_BBa0122N12
            3', genomic survey sequence.
ACCESSION   CL755178
VERSION     CL755178.1  GI:50706035
KEYWORDS    GSS.
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 707)
  AUTHORS   Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R.,
            Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
  TITLE     OMAP Project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Insert Length: 161   Std Error: 0.00
            Plate: 0122  row: N  column: 12
            Seq primer: CAC TCA TTA GGC ACC CCA
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .707
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
                     /clone="OR_BBa0122N12"
                     /tissue_type="young leaves"
                     /lab_host="DH10B-T1 phage resistant"
                     /clone_lib="OR_BBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

  Query Match             6.7%;  Score 331.6;  DB 9;  Length 707;
  Best Local Similarity   79.4%;  Pred. No. 8.7e-47;
  Matches  405;  Conservative    0;  Mismatches  104;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||| | |||||||||||||| | | | | | |||||||||||| | |||||||

Db        563 TATATATATCTTTATATATATATATATATATATATATATATATATATATATATATATA 504

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||  |||| || | |
Db        503 TATATATATATATATATATATATATATATATATATATATATATATATATCTATATTTATATA 444

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              ||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        443 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 384

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db        383 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 324

Qy        467 TACATATATACACATATATATAATATACACACATATATATACACATATATACACACATAT 526
              || |||||||| | |||||||| ||||| | | ||||||||| | ||||||| | | ||||
Db        323 TATATATATATATATATATAT-ATATATATATATATATATATATATATATATATATATAT 265

Qy        527 ATACACATATATACACATATATATACACATATATACACATATATACATATATACACATAT 586
              ||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | ||||
Db        264 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 205

Qy        587 ATATACATATATACACATATATACATATATACACATATATACATATATACACACATATAT 646
              ||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | ||||||
Db        204 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 145

Qy        647 ACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATAT 706
              | | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  |||||||
Db        144 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 85

Qy        707 ATGTATATATATATGCTCCAGAGTTCATAA 736
              || |||||||||||| | | ||  | || |
Db         84 ATATATATATATATGATGCTGATATGATTA 55


RESULT 26
CL934691
LOCUS       CL934691                 784 bp    DNA     linear   GSS 14-SEP-2004
DEFINITION  OA_ABa0044J13.f OA_ABa Oryza australiensis genomic clone
            OA_ABa0044J13 5', genomic survey sequence.
ACCESSION   CL934691
VERSION     CL934691.1  GI:52063483
KEYWORDS    GSS.
SOURCE      Oryza australiensis
  ORGANISM  Oryza australiensis
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 784)
  AUTHORS   Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R.,
            Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
  TITLE     OMAP Project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0044  row: J  column: 13
            Seq primer: TAA TAC GAC TCA CTA TAG GG
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .784
                     /organism="Oryza australiensis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4532"
                     /clone="OA_ABa0044J13"
                     /tissue_type="young leaves"
```

```
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OA_ABa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

  Query Match             6.7%;  Score 331.2;  DB 9;  Length 784;
  Best Local Similarity   78.9%;  Pred. No. 1e-46;
  Matches  407;  Conservative    0;  Mismatches  108;  Indels    1;  Gaps    1;

Qy        205 TCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACA 264
              ||||   | | | | |       | ||||||||||||||||||||||||| | | | | | |
Db        201 TCTCTCTATATATATATATATATATATATATATATATATATATATATATATATATATATA 260

Qy        265 CATATATATATACACATATATATACACATATATACATATATACACATATATACACATATA 324
               |||||||||||| | ||||||||| | ||||||| ||||||| | ||||||| | |||||
Db        261 TATATATATATATATATATATATATATATATATATATATATATATAGATATATATATATATA 320

Qy        325 TATACACATACATATGTACACATATATATACACATATGTATACACATATATACACATATA 384
              |||| | ||| |||| || | |||||||||| | |||| |||| | ||||||| | |||||
Db        321 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 380

Qy        385 TACACACATATATACACATATATACACACATATATACACATATATACACATATATACACA 444
              || | | ||||||| | ||||||| | | ||||||| | ||||||| | ||||||| | |
Db        381 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 440

Qy        445 TATACACATATATACACATATATACATATATACACATATATAATATACACACATATAT 504
              |||| | ||||||| | ||||||| ||||||| | |||||||| ||||| | | ||||||
Db        441 TATATATATATATATATATATATATATATATATATATATATAT-ATATATATATATAT 499

Qy        505 ATACACATATATACACACATATATACACATATATACACATATATATACACATATATACAC 564
              ||| | ||||| | | |||||||| | ||||||| | ||||||||| | ||||||| |
Db        500 ATAGATATATAGATATATATATATATATATATATATATATATATATATATATATATAT 559

Qy        565 ATATATACATATATACACATATATATACATATATACACATATATACATATATACACATAT 624
              ||||||| ||||||| | ||||||||| ||||||| | ||||||| ||||||| | ||||
Db        560 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 619

Qy        625 ATACATATATACACACATATATACACATACATATACACACACATAGATATACATATATAT 684
              ||| ||||||| | | ||||||| | ||| ||||| | | | ||||||||| ||||||||
Db        620 ATATATATATATATATATATATATATATATATATATATATATATAGATATATATATATAT 679

Qy        685 ACACATATATATACGTATATATATGTATATATATAT 720
              | | |||||||||  ||||||||  ||||||||||
Db        680 ATATATATATATATATATATATAAATATATATATAT 715


RESULT 27
BZ796807
LOCUS       BZ796807                 512 bp    DNA     linear   GSS 17-MAR-2003
DEFINITION  PUGDF25TD ZM_0.6_1.0_KB Zea mays genomic clone ZMMBTa341F02,
            genomic survey sequence.
ACCESSION   BZ796807
VERSION     BZ796807.1  GI:28998547
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 512)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
            Bennetzen,J.
  TITLE     Maize Genomics Consortium
  JOURNAL   Unpublished (2003)
COMMENT     Other_GSSs: PUGDF25TB
            Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .512
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBTa341F02"
                     /clone_lib="ZM_0.6_1.0_KB"
                     /note="Vector: pCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high
                     CoT selected genomic DNA library"
ORIGIN

  Query Match             6.7%;  Score 331;  DB 8;  Length 512;
  Best Local Similarity   80.6%;  Pred. No. 1.1e-46;
  Matches  399;  Conservative    0;  Mismatches   95;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | |||||||||||||||||||||||||| | | | |    | |||||||||||| | |||||||
Db          1 TATATATATATATATATATATATATATATATATATTTATATATATATATATATATATA 60

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db         61 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 120

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        121 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 180

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db        181 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 240

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db        241 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 300

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db        301 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 360

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db        361 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 420

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db        421 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 480

Qy        706 TATGTATATATATAT 720
              |||  ||||||||||||
Db        481 TATATATATATATAT 495


RESULT 28
BX145650/c
LOCUS       BX145650                 668 bp    DNA     linear   GSS 28-JAN-2003
DEFINITION  Danio rerio genomic clone DKEY-109N18, genomic survey sequence.
ACCESSION   BX145650
VERSION     BX145650.1  GI:27977003
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 668)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the T7 end of BAC 109N18. 109N18
```

```
           is part of the Daniokey BAC Library created by R. Plasterk and N.V.
           Keygene. Further details:
           http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .668
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-109N18"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             6.6%;  Score 330.4;  DB 9;  Length 668;
  Best Local Similarity   78.5%;  Pred. No. 1.4e-46;
  Matches  408;  Conservative    0;  Mismatches  111;  Indels    1;  Gaps    1;

Qy        229 TATATATATATATATATATATATACACACACACACACATATATATATACACATATATATA 288
              ||||||||||||||||||||||| | | | | | |||||||||||| | |||||||||
Db        668 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 609

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              | ||||||| ||||||| | |||||||| | ||||||||| | ||| |||| || | |||
Db        608 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 549

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              |||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||||
Db        548 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 489

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||||
Db        488 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 429

Qy        469 CATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATATA 527
              ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||||
Db        428 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 369

Qy        528 TACACATATATACACATATATATACACATATATACACATATATACATATATACACATATA 587
              || | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||||
Db        368 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 309

Qy        588 TATACATATATACACATATATACATATATACACATATATACATATATACACACATATATA 647
              |||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||||
Db        308 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 249

Qy        648 CACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATA 707
              | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||||
Db        248 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 189

Qy        708 TGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGT 747
              | |||| | |  |  ||   ||  || || |   | |||
Db        188 TATATAAACACTTCATCTCTAGACCAAAACACCCATGAGT 149


RESULT 29
CL755178
LOCUS       CL755178               707 bp    DNA     linear   GSS 27-JUL-2004
DEFINITION  OR_BBa0122N12.r OR_BBa Oryza rufipogon genomic clone OR_BBa0122N12
            3', genomic survey sequence.
ACCESSION   CL755178
VERSION     CL755178.1  GI:50706035
KEYWORDS    GSS.
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 707)
  AUTHORS   Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R.,
            Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
  TITLE     OMAP Project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Insert Length: 161   Std Error: 0.00
            Plate: 0122  row: N  column: 12
            Seq primer: CAC TCA TTA GGC ACC CCA
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .707
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
                     /clone="OR_BBa0122N12"
                     /tissue_type="young leaves"
                     /lab_host="DH10B-T1 phage resistant"
                     /clone_lib="OR_BBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

  Query Match             6.6%;  Score 330.4;  DB 9;  Length 707;
  Best Local Similarity   70.8%;  Pred. No. 1.4e-46;
  Matches  453;  Conservative    0;  Mismatches  186;  Indels    1;  Gaps    1;

Qy        228 TTATATATATATATATATATATATACACACACACACATATATATATACACATATATAT 287
              | |||||||||||||||||||||| | | | | | |||||||||||| | ||||||||
Db         69 TCATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 128

Qy        288 ACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACAT 347
              | | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | ||
Db        129 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 188

Qy        348 ATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATAT 407
              ||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | ||||||
Db        189 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 248

Qy        408 ACACACATATATACACATATATACACATATATACACATATACACATATATACACATATAT 467
              | | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | ||||||
Db        249 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 308

Qy        468 ACATATATACACATATATATAATATACACACATATATATACACATATATACACACATATA 527
              | ||||||| | ||||||||| ||||| | | |||||||||| | ||||||| | | |||||
Db        309 ATATATATATATATATATAT-ATATATATATATATATATATATATATATATATATATATA 367

Qy        528 TACACATATATACACATATATATACACATATATACACATATATACATATATACACATATA 587
              || | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||||
Db        368 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 427

Qy        588 TATACATATATACACATATATACATATATACACATATATACATATATACACACATATATA 647
              |||| ||||||| | ||||||| ||||| | | ||||||| ||||||| | | |||||||
Db        428 TATATATATATATATATATATATAAATATAGATATATATATATATATATATATATATATA 487

Qy        648 CACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATA 707
              | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||||
Db        488 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 547

Qy        708 TGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGA 767
              | || | ||||||          |  |||          |  |   ||  | |    ||| |
Db        548 TATAAAGATATATATAGATATATATATATATATATATATATATATATATAAATATATATA 607

Qy        768 GGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATA 827
              | | | |  | |  | | |||    | |  | || |      |          | |||
Db        608 TATATATATATATATATATATAAATATATATATATATATATATATATATATATATATA 667

Qy        828 GTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA 867
                |  ||||| | | || | ||   | |  ||| | ||
Db        668 TATAGACAAATAGAGATATATATATATATATATATATA 707
```

```
RESULT 30
CA916667/c
LOCUS       CA916667                496 bp    mRNA    linear   EST 27-DEC-2002
DEFINITION  PCSC21302 Scarlet Runner Bean Suspensor Region TriplEx2 Phaseolus
            coccineus cDNA 5' similar to No significant hits, mRNA sequence.
ACCESSION   CA916667
VERSION     CA916667.1  GI:27403659
KEYWORDS    EST.
SOURCE      Phaseolus coccineus
  ORGANISM  Phaseolus coccineus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Phaseolus.
REFERENCE   1  (bases 1 to 496)
  AUTHORS   Bui,A.Q., Le,B.H., Weterings,K., Bi,Y.-P., Choi,J.-S.,
            McElroy,K.E., Choi,P.S., Harada,J.J., Fischer,R.L. and
            Goldberg,R.B.
  TITLE     Gene Activity in Different Regions of a Post-Fertilization Plant
            Embryo by EST Analysis
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Goldberg, R.B.
            Department of Molecular, Cell, & Developmental Biology
            University of California, Los Angeles
            621 Charles E. Young Drive South, Los Angeles, CA 90095-1606, USA
            Tel: 310 825 3270
            Fax: 310 825 8201
            Email: bobg@ucla.edu
            Seq primer: 5' TriplEx
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .496
                     /organism="Phaseolus coccineus"
                     /mol_type="mRNA"
                     /cultivar="Hammond's Dwarf Scarlet"
                     /db_xref="taxon:3886"
                     /dev_stage="6-days post-pollination"
                     /clone_lib="Scarlet Runner Bean Suspensor Region TriplEx2"
                     /note="Organ: Suspensor Region of Globular-Stage Embryos;
                     Vector: TriplEx2; Site_1: SfiIA; Site_2: SfiIB; Suspensor
                     regions were micro-dissected from globular-stage embryos
                     six days after pollination from greenhouse-grown plants
                     [Weterings et al., Plant Cell 13, 2409-2425 (2001)].
                     Double-stranded cDNA was synthesized from suspensor mRNA
                     using the SMART cDNA Library Construction Kit according to
                     the manufacturer (Clontech). The suspensor cDNA fragments
                     were directionally ligated into the SfiI restriction site
                     of the lambda TriplEx2 vector (Clontech), and the
                     recombinant cDNAs were transformed into E. coli XL1-Blue
                     cells (Clontech). Suspensor cDNA plasmids used for
                     directional sequencing were obtained by in vivo excision
                     from the lambda TriplEx2 recombinants in E. coli BM25.8
                     cells (Clontech)."
ORIGIN

  Query Match             6.6%;  Score 330.2;  DB 6;  Length 496;
  Best Local Similarity   80.9%;  Pred. No. 1.6e-46;
  Matches  397;  Conservative    0;  Mismatches   93;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db        491 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 432

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db        431 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 372

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        371 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 312

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db        311 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 252

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db        251 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 192

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db        191 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 132

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db        131 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 72

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| |||||||||||||||||| |||||||||  |||| |
Db         71 TATATATATATATATATATATATATATATACATATATATACATATATATATATATACA 12

Qy        706 TATGTATATAT 716
              ||| |||||||
Db         11 TATCTATATAT 1


RESULT 31
CF924719/c
LOCUS       CF924719                751 bp    mRNA    linear   EST 10-NOV-2003
DEFINITION  DD005F11.ab1 Codonopsis lanceolata cDNA library Codonopsis
            lanceolata cDNA, mRNA sequence.
ACCESSION   CF924719
VERSION     CF924719.1  GI:38243245
KEYWORDS    EST.
SOURCE      Codonopsis lanceolata
  ORGANISM  Codonopsis lanceolata
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            asterids; campanulids; Asterales; Campanulaceae; Codonopsis.
REFERENCE   1  (bases 1 to 751)
  AUTHORS   Yang,D.C., In,J.G. and Lee,B.S.
  TITLE     Codonopsis lanceolata
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Deok Chun Yang
            BioPia
            1 Seochonri, Kihung, Yongin, Kyungki-do, 449-701, Korea
            Tel: 82 31 201 2688
            Fax: 82 31 205 6323
            Email: dcyang@ibiopia.com.
FEATURES             Location/Qualifiers
     source          1. .751
                     /organism="Codonopsis lanceolata"
                     /mol_type="mRNA"
                     /db_xref="taxon:103999"
                     /clone_lib="Codonopsis lanceolata cDNA library"
                     /note="Organ: root; Vector: pTriplEx2; Site_1: Sfi IA;
                     Site_2: Sfi IB"
ORIGIN

  Query Match             6.6%;  Score 329.8;  DB 7;  Length 751;
  Best Local Similarity   80.3%;  Pred. No. 1.8e-46;
  Matches  399;  Conservative    0;  Mismatches   97;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db        749 TATATATATATATATATATATATATAGATATATATATATATATATATATATATATATA 690

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db        689 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 630

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        629 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 570
```

```
Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db        569 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 510

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | |||||||| ||||| | | |||||||| | ||||||| | | |||
Db        509 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 450

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | |||||||| | ||||||| | ||||||| ||||||| | |||
Db        449 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 390

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db        389 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 330

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | ||| |||||  |||||
Db        329 TATATATATATATATATATATATATATATATATATATATATATATATATATAGATATATATATATA 270

Qy        706 TATGTATATATATATGC 722
              ||  ||| |||||||||
Db        269 TAGATATGTATATATGC 253
```

RESULT 32
BX145650
LOCUS BX145650 668 bp DNA linear GSS 28-JAN-2003
DEFINITION Danio rerio genomic clone DKEY-109N18, genomic survey sequence.
ACCESSION BX145650
VERSION BX145650.1 GI:27977003
KEYWORDS GSS.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 668)
AUTHORS Humphray,S.J., Huckle,E. and Durham,J.L.
TITLE Direct Submission
JOURNAL Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Unpublished
COMMENT This sequence was generated from the T7 end of BAC 109N18. 109N18 is part of the Daniokey BAC Library created by R. Plasterk and N.V. Keygene. Further details: http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES Location/Qualifiers
source 1. .668
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="DKEY-109N18"
/tissue_type="Testis"
/note="vector pIndigoBAC-536"
ORIGIN

Query Match 6.6%; Score 329; DB 9; Length 668;
Best Local Similarity 79.8%; Pred. No. 2.4e-46;
Matches 400; Conservative 0; Mismatches 100; Indels 1; Gaps 1;

```
Qy        217 AAAAAGGCATTTTATATATATATATATATATATACACACACACACACATATATATATA 276
              | | | | | | ||||||||||||||||||||||| | | | | | ||||||||||
Db        169 AGAGATGAAGTGTTTATATATATATATATATATATATATATATATATATATATATATA 228

Qy        277 CACATATATATACACATATATACATATATACACATATATACACATATATATACACATACA 336
              | ||||||||| | ||||||| ||||||| | ||||||| | ||||||||| | ||| |
Db        229 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 288

Qy        337 TATGTACACATATATATACACATATGTATACACATATATACACATATATACACACATATA 396
              ||| || | ||||||||| | |||| |||| | ||||||| | ||||||| | | |||||
Db        289 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 348
```

```
Qy        397 TACACATATATACACACATATATACACATATATACACATATATACACATATACACATATA 456
              || | ||||||| | | |||||||| | ||||||| | ||||||| | ||||| | |||||
Db        349 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 408

Qy        457 TACACATATATACATATATACACATATATATAATATACACACATATATATACACATATAT 516
              || | ||||||| ||||||| | |||||||| ||||| | | ||||||||| | ||||||
Db        409 TATATATATATATATATATATATATATATAT-ATATATATATATATATATATATATATAT 467

Qy        517 ACACACATATATACACATATATACACATATATATACACATATATACACATATATACACATAT 576
              | | | ||||||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||
Db        468 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 527

Qy        577 ATACACATATATATACATATATACACATATATACATATATACACATATATACATATATAC 636
              ||| | ||||||||| ||||||| | ||||||| ||||||| | ||||||| |||||||
Db        528 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 587

Qy        637 ACACATATATACACATACATATACACACACATAGATATACATATATATACACATATATAT 696
              | | ||||||| | ||| ||||| | | | ||| ||||| ||||||||| | ||||||||
Db        588 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 647

Qy        697 ACGTATATATATGTATATATA 717
              |  |||||||||  ||||||||
Db        648 ATATATATATATATATATATA 668
```

RESULT 33
CL934691/c
LOCUS CL934691 784 bp DNA linear GSS 14-SEP-2004
DEFINITION OA_ABa0044J13.f OA_ABa Oryza australiensis genomic clone OA_ABa0044J13 5', genomic survey sequence.
ACCESSION CL934691
VERSION CL934691.1 GI:52063483
KEYWORDS GSS.
SOURCE Oryza australiensis
ORGANISM Oryza australiensis
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 784)
AUTHORS Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
TITLE OMAP Project
JOURNAL Unpublished (2004)
COMMENT Contact: Rod A. Wing
Arizona Genomics Institute
University of Arizona
Forbes Building Room 303, Tucson, AZ 85721-0036, USA
Tel: 520 626 9595
Fax: 520 621 1259
Email: http://genome.arizona.edu
PCR PRimers
FORWARD: TAA TAC GAC TCA CTA TAG GG
BACKWARD: CAC TCA TTA GGC ACC CCA
Plate: 0044 row: J column: 13
Seq primer: TAA TAC GAC TCA CTA TAG GG
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .784
/organism="Oryza australiensis"
/mol_type="genomic DNA"
/db_xref="taxon:4532"
/clone="OA_ABa0044J13"
/tissue_type="young leaves"
/lab_host="DH10B T1 phage resistant"
/clone_lib="OA_ABa"
/note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

Query Match 6.6%; Score 328.4; DB 9; Length 784;
Best Local Similarity 79.0%; Pred. No. 3e-46;
Matches 403; Conservative 0; Mismatches 106; Indels 1; Gaps 1;

```
Qy        212 AAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACATATAT 271
```

```
Db        719 ATAAATATATATATATTTATATATATATATATATATATATATATATATATATATCTATAT 660

Qy        272 ATATACACATATATATACACATATATACATATATACACATATATACACATATATATACAC 331

Db        659 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 600

Qy        332 ATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACACAC 391

Db        599 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 540

Qy        392 ATATATACACATATATACACACATATATACACATATATACACATATATACACATATACAC 451

Db        539 ATATATATATATATATATATATATATATCTATATATCTATATATATATATATATATATAT 480

Qy        452 ATATATACACATATATACATATATACACATATATATA-ATATACACACATATATATACAC 510

Db        479 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 420

Qy        511 ATATATACACACATATATACACATATATACACATATATATACACATATATACACATATAT 570

Db        419 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 360

Qy        571 ACATATATACACATATATATACATATATACACATATATACATATATACACATATATACAT 630

Db        359 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATCTAT 300

Qy        631 ATATACACACATATATACACATACATATACACACACATAGATATACATATATATACACAT 690

Db        299 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 240

Qy        691 ATATATACGTATATATATGTATATATATAT 720

Db        239 ATATATATATATATATATATATATATATAT 210
```

RESULT 34
CL776144/c
LOCUS CL776144 587 bp DNA linear GSS 30-JUL-2004
DEFINITION OR_BBa0090B12.r OR_BBa Oryza rufipogon genomic clone OR_BBa0090B12 3', genomic survey sequence.
ACCESSION CL776144
VERSION CL776144.1 GI:50850530
KEYWORDS GSS.
SOURCE Oryza rufipogon
ORGANISM Oryza rufipogon
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 587)
AUTHORS Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
TITLE OMAP Project
JOURNAL Unpublished (2004)
COMMENT Contact: Rod A. Wing
Arizona Genomics Institute
University of Arizona
Forbes Building Room 303, Tucson, AZ 85721-0036, USA
Tel: 520 626 9595
Fax: 520 621 1259
Email: http://genome.arizona.edu
PCR PRimers
FORWARD: TAA TAC GAC TCA CTA TAG GG
BACKWARD: CAC TCA TTA GGC ACC CCA
Insert Length: 161 Std Error: 0.00
Plate: 0090 row: B column: 12
Seq primer: CAC TCA TTA GGC ACC CCA
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .587
/organism="Oryza rufipogon"
/mol_type="genomic DNA"
/db_xref="taxon:4529"
/clone="OR_BBa0090B12"
/tissue_type="young leaves"
/lab_host="DH10B-T1 phage resistant"
/clone_lib="OR_BBa"
/note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

Query Match 6.6%; Score 328.2; DB 9; Length 587;
Best Local Similarity 77.5%; Pred. No. 3.4e-46;
Matches 410; Conservative 0; Mismatches 118; Indels 1; Gaps 1;

```
Qy        209 AGAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACATA 268

Db        537 ATATATATATATATATATATAATATATATATATATATATATATATATATATATATATA 478

Qy        269 TATATATACACATATATATACACATATATACATATATACACATATATACACATATATATA 328

Db        477 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 418

Qy        329 CACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACA 388

Db        417 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 358

Qy        389 CACATATATACACATATATACACACATATATACACATATATACACATATATACACATATA 448

Db        357 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 298

Qy        449 CACATATATACACATATATACATATATACACATATATATA-ATATACACACATATATATA 507

Db        297 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 238

Qy        508 CACATATATACACACATATATACACATATATACACATATATATACACATATATACACATA 567

Db        237 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 178

Qy        568 TATACATATATACACATATATATACATATATACACATATATACATATATACACATATATA 627

Db        177 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 118

Qy        628 CATATATACACACATATATACACATACATATACACACACATAGATATACATATATATACA 687

Db        117 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 58

Qy        688 CATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAA 736

Db         57 TATATATATATATATATATGCATATGCGCCTGTGTATAAGAAATCATAA 9
```

RESULT 35
CF924719
LOCUS CF924719 751 bp mRNA linear EST 10-NOV-2003
DEFINITION DD005F11.ab1 Codonopsis lanceolata cDNA library Codonopsis lanceolata cDNA, mRNA sequence.
ACCESSION CF924719
VERSION CF924719.1 GI:38243245
KEYWORDS EST.
SOURCE Codonopsis lanceolata
ORGANISM Codonopsis lanceolata
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids; campanulids; Asterales; Campanulaceae; Codonopsis.
REFERENCE 1 (bases 1 to 751)
AUTHORS Yang,D.C., In,J.G. and Lee,B.S.
TITLE Codonopsis lanceolata
JOURNAL Unpublished (2003)
COMMENT Contact: Deok Chun Yang
BioPia
1 Seochonri, Kihung, Yongin, Kyungki-do, 449-701, Korea
Tel: 82 31 201 2688
Fax: 82 31 205 6323
Email: dcyang@ibiopia.com.
FEATURES Location/Qualifiers
source 1. .751

```
                     /organism="Codonopsis lanceolata"
                     /mol_type="mRNA"
                     /db_xref="taxon:103999"
                     /clone_lib="Codonopsis lanceolata cDNA library"
                     /note="Organ: root; Vector: pTriplEx2; Site_1: Sfi IA;
                     Site_2: Sfi IB"
ORIGIN

  Query Match             6.6%;  Score 327.2;  DB 7;  Length 751;
  Best Local Similarity   80.0%;  Pred. No. 4.9e-46;
  Matches  397;  Conservative    0;  Mismatches   98;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | |||| |||| |||||||||||||||    | | | | ||||||||||| | |||||||
Db        256 TATATACATATCTATATATATATATATCTATATATATATATATATATATATATATATATA 315

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| |||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db        316 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 375

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        376 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 435

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db        436 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 495

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db        496 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 555

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db        556 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 615

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db        616 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 675

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||| ||||  ||||||
Db        676 TATATATATATATATATATATATATATATATATATATATATATATATATATCTATATATATA 735

Qy        706 TATGTATATATATATG 721
              ||| ||||||||||||
Db        736 TATATATATATATATG 751


RESULT 36
CA916667
LOCUS       CA916667                496 bp    mRNA    linear   EST 27-DEC-2002
DEFINITION  PCSC21302 Scarlet Runner Bean Suspensor Region TriplEx2 Phaseolus
            coccineus cDNA 5' similar to No significant hits, mRNA sequence.
ACCESSION   CA916667
VERSION     CA916667.1  GI:27403659
KEYWORDS    EST.
SOURCE      Phaseolus coccineus
  ORGANISM  Phaseolus coccineus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Phaseolus.
REFERENCE   1  (bases 1 to 496)
  AUTHORS   Bui,A.Q., Le,B.H., Weterings,K., Bi,Y.-P., Choi,J.-S.,
            McElroy,K.E., Choi,P.S., Harada,J.J., Fischer,R.L. and
            Goldberg,R.B.
  TITLE     Gene Activity in Different Regions of a Post-Fertilization Plant
            Embryo by EST Analysis
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Goldberg, R.B.
            Department of Molecular, Cell, & Developmental Biology
            University of California, Los Angeles
            621 Charles E. Young Drive South, Los Angeles, CA 90095-1606, USA
            Tel: 310 825 3270
            Fax: 310 825 8201
            Email:-bobg@ucla.edu
            Seq primer: 5' TriplEx
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .496
                     /organism="Phaseolus coccineus"
                     /mol_type="mRNA"
                     /cultivar="Hammond's Dwarf Scarlet"
                     /db_xref="taxon:3886"
                     /dev_stage="6-days post-pollination"
                     /clone_lib="Scarlet Runner Bean Suspensor Region TriplEx2"
                     /note="Organ: Suspensor Region of Globular-Stage Embryos;
                     Vector: TriplEx2; Site_1: SfiIA; Site_2: SfiIB; Suspensor
                     regions were micro-dissected from globular-stage embryos
                     six days after pollination from greenhouse-grown plants
                     [Weterings et al., Plant Cell 13, 2409-2425 (2001)].
                     Double-stranded cDNA was synthesized from suspensor mRNA
                     using the SMART cDNA Library Construction Kit according to
                     the manufacturer (Clontech). The suspensor cDNA fragments
                     were directionally ligated into the SfiI restriction site
                     of the lambda TriplEx2 vector (Clontech), and the
                     recombinant cDNAs were transformed into E. coli XL1-Blue
                     cells (Clontech). Suspensor cDNA plasmids used for
                     directional sequencing were obtained by in vivo excision
                     from the lambda TriplEx2 recombinants in E. coli BM25.8
                     cells (Clontech)."
ORIGIN

  Query Match             6.6%;  Score 326.2;  DB 6;  Length 496;
  Best Local Similarity   80.0%;  Pred. No. 7.7e-46;
  Matches  396;  Conservative    0;  Mismatches   98;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | |||| |||| |||||||||||||||    | | | | ||||||||||| | |||||||
Db          2 TATATAGATATGTATATATATATATATATGTATATATATGTATATATATATATATATATA 61

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| |||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db         62 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 121

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        122 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 181

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db        182 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 241

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db        242 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 301

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db        302 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 361

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db        362 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 421

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db        422 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 481

Qy        706 TATGTATATATATAT 720
              ||| ||||||||||||
Db        482 TATATATATATATAT 496


RESULT 37
```

```
CL623140
LOCUS       CL623140                744 bp    DNA     linear   GSS 01-JUL-2004
DEFINITION  OR_BBa0018D11.r OR_BBa Oryza rufipogon genomic clone OR_BBa0018D11
            3', genomic survey sequence.
ACCESSION   CL623140
VERSION     CL623140.1  GI:49121028
KEYWORDS    GSS.
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 744)
  AUTHORS   Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R.,
            Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
  TITLE     OMAP Project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Insert Length: 161    Std Error: 0.00
            Plate: 0018  row: D  column: 11
            Seq primer: CAC TCA TTA GGC ACC CCA
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .744
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
                     /clone="OR_BBa0018D11"
                     /tissue_type="young leaves"
                     /lab_host="DH10B-T1 phage resistant"
                     /clone_lib="OR_BBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

  Query Match             6.5%;  Score 324.2;  DB 9;  Length 744;
  Best Local Similarity   79.2%;  Pred. No. 1.6e-45;
  Matches  397;  Conservative    0;  Mismatches  103;  Indels    1;  Gaps    1;

Qy        217 AAAAAGGCATTTTATATATATATATATATATATACACACACACACACATATATATATA 276
              |||||      |  ||| ||| ||||||||||||||| | | | | | ||||||||||
Db        183 AAAAAATACATACATACATACATATATATATATATATATATATATATATATATATATA 242

Qy        277 CACATATATATACACATATATACATATATACACATATATACACATATATATACACATACA 336
              | |||||||||| | ||||||| ||||||| | ||||||| | |||||||||| | ||| |
Db        243 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 302

Qy        337 TATGTACACATATATATACACATATGTATACACATATATACACATATATACACACATATA 396
              ||| || | |||||||||| | |||| |||| | ||||||| | ||||||| | | |||||
Db        303 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 362

Qy        397 TACACATATATACACACATATATACACATATATACACATATATACACATATACACATATA 456
              || | ||||||| | | ||||||| | ||||||| | ||||||| | ||||| | |||||
Db        363 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 422

Qy        457 TACACATATATACATATATACACATATATATAATATACACACATATATATACACATATAT 516
              || | ||||||| ||||||| | |||||||| ||||| | | |||||||||| | ||||||
Db        423 TATATATATATATATATATATATATATATATAT-ATATATATATATATATATATATATAT 481

Qy        517 ACACACATATATACACATATATACACATATATATACACATATATACACATATATACATAT 576
              | | | ||||||| | ||||||| | |||||||||| | ||||||| | ||||||| ||||
Db        482 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 541

Qy        577 ATACACATATATATACATATATACACATATATACATATATACACATATATACATATATAC 636
              ||| | ||||||||| ||||||| | ||||||| ||||||| | ||||||| |||||||
Db        542 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 601

Qy        637 ACACATATATACACATACATATACACACATAGATATACATATATATACACATATATAT 696
              | | ||||||| | ||| ||||| | | | ||| ||||| ||||||||| | ||||||||
Db        602 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 661

Qy        697 ACGTATATATATGTATATATA 717
              |  |||||||||  |||||||
Db        662 ATATATATATATAGATATATA 682


RESULT 38
CL776144
LOCUS       CL776144                587 bp    DNA     linear   GSS 30-JUL-2004
DEFINITION  OR_BBa0090B12.r OR_BBa Oryza rufipogon genomic clone OR_BBa0090B12
            3', genomic survey sequence.
ACCESSION   CL776144
VERSION     CL776144.1  GI:50850530
KEYWORDS    GSS.
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 587)
  AUTHORS   Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R.,
            Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
  TITLE     OMAP Project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Insert Length: 161    Std Error: 0.00
            Plate: 0090  row: B  column: 12
            Seq primer: CAC TCA TTA GGC ACC CCA
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .587
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
                     /clone="OR_BBa0090B12"
                     /tissue_type="young leaves"
                     /lab_host="DH10B-T1 phage resistant"
                     /clone_lib="OR_BBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

  Query Match             6.5%;  Score 323.8;  DB 9;  Length 587;
  Best Local Similarity   77.8%;  Pred. No. 1.9e-45;
  Matches  416;  Conservative    0;  Mismatches  117;  Indels    2;  Gaps    2;

Qy        188 ACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATAT 247
              |||   ||| |    ||| |    | | | |        | ||||||||||||||||||
Db         23 ACACAGGCGCATATGCATATATATATACACATATATATACACATATATACATATATAC 82

Qy        248 ATATACACACACACACACATATATATATACACATATATATACACATATATACATATATAC 307
              ||||| | | | | | | |||||||||| | |||||||| | ||||||| |||||||
Db         83 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 142

Qy        308 ACATATATACACATATATATACACATACATATGTACACATATATATACACATATGTATAC 367
              | ||||||| | |||||||||| | ||| |||| || | |||||||||| | |||| ||||
Db        143 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 202

Qy        368 ACATATATACACATATATACACACATATATACACATATATACACACATATATACACATAT 427
              | ||||||| | ||||||| | | ||||||| | ||||||| | | ||||||| | ||||
```

```
Db        203 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 262

Qy        428 ATACACATATATACACATATACACATATATACACATATATACATATATACACATATATAT 487
              ||| | ||||||| | ||||| | ||||||| | ||||||| ||||||| | ||||||||
Db        263 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 322

Qy        488 A-ATATACACACATATATATACACATATATACACACATATATACACATATATACACATAT 546
              | ||||| | | ||||||||| | ||||||| | | ||||||| | ||||||| | ||||
Db        323 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 382

Qy        547 ATATACACATATATACACATATATACATATATACACATATATATACATATATACACATAT 606
              ||||| | ||||||| | ||||||| ||||||| | ||||||||| ||||||| | ||||
Db        383 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 442

Qy        607 ATACATATATACACATATATACATATATACACACATATATACACATACATATACACACAC 666
              ||| ||||||| | ||||||| ||||||| | | ||||||| | ||| ||||| | | |
Db        443 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 502

Qy        667 ATAGATATACATA-TATATACACATATATATACGTATATATATGTATATATATAT 720
              ||| ||||| ||| |||||| | ||||||||| |||||||| ||||||||||
Db        503 ATATATATATATATTATATATATATATATATATATATATATATATATATATATATAT 557


RESULT 39
CL830207/c
LOCUS       CL830207                491 bp    DNA     linear   GSS 09-AUG-2004
DEFINITION  OR_CBa0052C06.f OR_CBa Oryza rufipogon genomic clone OR_CBa0052C06
            5', genomic survey sequence.
ACCESSION   CL830207
VERSION     CL830207.1  GI:51075808
KEYWORDS    GSS.
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 491)
  AUTHORS   Kim,H., Yu,Y., Wissotski,M., Yost,D., Stum,D., Rao,K., Luo,M.,
            Jetty,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and
            Wing,R.
  TITLE     OMAP project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0052  row: C  column: 06
            Seq primer: TAA TAC GAC TCA CTA TAG GG
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .491
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
                     /clone="OR_CBa0052C06"
                     /tissue_type="young leaves"
                     /dev_stage="2 week old seedlings"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OR_CBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII;
                     drk treated 36 hrs before harvest"
ORIGIN

  Query Match             6.5%;  Score 322.6;  DB 9;  Length 491;
  Best Local Similarity   80.2%;  Pred. No. 3.2e-45;
  Matches  392;  Conservative    0;  Mismatches   94;  Indels    3;  Gaps    1;

Qy        232 ATATATATATATATATATATACACACACACACATATATATATACACATATATATACAC 291
              ||||| ||||||||||||||| | | | | | |||||||||| |  ||| |||||||
Db        491 ATATACATATATATATATATATATATATATATATATATATATAGATCTATGTATACAC 432

Qy        292 ATATATACATATATACACATATATACACATATATATACACATACATATGTACACATATAT 351
              ||||||| ||||||| | ||| | | |||| |||  | ||| ||| ||| ||||||||
Db        431 ATATATATATATATATATATACAGATAGATATGTATGTATATATATAGGTATACATATAT 372

Qy        352 ATACACATATGTATACACATATATACACATATATACACACATATATACACATATATACAC 411
              ||| |||||| |||| | ||||||| | ||||||  | | ||||||| | ||||||| |
Db        371 ATATACATATATATATATATATATATATATATATCTATATATATATAGATATATATAT 312

Qy        412 ACATATATACACATATATACACATATATACACATATACACATATATACACATATATACAT 471
              | ||||| | | ||||||| | ||||||| | ||||| | ||||||| | ||||||| ||
Db        311 ATATATAGAGATATATATATATATATATATATATATATATATATATATATATATATAT 252

Qy        472 ATATACACATATATATAATATACACACATATATATACACATATATACACACATATATACA 531
              ||||| | ||||||   ||||| | | ||||||||| | ||||||| | | ||||||| |
Db        251 ATATATATATATAT---ATATATATATATATATATATATATATATAGATATATATATATA 195

Qy        532 CATATATACACATATATATACACATATATACACATATATACATATATACACATATATATA 591
              ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||||||||
Db        194 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 135

Qy        592 CATATATACACATATATACATATATACACATATATACATATATACACACATATATACACA 651
              ||||||| | ||||||| ||||||| |||||||||||||||||||||| ||||||| | |
Db        134 TATATATATATATATATATATATATATACATATATACATATATACACATATATATATATA 75

Qy        652 TACATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTA 711
              || ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||||| ||
Db         74 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 15

Qy        712 TATATATAT 720
              |||||||||
Db         14 TATATATAT 6


RESULT 40
CC099198
LOCUS       CC099198                845 bp    DNA     linear   GSS 16-APR-2003
DEFINITION  CSU-K34.116B12.T7 CSU-K34 Aedes aegypti genomic clone
            CSU-K34.116B12, genomic survey sequence.
ACCESSION   CC099198
VERSION     CC099198.1  GI:29965572
KEYWORDS    GSS.
SOURCE      Aedes aegypti (yellow fever mosquito)
  ORGANISM  Aedes aegypti
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Aedes;
            Stegomyia.
REFERENCE   1  (bases 1 to 845)
  AUTHORS   Loftus,B., Shetty,J., Severson,D., Brown,S. and Knudson,D.
  TITLE     End sequencing of Aedes aegypti BACs
  JOURNAL   Unpublished (2003)
COMMENT     Other_GSSs: CSU-K34.116B12.SP6
            Contact: Brendan Loftus
            Department of Eukaryotic Genomics
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-3543
            Fax: 301-838-0208
            Email: enta@tigr.org
            Library was provided by Susan Brown and Dennis Knudson at Colorado
            State University.
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .845
                     /organism="Aedes aegypti"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7159"
                     /clone="CSU-K34.116B12"
                     /clone_lib="CSU-K34"
```

/note="Vector: pBACe3.6; Site_1: EcoRI; Source DNA: Aedes aegypti; strain unknown [derived from freshly hatched larvae at the Virus Research Centre, Poona, India. Reference: SINGH, K. R. P., 1967 Cell cultures derived from larvae of Aedes albopictus (Skuse) and Aedes aegypti (L.). Current Science 36: 506-508]; ATC-10 cell line ATCC CCL-125"

ORIGIN

Query Match 6.4%; Score 318.2; DB 8; Length 845;
Best Local Similarity 79.7%; Pred. No. 1.7e-44;
Matches 388; Conservative 0; Mismatches 98; Indels 1; Gaps 1;

```
Qy        211 AAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACATATA 270
              |||  || |||   | | |||||||||||||||||||||||| | | | | | | |||||
Db        360 AAACTAACAAATTAAATATATATATATATATATATATATATATATATATATATATATA 419

Qy        271 TATATACACATATATATACACATATATACATATATACACATATATACACATATATATACA 330
              |||||| | |||||||||| | ||||||| ||||||| | ||||||| | |||||||||| |
Db        420 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 479

Qy        331 CATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACACA 390
              ||| |||| || | ||||||||| | |||| |||| | ||||||| | ||||||| | |
Db        480 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 539

Qy        391 CATATATACACATATATACACACATATATACACATATATACACATATATACACATATACA 450
              ||||||| | ||||||| | | ||||||| | ||||||| | ||||||| | ||||| |
Db        540 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 599

Qy        451 CATATATACACATATATACATATATACACATATATATAATATACACACATATATATACAC 510
              ||||||| | ||||||| ||||||| | |||||||| ||||| | | ||||||||| |
Db        600 TATATATATATATATATATATATATATATATATATAT-ATATATATATATATATATATAT 658

Qy        511 ATATATACACACATATATACACATATATACACATATATATACACATATATACACATATAT 570
              ||||||| | | ||||||| | ||||||| | ||||||||| | ||||||| | ||||||
Db        659 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 718

Qy        571 ACATATATACACATATATATACATATATACACATATATACATATATACACATATATACAT 630
              | ||||||| | ||||||||| ||||||| | ||||||| ||||||| | ||||||| ||
Db        719 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 778

Qy        631 ATATACACACATATATACACATACATATACACACACATAGATATACATATATATACACAT 690
              ||||| | | ||||||| | ||| ||||| | | | ||| ||||| ||||||||| | ||
Db        779 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 838

Qy        691 ATATATA 697
              |||||||
Db        839 ATATATA 845
```

RESULT 41
CC099198/c
LOCUS CC099198 845 bp DNA linear GSS 16-APR-2003
DEFINITION CSU-K34.116B12.T7 CSU-K34 Aedes aegypti genomic clone CSU-K34.116B12, genomic survey sequence.
ACCESSION CC099198
VERSION CC099198.1 GI:29965572
KEYWORDS GSS.
SOURCE Aedes aegypti (yellow fever mosquito)
ORGANISM Aedes aegypti
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Aedes; Stegomyia.
REFERENCE 1 (bases 1 to 845)
AUTHORS Loftus,B., Shetty,J., Severson,D., Brown,S. and Knudson,D.
TITLE End sequencing of Aedes aegypti BACs
JOURNAL Unpublished (2003)
COMMENT Other_GSSs: CSU-K34.116B12.SP6
Contact: Brendan Loftus
Department of Eukaryotic Genomics
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-3543
Fax: 301-838-0208
Email: enta@tigr.org
Library was provided by Susan Brown and Dennis Knudson at Colorado State University.
Seq primer: T7
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .845
/organism="Aedes aegypti"
/mol_type="genomic DNA"
/db_xref="taxon:7159"
/clone="CSU-K34.116B12"
/clone_lib="CSU-K34"
/note="Vector: pBACe3.6; Site_1: EcoRI; Source DNA: Aedes aegypti; strain unknown [derived from freshly hatched larvae at the Virus Research Centre, Poona, India. Reference: SINGH, K. R. P., 1967 Cell cultures derived from larvae of Aedes albopictus (Skuse) and Aedes aegypti (L.). Current Science 36: 506-508]; ATC-10 cell line ATCC CCL-125"

ORIGIN

Query Match 6.4%; Score 317; DB 8; Length 845;
Best Local Similarity 78.0%; Pred. No. 2.7e-44;
Matches 394; Conservative 0; Mismatches 110; Indels 1; Gaps 1;

```
Qy        229 TATATATATATATATATATATACACACACACACATATATATATACACATATATATA 288
              |||||||||||||||||||||||| | | | | | |||||||||| | |||||||||
Db        845 TATATATATATATATATATATATATATATATATATATATATATATATATATATATA 786

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |||
Db        785 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 726

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              |||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||||
Db        725 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 666

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||||
Db        665 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 606

Qy        469 CATATATACACATATATATAATATACACACATATATATACACATATATACACACATATAT 528
              ||||||| | |||||||| ||||| | | ||||||||| | ||||||| | | ||||||
Db        605 TATATATATATATATATAT-ATATATATATATATATATATATATATATATATATATAT 547

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACATATAT 588
              | | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | ||||||
Db        546 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 487

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
              ||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||||
Db        486 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 427

Qy        649 ACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATAT 708
              | ||| ||||| | | | ||| ||||| ||||||||| | ||||||||  | ||  |
Db        426 ATATATATATATATATATATATATATATATATATATATATATATATATATTTAATTTG 367

Qy        709 GTATATATATATGCTCCAGAGTTCA 733
              ||  |  | ||  |  |  || ||
Db        366 TTAGTTTGAAATCATTGACTGTGCA 342
```

RESULT 42
CL855754
LOCUS CL855754 816 bp DNA linear GSS 16-AUG-2004
DEFINITION OR_CBa0085L12.r OR_CBa Oryza rufipogon genomic clone OR_CBa0085L12 3', genomic survey sequence.
ACCESSION CL855754
VERSION CL855754.1 GI:51264405
KEYWORDS GSS.
SOURCE Oryza rufipogon
ORGANISM Oryza rufipogon

```
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 816)
  AUTHORS   Kim,H., Yu,Y., Wissotski,M., Yost,D., Stum,D., Rao,K., Luo,M.,
            Jetty,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and
            Wing,R.
  TITLE     OMAP project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0085  row: L  column: 12
            Seq primer: CAC TCA TTA GGC ACC CCA
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .816
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
                     /clone="OR_CBa0085L12"
                     /tissue_type="young leaves"
                     /dev_stage="2 week old seedlings"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OR_CBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII;
                     drk treated 36 hrs before harvest"
ORIGIN

  Query Match             6.3%;  Score 315;  DB 9;  Length 816;
  Best Local Similarity   79.2%;  Pred. No. 6.1e-44;
  Matches  411;  Conservative    0;  Mismatches  105;  Indels    3;  Gaps    3;

Qy        204 ATCTCAGAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACAC 263
              || | |  | | || | |   || | ||||||||||||||||||||||| | | | | |
Db        250 ATATAATCATATAATATATATATATCATATATATATATATATATATATATATATATAT 309

Qy        264 ACATATATATATACACATATATATACACATATATACATATATACACATATATACACATAT 323
              | |||||||||||| | ||||||||| | ||||||| ||||||| | ||||||| | ||||
Db        310 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 369

Qy        324 ATATACACATACATATGTACACATATATATACACATATGTATACACATATATACACATAT 383
              ||||| | ||| |||| || | ||||||||| | |||| |||| | |||||||| | ||||
Db        370 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 429

Qy        384 ATACACACATATATACACATATATACACA-CATATATACACATATATACACATATATACA 442
              ||| | | ||||||| | ||||||| | | |||||||| | ||||||| | ||||||| |
Db        430 ATATATATATATATATATATATATATATATCATATATATATATATATATATATATATA 489

Qy        443 CATATACACATATATACACATATATACATATATACACATATATATA-ATATACACACATA 501
              |||||  | ||||||| | ||||||| ||||||| | ||||||||| |||||  | | |||
Db        490 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 549

Qy        502 TATATACACATATATACACACATATATACACATATATACACATATATATACACATATATA 561
              |||||| | ||||||| | | ||||||| | ||||||| | ||||||||| | |||||||
Db        550 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 609

Qy        562 CACATATATACATATATACACATATATATACATATATACACATATATACATATATACACA 621
              | ||||||| ||||||| | ||||||||| ||||||| | ||||||| ||||||| | |
Db        610 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 669

Qy        622 TATATACATATATACACACATATATACACATACATATACACACACATAGATATACATATA 681
              |||||| ||||||| | | ||||||| | ||| ||||| | | | ||| ||||| |||||
Db        670 TATATATATATATATA-ATATATATATATATATATATATATATATATATATATATATA 728

Qy        682 TATACACATATATATACGTATATATATGTATATATATAT 720
```

```
              |||| | |||||||||  |||||||| |||||||| ||
Db        729 TATATATATATATATATATATATATATATATATATAAAT 767


RESULT 43
BI506858
LOCUS       BI506858                570 bp    mRNA    linear   EST 08-APR-2002
DEFINITION  BB170024B20C02.5 Bee Brain Normalized/Subtracted Library, BB17 Apis
            mellifera cDNA clone BB170024B20C02 5', mRNA sequence.
ACCESSION   BI506858
VERSION     BI506858.1  GI:15357232
KEYWORDS    EST.
SOURCE      Apis mellifera (honey bee)
  ORGANISM  Apis mellifera
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Apoidea;
            Apidae; Apis.
REFERENCE   1  (bases 1 to 570)
  AUTHORS   Whitfield,C.W., Band,M.R., Bonaldo,M.F., Kumar,C.G., Liu,L.,
            Pardinas,J., Robertson,H.M., Soares,B. and Robinson,G.E.
  TITLE     Annotated expressed sequence tags and cDNA microarrays for studies
            of brain and behavior in the honey bee
  JOURNAL   Genome Res. 12 (4), 555-566 (2002)
  MEDLINE   21929762
   PUBMED   11932240
COMMENT     Contact: Gene E. Robinson
            Department of Entomology
            University of Illinois
            505 S. Goodwin Ave., Urbana, IL 61801, USA
            Tel: 217 265 0309
            Fax: 217 244 3499
            Email: generobi@life.uiuc.edu
            This research was funded by the University of Illinois Critical
            Research Initiatives Fund and a Burroughs-Wellcome Trust Innovation
            Award in Functional Genomics to G.E. Robinson and an NSF
            Postdoctoral Fellowship in Bioinformatics to C.W. Whitfield.
            REPEAT IN THE SEQUENCE
            Simple_repeat STRAND (+) ELEMENT (TA)n LOCATION [90,570].
            PCR PRimers
            FORWARD: TAATACGACTCACTATAGGG
            BACKWARD: ATTAACCCTCACTAAAG
            Plate: BB170024B20  row: C  column: 02
            Seq primer: AGCGGATAACAATTTCACACAGGA
            High quality sequence stop: 570.
FEATURES             Location/Qualifiers
     source          1. .570
                     /organism="Apis mellifera"
                     /mol_type="mRNA"
                     /strain="mixed strains of European bees, predominantly
                     A.m. ligustica"
                     /db_xref="taxon:7460"
                     /clone="BB170024B20C02"
                     /sex="female"
                     /tissue_type="brain"
                     /dev_stage="adult worker honey bee"
                     /lab_host="DH10B"
                     /clone_lib="Bee Brain Normalized/Subtracted Library, BB17"
                     /note="Organ: brain; Vector: pT7T3-Pac; Site_1: EcoR1;
                     Site_2: Not1; This BB17 cDNA library was generated by
                     subtraction of the BB16  library with 4000 previously
                     sequenced clones. The BB16 library was  contributed by the
                     Soares laboratory and it was constructed and  normalized
                     as described by Bonaldo, M.F., Lennon, G. and  Soares,
                     M.B. (1996), Genome Research 6(9): 791-806. RNA was
                     prepared  from dissected brains of adult worker bees of
                     various ages and  various behavioral groups. "
ORIGIN

  Query Match             6.3%;  Score 314.6;  DB 4;  Length 570;
  Best Local Similarity   77.7%;  Pred. No. 7.4e-44;
  Matches  405;  Conservative    0;  Mismatches  114;  Indels    2;  Gaps    2;

Qy        201 GTCATCTCAGAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACA 260
```

```
Db         48 GTCTCCACAGATAGAGAGGGGTGTAACTCAATTAAAATGATTTATATATATATATATATA 107

Qy        261 CACACATATATATATACACATATATATACACATATATACATATATACACATATATACACA 320
Db        108 -ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 166

Qy        321 TATATATACACATACATATGTACACATATATATACACATATGTATACACATATATACACA 380
Db        167 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 226

Qy        381 TATATACACACATATATACACATATATACACACATATATACACATATATACACATATATA 440
Db        227 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 286

Qy        441 CACATATACACATATATACACATATATACATATATACACATATATATA-ATATACACACA 499
Db        287 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 346

Qy        500 TATATATACACATATATACACACATATATACACATATATACACATATATATACACATATA 559
Db        347 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 406

Qy        560 TACACATATATACATATATACACATATATATACATATATACACATATATACATATATACA 619
Db        407 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 466

Qy        620 CATATATACATATATACACACATATATACACATACATATACACACACATAGATATACATA 679
Db        467 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 526

Qy        680 TATATACACATATATATACGTATATATATGTATATATATAT 720
Db        527 TATATATATATATATATATATATATATATATATATATAT 567
```

```
RESULT 44
BI506858/c
LOCUS       BI506858                570 bp    mRNA    linear   EST 08-APR-2002
DEFINITION  BB170024B20C02.5 Bee Brain Normalized/Subtracted Library, BB17 Apis
            mellifera cDNA clone BB170024B20C02 5', mRNA sequence.
ACCESSION   BI506858
VERSION     BI506858.1  GI:15357232
KEYWORDS    EST.
SOURCE      Apis mellifera (honey bee)
  ORGANISM  Apis mellifera
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Apoidea;
            Apidae; Apis.
REFERENCE   1  (bases 1 to 570)
  AUTHORS   Whitfield,C.W., Band,M.R., Bonaldo,M.F., Kumar,C.G., Liu,L.,
            Pardinas,J., Robertson,H.M., Soares,B. and Robinson,G.E.
  TITLE     Annotated expressed sequence tags and cDNA microarrays for studies
            of brain and behavior in the honey bee
  JOURNAL   Genome Res. 12 (4), 555-566 (2002)
  MEDLINE   21929762
   PUBMED   11932240
COMMENT     Contact: Gene E. Robinson
            Department of Entomology
            University of Illinois
            505 S. Goodwin Ave., Urbana, IL 61801, USA
            Tel: 217 265 0309
            Fax: 217 244 3499
            Email: generobi@life.uiuc.edu
            This research was funded by the University of Illinois Critical
            Research Initiatives Fund and a Burroughs-Wellcome Trust Innovation
            Award in Functional Genomics to G.E. Robinson and an NSF
            Postdoctoral Fellowship in Bioinformatics to C.W. Whitfield.
            REPEAT IN THE SEQUENCE
            Simple_repeat STRAND (+) ELEMENT (TA)n LOCATION [90,570].
            PCR PRimers
            FORWARD: TAATACGACTCACTATAGGG
            BACKWARD: ATTAACCCTCACTAAAG
            Plate: BB170024B20  row: C  column: 02
            Seq primer: AGCGGATAACAATTTCACACAGGA
            High quality sequence stop: 570.
FEATURES             Location/Qualifiers
     source          1. .570
                     /organism="Apis mellifera"
                     /mol_type="mRNA"
                     /strain="mixed strains of European bees, predominantly
                     A.m. ligustica"
                     /db_xref="taxon:7460"
                     /clone="BB170024B20C02"
                     /sex="female"
                     /tissue_type="brain"
                     /dev_stage="adult worker honey bee"
                     /lab_host="DH10B"
                     /clone_lib="Bee Brain Normalized/Subtracted Library, BB17"
                     /note="Organ: brain; Vector: pT7T3-Pac; Site_1: EcoR1;
                     Site_2: Not1; This BB17 cDNA library was generated by
                     subtraction of the BB16  library with 4000 previously
                     sequenced clones. The BB16 library was  contributed by the
                     Soares laboratory and it was constructed and  normalized
                     as described by Bonaldo, M.F., Lennon, G. and  Soares,
                     M.B. (1996), Genome Research 6(9): 791-806. RNA was
                     prepared  from dissected brains of adult worker bees of
                     various ages and  various behavioral groups. "
ORIGIN
```

```
  Query Match             6.3%;  Score 313.2;  DB 4;  Length 570;
  Best Local Similarity   80.5%;  Pred. No. 1.3e-43;
  Matches  391;  Conservative    0;  Mismatches   93;  Indels    2;  Gaps    2;

Qy        235 TATATATATATATATATACACACACACACATATATATATACACATATATATACACATA 294
Db        570 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 511

Qy        295 TATACATATATACACATATATACACATATATATACACATACATATGTACACATATATATA 354
Db        510 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 451

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
Db        450 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 391

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
Db        390 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 331

Qy        475 TACACATATATATAATATACACACATATATATACACATATATACACACATATATACACAT 534
Db        330 TATATATATATAT-ATATATATATATATATATATATATATATATATATATATATATAT 272

Qy        535 ATATACACATATATATACACATATATACACATATATACATATATACACATATATATACAT 594
Db        271 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 212

Qy        595 ATATACACATATATACATATATACACATATATACATATATACACACATATATACACATAC 654
Db        211 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 152

Qy        655 ATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATAT 714
Db        151 ATATATATATATATATATATATATATATATATATATATA-TTATATATATATATAT 93

Qy        715 ATATAT 720
Db         92 ATAAAT 87
```

```
RESULT 45
CL855754/c
LOCUS       CL855754                816 bp    DNA     linear   GSS 16-AUG-2004
DEFINITION  OR_CBa0085L12.r OR_CBa Oryza rufipogon genomic clone OR_CBa0085L12
            3', genomic survey sequence.
```

```
ACCESSION   CL855754
VERSION     CL855754.1  GI:51264405
KEYWORDS    GSS.
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 816)
  AUTHORS   Kim,H., Yu,Y., Wissotski,M., Yost,D., Stum,D., Rao,K., Luo,M.,
            Jetty,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and
            Wing,R.
  TITLE     OMAP project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0085  row: L  column: 12
            Seq primer: CAC TCA TTA GGC ACC CCA
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .816
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
                     /clone="OR_CBa0085L12"
                     /tissue_type="young leaves"
                     /dev_stage="2 week old seedlings"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OR_CBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII;
                     drk treated 36 hrs before harvest"
ORIGIN

  Query Match            6.3%;  Score 311.4;  DB 9;  Length 816;
  Best Local Similarity   77.6%;  Pred. No. 2.5e-43;
  Matches  413;  Conservative    0;  Mismatches  116;  Indels    3;  Gaps    3;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||| ||||||||||||||||||| | | | | | |||||||||| | |||||||
Db        770 TATATTTATATATATATATATATATATATATATATATATATATATATATATATATATATA 711

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | || |||| | ||||||||| | ||| |||| || | |
Db        710 TATATATATATATATATATATATAT-TATATATATATATATATATATATATATATATATA 652

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        651 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 592

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db        591 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 532

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db        531 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 472

Qy        526 TATACACATAT-ATACACATATATATACACATATATACACATATATACATATATACACAT 584
              |||| | |||| ||| | ||||||||| | ||||||| | ||||||| ||||||| | ||
Db        471 TATATATATATGATATATATATATATATATATATATATATATATATATATATATATATAT 412

Qy        585 ATATATACATATATACACATATATACATATATACACATATATACATATATACACACATAT 644
              ||||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | ||||
Db        411 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 352

Qy        645 ATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATAT 704
              ||| | ||| ||||| | | | ||| ||||| ||||||||| | ||||||||  |||||
Db        351 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 292

Qy        705 ATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCA 756
              |||| ||||||||||||| |  | |  |  ||   | |   |  |   || ||
Db        291 ATATATATATATATATGATATATATATATTATATGATTATATATATATATCA 240


RESULT 46
CL858220/c
LOCUS       CL858220                546 bp    DNA     linear   GSS 16-AUG-2004
DEFINITION  OR_CBa0089A17.f OR_CBa Oryza rufipogon genomic clone OR_CBa0089A17
            5', genomic survey sequence.
ACCESSION   CL858220
VERSION     CL858220.1  GI:51269336
KEYWORDS    GSS.
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 546)
  AUTHORS   Kim,H., Yu,Y., Wissotski,M., Yost,D., Stum,D., Rao,K., Luo,M.,
            Jetty,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and
            Wing,R.
  TITLE     OMAP project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0089  row: A  column: 17
            Seq primer: TAA TAC GAC TCA CTA TAG GG
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .546
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
                     /clone="OR_CBa0089A17"
                     /tissue_type="young leaves"
                     /dev_stage="2 week old seedlings"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OR_CBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII;
                     drk treated 36 hrs before harvest"
ORIGIN

  Query Match            6.2%;  Score 310.6;  DB 9;  Length 546;
  Best Local Similarity   76.0%;  Pred. No. 3.6e-43;
  Matches  396;  Conservative    0;  Mismatches  124;  Indels    1;  Gaps    1;

Qy        242 ATATATATATACACACACACACATATATATATACACATATATATACACATATATACAT 301
              |||||||||||| | | | |   | |||||          ||||||| |||||||| ||
Db        546 ATATATATATATATATATATCTATATATAATATATATATCATATATATACATATATATAT 487

Qy        302 ATATACACATATATACACATATATATACACATACATATGTACACATATATATACACATAT 361
              ||||| | ||||||| | ||||||||| | |||  ||| || | | ||||||||| | ||||
Db        486 ATATATATATATATATATATATATATATATATATCTATATATATATATATATATATAT 427

Qy        362 GTATACACATATATACACATATATACACACATATATACACATATATACACACATATATAC 421
              |||| | ||||||| | ||||||||| | ||||| | | ||||||  | | |||||||
Db        426 ATATATATATATATATATATATATACATATATATACATATATATATCTATATATATAT 367

Qy        422 ACATATATACACATATATACACATATACACATATATACACATATATACATATATACACAT 481
              | ||||||| | ||||||| | ||||| | ||||||| | ||||||| ||||||| | ||
```

```
Db        366 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 307

Qy        482 ATATATAATATACACACATATATATACACATATATACACACATATATACACATATATACA 541
              |||||| ||||| | | |||||||||| | |||||||| | | |||||||| | |||||||| |
Db        306 ATATAT-ATATATATATATATATATATATATATATATATATATATATATATATATATATA 248

Qy        542 CATATATATACACATATATACACATATATACATATATACACATATATATACATATATACA 601
              ||||||||| | |||||||| | |||||||| |||||||| | ||||||||| ||||||| |
Db        247 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 188

Qy        602 CATATATACATATATACACATATATACATATATACACACATATATACACATACATATACA 661
              |||||||| |||||||| | |||||||| |||||||| | | |||||||| | ||| ||||| |
Db        187 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 128

Qy        662 CACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATATG 721
              | | ||| ||||| ||||||||| | |||||||||  |||||||| ||||||||||||
Db        127 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATG 68

Qy        722 CTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGG 762
                  |  ||  || |||  |  | |  ||||||| ||||
Db         67 AGGTGAAACTCTAAAAAGGATGACGATCCATACCACAGGGG 27


RESULT 47
BX145980
LOCUS       BX145980                709 bp    DNA     linear   GSS 28-JAN-2003
DEFINITION  Danio rerio genomic clone DKEY-104E13, genomic survey sequence.
ACCESSION   BX145980
VERSION     BX145980.1  GI:27977335
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 709)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the T7 end of BAC 104E13. 104E13
            is part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .709
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-104E13"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             6.2%;  Score 308;  DB 9;  Length 709;
  Best Local Similarity   79.7%;  Pred. No. 9.8e-43;
  Matches  376;  Conservative    0;  Mismatches   95;  Indels    1;  Gaps    1;

Qy        250 ATACACACACACACATATATATATACACATATATATACACATATATACATATATACAC 309
              | |  || |    |||||||||||| |||||||||| |||||||| ||||||| |
Db        238 AAAGTCAAATTTATGCATATATATATATATATATATATATATATATATATATATATATAT 297

Qy        310 ATATATACACATATATATACACATACATATGTACACATATATATACACATATGTATACAC 369
              ||||||| | ||| | ||| | ||| |||| |||||||||||||| | |||| |||||||
Db        298 ATATATATATATACACATATATATATATATACACATATATATATATATATATATACAC 357

Qy        370 ATATATACACATATATACACACATATATACACATATATACACACATATATACACATATAT 429
              | ||||| | ||||||| | | ||||||| | ||||||| | | ||||||| | ||||||
Db        358 ACATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 417

Qy        430 ACACATATATACACATATACACATATATACACATATATACATATATACACATATATATA- 488
              | | |||||| | | |||||| | ||||||| | ||| ||| ||| ||| | |||||||||
Db        418 ATATATATACATATATATATATATATATATATATACATATATACATATATATATATAT 477

Qy        489 ATATACACACATATATATACACATATATACACACATATATACACATATATACACATATAT 548
              ||||||| |||||||||| | ||||||| | | |||||||| | |||||||| | ||||||
Db        478 ATATACATACATATATATATATATATATATATATATATATATATATATATATATATATAT 537

Qy        549 ATACACATATATACACATATATACATATATACACATATATATACATATATACACATATAT 608
              ||| | ||||||| | ||||||| ||||||| ||| ||||||| ||||||| ||||||||
Db        538 ATATATATATATATATATATATATATATATATACACATATATATATATATATACATATAT 597

Qy        609 ACATATATACACATATATACATATATACACACATATATACACATACATATACACACACAT 668
              | ||||||| | ||||||| ||||||| | | |||||||| | ||| ||||| | | |  |
Db        598 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATCT 657

Qy        669 AGATATACATATATATACACATATATATACGTATATATATGTATATATATAT 720
              | ||||| |||||||||| | |||||||||  |||||||| ||||||||||
Db        658 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 709


RESULT 48
CR177142
LOCUS       CR177142                648 bp    DNA     linear   GSS 06-JUL-2004
DEFINITION  Reverse strand read from insert in 5'HPRT insertion targeting and
            chromosome engineering clone MHPN105k12, genomic survey sequence.
ACCESSION   CR177142
VERSION     CR177142.1  GI:49955991
KEYWORDS    GSS; genome survey sequence; MICER.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 648)
  AUTHORS   Adams,D.J., Biggs,P.J., Cox,A.V., Davies,R.M., van der Weyden,L.,
            Jonkers,J., Smith,J., Plumb,R.W., Taylor,R.G., Nishijima,I., Yu,Y.,
            Rogers,J. and Bradley,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-FEB-2004) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. http://www.sanger.ac.uk/MICER
FEATURES             Location/Qualifiers
     source          1. .648
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="MHPN105k12"
                     /clone_lib="MHPN"
ORIGIN

  Query Match             6.2%;  Score 307.2;  DB 9;  Length 648;
  Best Local Similarity   75.4%;  Pred. No. 1.4e-42;
  Matches  395;  Conservative    0;  Mismatches  128;  Indels    1;  Gaps    1;

Qy        197 AAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACA 256
              |||   ||   | |||||  |||   |||| |||||||||||| | ||||||| | | |
Db        107 AAACTACAAGCAATAAAAATCAAATAACATTATATATATATATACACATATATACATATA 166

Qy        257 CACACACACATATATATATACACATATATATACACATATATACATATATACACATATATA 316
              ||| | | | ||||| | | | | | ||||| |||||||| |||||||| ||| ||||| |
Db        167 TACATATATACATATACACATATACACATATATACATATACACATATACACATATATACA 226

Qy        317 CACATATATATACACATACATATGTACACATATATATACACATATGTATACACATATATA 376
              | | | |||||||||| | |  || ||| ||||| ||| |||   ||| ||| |||||
Db        227 TATACACATATACACATATACACATATACACATATACACATATACACATATACACATATA 286

Qy        377 CACATATATACACACATATATACACATATATACACACATATATACACATATATACACATA 436
              |||||||| ||| | | | ||| ||| ||||||||| ||| | | | | | ||| ||| |
Db        287 CACATATACACATATACACATATACACATATACACATATACACATATACACATATACACA 346

Qy        437 TATACACATATACACATATATACACATATATACATATATACACATATATATAATATACAC 496
              |||||||||||||||||||| ||| ||| | | ||| | | | | | |||| | | | |
Db        347 TATACACATATACACATATACACATATACACATATACACATATACACATAT-ACACATAT 405

Qy        497 ACATATATATACACATATATACACACATATATACACATATATACACATATATATACACAT 556
              ||| ||||| ||| ||| | | | ||| |||||||||||||| ||| ||||| ||| | |
```

```
Db        406 ACACATATACACATATACACATATACACATATACACATATACACATATATACATATATAC 465

Qy        557 ATATACACATATATACATATATACACATATATATACATATATACACATATATACATATAT 616
              ||||| ||||||||||||||||||||||| | | | ||| ||||| | ||| ||| ||| ||
Db        466 ATATATACATATATACATATATACACACACACACACACATATATATATACATATATACAT 525

Qy        617 ACACATATATACATATATACACACATATATACACATACATATACACACACATAGATATAC 676
              | | ||||||||| | | ||||||||||||||| | | | ||||||| | | ||| ||||||
Db        526 ATATATATATACACACACACACACATATATATATACATATATACATATATATATATATAC 585

Qy        677 ATATATATACACATATATATACGTATATATATGTATATATATAT 720
              | | | | ||||||||||||||  ||||||||| ||||||||||
Db        586 ACACACACACACATATATATATATATATATATATATATATAT 629


RESULT 49
CL626353/c
LOCUS       CL626353                581 bp    DNA     linear   GSS 01-JUL-2004
DEFINITION  OR_BBa0022K14.f OR_BBa Oryza rufipogon genomic clone OR_BBa0022K14
            5', genomic survey sequence.
ACCESSION   CL626353
VERSION     CL626353.1  GI:49128388
KEYWORDS    GSS.
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 581)
  AUTHORS   Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R.,
            Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
  TITLE     OMAP Project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Insert Length: 161   Std Error: 0.00
            Plate: 0022  row: K  column: 14
            Seq primer: TAA TAC GAC TCA CTA TAG GG
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .581
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
                     /clone="OR_BBa0022K14"
                     /tissue_type="young leaves"
                     /lab_host="DH10B-T1 phage resistant"
                     /clone_lib="OR_BBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

  Query Match             6.2%;  Score 306.4;  DB 9;  Length 581;
  Best Local Similarity   77.4%;  Pred. No. 1.9e-42;
  Matches  384;  Conservative    0;  Mismatches  111;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | || |||||||||||| ||| | ||| |||   || ||| ||||| ||| ||| ||| |||
Db        574 TATACATATATATATACATACACATATACATTCATACATATATACATACACATATACATA 515

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              | |||||||||||||| |||||  | ||| | |||||||||||| |||||||||  |||| |
Db        514 CATACATATATACATACATACATACATACATACATATATATATACATACATACATACATA 455

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||  | | ||  ||||| | ||||| | ||||||||| ||||| ||||| |||||
```

```
Db        454 CATATATATATACATACATACATACATATATATATATATACATACATACATACATATATA 395

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | | ||| | ||||| ||| | ||||||| ||||| | ||||| ||| | |||||
Db        394 TATATATACATACATACATACATATATATATATATACATACATACATACATATATATATA 335

Qy        467 TACATATATACACATATATATAATATACACACATATATATACACATATATACACACATAT 526
              |||||| ||||| | ||||||| ||||| | ||||| ||| | |||||||||| | | ||||
Db        334 TACATACATACATACATATAT-ATATATATACATACATACATACATATATATATATATAT 276

Qy        527 ATACACATATATACACATATATATACACATATATACACATATATACATATATACACATAT 586
              | | ||||| ||| | ||||||||| ||||| ||||| ||||||| ||||||| | || |
Db        275 ACATACATACATATATATATATATACATACATACATATATATATATATATATATGT 216

Qy        587 ATATACATATATACACATATATACATATATACACATATATACATATATACACACATATAT 646
              ||||| ||||||  | ||||| | | ||| | | || |||| ||||||| |   ||||||
Db        215 ATATATATATATGTATATATACATACATACATATGTATATATATATATATGTATATAT 156

Qy        647 ACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATAT 706
              | | ||| ||||    | | ||| |||| |||| |||| | |||||||||  |||||||
Db        155 ATATATATATATATGTATATATATATATATATATGTATATATATATATATATATATAT 96

Qy        707 ATGTATATATATATGC 722
              || |||||||  | ||
Db         95 ATATATATATGGACGC 80


RESULT 50
CA255084
LOCUS       CA255084                613 bp    mRNA    linear   EST 26-SEP-2003
DEFINITION  SCEPFL4175C11.g FL4 Saccharum officinarum cDNA clone SCEPFL4175C11
            5', mRNA sequence.
ACCESSION   CA255084
VERSION     CA255084.1  GI:35939612
KEYWORDS    EST.
SOURCE      Saccharum officinarum
  ORGANISM  Saccharum officinarum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Saccharum; Saccharum officinarum
            complex.
REFERENCE   1  (bases 1 to 613)
  AUTHORS   Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.
  TITLE     The libraries that made SUCEST
  JOURNAL   Genet. Mol. Biol. 24 (1-4), 1-7 (2001)
COMMENT     Contact: Arruda P
            Centro de Biologia Molecular e Engenharia Genetica
            Universidade Estadual de Campinas
            Caixa Postal 6010, 13083-970, Campinas SP, Brazil
            Tel: 55 19 3788 1137
            Fax: 55 19 3788 1089
            Email: parruda@unicamp.br
            Clone distribution: clone distribution information can be found
            through the Brazilian Clone Collection Center (BCCC) at
            http://www.bcccenter.fcav.unesp.br
            Plate: 175  row: C  column: 11
            Seq primer: T7 Promoter Primer.
FEATURES             Location/Qualifiers
     source          1. .613
                     /organism="Saccharum officinarum"
                     /mol_type="mRNA"
                     /db_xref="taxon:4547"
                     /clone="SCEPFL4175C11"
                     /lab_host="DH10B"
                     /clone_lib="FL4"
                     /note="Organ: Developed inflorescence and rachis
                     (20cm-long); Vector: pSport1; Site_1: SalI; Site_2: NotI;
                     An unidirectional cDNA library generated from [Developed
                     inflorescence  and rachis (20cm-long)]. cDNA was prepared
                     from polyA+ mRNA using  SuperScript Plasmid System Kit
                     (Invitrogen). The double-strand cDNAs were  fractionated
                     in a sepharose CL-2B 40cm-columns and fragments sizing
                     between 0.8 and 1.5 Kb were directionally cloned into the
```

vector. Details of each source of RNA and library construction can be obtained at http://sucest.lad.ic.unicamp.br/public"

ORIGIN

Query Match 6.2%; Score 306; DB 6; Length 613;
Best Local Similarity 76.9%; Pred. No. 2.2e-42;
Matches 386; Conservative 0; Mismatches 115; Indels 1; Gaps 1;

```
Qy        220 AAGGCATTTTATATATATATATATATATATACACACACACACACATATATATATACAC 279
              | |||  | | |||  |||| | ||| ||| |  |      | | | ||||||| ||  |
Db        109 ATGGCTCTCTGTATGCATATGTCTATCTATGTGGAGCGCTAGATAGATATATACATCTAT 168

Qy        280 ATATATATACACATATATACATATATACACATATATACACATATATATACACATACATAT 339
              ||||||||| | ||||||  ||||||| | ||||||| | ||||||||| | ||| || |
Db        169 ATATATATATATATATATCGATATATATATATATATATATATATATATATATATATGT 228

Qy        340 GTACACATATATATACACATATGTATACACATATATACACATATATACACACATATATAC 399
              || | |||||||||| | |||| |||| | ||||||| | ||||||| | | |||||||
Db        229 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 288

Qy        400 ACATATATACACACATATATACACATATATACACATATATACACATATACACATATATAC 459
              | ||||||| | | ||||||| | ||||||| | ||||||| | ||||| | |||||||
Db        289 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 348

Qy        460 ACATATATACATATATACACATATATATAATATACACACATATATATACACATATATACA 519
              | ||||||| ||||||| | |||||||| ||||| | | ||||||||| | ||||||| |
Db        349 ATATATATATATATATATATATATATAT-ATATATATATATATATATATATATATATA 407

Qy        520 CACATATATACACATATATACACATATATATACACATATATACACATATATACATATATA 579
              | |||||||| | ||||||| | ||||||||| | ||||||| | ||||||| |||||||
Db        408 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 467

Qy        580 CACATATATATACATATATACACATATATACATATATACACATATATACATATATACACA 639
              | ||||||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | |
Db        468 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 527

Qy        640 CATATATACACATACATATACACACACATAGATATACATATATATACACATATATATACG 699
              |||||||  | ||| |||||  | | | ||| ||||| |||||||||  | |||||  |||
Db        528 TATATATATATATATATATATATATATATATATATATATATATATATATATATAGATATA 587

Qy        700 TATATATATGTATATATATATG 721
              |||||  |||  |||||||   |||
Db        588 TATATCTATATATATATCGATG 609
```

RESULT 51
CA255084/c
LOCUS CA255084 613 bp mRNA linear EST 26-SEP-2003
DEFINITION SCEPFL4175C11.g FL4 Saccharum officinarum cDNA clone SCEPFL4175C11 5', mRNA sequence.
ACCESSION CA255084
VERSION CA255084.1 GI:35939612
KEYWORDS EST.
SOURCE Saccharum officinarum
ORGANISM Saccharum officinarum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Saccharum; Saccharum officinarum complex.
REFERENCE 1 (bases 1 to 613)
AUTHORS Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.
TITLE The libraries that made SUCEST
JOURNAL Genet. Mol. Biol. 24 (1-4), 1-7 (2001)
COMMENT Contact: Arruda P
Centro de Biologia Molecular e Engenharia Genetica
Universidade Estadual de Campinas
Caixa Postal 6010, 13083-970, Campinas SP, Brazil
Tel: 55 19 3788 1137
Fax: 55 19 3788 1089
Email: parruda@unicamp.br
Clone distribution: clone distribution information can be found through the Brazilian Clone Collection Center (BCCC) at http://www.bcccenter.fcav.unesp.br
Plate: 175 row: C column: 11
Seq primer: T7 Promoter Primer.
FEATURES Location/Qualifiers
source 1. .613
/organism="Saccharum officinarum"
/mol_type="mRNA"
/db_xref="taxon:4547"
/clone="SCEPFL4175C11"
/lab_host="DH10B"
/clone_lib="FL4"
/note="Organ: Developed inflorescence and rachis (20cm-long); Vector: pSport1; Site_1: SalI; Site_2: NotI; An unidirectional cDNA library generated from (Developed inflorescence and rachis (20cm-long)). cDNA was prepared from polyA+ mRNA using SuperScript Plasmid System Kit (Invitrogen). The double-strand cDNAs were fractionated in a sepharose CL-2B 40cm-columns and fragments sizing between 0.8 and 1.5 Kb were directionally cloned into the vector. Details of each source of RNA and library construction can be obtained at http://sucest.lad.ic.unicamp.br/public"

ORIGIN

Query Match 6.1%; Score 304.4; DB 6; Length 613;
Best Local Similarity 77.3%; Pred. No. 4.2e-42;
Matches 382; Conservative 0; Mismatches 111; Indels 1; Gaps 1;

```
Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | |||||| |||||||||| |||||| | | | | | | |||||||||||| | |||||||
Db        601 TATATATAGATATATATATCTATATATATATATATATATATATATATATATATATATA 542

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db        541 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 482

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        481 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 422

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db        421 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 362

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db        361 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 302

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db        301 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 242

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||||| ||||||| ||||||| | ||||||| ||||||| |    |||||
Db        241 TATATATATATATACATATATATATATATATATATATATATATATATATATCGATATA 182

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | |   ||| || ||  ||     | |||||| ||| || |||  |
Db        181 TATATATATATATATAGATGTATATATCTATCTAGCGCTCCACATAGATAGACATATGCA 122

Qy        706 TATGTATATATATA 719
              ||   | |   |||
Db        121 TACAGAGAGCCATA 108
```

RESULT 52
CL830207
LOCUS CL830207 491 bp DNA linear GSS 09-AUG-2004
DEFINITION OR_CBa0052C06.f OR_CBa Oryza rufipogon genomic clone OR_CBa0052C06 5', genomic survey sequence.
ACCESSION CL830207
VERSION CL830207.1 GI:51075808
KEYWORDS GSS.

```
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 491)
  AUTHORS   Kim,H., Yu,Y., Wissotski,M., Yost,D., Stum,D., Rao,K., Luo,M.,
            Jetty,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and
            Wing,R.
  TITLE     OMAP project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0052  row: C  column: 06
            Seq primer: TAA TAC GAC TCA CTA TAG GG
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .491
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
                     /clone="OR_CBa0052C06"
                     /tissue_type="young leaves"
                     /dev_stage="2 week old seedlings"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OR_CBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII;
                     drk treated 36 hrs before harvest"
ORIGIN

  Query Match             6.1%;  Score 302;  DB 9;  Length 491;
  Best Local Similarity   77.3%;  Pred. No. 1.1e-41;
  Matches  379;  Conservative    0;  Mismatches  110;  Indels    1;  Gaps    1;

Qy        230 ATATATATATATATATATATATACACACACACACATATATATATACACATATATATAC 289
              ||||||||||||||||||||||| | | | | | |||||||||||| | |||||||||
Db          2 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 61

Qy        290 ACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATAT 349
              | ||||||| ||||||| | ||||||    |||||| || | ||  |||| || | ||||
Db         62 ATATATATATATATATATATATATATATGTGTATATATGTATATATGTATATATATATATAT 121

Qy        350 ATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATAC 409
              ||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||||
Db        122 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 181

Qy        410 ACACATATATACACATATATACACATATATACACATATACACATATATACACATATATAC 469
              | | ||||||| | ||||||| | || |||| | ||||| |.||||||| | |||||||
Db        182 ATATATATATATATATATATATATATATCTATATATATATATATATATATATATATATAT 241

Qy        470 ATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATATAT 528
              ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | ||||||
Db        242 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 301

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACACATATAT 588
              |||||||| | |||||| || | ||||||| | ||||||| ||||||| | ||||||
Db        302 CTCTATATATATATATATCTATATATATATATAGATATATATATATATATATATATAT 361

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
              |||  |||||| | || ||||| |||||| | | ||| | || |||     | |||||||
Db        362 ATATGTATATATATATGTATACCTATATATATACATACATATCTATCTGTATATATATAT 421

Qy        649 ACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATAT 708
              | ||| ||||    | |||||||| || ||||||||| | |||||||||  |||||||||
Db        422 ATATATATATGTGTATACATAGATCTATATATATATATATATATATATATATATATAT 481

Qy        709 GTATATATAT 718
              ||| |||||
Db        482 ATATGTATAT 491


RESULT 53
CL858220
LOCUS       CL858220                546 bp    DNA     linear   GSS 16-AUG-2004
DEFINITION  OR_CBa0089A17.f OR_CBa Oryza rufipogon genomic clone OR_CBa0089A17
            5', genomic survey sequence.
ACCESSION   CL858220
VERSION     CL858220.1  GI:51269336
KEYWORDS    GSS.
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 546)
  AUTHORS   Kim,H., Yu,Y., Wissotski,M., Yost,D., Stum,D., Rao,K., Luo,M.,
            Jetty,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and
            Wing,R.
  TITLE     OMAP project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0089  row: A  column: 17
            Seq primer: TAA TAC GAC TCA CTA TAG GG
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .546
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
                     /clone="OR_CBa0089A17"
                     /tissue_type="young leaves"
                     /dev_stage="2 week old seedlings"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OR_CBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII;
                     drk treated 36 hrs before harvest"
ORIGIN

  Query Match             6.0%;  Score 300.2;  DB 9;  Length 546;
  Best Local Similarity   77.5%;  Pred. No. 2.2e-41;
  Matches  376;  Conservative    0;  Mismatches  108;  Indels    1;  Gaps    1;

Qy        224 CATTTTATATATATATATATATATATACACACACACACATATATATATACACATAT 283
              ||  | |||||||||||||||||||||| | | | | | ||||||||||| | ||||
Db         63 CACCTCATATATATATATATATATATATATATATATATATATATATATATATATATAT 122

Qy        284 ATATACACATATATACATATATACACATATATACACATATATATACACATACATATGTAC 343
              ||||| | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| ||
Db        123 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 182

Qy        344 ACATATATATACACATATGTATACACATATATACACATATATACACACATATATACACAT 403
              | ||||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | ||
Db        183 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 242

Qy        404 ATATACACACATATATACACATATATACACATATATACACATATACACATATATACACAT 463
              ||||| | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | ||
Db        243 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 302

Qy        464 ATATACATATATACACATATATATAATATACACACATATATATACACATATATACACACA 523
```

```
Db        303 ATATATATATATATATATATATAT-ATATATATATATATATATATATATATATATATA 361

Qy        524 TATATACACATATATACACATATATATACACATATATACACATATATACATATATACACA 583
              |||||| | ||||||| | |||||||||    |||||||   ||||||| ||||||| | |
Db        362 TATATATATATATATATAGATATATATATGTATATATATGTATATATATATATATATATA 421

Qy        584 TATATATACATATATACACATATATACATATATACACATATATACATATATACACACATA 643
              |||||||| ||||||| | ||||||| ||| ||| | ||||||| ||||||| | | |||
Db        422 TATATATATATATATATATATATATATATAGATATATATATATATATATATATATATATA 481

Qy        644 TATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATA 703
              |||| | ||| |||| | | | ||          |||||| | |||||||||  ||||
Db        482 TATATATATATATATGTATATATATGATATATATATTATATATAGATATATATATATATA 541

Qy        704 TATAT 708
              |||||
Db        542 TATAT 546


RESULT 54
BX180548/c
LOCUS       BX180548                592 bp    DNA     linear   GSS 28-JAN-2003
DEFINITION  Danio rerio genomic clone DKEY-180C22, genomic survey sequence.
ACCESSION   BX180548
VERSION     BX180548.1  GI:28012355
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 592)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the T7 end of BAC 180C22. 180C22
            is part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .592
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-180C22"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             6.0%;  Score 298.6;  DB 9;  Length 592;
  Best Local Similarity   78.0%;  Pred. No. 4.1e-41;
  Matches  372;  Conservative    0;  Mismatches  104;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | |||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db        590 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 531

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | |||||||| | ||| |||| || | |
Db        530 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 471

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        470 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 411

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db        410 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 351

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | ||||| | | ||||||| | ||||||| |
Db        350 TATATATATATATATATATATATATATATATACATACACACACACATACACACACACA 291

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              | ||||||| | ||||| | | | |||||||||||| ||||| ||||||| ||||| |||
Db        290 CACACACATACACACACACACACACACACATATATATACATACATACATACATACATATA 231

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db        230 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 171

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTAT 702
              || | ||| ||||| | | | ||| |||||| | | |         || | | |  |||
Db        170 TATATATATATATATATATATATATATATACTTTTTTCTGTGCTAAACAAATTTAT 114


RESULT 55
BX134913
LOCUS       BX134913                696 bp    DNA     linear   GSS 28-JAN-2003
DEFINITION  Danio rerio genomic clone DKEY-89F23, genomic survey sequence.
ACCESSION   BX134913
VERSION     BX134913.1  GI:27966206
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 696)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the T7 end of BAC 89F23. 89F23 is
            part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .696
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-89F23"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             6.0%;  Score 297.4;  DB 9;  Length 696;
  Best Local Similarity   79.8%;  Pred. No. 6.5e-41;
  Matches  363;  Conservative    0;  Mismatches   91;  Indels    1;  Gaps    1;

Qy        231 TATATATATATATATATATATACACACACACACATATATATATACACATATATATACA 290
              ||||||| |||||||||||||| | | | | | |||||||||||| | |||||||| |
Db        243 TATATATTTATATATATATATATATATATATATATATATATATATATATATATATA 302

Qy        291 CATATATACATATATACACATATATACACATATATATACACATACATATGTACACATATA 350
              ||||||| ||||||| | |||||||     |||||| | | ||| |||| || | |||||
Db        303 TATATATATATATATATATATATATAGTAGTATATAGATAGATATATATATATATATATA 362

Qy        351 TATACACATATGTATACACATATATACACATATATACACACATATATACACATATATACA 410
              |||| | ||||||||| | ||||||| | ||||||| | | ||||||| | ||||||| |
Db        363 TATATATATATGTATATATATATATATATATATATATATATATATATATATATATA 422

Qy        411 CACATATATACACATATATACACATATATACACATATACACATATATACACATATATACA 470
              | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | ||||||| |
Db        423 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 482

Qy        471 TATATACACATATATATAATATACACACATATATATACACATATATACACACATATATAC 530
              |||||| | |||||||| ||||| | | |||||||||| | ||||||| | | |||||||
Db        483 TATATATATATATATAT-ATATATATATATATATATATATATATATATATATATAT 541

Qy        531 ACATATATACACATATATATACACATATATACACATATATACATATATACACATATATAT 590
```

```
Db        542 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 601

Qy        591 ACATATATACACATATATACATATATACACATATATACATATATACACACATATATACAC 650
              | ||||||| | |||||||| | ||||||| | ||||||| ||||||| | ||||||||
Db        602 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 661

Qy        651 ATACATATACACACACATAGATATACATATATATA 685
              ||| ||||| | | | ||| ||||| |||||||||
Db        662 ATATATATATATATATATATATATATATATATATATA 696


RESULT 56
BX163144
LOCUS       BX163144                700 bp    DNA     linear   GSS 13-MAR-2003
DEFINITION  Danio rerio genomic clone DKEY-154G8, genomic survey sequence.
ACCESSION   BX163144
VERSION     BX163144.1  GI:27994657
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 700)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAR-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the T7 end of BAC 154G8. 154G8 is
            part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .700
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-154G8"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             6.0%;  Score 297.4;  DB 9;  Length 700;
  Best Local Similarity   79.0%;  Pred. No. 6.5e-41;
  Matches  391;  Conservative    0;  Mismatches  101;  Indels    3;  Gaps    3;

Qy        227 TTTATATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | |||||||||||||||||||||||||||||||||||||||||||| |||||||| | |||||||
Db        198 TATATATATATATATATATATATATATACACACACACACACAT-TATATATATATATATATA 256

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | | | | | ||||| ||| |||||||||||  |||| |
Db        257 TATATATATATATATATATATATATATACACACATACATACATACACACATACATACATACATA 316

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
               ||| | ||||| |   | ||||| | ||||| | ||| | ||| | ||| | ||||| |
Db        317 CATACACACACACACACACACACACACATACATACATACATACATACATACATACATACA 376

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              |||| ||||| ||||| | ||| | |||| |||| | ||||| | ||| | ||||| |
Db        377 TACATACATACATACATACATACATACATACATACATACATACATACATACATACATACA 436

Qy        467 TACATATATACACATATATATAATATACACACATATATATACACATATATACACACATAT 526
              |||||| ||||| | ||| || | ||||| ||||| ||||| | |||||||| | | ||||
Db        437 TACATACATACATACATACAT-ACATACATACATACATATATATATATATATATATATATAT 495

Qy        527 ATACACATATATACACATATATATACACATATATACACATATATACATATATACACATAT 586
              |||     ||||| | | ||||||||| | ||| | ||||||||| ||||||| | ||||
Db        496 ATAATATCATATATATACATATATACATACATACATACATATATATATATATATATATAT 555

Qy        587 ATATACATATATACACATATATACATATATACACATATATACATATATACACACATATAT 646
```

```
              ||||| ||||||| | ||||||| ||||| ||||| | | ||||| ||||| ||||| ||
Db        556 ATATATATATATATATATATATATATATATACACACACACACACATACATACATACATACAT 615

Qy        647 ACACATACATATACACAC-ACATAGATATACATATATATACACATATATATACGTATATA 705
              ||||| ||| | |||||     || |||| ||||||||| | |||||||||  ||||||
Db        616 ACACACACACACACACGTGTTATATATATATATATATATATATATATATATATATATA 675

Qy        706 TATGTATATATATAT 720
              ||| ||||||||||||
Db        676 TATATATATATATAT 690


RESULT 57
BX180768/c
LOCUS       BX180768                701 bp    DNA     linear   GSS 28-JAN-2003
DEFINITION  Danio rerio genomic clone DKEY-176A10, genomic survey sequence.
ACCESSION   BX180768
VERSION     BX180768.1  GI:28012577
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 701)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the SP6 end of BAC 176A10. 176A10
            is part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .701
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-176A10"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             6.0%;  Score 297;  DB 9;  Length 701;
  Best Local Similarity   70.1%;  Pred. No. 7.6e-41;
  Matches  413;  Conservative    0;  Mismatches  175;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | |||||||||||||||||||||||||| | | | | | |||||||||| | |||||||
Db        701 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 642

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db        641 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 582

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        581 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 522

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | |||||||| | |||||||| | ||||||| | ||||| | ||||||| | |||||
Db        521 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 462

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db        461 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 402

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db        401 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 342

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
```

```
Db        341 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 282

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705

Db        281 TATATATATATATATTCTCGAAAACCTCTGATCAAACTCTTACCTTCATTTAAAACATTT 222

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765

Db        221 TCATTTAAGCCATACACAGCTTAACATAAATGCGCCATCATTAAGTCATCGCCAAGGAAA 162

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGG 814

Db        161 GTAAATAAATACATATCAGCGCATACCTTTCTAGAAAACTGAGTTCAGG 113


RESULT 58
CL763005
LOCUS       CL763005                617 bp    DNA     linear   GSS 27-JUL-2004
DEFINITION  OR_BBa0133F16.f OR_BBa Oryza rufipogon genomic clone OR_BBa0133F16
            5', genomic survey sequence.
ACCESSION   CL763005
VERSION     CL763005.1  GI:50720952
KEYWORDS    GSS.
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 617)
  AUTHORS   Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R.,
            Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
  TITLE     OMAP Project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Insert Length: 161   Std Error: 0.00
            Plate: 0133  row: F  column: 16
            Seq primer: TAA TAC GAC TCA CTA TAG GG
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .617
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
                     /clone="OR_BBa0133F16"
                     /tissue_type="young leaves"
                     /lab_host="DH10B-T1 phage resistant"
                     /clone_lib="OR_BBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

  Query Match             5.9%;  Score 295.4;  DB 9;  Length 617;
  Best Local Similarity   76.3%;  Pred. No. 1.5e-40;
  Matches  376;  Conservative    0;  Mismatches  116;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286

Db         85 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 144

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346

Db        145 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 204

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406

Db        205 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 264

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466

Db        265 TATATATATATATATATATATATATATATATATATATATATATATATATATATAAATATATA 324

Qy        467 TACATATATACACATATATATAATATACACACATATATATACACATATATACACACATAT 526

Db        325 TAAATATATATATATATATATATTATATA-ATATATATATTATTTTATTTTTTTTAATAT 383

Qy        527 ATACACATATATACACATATATATACACATATATACACATATATACATATATACACATAT 586

Db        384 ATATATATATATATTTATATATATATATATATATATATATATATAAAAAAATATATTTTT 443

Qy        587 ATATACATATATACACATATATACATATATACACATATATACATATATACACACATATAT 646

Db        444 ATATATATAGATACATATATATAAAAATATAGAGAGAGATATATATAAATATAAAAAAAG 503

Qy        647 ACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATAT 706

Db        504 ATATATATATATATATATGCAGATACATAGATATATATACATATAGATATATATATAGTT 563

Qy        707 ATGTATATATATA 719

Db        564 ATGTATATATATA 576


RESULT 59
BX134913/c
LOCUS       BX134913                696 bp    DNA     linear   GSS 28-JAN-2003
DEFINITION  Danio rerio genomic clone DKEY-89F23, genomic survey sequence.
ACCESSION   BX134913
VERSION     BX134913.1  GI:27966206
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 696)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the T7 end of BAC 89F23. 89F23 is
            part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .696
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-89F23"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             5.9%;  Score 294.4;  DB 9;  Length 696;
  Best Local Similarity   79.6%;  Pred. No. 2.1e-40;
  Matches  360;  Conservative    0;  Mismatches   91;  Indels    1;  Gaps    1;

Qy        269 TATATATACACATATATATACACATATATACATATATACACATATATACACATATATATA 328

Db        696 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 637

Qy        329 CACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACA 388

Db        636 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 577

Qy        389 CACATATATACACATATATACACACATATATACACATATATACACATATATACACATATA 448
```

```
Db        576 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 517

Qy        449 CACATATATACACATATATACATATATACACATATATATAATATACACACATATATATAC 508
Db        516 TATATATATATATATATATATATATATATATATATATAT-ATATATATATATATATATAT 458

Qy        509 ACATATATACACACATATATACACATATATACACATATATATACACATATATACACATAT 568
Db        457 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 398

Qy        569 ATACATATATACACATATATATACATATATACACATATATACATATATACACATATATAC 628
Db        397 ATATATATATATATATATATATATACATATATATATATATATATATATATATATCTATCT 338

Qy        629 ATATATATACACACATATATACACATACATATACACACACATAGATATACATATATATACAC 688
Db        337 ATATACTACTATATATATATATATATATATATATATATATATATATATATATATATATAT 278

Qy        689 ATATATATACGTATATATATGTATATATATAT 720
Db        277 ATATATATATATATATATATATATATAAATAT 246
```

RESULT 60
CF924712/c

```
LOCUS       CF924712                760 bp    mRNA    linear   EST 10-NOV-2003
DEFINITION  DD005E11.ab1 Codonopsis lanceolata cDNA library Codonopsis
            lanceolata cDNA, mRNA sequence.
ACCESSION   CF924712
VERSION     CF924712.1  GI:38243228
KEYWORDS    EST.
SOURCE      Codonopsis lanceolata
  ORGANISM  Codonopsis lanceolata
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            asterids; campanulids; Asterales; Campanulaceae; Codonopsis.
REFERENCE   1  (bases 1 to 760)
  AUTHORS   Yang,D.C., In,J.G. and Lee,B.S.
  TITLE     Codonopsis lanceolata
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Deok Chun Yang
            BioPia
            1 Seochonri, Kihung, Yongin, Kyungki-do, 449-701, Korea
            Tel: 82 31 201 2688
            Fax: 82 31 205 6323
            Email: dcyang@ibiopia.com.
FEATURES             Location/Qualifiers
     source          1. .760
                     /organism="Codonopsis lanceolata"
                     /mol_type="mRNA"
                     /db_xref="taxon:103999"
                     /clone_lib="Codonopsis lanceolata cDNA library"
                     /note="Organ: root; Vector: pTriplEx2; Site_1: Sfi IA;
                     Site_2: Sfi IB"
ORIGIN
```

```
  Query Match             5.9%;  Score 294.2;  DB 7;  Length 760;
  Best Local Similarity   79.8%;  Pred. No. 2.3e-40;
  Matches  347;  Conservative    0;  Mismatches   88;  Indels    0;  Gaps    0;

Qy        227 TTTATATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
Db        748 TATGTATATATAGATATATATATATATATAGATATATATATATATATATATATATATATA 689

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
Db        688 TATATATATATATATATATATATATATATATCTATATATATATATATATATATATATATATA 629

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
Db        628 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 569

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
Db        568 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 509

Qy        467 TACATATATACACATATATATAATATACACACATATATATACACATATATACACACATAT 526
Db        508 TATATATATATATATATATATCATATATATATATATATATATATATATATATATATATAT 449

Qy        527 ATACACATATATACACATATATATACACATATATACACATATATACACATATATACACATAT 586
Db        448 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 389

Qy        587 ATATACATATATACACATATATACATATATACACATATATACATATATACACACATATAT 646
Db        388 ATATATATATATATATATATATATATATATATATATATATATATATATATATGCATATAT 329

Qy        647 ACACATACATATACA 661
Db        328 ATATATATATATATA 314
```

RESULT 61
BX197183

```
LOCUS       BX197183                771 bp    DNA     linear   GSS 28-JAN-2003
DEFINITION  Danio rerio genomic clone DKEY-207A10, genomic survey sequence.
ACCESSION   BX197183
VERSION     BX197183.1  GI:28029069
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 771)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the SP6 end of BAC 207A10. 207A10
            is part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .771
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-207A10"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN
```

```
  Query Match             5.9%;  Score 293.6;  DB 9;  Length 771;
  Best Local Similarity   77.1%;  Pred. No. 2.9e-40;
  Matches  370;  Conservative    0;  Mismatches  109;  Indels    1;  Gaps    1;

Qy        168 ATTGCACTCCAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATT 227
Db        292 ATTCAATTCAAAAGCATAAATAAATTGGGGAAAAAAAAATAAATAAAAAAAAAAAAAATAT 351

Qy        228 TTATATATATATATATATATATATACACACACACACATATATATATACACATATATAT 287
Db        352 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 411

Qy        288 ACACATATATACACATATATACACATATATACACATATATATACACATACATATGTACACAT 347
Db        412 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 471

Qy        348 ATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATAT 407
Db        472 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 531
```

```
Qy        408 ACACACATATATACACATATATACACATATATACACATATACACATATATACACATATAT 467
              | | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | ||||||
Db        532 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 591

Qy        468 ACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATAT 526
              | ||||||| | |||||||||| ||||| | | |||||||||| | ||||||| | | ||||
Db        592 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 651

Qy        527 ATACACATATATACACATATATATACACATATATACACATATATACATATATACACACATAT 586
              ||| | ||||||| | |||||||||| | ||||||| | ||||||| ||||||| | ||||
Db        652 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 711

Qy        587 ATATACATATATACACATATATACATATATACACATATATACATATATACACACATATAT 646
              ||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | ||||||||
Db        712 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATACATATAT 771
```

RESULT 62
BZ215061
LOCUS BZ215061 643 bp DNA linear GSS 11-OCT-2002
DEFINITION CH230-280D4.TV CHORI-230 Segment 2 Rattus norvegicus genomic clone CH230-280D4, genomic survey sequence.
ACCESSION BZ215061
VERSION BZ215061.1 GI:23873314
KEYWORDS GSS.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 643)
AUTHORS Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K., Shvartsbeyn,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D., Riggs,F., de Jong,P. and Fraser,C.M.
TITLE Rat BAC End Sequences from Library CHORI-230 MboI segment
JOURNAL Unpublished (1999)
COMMENT Other_GSSs: CH230-280D4.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the rat BAC library CHORI-230 (http://www.chori.org/bacpac/rat230.htm). For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (http://www.chori.org/bacpac/or ering_information.htm). BAC end page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
Plate: 280 row: D column: 4
Seq primer: T7
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .643
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/strain="BN/SsNHsd/MCW"
/db_xref="taxon:10116"
/clone="CH230-280D4"
/sex="Female"
/cell_type="Brain"
/clone_lib="CHORI-230 Segment 2"
/note="Vector: pTARBAC1.3; Site_1: MboI; Site_2: MboI; CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by Pieter de Jong"
ORIGIN

Query Match 5.9%; Score 293; DB 8; Length 643;
Best Local Similarity 75.7%; Pred. No. 3.8e-40;
Matches 376; Conservative 0; Mismatches 120; Indels 1; Gaps 1;

```
Qy        229 TATATATATATATATATATATATACACACACACACATATATATATACACATATATATA 288
              || | | | | ||| | | |||||||||| |||||||||| ||||||||| ||| | |
Db        118 TACAAACACACATACACACATATACACACATACACACATATACATATACACACATACACA 177

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              |||||||| ||||| ||||||| ||||| ||| | |||||||||||||| || ||||    |
Db        178 CACATATACACATACATACACACATATAAACACACATATACACATACACATATACATGCA 237

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              | ||||||| || || ||||| ||| ||    |||||||||||||||| ||| | ||| |
Db        238 CACATACACACATATACACACACATACACTTCCATATACACACATATACACAGACATACA 297

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              |||| ||| | | ||||| | ||| ||| | | ||||| ||||||||| ||||| ||| |
Db        298 CACATATACACAGACATACACACATATACACAGACATACACACATATAGACACACATACA 357

Qy        469 CATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATATA 527
              | ||  || ||| | ||||| | | ||| ||| | ||||||||| ||||||||||||||||
Db        358 CTTACTTATACACACATATACACAGACATACACACATATACACAGATATACACACATATA 417

Qy        528 TACACATATATACACATATATATACACATATATACACATATATACATATATACACATATA 587
              ||| | ||| ||||||||| | ||||||| |   ||||||| ||| |||||||||| | |
Db        418 CACAGACATACACACATATAGACACACATACACTTACATATACACACATATACACAGACA 477

Qy        588 TATACATATATACACATATATACATATATACACATATATACATATATACACACATATATA 647
              || ||| |||||||||| ||||||| ||| ||||| ||| ||| |||||||||| | | | |
Db        478 TACACACATATACACACATATACACATACACACACATACACACATATACACATACACACA 537

Qy        648 CACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATA 707
              || | ||| ||| ||||| | | ||| ||||| | | | ||| | ||| ||  ||| | |
Db        538 CATACACACATATACACATACATATACACATACACACAAACACACATACACACATACACA 597

Qy        708 TGTATATATATATGCTC 724
                || . | | |  | |
Db        598 CATACGAACACACACAC 614
```

RESULT 63
BM906116
LOCUS BM906116 1156 bp mRNA linear EST 12-MAR-2002
DEFINITION AGENCOURT_6621231 NIH_MGC_125 Homo sapiens cDNA clone IMAGE:5589966 5', mRNA sequence.
ACCESSION BM906116
VERSION BM906116.1 GI:19356495
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1156)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Invitrogen
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate: LLAM12363 row: i column: 07
High quality sequence start: 323
High quality sequence stop: 613.
FEATURES Location/Qualifiers
source 1. .1156
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5589966"
/lab_host="DH10B"
/clone_lib="NIH_MGC_125"
/note="Organ: ovary (pool of 3); Vector: pCMV-SPORT6;

Site_1: EcoRV (destroyed); Site_2: NotI; RNA source pool of three ovaries, from females ranging in age from 38 to 49 yo. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 2.1 kb, insert size range 1-3.5 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 036."

ORIGIN

Query Match 5.9%; Score 292.6; DB 5; Length 1156;
Best Local Similarity 76.5%; Pred. No. 4.1e-40;
Matches 385; Conservative 0; Mismatches 114; Indels 4; Gaps 2;

```
Qy        210 GAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACACATAT 269
              || | | |   |||    |  | || | |||||| |||||||| | | | ||| | ||||
Db        223 GACATAGATGGAGGAGGATCCTCTAGACATATATGTATATATAAATAGATACATATATAT 282

Qy        270 ATATATACACATATATATACACATATATACATATATACACATATATACACATATATATAC 329
              | |||||   | ||  ||| | ||||||||||||||| | ||||||||||||||||||
Db        283 AGATATATCTACATGCATATATATATATACATATATATAGATATATACACATATATATAT 342

Qy        330 ACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACAC 389
              | ||| |||| || | ||||||||| | |||| |||| | ||||||| | ||||||| |
Db        343 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 402

Qy        390 ACATATATACACATATATACACACATATATACACATATATACACATATATACACATATAC 449
              | ||||||| | ||||||| | | ||||||| | ||||||| | ||||||| | |||||
Db        403 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 462

Qy        450 ACATATATACACATATATACATATATACACATATATATA-ATATACACACATATATATAC 508
              | ||||||| | ||||||| ||||||| | ||||||||| ||||| | | |||||||||
Db        463 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 522

Qy        509 ACATATATACACACATATATACACATATATACACATATATATACACATATATACACATAT 568
              | ||||||| | | ||||||| | ||||||| | ||||||||| | ||||||| | ||||
Db        523 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 582

Qy        569 ATACATATATACACATATATATACATATATACACATATATACATATATACACATATATAC 628
              ||| ||||||| | ||||||||| ||||||| | ||||||| ||||||| ||||||||||
Db        583 ATATATATATATATATATATATATATATATATATATATATATATATATATATACATATAC 642

Qy        629 ATATATACACACATATATACACATACATATACACACACATAGATATACATATATATACAC 688
              ||||||| | | ||||||| | ||| ||||| | | | ||| |||| |    | ||||
Db        643 ATATATATATATATATATATATATATATATATATATATATACATATGGA---CTGTACAT 699

Qy        689 ATATATATACGTATATATATGTA 711
              ||||||| ||  | ||   ||||
Db        700 ATATATACACCCACATGACTGTA 722
```

RESULT 64
CL907900/c
LOCUS CL907900 714 bp DNA linear GSS 14-SEP-2004
DEFINITION OA_ABa0007A21.r OA_ABa Oryza australiensis genomic clone OA_ABa0007A21 3', genomic survey sequence.
ACCESSION CL907900
VERSION CL907900.1 GI:52016779
KEYWORDS GSS.
SOURCE Oryza australiensis
ORGANISM Oryza australiensis
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 714)
AUTHORS Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
TITLE OMAP Project
JOURNAL Unpublished (2004)
COMMENT Contact: Rod A. Wing
Arizona Genomics Institute
University of Arizona
Forbes Building Room 303, Tucson, AZ 85721-0036, USA
Tel: 520 626 9595
Fax: 520 621 1259
Email: http://genome.arizona.edu
PCR PRimers
FORWARD: TAA TAC GAC TCA CTA TAG GG
BACKWARD: CAC TCA TTA GGC ACC CCA
Plate: 0007 row: A column: 21
Seq primer: CAC TCA TTA GGC ACC CCA
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .714
/organism="Oryza australiensis"
/mol_type="genomic DNA"
/db_xref="taxon:4532"
/clone="OA_ABa0007A21"
/tissue_type="young leaves"
/lab_host="DH10B T1 phage resistant"
/clone_lib="OA_ABa"
/note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"

ORIGIN

Query Match 5.8%; Score 290.2; DB 9; Length 714;
Best Local Similarity 77.7%; Pred. No. 1.1e-39;
Matches 363; Conservative 0; Mismatches 103; Indels 1; Gaps 1;

```
Qy        286 ATACACATATATACATATATACACATATATACACATATATATACACATACATATGTACAC 345
              ||| | ||||||| ||||||| | ||||||  | |||| |||| | ||| |||| || |
Db        714 ATATATATATATATATATATATATATATATTTATATATTTATATATATATATATATAT 655

Qy        346 ATATATATACACATATGTATACACATATATACACATATATACACACATATATACACATAT 405
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | ||||
Db        654 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 595

Qy        406 ATACACACATATATACACATATATACACATATATACACATATACACATATATACACATAT 465
              ||| | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | ||||
Db        594 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 535

Qy        466 ATACATATATACACATATATATAATATACACACATATATATACACATATATACACACATA 525
              ||| ||||||| | |||||||| ||||| | | ||||||||| | ||||||| | | |||
Db        534 ATATATATATATATATATATAT-ATATATATATATATATATATATATATATATATATA 476

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db        475 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 416

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db        415 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 356

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | |||||||||| | | | ||| ||||| |||||||||| | ||||||||   ||||||
Db        355 TATATATACATATATATATATATATATATATATATATATATATATATATATATATATA 296

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATT 752
              ||| |||||||| | || |  | || |  | |   | ||  |  |||
Db        295 TATATATATATACACGCGCGCGTGTGCGCACGTTATTGCGTTGAATT 249
```

RESULT 65
BX180768
LOCUS BX180768 701 bp DNA linear GSS 28-JAN-2003
DEFINITION Danio rerio genomic clone DKEY-176A10, genomic survey sequence.
ACCESSION BX180768
VERSION BX180768.1 GI:28012577
KEYWORDS GSS.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 701)
AUTHORS Humphray,S.J., Huckle,E. and Durham,J.L.
TITLE Direct Submission

```
  JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the SP6 end of BAC 176A10. 176A10
            is part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .701
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-176A10"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             5.8%;  Score 290;  DB 9;  Length 701;
  Best Local Similarity   76.1%;  Pred. No. 1.2e-39;
  Matches  370;  Conservative    0;  Mismatches  115;  Indels    1;  Gaps    1;

Qy        213 AAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACACATATATA 272
              ||| |||| |  ||  ||  |  ||    | |    | |      |    | |||||||
Db        216 AAATGAAAATGTTTTAAATGAAGGTAAGAGTTTGATCAGAGGTTTTCGAGAATATATATA 275

Qy        273 TATACACATATATATACACATATATACATATATACACATATATACACATATATATACACA 332
              ||||  | |||||||||  | |||||||  |||||||  | |||||||  | |||||||||  | |
Db        276 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 335

Qy        333 TACATATGTACACATATATATACACATATGTATACACATATATACACATATATACACACA 392
              ||  |||| || | |||||||||  | ||||  ||||  | |||||||  | |||||||  | | |
Db        336 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 395

Qy        393 TATATACACATATATACACACATATATACACATATATACACATATATACACATATACACA 452
              ||||||  | |||||||  | |  |||||||  | |||||||  | |||||||  | |||||  | |
Db        396 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 455

Qy        453 TATATACACATATATACATATATACACATATATATA-ATATACACACATATATATACACA 511
              ||||||  | |||||||  |||||||  | |||||||||  |||||  | |  |||||||||  | |
Db        456 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 515

Qy        512 TATATACACACATATATACACATATATACACATATATATACACATATATACACATATATA 571
              ||||||  | |  |||||||  | |||||||  | |||||||||  | |||||||  | |||||||
Db        516 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 575

Qy        572 CATATATACACATATATATACATATATACACATATATACATATATACACATATATACATA 631
              |  |||||||  | |||||||||  |||||||  | |||||||  |||||||  | |||||||  |||
Db        576 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 635

Qy        632 TATACACACATATATACACATACATATACACACACATAGATATACATATATATACACATA 691
              ||||  | |  |||||||  | |||  |||||  | |  | |||  |||||  |||||||||  | |||
Db        636 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 695

Qy        692 TATATA 697
              ||||||
Db        696 TATATA 701


RESULT 66
CL907900
LOCUS       CL907900                 714 bp    DNA     linear   GSS 14-SEP-2004
DEFINITION  OA_ABa0007A21.r OA_ABa Oryza australiensis genomic clone
            OA_ABa0007A21 3', genomic survey sequence.
ACCESSION   CL907900
VERSION     CL907900.1  GI:52016779
KEYWORDS    GSS.
SOURCE      Oryza australiensis
  ORGANISM  Oryza australiensis
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 714)
  AUTHORS   Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R.,
            Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
  TITLE     OMAP Project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0007  row: A  column: 21
            Seq primer: CAC TCA TTA GGC ACC CCA
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .714
                     /organism="Oryza australiensis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4532"
                     /clone="OA_ABa0007A21"
                     /tissue_type="young leaves"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OA_ABa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

  Query Match             5.8%;  Score 289.4;  DB 9;  Length 714;
  Best Local Similarity   79.4%;  Pred. No. 1.5e-39;
  Matches  355;  Conservative    0;  Mismatches   91;  Indels    1;  Gaps    1;

Qy        251 TACACACACACACACATATATATATACACATATATATACACATATATACATATATACACA 310
              | | ||||| | | | | ||||||||  | |||||||||  | |||||||  |||||||  | |
Db        266 TGCGCACACGCGCGCGTGTATATATATATATATATATATATATATATATATATATATATA 325

Qy        311 TATATACACATATATATACACATACATATGTACACATATATATACACATATGTATACACA 370
              ||||||  | |||||||||  |   || ||||  ||  | |||||||||  | ||||  ||||  | |
Db        326 TATATATATATATATATATATGTATATATATATATATATATATATATATATATATATATA 385

Qy        371 TATATACACATATATACACACATATATACACATATATACACACATATATACACATATATA 430
              ||||||  | |||||||  | |  |||||||  | |||||||  | |  |||||||  | |||||||
Db        386 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 445

Qy        431 CACATATATACACATATACACATATATACACATATATACATATATACACATATATAAT 490
              |  |||||||  | |||||  | |||||||  | |||||||  |||||||  | |||||||  ||
Db        446 TATATATATATATATATATATATATATATATATATATATATATATATATATATATAT-AT 504

Qy        491 ATACACACATATATATACACATATATACACACATATATACACATATATACACATATATAT 550
              |||  | |  |||||||||  | |||||||  | |  |||||||  | |||||||  | |||||||||
Db        505 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 564

Qy        551 ACACATATATACACATATATACATATATACACATATATATACATATATACACATATATAC 610
              | |  |||||||  | |||||||  |||||||  | |||||||||  |||||||  | |||||||
Db        565 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 624

Qy        611 ATATATACACATATATACATATATACACACATATATACACATACATATACACACACATAG 670
              |||||||  | |||||||  |||||||  | |  |||||||  | |||  |||||  | |  | |||
Db        625 ATATATATATATATATATATATATATATATATATATATATATATATATATATATAAATATATAA 684

Qy        671 ATATACATATATATACACATATATATA 697
              |||||  |||||||||  | |||||||||
Db        685 ATATATATATATATATATATATATATA 711


RESULT 67
DR42I22S
LOCUS       DR42I22S                 673 bp    DNA     linear   GSS 22-NOV-2002
DEFINITION  Danio rerio genomic clone DKEY-42I22, genomic survey sequence.
ACCESSION   AL983929
VERSION     AL983929.1  GI:25183962
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
```

```
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 673)
  AUTHORS   Humphray,S.J., Huckle,E. and Hunt,S.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-NOV-2002) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail contact:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the SP6 end of BAC 42I22. 42I22 is
            part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene.
            Further details: http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .673
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-42I22"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             5.8%;  Score 288.8;  DB 9;  Length 673;
  Best Local Similarity   80.0%;  Pred. No. 2e-39;
  Matches  352;  Conservative    0;  Mismatches   87;  Indels    1;  Gaps    1;

Qy        281 TATATATACACATATATACATATATACACATATATACACATATATATACACATACATATG 340
              | | | || | ||||||| ||||||| | ||||||| | ||||||||| | ||| ||||
Db        234 TGTGTGTATATATATATATATATATATATATATATATATATATATATATATATATATA 293

Qy        341 TACACATATATATACACATATGTATACACATATATACACATATATACACACATATATACA 400
              || | |||||||||| | |||| |||||||| | | ||||| | | ||||| ||| | |||
Db        294 TATATATATATATATATATATATATATACACACACACACACACACACACACATATACACACA 353

Qy        401 CATATATACACACATATATACACATATATACACATATATACACATATACACATATATACA 460
              |||||||| | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| |
Db        354 CATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 413

Qy        461 CATATATACATATATACACATATATATA-ATATACACACATATATATACACATATATACA 519
              ||||||| ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| |
Db        414 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 473

Qy        520 CACATATATACACATATATACACATATATATACACATATATACACATATATACATATATA 579
              | ||||||| | ||||||| | ||||||||| | ||||||| | ||||||| |||||||
Db        474 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 533

Qy        580 CACATATATATACATATATACACATATATACATATATACACATATATACATATATACACA 639
              | ||||||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | |
Db        534 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 593

Qy        640 CATATATACACATACATATACACACACATAGATATACATATATATACACATATATATACG 699
              ||||||| | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||
Db        594 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 653

Qy        700 TATATATATGTATATATATA 719
              ||||||||| ||||||||||
Db        654 TATATATATATATATATATA 673


RESULT 68
CW007564
LOCUS       CW007564               569 bp    DNA     linear   GSS 23-SEP-2004
DEFINITION  ZMMBLa0009D17.r ZMMBLa Zea mays genomic clone ZMMBLa0009D17 3',
            genomic survey sequence.
ACCESSION   CW007564
VERSION     CW007564.1  GI:52588327
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 569)
  AUTHORS   Wing,R., Luo,M., Soderlund,C. and Haller,K.
  TITLE     ZMMBL sequences
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            Plate: 0009  row: D  column: 17
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .569
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /cultivar="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBLa0009D17"
                     /tissue_type="immature ears"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="ZMMBLa"
                     /note="Vector: pAGIBAC1; Site_1: SalI; Site_2: SalI"
ORIGIN

  Query Match             5.8%;  Score 288.4;  DB 9;  Length 569;
  Best Local Similarity   74.5%;  Pred. No. 2.3e-39;
  Matches  389;  Conservative    0;  Mismatches  131;  Indels    2;  Gaps    2;

Qy        200 TGTCATCTCAGAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACAC 259
              ||| |  || ||||| ||  | |||  ||   |||||  |     |  |    ||
Db         48 TGTAAAGTCTGAAAACAACCACGGCCATTGTAATATAATTTCCCAACTGCTCACGATGCT 107

Qy        260 ACA-CACATATATATATACACATATATATACACATATATACATATATACACATATATACA 318
              ||| | || |  || | |  |   |  |||| ||||||| ||||||| | ||||||| |
Db        108 ACATCTCAAAACTACAAAACCTCGTTGCTACAAATATATATATATATATATATATATATA 167

Qy        319 CATATATATACACATACATATGTACACATATATATACACATATGTATACACATATATACA 378
              |||||||||| | ||| |||| || | |||||||||| | |||| |||| | ||||||| |
Db        168 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 227

Qy        379 CATATATACACACATATATACACATATATACACACATATATACACATATATACACATATA 438
              ||||||| | | ||||||| | ||||||| | | ||||||| | ||||||| | |||||
Db        228 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 287

Qy        439 TACACATATACACATATATACACATATATACATATATACACATATATATA-ATATACACA 497
              || | ||||| | ||||||| | ||||||| ||||||| | ||||||||| ||||| | |
Db        288 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 347

Qy        498 CATATATATACACATATATACACACATATATACACATATATACACATATATATACACATA 557
              ||||||||| | ||||||| | | ||||||| | ||||||| | ||||||||| | |||
Db        348 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 407

Qy        558 TATACACATATATACATATATACACATATATATACATATATACACATATATACATATATA 617
              |||| | ||||||| ||||||| | ||||||||| ||||||| | ||||||| |||||||
Db        408 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 467

Qy        618 CACATATATACATATATACACACATATATACACATACATATACACACACATAGATATACA 677
              | ||||||| ||||||| | | ||||||| | ||| ||||| | | | ||| ||||| |
Db        468 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 527

Qy        678 TATATATACACATATATATACGTATATATATGTATATATATA 719
              |||||||| | |||||||||  |||| |||| |||| | |||
Db        528 TATATATATATATATATATATATATAAATATATATAGAAATA 569


RESULT 69
DR42I22S/c
LOCUS       DR42I22S               673 bp    DNA     linear   GSS 22-NOV-2002
DEFINITION  Danio rerio genomic clone DKEY-42I22, genomic survey sequence.
ACCESSION   AL983929
```

```
VERSION     AL983929.1  GI:25183962
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 673)
  AUTHORS   Humphray,S.J., Huckle,E. and Hunt,S.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-NOV-2002) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail contact:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the SP6 end of BAC 42I22. 42I22 is
            part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene.
            Further details: http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .673
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-42I22"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             5.8%;  Score 288;  DB 9;  Length 673;
  Best Local Similarity   75.4%;  Pred. No. 2.7e-39;
  Matches  371;  Conservative    0;  Mismatches  120;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db        673 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 614

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db        613 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 554

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | |||||||| | ||||||| | | ||||||| | |||||
Db        553 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 494

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db        493 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 434

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db        433 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 374

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | || | | | | || | |     | | |   | | |  | |||
Db        373 TATATATATATATATATATATGTGTGTGTATATGTGTGTGTGTGTGTGTGTGTGTATATA 314

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db        313 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 254

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||||||||||| | | | ||| | | | ||||| | | | ||  | | |
Db        253 TATATATATATATACACACACACACACACACACACACACACACACACACACACACACACA 194

Qy        706 TATGTATATATA 717
                |  || ||||
Db        193 GTTTCATTTATA 182


RESULT 70
CL917654
LOCUS       CL917654                720 bp    DNA     linear   GSS 14-SEP-2004
DEFINITION  OA_ABa0020N13.r OA_ABa Oryza australiensis genomic clone
            OA_ABa0020N13 3', genomic survey sequence.
ACCESSION   CL917654
VERSION     CL917654.1  GI:52029479
KEYWORDS    GSS.
SOURCE      Oryza australiensis
  ORGANISM  Oryza australiensis
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 720)
  AUTHORS   Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R.,
            Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
  TITLE     OMAP Project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0020  row: N  column: 13
            Seq primer: CAC TCA TTA GGC ACC CCA
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .720
                     /organism="Oryza australiensis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4532"
                     /clone="OA_ABa0020N13"
                     /tissue_type="young leaves"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OA_ABa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

  Query Match             5.8%;  Score 288;  DB 9;  Length 720;
  Best Local Similarity   75.2%;  Pred. No. 2.7e-39;
  Matches  385;  Conservative    0;  Mismatches  125;  Indels    2;  Gaps    2;

Qy        196 GAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACAC 255
              || || ||||   || ||||||  ||    | | ||  || |  ||   ||| |
Db        200 GACATATCATAGAAGGAAAAAGTTAAATTTTGTAATTAGTAAAACAATTGGAATACATTT 259

Qy        256 ACACACACACATATATATATACACATATATATACACATATATACATATATACACATATAT 315
              ||       |||||||||||| | ||||||||| | ||||||| ||||||| | ||||||
Db        260 ACTGTGTGGAATATATATATATATATATATATATATATATATATATATATATATATATAT 319

Qy        316 ACACATATATATACACATACATATGTACACATATATATACACATATGTATACACATATAT 375
              | | |||||||||| | ||| |||| || | ||||||||| | |||| |||| | ||||||
Db        320 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 379

Qy        376 ACACATATATACACACATATATACACATATATACACACATATATACACATATATACACAT 435
              | | ||||||| | | ||||||| | ||||||| | | ||||||| | ||||||| | ||
Db        380 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 439

Qy        436 ATATACACATATACACATATATACACATATATACATATATACACATATATATAATATACA 495
              ||||| | ||||| | |||||||| | |||||||| ||||||| | |||||||| ||||| |
Db        440 ATATATATATATATATATATATATATATATATATATATATATATATATATATATAT-ATATATA 498

Qy        496 CACATATATATACACATATATACACACATATATACACATATATACACATATATATACACA 555
              | |||||||||| | ||||||| | | ||||||| | ||||||| | |||||||||| | |
Db        499 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 558

Qy        556 TATATACACATATATACATATATACACATATATATACATATATACACATATATACATATA 615
              |||||| | ||||||| ||||||| | ||||||||| ||||||| | ||||||| |||||
Db        559 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 618

Qy        616 TACACATATATACATATATACACACATATATACACATACATATACACACACATAGATATA 675
              || | ||||||| |||||||    | ||||||| | ||| ||||| | |   ||| |||||
Db        619 TATATATATATATATATATATTTATATATATATATATATATATATATATTTATATATATA 678
```

```
Qy        676 CATATATATACACATATATATACGTATATATA 707
              |||| ||||   |||||||| ||||||||
Db        679 TATATGTATA-GTATATATATATGTATATATA 709


RESULT 71
BX157670
LOCUS       BX157670                668 bp    DNA     linear   GSS 28-JAN-2003
DEFINITION  Danio rerio genomic clone DKEY-125M16, genomic survey sequence.
ACCESSION   BX157670
VERSION     BX157670.1  GI:27989269
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 668)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the T7 end of BAC 125M16. 125M16
            is part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .668
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-125M16"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             5.8%;  Score 287.4;  DB 9;  Length 668;
  Best Local Similarity   77.5%;  Pred. No. 3.4e-39;
  Matches  373;  Conservative    0;  Mismatches  106;  Indels    2;  Gaps    2;

Qy        242 ATATATATATACACACACACACACATATATATATACACATATATATACACATATATACAT 301
              | | | ||  ||| | | | | | ||| ||| ||| |  |||||| || |  |||||| ||
Db        185 AAAAAAATGAACATAAATAAAGAAATAAATAAATAAATTTATATCTATATCTATATATAT 244

Qy        302 ATATACACATATATACACATATATATACACATACATATGTACACA-TATATATACACATA 360
              ||||| | |||||||| | ||||||||| | ||| |||| || ||| |||||| | ||| |
Db        245 ATATATATATATATATATATATATATATATATATATATATATACAGTATATACATACACA 304

Qy        361 TGTATACACATATATACACATATATACACACATATATACACATATATACACACATATATA 420
                || | |||||||| ||||| ||| | | ||||||| ||||| ||||| |||||| |||
Db        305 CATACATACATATATATACATACATATATATATATATATACATACATACATACATACATA 364

Qy        421 CACATATATACACATATATACACATATACACATATATACACATATATACATATATACACA 480
              || | ||| | ||||| ||||| | ||||| | ||| | ||||| ||||||| ||||| |
Db        365 CATACATACATACATACATACATACATACATACATACATACATACATACATACATACATA 424

Qy        481 TATATATA-ATATACACACATATATATACACATATATACACACATATATACACATATATA 539
              |||||||| |||||  | | |||||||| | ||||||| | | ||||||| | |||||||
Db        425 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 484

Qy        540 CACATATATATACACATATATACACATATATACATATATACACATATATATACATATATA 599
              | |||||||| | ||||||| | ||||||| ||||||| | | ||||||| |||||||
Db        485 TATATATATATATATATATATATATATATATATATATATATATATAGATATATATATATATA 544

Qy        600 CACATATATACATATATACACATATATACATATATACACACATATATACACATACATATA 659
              | ||||||| ||||||| | ||||||| ||||||| | | ||||||| | ||| |||||
Db        545 TATATATATATATATATATAGATATATATATATATATATATATATATATATATATATATA 604

Qy        660 CACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATA 719
              | | | ||| ||||| |||||||| | |||||||||  ||||| ||  ||||||||||
Db        605 TATATATATATATATATATATATATATATATATATATATATATATATGTAGATATATATATA 664
```

```
Qy        720 T 720
              |
Db        665 T 665


RESULT 72
BX149919
LOCUS       BX149919                737 bp    DNA     linear   GSS 13-MAR-2003
DEFINITION  Danio rerio genomic clone DKEY-112P5, genomic survey sequence.
ACCESSION   BX149919
VERSION     BX149919.1  GI:27981385
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 737)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAR-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the T7 end of BAC 112P5. 112P5 is
            part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .737
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-112P5"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             5.8%;  Score 287.4;  DB 9;  Length 737;
  Best Local Similarity   76.5%;  Pred. No. 3.4e-39;
  Matches  365;  Conservative    0;  Mismatches  111;  Indels    1;  Gaps    1;

Qy        242 ATATATATATACACACACACACACATATATATATACACATATATATACACATATATACAT 301
              | || | |||||        ||  | | | | | |      | | | |    | |  || ||
Db        262 AGATTTCTATACAGTTTGGCAACCTTTTCTGTTTTTTGGTTTTTTTTTTAAAAGCTATAT 321

Qy        302 ATATACACATATATACACATATATATACACATACATATGTACACATATATATACACATAT 361
              ||||| | ||||||| | ||||||||| | ||| |||| || | |||||||||| | ||||
Db        322 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 381

Qy        362 GTATACACATATATACACATATATACACACATATATACACATATATACACACATATATAC 421
              |||| | ||||||| | ||||||| | | ||||||| | ||||||| | | |||||||
Db        382 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 441

Qy        422 ACATATATACACATATATACACATATACACATATATACACATATATACATATATACACAT 481
              | ||||||| | ||||||| | ||||| | ||||||| | ||||||| |||||||| | ||
Db        442 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 501

Qy        482 ATATATAATATACACACATATATATACACATATATACACACATATATACACATATATACA 541
              ||||| |||||||| |||||||| | ||||||| | | ||||||| | ||||||| |
Db        502 ATATAT-ATATACACATATATATATATATATATATATATATATATATATATATATATATA 560

Qy        542 CATATATATACACATATATACACATATATACATATATACACATATATATACATATATACA 601
              |||||||| | ||||||| | ||||||| ||||||| | |||||||| ||||||| |
Db        561 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 620

Qy        602 CATATATACATATATACACATATATACATATATACACACATATATACACATACATATACA 661
              ||||||| ||||||| | ||||||| ||||||| | | ||||||| | ||| |||||
Db        621 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 680

Qy        662 CACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATAT 718
              | | ||| ||||||| ||||||| | |||||||||  |||||||| |||||||||
Db        681 TATATATATATATACACATATATATATATATATATATATATATATATATATATATATAT 737
```

```
RESULT 73
BX197183/c
LOCUS       BX197183                771 bp    DNA     linear   GSS 28-JAN-2003
DEFINITION  Danio rerio genomic clone DKEY-207A10, genomic survey sequence.
ACCESSION   BX197183
VERSION     BX197183.1  GI:28029069
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 771)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the SP6 end of BAC 207A10. 207A10
            is part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .771
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-207A10"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             5.8%;  Score 286.6;  DB 9;  Length 771;
  Best Local Similarity   78.1%;  Pred. No. 4.6e-39;
  Matches  357;  Conservative    0;  Mismatches   99;  Indels    1;  Gaps    1;

Qy        270 ATATATACACATATATATACACATATATACATATATACACATATATACACATATATATAC 329
              ||||||   | |||||||||| | ||||||| ||||||| | ||||||| | |||||||||
Db        771 ATATATGTATATATATATATATATATATATATATATATATATATATATATATATATAT 712

Qy        330 ACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACAC 389
              | ||| |||| || | |||||||||| | |||| |||| | ||||||| | ||||||| |
Db        711 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 652

Qy        390 ACATATATACACATATATACACACATATATACACATATATACACATATATACACATATAC 449
              | ||||||| | ||||||| | | ||||||| | ||||||| | ||||||| | |||||
Db        651 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 592

Qy        450 ACATATATACACATATATACATATATACACATATATATAATATACACACATATATATACA 509
              | ||||||| | ||||||| ||||||| | |||||||| ||||| | | |||||||||
Db        591 ATATATATATATATATATATATATATATATATATATAT-ATATATATATATATATATATA 533

Qy        510 CATATATACACACATATATACACATATATACACATATATATACACATATATACACATATA 569
               ||||||| | | ||||||| | ||||||| | ||||||||| | ||||||| | |||||
Db        532 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 473

Qy        570 TACATATATACACATATATATACATATATACACATATATACATATATACACATATATACA 629
              || ||||||| | ||||||||| ||||||| | ||||||| ||||||| | ||||||| |
Db        472 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 413

Qy        630 TATATACACACATATATACACATACATATACACACACATAGATATACATATATATACACA 689
              |||||| | | ||||||| | ||| ||||| | | | ||| ||||| |||||||||| | |
Db        412 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 353

Qy        690 TATATATACGTATATATATGTATATATATATGCTCCA 726
              ||||| |   | | | ||| ||| | | | | | | |
Db        352 TATATTTTTTTTTTTTTATTTATTTTTTTTTCCCCAA 316


RESULT 74
CW007564/c
LOCUS       CW007564                569 bp    DNA     linear   GSS 23-SEP-2004
DEFINITION  ZMMBLa0009D17.r ZMMBLa Zea mays genomic clone ZMMBLa0009D17 3',
            genomic survey sequence.
ACCESSION   CW007564
VERSION     CW007564.1  GI:52588327
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 569)
  AUTHORS   Wing,R., Luo,M., Soderlund,C. and Haller,K.
  TITLE     ZMMBL sequences
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            Plate: 0009  row: D  column: 17
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .569
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /cultivar="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBLa0009D17"
                     /tissue_type="immature ears"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="ZMMBLa"
                     /note="Vector: pAGIBAC1; Site_1: SalI; Site_2: SalI"
ORIGIN

  Query Match             5.8%;  Score 286;  DB 9;  Length 569;
  Best Local Similarity   77.3%;  Pred. No. 6.1e-39;
  Matches  360;  Conservative    0;  Mismatches  105;  Indels    1;  Gaps    1;

Qy        281 TATATATACACATATATACATATATACACATATATACACATATATATACACATACATATG 340
              ||| | || | |||| || ||||||| | |||||||| | ||||||||| | ||| ||||
Db        569 TATTTCTATATATATTTATATATATATATATATATATATATATATATATATATATATA 510

Qy        341 TACACATATATATACACATATGTATACACATATATACACATATATACACACATATATACA 400
              || | |||||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| |
Db        509 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 450

Qy        401 CATATATACACACATATATACACATATATACACATATATACACATATACACATATATACA 460
              ||||||| | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| |
Db        449 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 390

Qy        461 CATATATACATATATACACATATATATA-ATATACACACATATATATACACATATATACA 519
              ||||||| ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| |
Db        389 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 330

Qy        520 CACATATATACACATATATACACATATATATACACATATATACACATATATACATATATA 579
              | ||||||| | ||||||| | ||||||||| | ||||||| | ||||||| |||||||
Db        329 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 270

Qy        580 CACATATATATACATATATACACATATATACATATATACACATATATACATATATACACA 639
              | ||||||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | |
Db        269 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 210

Qy        640 CATATATACACATACATATACACACACATAGATATACATATATATACACATATATATACG 699
              ||||||| | ||| ||||| | | | ||| ||||| |||||||||| | |||||||||
Db        209 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 150

Qy        700 TATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCA 745
              ||||||||| | ||   |  | |   |   |  | ||| ||||||
Db        149 TATATATATTTGTAGCAACGAGGTTTTGTAGTTTTGAGATGTAGCA 104
```

```
RESULT 75
CL917654/c
LOCUS       CL917654               720 bp    DNA     linear   GSS 14-SEP-2004
DEFINITION  OA_ABa0020N13.r OA_ABa Oryza australiensis genomic clone
            OA_ABa0020N13 3', genomic survey sequence.
ACCESSION   CL917654
VERSION     CL917654.1  GI:52029479
KEYWORDS    GSS.
SOURCE      Oryza australiensis
  ORGANISM  Oryza australiensis
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 720)
  AUTHORS   Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R.,
            Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
  TITLE     OMAP Project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0020  row: N  column: 13
            Seq primer: CAC TCA TTA GGC ACC CCA
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .720
                     /organism="Oryza australiensis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4532"
                     /clone="OA_ABa0020N13"
                     /tissue_type="young leaves"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OA_ABa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

  Query Match             5.8%;  Score 286;  DB 9;  Length 720;
  Best Local Similarity   79.7%;  Pred. No. 5.9e-39;
  Matches  362;  Conservative    0;  Mismatches   90;  Indels    2;  Gaps    2;

Qy        283 TATATACACATATATACATA-TATACACATATATACACATATATATACACATACATATGT 341
              ||||||| ||||||||| ||| |||||| ||||||| | ||| ||||| | ||| |||| |
Db        709 TATATATACATATATATATACTATACATATATATATATATAAATATATATATATATATAT 650

Qy        342 ACACATATATATACACATATGTATACACATATATACACATATATACACACATATATACAC 401
              | | ||||| ||| | |||| |||| | ||||||| | ||||||| | | ||||||| |
Db        649 ATATATATAAATATATATATATATATATATATATATATATATATATATATATATATATAT 590

Qy        402 ATATATACACACATATATACACATATATACACATATATACACATATACACATATATACAC 461
              ||||||| | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| |
Db        589 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 530

Qy        462 ATATATACATATATACACATATATATA-ATATACACACATATATATACACATATATACAC 520
              ||||||| ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| |
Db        529 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 470

Qy        521 ACATATATACACATATATACACATATATATACACATATATACACATATATACATATATAC 580
              | ||||||| | ||||||| | ||||||||| | ||||||| | ||||||| |||||||
Db        469 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 410

Qy        581 ACATATATATACATATATACACATATATACATATATACACATATATACATATATACACAC 640
              | ||||||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | |
Db        409 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 350

Qy        641 ATATATACACATACATATACACACACATAGATATACATATATATACACATATATATACGT 700
              ||||||| | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  |
Db        349 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 290

Qy        701 ATATATATGTATATATATATGCTCCAGAGTTCAT 734
              |||||||| |||||||||| |  || |||  ||
Db        289 ATATATATATATATATATATTCCACACAGTAAAT 256


RESULT 76
CF924712
LOCUS       CF924712               760 bp    mRNA    linear   EST 10-NOV-2003
DEFINITION  DD005E11.ab1 Codonopsis lanceolata cDNA library Codonopsis
            lanceolata cDNA, mRNA sequence.
ACCESSION   CF924712
VERSION     CF924712.1  GI:38243228
KEYWORDS    EST.
SOURCE      Codonopsis lanceolata
  ORGANISM  Codonopsis lanceolata
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            asterids; campanulids; Asterales; Campanulaceae; Codonopsis.
REFERENCE   1  (bases 1 to 760)
  AUTHORS   Yang,D.C., In,J.G. and Lee,B.S.
  TITLE     Codonopsis lanceolata
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Deok Chun Yang
            BioPia
            1 Seochonri, Kihung, Yongin, Kyungki-do, 449-701, Korea
            Tel: 82 31 201 2688
            Fax: 82 31 205 6323
            Email: dcyang@ibiopia.com.
FEATURES             Location/Qualifiers
     source          1. .760
                     /organism="Codonopsis lanceolata"
                     /mol_type="mRNA"
                     /db_xref="taxon:103999"
                     /clone_lib="Codonopsis lanceolata cDNA library"
                     /note="Organ: root; Vector: pTriplEx2; Site_1: Sfi IA;
                     Site_2: Sfi IB"
ORIGIN

  Query Match             5.8%;  Score 285.8;  DB 7;  Length 760;
  Best Local Similarity   80.0%;  Pred. No. 6.3e-39;
  Matches  360;  Conservative    0;  Mismatches   88;  Indels    2;  Gaps    2;

Qy        268 ATATATATACACATATATATACACATATATACATATATACACATATATACACATATATAT 327
              ||||||||| | |||||||||  |||||||| |||||||  | |||||||| | ||||||||
Db        313 ATATATATATATATATATATATGCATATATATATATATATATATATATATATATATATAT 372

Qy        328 ACACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATAC 387
              | | ||| |||| || | ||||||||| | |||| |||| | ||||||| | |||||||
Db        373 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 432

Qy        388 ACACATATATACACATATATACACACATATATACACATATATACACATATATACACATAT 447
              | | ||||||| | ||||||| | | ||||||| | ||||||| | ||||||| ||||
Db        433 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATA-TGATAT 491

Qy        448 ACACATATATACACATATATACATATATACACATATATATAATATACACACATATATATA 507
              | | ||||||| | ||||||| ||||||| | |||||||| ||||| | | |||||||||
Db        492 ATATATATATATATATATATATATATATATATATATATAT-ATATATATATATATATA 550

Qy        508 CACATATATACACACATATATACACATATATACACATATATATACACATATATACACATA 567
              | ||||||| | | ||||||| | ||||||| | ||||||||| | ||||||| | |||
Db        551 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 610

Qy        568 TATACATATATACACATATATATACATATATACACATATATACATATATACACATATATA 627
              |||| ||||||| | ||||||||| ||||||| | ||||||| ||||||| | |||||||
Db        611 TATATATATATATATATATATATATATATATATATATATATATATATATATATAGATATATATA 670

Qy        628 CATATATACACACATATATACACATACATATACACACACATAGATATACATATATATACA 687
              ||||||| | | ||||||| | ||| ||||| | | | ||| ||||  |||||||||  |
Db        671 TATATATATATATATATATATATATATATATATATATATATATATATATATCTATATATATA 730
```

```
Qy        688 CATATATATACGTATATATATGTATATATA 717
              |||| ||||  || ||||||||||||||||
Db        731 TATATCTATATATACATATATGTATATATA 760


RESULT 77
BX145980/c
LOCUS       BX145980                709 bp    DNA     linear   GSS 28-JAN-2003
DEFINITION  Danio rerio genomic clone DKEY-104E13, genomic survey sequence.
ACCESSION   BX145980
VERSION     BX145980.1  GI:27977335
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 709)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the T7 end of BAC 104E13. 104E13
            is part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .709
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-104E13"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             5.7%;  Score 285.4;  DB 9;  Length 709;
  Best Local Similarity   74.0%;  Pred. No. 7.5e-39;
  Matches  375;  Conservative    0;  Mismatches  131;  Indels    1;  Gaps    1;

Qy        230 ATATATATATATATATATATATACACACACACACACATATATATATACACATATATATAC 289
              |||||||||||||||||||||||| | | | | | |||||||||||| | ||| |||||
Db        709 ATATATATATATATATATATATATATATATATATATATATATATATATATATAGATATAT 650

Qy        290 ACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATAT 349
              | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | || |
Db        649 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATGT 590

Qy        350 ATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATAC 409
              ||||| | |||| |||    ||||||| | ||||||| | | ||||||| | |||||||
Db        589 ATATATATATATATATGTGTATATATATATATATATATATATATATATATATATATATAT 530

Qy        410 ACACATATATACACATATATACACATATATACACATATACACATATATACACATATATAC 469
              | | ||||||| | ||||||| | ||||||| | ||||| |   ||| || | |||||||
Db        529 ATATATATATATATATATATATATATATATATATATATATATGTATGTATATATATATAT 470

Qy        470 ATATATACACATATATATAATATACACACATATATATACACATATATACACACATATATA 529
              |||||||   |||||| | ||||| | | ||||||||| | || |||| | | |||||||
Db        469 ATATATATGTATATATGT-ATATATATATATATATATATATGTATATATATATATATA 411

Qy        530 CACATATATACACATATATATACACATATATACACATATATACATATATACACATATATA 589
              | ||||||| | |||||||| | ||||||| | ||||||| |||||||    | ||||
Db        410 TATATATATATATATATATATATATATATATATATATATATATATATATGTGTGTATA 351

Qy        590 TACATATATACACATATATACATATATACACATATATACATATATACACACATATATACA 649
              || ||||||| | |||| |  ||||||| | ||||||   |||||| | | ||||||| |
Db        350 TATATATATATATATATGTGTATATATATATATATATGTGTATATATATATATATATATA 291

Qy        650 CATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATATG 709
              |||  ||||| | | | ||| ||||| |||||||||  |||| || |   | |
Db        290 TATATATATATATATATATATATATATATATATATATGCATAAATTTGACTTTTGCATGC 231
```

```
Qy        710 TATATATATATGCTCCAGAGTTCATAA 736
              || |  ||  || |  ||| ||  | |
Db        230 TAAAATTACTTGATTAAGATTTTGTCA 204


RESULT 78
BU507480/c
LOCUS       BU507480                1836 bp    mRNA    linear   EST 12-SEP-2002
DEFINITION  AGENCOURT_10309824 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:6501420
            5', mRNA sequence.
ACCESSION   BU507480
VERSION     BU507480.1  GI:22813713
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1836)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM14055  row: n  column: 13
            High quality sequence start: 74
            High quality sequence stop: 447.
FEATURES             Location/Qualifiers
     source          1. .1836
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6501420"
                     /tissue_type="leiomyosarcoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_71"
                     /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 2.1 kb.  "
ORIGIN

  Query Match             5.7%;  Score 284.6;  DB 5;  Length 1836;
  Best Local Similarity   75.5%;  Pred. No. 9.2e-39;
  Matches  361;  Conservative    0;  Mismatches  116;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | | | | | | | | | | | |||||||||||||||||| |         | | | | |
Db        548 TCTNTTTCTCTNTTTCTTTTTTTGTACACACACACACACACTCTCTCTCTCTTTTTTC 489

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | | | | ||| | | ||| ||||||| | |||||||||| ||| |||| || | |
Db        488 TATATACACACACACACACATATATATANATATATATATACATATATATATATATATA 429

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        428 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 369

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db        368 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 309

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db        308 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 249

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
```

```
Db        248 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 189

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db        188 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 129

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATA 703
              || | ||| ||||| | | | ||| ||||| |||||||||| | ||| || | ||   |
Db        128 TATATATATATATATATATATATATATATATATATATATATATATANATNTCCGNGNA 71


RESULT 79
CA364719/c
LOCUS       CA364719                683 bp    mRNA    linear   EST 06-NOV-2002
DEFINITION  639681 NCCCWA 1RT Oncorhynchus mykiss cDNA clone 1RT70B09_B_A05 5',
            mRNA sequence.
ACCESSION   CA364719
VERSION     CA364719.1  GI:24674188
KEYWORDS    EST.
SOURCE      Oncorhynchus mykiss (rainbow trout)
  ORGANISM  Oncorhynchus mykiss
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei;
            Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
REFERENCE   1  (bases 1 to 683)
  AUTHORS   Rexroad,C.E. 3rd, Lee,Y., Keele,J.W., Karamycheva,S., Brown,G.,
            Koop,B., Gahr,S.A., Palti,Y. and Quackenbush,J.
  TITLE     Sequence analysis of a rainbow trout cDNA library and creation of a
            gene index
  JOURNAL   Cytogenet. Genome Res. 102 (1-4), 347-354 (2003)
COMMENT     Contact: Rexroad CE
            USDA, ARS, National Center for Cool and Cold Water Aquaculture
            11876 Leetown Road, Kearneysville, WV 25430, USA
            Tel: 304 724 8340 x2129
            Fax: 304 725 0351
            Email: crexroad@ncccwa.ars.usda.gov
            Single pass sequencing. Bases called with phred v0.020425.c and
            trimmed with the aid of the trim_alt option. Vector identified by
            cross_match v0.990329.
            Seq primer: AGCGGATAACAATTTCACACAGGA.
FEATURES             Location/Qualifiers
     source          1. .683
                     /organism="Oncorhynchus mykiss"
                     /mol_type="mRNA"
                     /db_xref="taxon:8022"
                     /clone="1RT70B09_B_A05"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /clone_lib="NCCCWA 1RT"
                     /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
                     Library made from pooled tissue from brain, gill, liver,
                     spleen, muscle, and kidney."
ORIGIN

  Query Match             5.7%;  Score 282.2;  DB 6;  Length 683;
  Best Local Similarity   74.5%;  Pred. No. 2.7e-38;
  Matches  369;  Conservative    0;  Mismatches  123;  Indels    3;  Gaps    1;

Qy        229 TATATATATATATATATATATATATACACACACACACACATATATATATACACATATATATA 288
              |||||||||||||| |||||||||| ||| ||| | |||||||| ||    | ||||||
Db        674 TATATATATATATACATATATATACATACATACATATATATATATGTATGTGTGTATATA 615

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
                || | || |||| ||  | |||||| ||| || ||||||||| || |   | |||||
Db        614 TGTATGTGTATATATTTATGCGTATATATACACATGTATACACATGCACACACATACATA 555

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              | |||  | |||||||| |||||||| | | | |||   |  |||||||| |||||||
Db        554 TGTATGTATATATGTATATACATATATATATACACATATGTATGTATATACATATATATA 495

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              ||||||  |||  | ||| ||| | ||||||| ||| | |||||  |||| | |||||||
Db        494 CACACACGTATGTATATACATATATATATATACACACACACGTATATATATATATA 435

Qy        469 CATATATACACATATATATA---ATATACACACATATATATACACATATATACACACATA 525
              ||| |   ||||||| |||   ||||| | | ||| | |||||||  |||| | | |||
Db        434 TATACACGTACATATACATACACATATATATATATACACATACACACGTATATATATATA 375

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||||||||| ||| | ||||||||||   | |||    ||| ||| ||||||| ||| |
Db        374 TATACACATACATATATATATATATACGTGTGTATGTGTATACATATATATATATACACA 315

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              || | || |||| | | |||||||| ||||| ||  |||||||| ||||| | | | |||||
Db        314 TACACACGTATACATATATATATATATATACACGTATATATATATATACATATATATATA 255

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              |||||||||| |  |||||||||| ||||| |||||||||   || | ||||  ||||||
Db        254 TACACATACACACGCACACACATATATATATATATATATATGTATGTGTATATATATATA 195

Qy        706 TATGTATATATATAT 720
              |||  ||||  |||
Db        194 TATACATATGCGTAT 180


RESULT 80
CW518121
LOCUS       CW518121                663 bp    DNA     linear   GSS 08-OCT-2004
DEFINITION  OP__Ba0025C13.f OP__Ba Oryza punctata genomic clone OP__Ba0025C13
            5', genomic survey sequence.
ACCESSION   CW518121
VERSION     CW518121.1  GI:53992343
KEYWORDS    GSS.
SOURCE      Oryza punctata
  ORGANISM  Oryza punctata
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 663)
  AUTHORS   SanMiguel,P., Westerman,R., Kim,H., Yu,Y., Wissotski,M., Yost,D.,
            Stum,D., Rao,K., Luo,M., Jetty,R., Kudrna,D., Muller,C.,
            Hatfield,J., Soderlund,C., Wing,R. and Jackson,S.A.
  TITLE     OMAP Project - Purdue University
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Scott A. Jackson
            Jackson Laboratory
            Purdue University
            915 W. State St., West Lafayette, IN 47907, USA
            Tel: 7654963621
            Fax: 7654967255
            Email: sjackson@purdue.edu
            Basecalling by phred version 0.020425.c. This sequence was derived
            from the raw sequence read by clipping with lucy version 1.19s.
            Bases 50-712 of the raw sequence (length 1364) were retained after
            clipping.
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Insert Length: 161000   Std Error: 0.00
            Plate: 0025  row: C  column: 13
            Seq primer: TAA TAC GAC TCA CTA TAG GG
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .663
                     /organism="Oryza punctata"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4537"
                     /clone="OP__Ba0025C13"
                     /tissue_type="young leaves"
                     /lab_host="DH10B-T1 phage resistant"
                     /clone_lib="OP__Ba"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

  Query Match             5.6%;  Score 280.4;  DB 9;  Length 663;
  Best Local Similarity   77.2%;  Pred. No. 5.5e-38;
```

```
  Matches  366;  Conservative    0;  Mismatches  106;  Indels    2;  Gaps    2;

Qy        208 CAGAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACACAT 267
              ||  ||||||||||     |||     ||  |   ||  | | | |||||  | | | | | | ||
Db        191 CAAAAAAAAAAAATCCAATTCACCATGTCAATGCAAAAAGATATATATATATATATATAT 250

Qy        268 ATATATATACACATATATATACACATATATACATATATACACATATATACACATATATAT 327
              |||||||||  | |||||||||  | |||||||  |||||||  | |||||||  | ||||||||
Db        251 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 310

Qy        328 ACACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATAC 387
              |  | |||  ||||  ||  |  |||||||||  |  ||||  ||||  |  |||||||  |  |||||||
Db        311 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 370

Qy        388 ACACATATATACACATATATACACACATATATACACATATATACACATATATACACATAT 447
              |  |  |||||||  |  |||||||  |  |  |||||||  |  |||||||  |  |||||||  |  |||||||
Db        371 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 430

Qy        448 ACACATATATACACATATATACATATATACACATATATATA-ATATACACACATATATAT 506
              |  |  |||||||  |  |||||||  |||||||  |  |||||||||  |||||  |  |  |||||||||
Db        431 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 490

Qy        507 ACACATATATACACACATATATACACATATATACACATATATATACACATATATACACAT 566
              |  |  |||||||  |  |  |||||||||||||||  |  |  |||||  |||  |  |||||  |  |  |
Db        491 ATATATATATATATATATATATATACACATATAAATAAATATAGATAGATATATA-ATATAA 549

Qy        567 ATATACATATATACACATATATATACATATATACACATATATACATATATACACATATAT 626
              |||||  |||||||  |  |||||||||  |  |||||||  |||||||  |  |||||||
Db        550 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 609

Qy        627 ACATATATACACACATATATACACATACATATACACACACATAGATATACATAT 680
              |  |||||||  |  |  |||||||  |  |||  |||||  |  |  |||  |||||  ||||
Db        610 ATATATATATATATATATATATATATATATATATATATATATATATATATATATAAATAT 663


RESULT 81
BX149919/c
LOCUS       BX149919                737 bp    DNA     linear   GSS 13-MAR-2003
DEFINITION  Danio rerio genomic clone DKEY-112P5, genomic survey sequence.
ACCESSION   BX149919
VERSION     BX149919.1  GI:27981385
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 737)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAR-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the T7 end of BAC 112P5. 112P5 is
            part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .737
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-112P5"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             5.6%;  Score 280;  DB 9;  Length 737;
  Best Local Similarity   75.6%;  Pred. No. 6.3e-38;
  Matches  360;  Conservative    0;  Mismatches  115;  Indels    1;  Gaps    1;

Qy        230 ATATATATATATATATATATATACACACACACACACATATATATATACACATATATATAC 289
              |||||||||||||||||||||||  | | | | | ||||  | ||||  | |||||||||

Db        737 ATATATATATATATATATATATATATATATATATATATATATGTGTATATATATATATATAT 678

Qy        290 ACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATAT 349
              |  |||||||  |||||||  |  |||||||  |||||||||  |  |||  ||||  ||  |  ||||
Db        677 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 618

Qy        350 ATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATAC 409
              |||||  |  ||||  ||||  |  |||||||  |  |||||||  |  |  |||||||  |  |||||||
Db        617 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 558

Qy        410 ACACATATATACACATATATACACATATATACACATATACACATATATACACATATATAC 469
              |  |  |||||||  |  |||||||  |  |||||||  |  |||||  |  |  ||||  |  |||||||
Db        557 ATATATATATATATATATATATATATATATATATATATATATATATGTGTATATATATATAT 498

Qy        470 ATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATATAT 528
              |||||||  |  |||||||||  |||||  |  |  |||||||||  |  |||||||  |  |  |||||||
Db        497 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 438

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACACATATAT 588
              |  |  |||||||  |  |||||||||  |  |||||||  |  |||||||  |||||||  |  |||||||
Db        437 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 378

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
              |||  |||||||  |  |||||||  |||||||  |  |||||||  |||||||  |  |  |||||||
Db        377 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 318

Qy        649 ACATACATATATACACACACATAGATATACATATATATACACATATATATACGTATAT 704
              |  |  |  |  |  |  |  ||  |  |  |  |  |  |  |  |  |  ||||  ||  |
Db        317 AGCTTTTAAAAAAAAAACCAAAAAACAGAAAAGGTTGCCAAACTGTATAGAAATCT 262


RESULT 82
CL044834
LOCUS       CL044834                1031 bp    DNA     linear   GSS 31-DEC-2003
DEFINITION  CH216-61O18_RM1.1 CH216 Xenopus tropicalis genomic clone
            CH216-61O18, genomic survey sequence.
ACCESSION   CL044834
VERSION     CL044834.1  GI:40500747
KEYWORDS    GSS.
SOURCE      Xenopus tropicalis (western clawed frog)
  ORGANISM  Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 1031)
  AUTHORS   Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
            Mardis,E. and Wilson,R.
  TITLE     A physical map of the xenopus tropicalis genome
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 175000   Std Error: 0.00
            Seq primer: RM1 TACGACTCACTATAGGGAGA
            Class: BAC ends
            High quality sequence start: 47
            High quality sequence stop: 682.
FEATURES             Location/Qualifiers
     source          1. .1031
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /strain="Nigerian frog"
                     /db_xref="taxon:8364"
                     /clone="CH216-61O18"
                     /sex="male"
                     /cell_line="Stock 248 F7A2, inbred N7"
                     /clone_lib="CH216"
                     /note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis
                     BAC library"
ORIGIN

  Query Match             5.6%;  Score 280;  DB 9;  Length 1031;
```

```
  Best Local Similarity   73.4%;  Pred. No. 6.1e-38;
  Matches  373;  Conservative    0;  Mismatches  130;  Indels    5;  Gaps    1;

Qy        215 AAAAAAAGGCATTTTATATATATATATATATATATACACACACACACACATATATATA 274
              |||| |       |  ||| | | | | ||| ||| |||||||||| ||| ||||| ||| |
Db        203 AAAATACATACATACATACACACACACATACATACATACACACATACATACATACATACA 262

Qy        275 TACACATATATATACACATATATACATATATACACATATATACACATATATATACACATA 334
              |||| | ||| ||||| | ||| | | ||| ||||| ||| ||||||| ||| | ||| |
Db        263 TACATACATACATACATACATAAAGACATACACACACATACACACATACATACATACACA 322

Qy        335 CATATGTACACATATATATACACATATGTATACACATATATACACATATATACACACATA 394
              ||||  |||||||| ||| | |||||  || | ||| | ||||||| ||| | ||| | |
Db        323 CATACATACACATACATACATACATACATACACACACATACACACATACATACATACA 382

Qy        395 TATACACATATATACACACATATATACACATATATACACATATATACACATATACACATA 454
              | | ||||| ||||| ||||| ||||| | | | | ||||| ||||||| ||||| | |
Db        383 CACATACATACATACAGACATACATACATACACACATACATACATACACACATACATACA 442

Qy        455 TATACACATATATACATATATACACATATATATAATATACACACATATATATACACATAT 514
              || ||||||| ||||||| | ||||| ||     | | ||| ||||| ||| |||||||
Db        443 TACACACATACATACATACACACACACAT-----ACACACATACATACATACACACATAC 497

Qy        515 ATACACACATATATACACATATATACACATATATATACACATATATACACATATATACAT 574
              |||||||| | ||||| | ||| | ||||| ||| | ||||| ||||||| ||| | |
Db        498 ATACACACACACATACATACATACATACATACATACATACATACACACACATACATAC 557

Qy        575 ATATACACATATATATACATATATACACATATATACATATATACACATATATACATATAT 634
              ||| |||||||| ||| | | ||| ||||||| ||||||| | ||| | ||| | | |||
Db        558 ATACACACATACATACATACATACACACATACATACATACAGACATACATACATACATAC 617

Qy        635 ACACACATATATACACATACATATACACACACATAGATATACATATATATACACATATAT 694
              | ||||| ||| |||||||||||| | ||| | | | ||| ||| ||| ||||||| |||
Db        618 ATACACACATACACACATACATATACATACATAAAGACATACACACATACATACACACATAC 677

Qy        695 ATACGTATATATATGTATATATATATGC 722
              |||| || ||  ||  ||| | | || |
Db        678 ATACATACATTCATACATACACACATAC 705


RESULT 83
AQ506817/c
LOCUS       AQ506817                537 bp    DNA     linear   GSS 29-APR-1999
DEFINITION  RPCI-11-281J17.TV RPCI-11 Homo sapiens genomic clone
            RPCI-11-281J17, genomic survey sequence.
ACCESSION   AQ506817
VERSION     AQ506817.1  GI:4711564
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 537)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: RPCI-11-281J17.TJ
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genet cs (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: T7
            Class: BAC ends.
FEATURES            Location/Qualifiers
     source         1. .537
                    /organism="Homo sapiens"
                    /mol_type="genomic DNA"
                    /db_xref="GDB:7607752"
                    /db_xref="taxon:9606"
                    /clone="RPCI-11-281J17"
                    /sex="Male"
                    /cell_type="Lymphocytes"
                    /clone_lib="RPCI-11"
                    /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                    RPCI11 Human Male BAC Library"
ORIGIN

  Query Match             5.6%;  Score 278.8;  DB 8;  Length 537;
  Best Local Similarity   76.0%;  Pred. No. 1.1e-37;
  Matches  357;  Conservative    0;  Mismatches  112;  Indels    1;  Gaps    1;

Qy        230 ATATATATATATATATATATATACACACACACACATATATATATACACATATATATAC 289
              |||||||||||||||||||| ||| | | ||| | | | ||||||||| | ||||| |||
Db        530 ATATATATATATATATATACATATATACATATATACATATATATAAATATATACATAT 471

Qy        290 ACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATAT 349
              | ||||| | ||||||| | ||||| | | ||||| ||| | |||||||| || | ||||
Db        470 ATATATAAATATATATATATATATAAATATATATAAATATATATACATATATAAATATAT 411

Qy        350 ATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATAC 409
              ||||| | |||| || | | ||| ||| | ||||||| | | | ||||| | |||||||
Db        410 ATATAAATATATATAAATATATACATATATATATATATATATAAATATATAAATATATAT 351

Qy        410 ACACATATATACACATATATACACATATATACACATATACACATATATACACATATATAC 469
              | | ||||||||| |||||| || ||||||| | ||||||| ||||||| | |||||||
Db        350 ATAAATATATACATATATATCCATATATATAAATATATACATATATATAAATATATATAA 291

Qy        470 ATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATATAT 528
              ||||||| | ||||||| | ||||| ||| ||||  ||| | ||  ||| | | | ||||
Db        290 ATATATAAATATATATAAATATATATACATATATCAATATATATCAATATATATAAATAT 231

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACATATAT 588
              | |   ||||| | |||||||| | ||||||| | ||||||| ||||||| | ||||||
Db        230 ATATCAATATATAAATATATATAAATATATATAAATATATATAAATATATAAATATATAT 171

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
              | | ||||||| | ||||||| ||||||| | ||||||| | ||||| | | ||||  |
Db        170 ACATATATATAAATATATATAAATATATAAATATATATATAAATATAAATATATATCAAT 111

Qy        649 ACATACATATACACACACATAGATATACATATATATACACATATATATAC 698
              | ||||||||| | | | ||| ||||| |||    | ||  ||||| | |
Db        110 ATATACATATATATATAAATATATATATATACTGTTTCATTTATATCTGC 61


RESULT 84
BX159146
LOCUS       BX159146                789 bp    DNA     linear   GSS 28-JAN-2003
DEFINITION  Danio rerio genomic clone DKEY-152P14, genomic survey sequence.
ACCESSION   BX159146
VERSION     BX159146.1  GI:27990759
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 789)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the SP6 end of BAC 152P14. 152P14
            is part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
```

```
          http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .789
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-152P14"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             5.6%;  Score 278.8;  DB 9;  Length 789;
  Best Local Similarity   75.5%;  Pred. No. 1e-37;
  Matches  373;  Conservative    0;  Mismatches  117;  Indels    4;  Gaps    2;

Qy        213 AAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACACATATATA 272
              ||| || |   |||||||||||||||| | | | ||| |   |||  |||||||| |||||
Db        300 AAATAAGATACCATTTTATATATATACACACACATACA-GTACATGCACACACACATATA 358

Qy        273 TATACACATATATATACACATATATACATATATACACATATATACACATATATATACACA 332
              |||||||| | | | |||||||||||| |||| |||  ||||| | | |||||| |
Db        359 TATACACACACACAAACACATATATATATTTACGTATATACATATATATATGTGTGTG 418

Qy        333 TACATATGTACACATATATATACACATATGTATACACATATATACACATATATACACACA 392
              || || ||||    ||| |||| | |||| |||| | |||||||| | ||||| |||||||
Db        419 TATATGTGTATGTGTATGTATATATATATATATATATATATATATATATACACACACA 478

Qy        393 TATATACACATATATACACACATATATACACATATATACACATATATACACATATACACA 452
              | ||| | ||||||| | | ||||||| | ||||||| | ||||||| ||| | |||||
Db        479 CACATATATATATATATATATATATATATATATATATATATATATATATACACACACACA 538

Qy        453 TATATACACATATATACATATATACACATATATATAATATACACACATATATATACACAT 512
              |||||| | ||||||| |||| || | ||||||   ||||| | | |||||||||| | ||
Db        539 TATATATATATATATATATATGTATATATATAT---ATATATATATATATATATATATAT 595

Qy        513 ATATACACACATATATACACATATATACACATATATATACACATATATACACATATATAC 572
              ||||| | | |||||||||| | | ||| ||||||||| | |||||||| | |||||||
Db        596 ATATATATATATATATATACACACACACATATATATATATATATATATATATATATAT 655

Qy        573 ATATATACACATATATATACATATATACACATATATACATATATACACATATATACATAT 632
              ||||||| | ||||||||| ||||| ||||| | | ||| | | ||||| | | ||| |
Db        656 ATATATATATATATATATATATATATACACACACACACACACACACACACACACACAC 715

Qy        633 ATACACACATATATACACATACATATACACACACATAGATATACATATATATACACATAT 692
              ||| | | ||||||| | ||| ||||| | | |   |||||||||| ||| | ||||
Db        716 ATATATATATATATATATATATATATATATATATGTGTATATACATATACATATATATAT 775

Qy        693 ATATACGTATATAT 706
              |||||  |||||||
Db        776 ATATATATATATAT 789


RESULT 85
BX240321
LOCUS       BX240321                446 bp    DNA     linear   GSS 29-JAN-2003
DEFINITION  Danio rerio genomic clone DKEY-276D22, genomic survey sequence.
ACCESSION   BX240321
VERSION     BX240321.1  GI:28162655
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 446)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the SP6 end of BAC 276D22. 276D22
            is part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
```

```
          http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .446
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-276D22"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             5.6%;  Score 277.6;  DB 9;  Length 446;
  Best Local Similarity   79.5%;  Pred. No. 1.7e-37;
  Matches  353;  Conservative    0;  Mismatches   89;  Indels    2;  Gaps    2;

Qy        268 ATATATATACACATATATATACACATATATACATATATACACATATATACACATATATAT 327
              ||||||||| | |||||||||  | ||||||| ||||||| | ||||||| | ||||||||
Db          4 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 63

Qy        328 ACACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATAC 387
              | | ||| | || |||| | ||| ||| | |||  |||| | ||||||| |||||||||
Db         64 ATATATATAAATATACATA-ATAAATAGAAATAGTTATATATATATATAGACATATATAT 122

Qy        388 ACACATATATACACATATATACACACATATATACACATATATACACATATATACACATAT 447
              | | ||||||| | | ||||| | | ||||||| | ||||| | | | ||||| | |||
Db        123 ATATATATATATATAGATATATATATATATATATATATATAGATAGACATATATATATAG 182

Qy        448 ACACATATATACACATATATACATATATACACATATATATA-ATATACACACATATATAT 506
              | | ||||||| | ||||||| ||||||| | ||||||||| ||||| | | |||||||||
Db        183 ATATATATATATATATATATATATATATATATATATATATAGATATATATATATATATAT 242

Qy        507 ACACATATATACACACATATATACACATATATACACATATATATACACATATATACACAT 566
              | | ||||||| | | ||||||| | ||||||| | ||||||||| | ||||||| | ||
Db        243 ATATATATATATATATATATATATATATATATATAGATATATATATATATATATATATAT 302

Qy        567 ATATACATATATACACATATATATACATATATACACATATATACATATATACACATATAT 626
              ||||| ||||||| | ||||||||| ||||||||| ||||||| ||||||| | ||||||
Db        303 ATATATATATATATATATATATATATATATATATACATATATATATATATATATATATAT 362

Qy        627 ACATATATACACACATATATACACATACATATACACACACATAGATATACATATATATAC 686
              | ||||||| | | ||||||| | ||| ||||| | | |||||  ||||| |||||||||
Db        363 ATATATATATATATATATATATATATATATATATATATATACATATATATATATATATAT 422

Qy        687 ACATATATATACGTATATATATGT 710
              | |||||||||  |||||||| |
Db        423 ATATATATATATATATATATATAT 446


RESULT 86
BX241290
LOCUS       BX241290                722 bp    DNA     linear   GSS 29-JAN-2003
DEFINITION  Danio rerio genomic clone DKEY-282A22, genomic survey sequence.
ACCESSION   BX241290
VERSION     BX241290.1  GI:28163624
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 722)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the T7 end of BAC 282A22. 282A22
            is part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .722
                     /organism="Danio rerio"
```

```
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-282A22"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             5.6%;  Score 276.6;  DB 9;  Length 722;
  Best Local Similarity   77.6%;  Pred. No. 2.4e-37;
  Matches  350;  Conservative    0;  Mismatches   94;  Indels    7;  Gaps    1;

Qy        250 ATACACACACACACACATATATATATACACATATATATACACATATATACATATATACAC 309
              ||||| |     |   |||| |||||| | |||||||| | ||||||| ||||||| |
Db         13 ATACATATTATTATTTATATTTATATATATATATATATATATATATATATATATATAT 72

Qy        310 ATATATACACATATATATACACATACATATGTACACATATATATACACATATGTATACAC 369
              ||||||| ||||||| ||||||| | | || |||| ||||||| | ||| || |||||||
Db         73 ATATATATACATATACATACACACATACATATACATATATATACATACACATATATACAC 132

Qy        370 ATATATACACATATATACACACATATATACACATATATACACACATATATACACATATAT 429
              ||| ||| ||||| ||| | | | ||| | ||||| ||||| ||||| ||||| | ||
Db        133 ATACATATACATACATATATATACATACATACATACATACATACATACATACATACATGC 192

Qy        430 ACACATATATACACATATACACATATATACACATATATACATATATACACATATATATA- 488
              | ||||  ||||| ||||| | |||||||||||||| | ||||| ||| ||||||||||
Db        193 ATACATGCATACATATATATATATATATACACATACACACATACATATACATATATATAC 252

Qy        489 ------ATATACACACATATATATACACATATATACACACATATATACACATATATACAC 542
                    ||||| ||||||| ||||||| | ||||| | | ||||| | ||||| |||||
Db        253 ATACACATATATACACATACATATACATACATATATATATATATACATACATACATACAT 312

Qy        543 ATATATATACACATATATACACATATATACATATATACACATATATATACATATATACAC 602
              | ||| ||||| | ||| | ||||| ||||||| ||||| | ||| ||||||| ||| |
Db        313 ACATACATACATACATACATACATACATACATACATACATACATACATACATACATATAT 372

Qy        603 ATATATACATATATACACATATATACATATATACACACATATATACACATACATATACAC 662
              |||||||||||||||||||| ||||||||| | | | ||||||||||||||||||| | ||
Db        373 ATATATACATATATACACATACATACATATACATATATATATATACACATACATACATAC 432

Qy        663 ACACATAGATATACATATATATACACATATA 693
              | ||||| ||| | | ||| ||||| | |||
Db        433 ATACATACATACATACATACATACATACATA 463


RESULT 87
CL763005/c
LOCUS       CL763005                617 bp    DNA     linear   GSS 27-JUL-2004
DEFINITION  OR_BBa0133F16.f OR_BBa Oryza rufipogon genomic clone OR_BBa0133F16
            5', genomic survey sequence.
ACCESSION   CL763005
VERSION     CL763005.1  GI:50720952
KEYWORDS    GSS.
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 617)
  AUTHORS   Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R.,
            Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
  TITLE     OMAP Project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Insert Length: 161   Std Error: 0.00
            Plate: 0133  row: F  column: 16
            Seq primer: TAA TAC GAC TCA CTA TAG GG
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .617
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
                     /clone="OR_BBa0133F16"
                     /tissue_type="young leaves"
                     /lab_host="DH10B-T1 phage resistant"
                     /clone_lib="OR_BBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

  Query Match             5.6%;  Score 276;  DB 9;  Length 617;
  Best Local Similarity   73.6%;  Pred. No. 3.1e-37;
  Matches  365;  Conservative    0;  Mismatches  130;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | |||| ||| |||||||||||| || |   | | | |  ||| ||| | || |||||||
Db        572 TATATACATAACTATATATATATCTATATGTATATATATCTATGTATCTGCATATATATA 513

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | |||| |   | ||||   |||||||    | | ||||    || |||| |  |
Db        512 TATATATATCTTTTTTTATATTTATATATATCTCTCTCTATATTTTTATATATATGTATC 453

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | | || |||      |||||| | ||||||| | | ||||||| | |||||
Db        452 TATATATATAAAAATATATTTTTTTTATATATATATATATATATATATATATAAATATA 393

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||  | | | ||| |   ||||| | ||       ||||||| | |||||
Db        392 TATATATATATATTAAAAAAAATAAAATAATATATATATTATATAATATATATATATATA 333

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || || |||| | || |||||| ||||| | | ||||||||| | ||||||| | | |||
Db        332 TATATTTATATATATTTATATATATATATATATATATATATATATATATATATATATATA 273

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db        272 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 213

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db        212 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 153

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db        152 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 93

Qy        706 TATGTATATATATATG 721
              ||| ||||||||||||
Db         92 TATATATATATATATG 77


RESULT 88
CL800151
LOCUS       CL800151                715 bp    DNA     linear   GSS 06-AUG-2004
DEFINITION  OR_CBa0011K05.r OR_CBa Oryza rufipogon genomic clone OR_CBa0011K05
            3', genomic survey sequence.
ACCESSION   CL800151
VERSION     CL800151.1  GI:51024771
KEYWORDS    GSS.
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 715)
  AUTHORS   Kim,H., Yu,Y., Wissotski,M., Yost,D., Stum,D., Rao,K., Luo,M.,
            Jetty,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and
            Wing,R.
```

```
  TITLE     OMAP project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0011  row: K  column: 05
            Seq primer: CAC TCA TTA GGC ACC CCA
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .715
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
                     /clone="OR_CBa0011K05"
                     /tissue_type="young leaves"
                     /dev_stage="2 week old seedlings"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OR_CBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII;
                     drk treated 36 hrs before harvest"
ORIGIN

  Query Match             5.5%;  Score 275.4;  DB 9;  Length 715;
  Best Local Similarity   80.3%;  Pred. No. 3.9e-37;
  Matches  335;  Conservative    0;  Mismatches   81;  Indels    1;  Gaps    1;

Qy        303 TATACACATATATACACATATATATACACATACATATGTACACATATATATACACATATG 362
              ||||  |||||||| | ||||||||| | ||| |||| || | ||||||||| | ||||
Db        300 TATAATCATATATATATATATATATATATATATATATATATATATATATATATATATA 359

Qy        363 TATACACATATATACACATATATACACACATATATACACATATATACACACATATATACA 422
              |||| | ||||||| | ||||||| | | ||||||| | ||||||| | | ||||||| |
Db        360 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 419

Qy        423 CATATATACACATATATACACATATACACATATATACACATATATACATATATACACATA 482
              ||||||| | ||||||| | ||||| | ||||||| | ||||||| ||||||| | |||
Db        420 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 479

Qy        483 TATATAATATACACACATATATATACACATATATACACACATATATACACATATATACAC 542
              ||||| ||||| | | ||||||||| | ||||||| | | ||||||| | ||||||| |
Db        480 TATAT-ATATATATATATATATATATATATATATATATATATATATATATATATATAT 538

Qy        543 ATATATATACACATATATACACATATATACATATATACACATATATATACATATATACAC 602
              ||||||||| | ||||||| | ||||||| ||||||| | ||||||||| ||||||| |
Db        539 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 598

Qy        603 ATATATACATATATACACATATATACATATATACACACATATATACACATACATATACAC 662
              ||||||| ||||||| | ||||||| ||||||| | | ||||||| | ||| ||||| |
Db        599 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 658

Qy        663 ACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATA 719
              | | ||| ||||| ||||||||| | |||||||||  ||||||||| ||||||||||
Db        659 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 715


RESULT 89
CL800151/c
LOCUS       CL800151                715 bp    DNA     linear   GSS 06-AUG-2004
DEFINITION  OR_CBa0011K05.r OR_CBa Oryza rufipogon genomic clone OR_CBa0011K05
            3', genomic survey sequence.
ACCESSION   CL800151
VERSION     CL800151.1  GI:51024771
KEYWORDS    GSS.
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 715)
  AUTHORS   Kim,H., Yu,Y., Wissotski,M., Yost,D., Stum,D., Rao,K., Luo,M.,
            Jetty,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and
            Wing,R.
  TITLE     OMAP project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0011  row: K  column: 05
            Seq primer: CAC TCA TTA GGC ACC CCA
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .715
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
                     /clone="OR_CBa0011K05"
                     /tissue_type="young leaves"
                     /dev_stage="2 week old seedlings"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OR_CBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII;
                     drk treated 36 hrs before harvest"
ORIGIN

  Query Match             5.5%;  Score 275.4;  DB 9;  Length 715;
  Best Local Similarity   80.3%;  Pred. No. 3.9e-37;
  Matches  335;  Conservative    0;  Mismatches   81;  Indels    1;  Gaps    1;

Qy        229 TATATATATATATATATATATATATACACACACACACACATATATATATACACATATATATA 288
              |||||||||||||||||||||||| | | | | | ||||||||||| | |||||||||
Db        715 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 656

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |||
Db        655 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 596

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              |||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||||
Db        595 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 536

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||||
Db        535 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 476

Qy        469 CATATATACACATATATATAATATACACACATATATATACACATATATACACACATATAT 528
              ||||||| | |||||||| ||||| | | ||||||||| | ||||||| | | ||||||
Db        475 TATATATATATATATATAT-ATATATATATATATATATATATATATATATATATATAT 417

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACATATAT 588
              | | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | ||||||
Db        416 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 357

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              ||| ||||||| | ||||||| ||||||| | ||||||| ||||||| |    ||||
Db        356 ATATATATATATATATATATATATATATATATATATATATATATATATATGATTATA 300


RESULT 90
BX158149
LOCUS       BX158149                705 bp    DNA     linear   GSS 28-JAN-2003
DEFINITION  Danio rerio genomic clone DKEY-151O11, genomic survey sequence.
ACCESSION   BX158149
```

```
VERSION     BX158149.1  GI:27989752
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 705)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the SP6 end of BAC 151O11. 151O11
            is part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .705
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-151O11"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             5.5%;  Score 275;  DB 9;  Length 705;
  Best Local Similarity   78.9%;  Pred. No. 4.6e-37;
  Matches  343;  Conservative    0;  Mismatches   85;  Indels    7;  Gaps    1;

Qy        266 ATATATATATACACATATATATACACATATATACATATATACACATATATACACATATAT 325
              |||| |||||| | |||||||| | ||||||| ||||||| | ||||||| |||||||
Db         12 ATATTTATATATATATATATATATATATATATATATATATATATATATATACATATAC 71

Qy        326 ATACACATACATATGTACACATATATATACACATATGTATACACATATATACACATATAT 385
              ||||||| | | || |||| ||||||| | ||| || |||||||||| ||| ||||| ||
Db         72 ATACACACATACATATACATATATATACATACACATATATACACATACATATACATACAT 131

Qy        386 ACACACATATATACACATATATACACACATATATACACATATATACACATATATACACAT 445
              | | | | ||| | ||||| ||||| ||||| ||||| | ||   | ||||  ||||| ||
Db        132 ATATATACATACATACATACATACATACATACATACATGCATACATGCATACATAT 191

Qy        446 ATACACATATATACACATATATACATATATACACATATATATA-------ATATACACAC 498
              ||| | |||||||||||||| | ||||| ||| ||||||||||       ||||| ||||
Db        192 ATATATATATATACACATACACACATACATATACATATATATACATACACATATATACAC 251

Qy        499 ATATATATACACATATATACACACATATATACACATATATACACATATATATACACATAT 558
              ||| ||||||| | ||||| | | ||||| | ||||| ||||| | ||| |||||  | ||
Db        252 ATACATATACATACATATATATATATATACATACATACATACATACATACATACATAT 311

Qy        559 ATACACATATATACATATATACACATATATATACATATATACACATATATACATATATAC 618
              | | ||||| ||||||| ||||| | ||| ||||||| ||| | ||||||||||||||||
Db        312 ACATACATACATACATACATACATACATACATACATATATATATATACATATATAC 371

Qy        619 ACATATATACATATATACACACATATATACACATACATATACACACACATAGATATACAT 678
              ||||| |||||||||| | | | ||||||||||||||||||| | ||| ||||| ||| | |
Db        372 ACATACATACATATACATATATATATATACACATACATACATACATACATACATACATAC 431

Qy        679 ATATATACACATATA 693
              ||| ||||| | |||
Db        432 ATACATACATACATA 446


RESULT 91
BM906116/c
LOCUS       BM906116               1156 bp    mRNA    linear   EST 12-MAR-2002
DEFINITION  AGENCOURT_6621231 NIH_MGC_125 Homo sapiens cDNA clone IMAGE:5589966
            5', mRNA sequence.
ACCESSION   BM906116
VERSION     BM906116.1  GI:19356495
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1156)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Invitrogen
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12363  row: i  column: 07
            High quality sequence start: 323
            High quality sequence stop: 613.
FEATURES             Location/Qualifiers
     source          1. .1156
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5589966"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_125"
                     /note="Organ: ovary (pool of 3); Vector: pCMV-SPORT6;
                     Site_1: EcoRV (destroyed); Site_2: NotI; RNA source pool
                     of three ovaries, from females ranging in age from 38 to
                     49 yo. Library is oligo-dT primed and directionally cloned
                     (EcoRV site is destroyed upon cloning). Average insert
                     size 2.1 kb, insert size range 1-3.5 kb. Library is
                     normalized and enriched for full-length clones and was
                     constructed by C. Gruber (Invitrogen). Research Genetics
                     tracking code 036."
ORIGIN

  Query Match             5.5%;  Score 274.2;  DB 5;  Length 1156;
  Best Local Similarity   76.3%;  Pred. No. 5.9e-37;
  Matches  350;  Conservative    0;  Mismatches  108;  Indels    1;  Gaps    1;

Qy        268 ATATATATACACATATATATACACATATATACATATATACACATATATACACATATATAT 327
              |||| |||| | |||||||| | ||||||| ||||||| | |||| || | || |||||
Db        688 ATATGTATATATATATATATATATATATATATATATATATATATATGTATATATGTATAT 629

Qy        328 ACACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATAC 387
              | | ||| |||| || | ||||||||| | |||| |||| | ||||||| | |||||||
Db        628 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 569

Qy        388 ACACATATATACACATATATACACACATATATACACATATATACACATATATACACATAT 447
              | | ||||||| | ||||||| | | ||||||| | ||||||| | ||||||| | ||||
Db        568 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 509

Qy        448 ACACATATATACACATATATACATATATACACATATATATAATATACACACATATATATA 507
              | | ||||||| | ||||||| ||||||| | |||||||| ||||| | | |||||||||
Db        508 ATATATATATATATATATATATATATATATATATATATAT-ATATATATATATATATATA 450

Qy        508 CACATATATACACACATATATACACATATATACACATATATATACACATATATACACATA 567
              | ||||||| | | ||||||| | ||||||| | |||||||| | ||||||| | |||
Db        449 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 390

Qy        568 TATACATATATACACATATATATACATATATACACATATATACATATATACACATATATA 627
              |||| ||||||| | |||||||| ||||||| | ||||||| ||||||| | |||| |
Db        389 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATGTG 330

Qy        628 CATATATACACACATATATACACATACATATACACACACATAGATATACATATATATACA 687
              ||||||  | | |||| || | ||| |||| ||    | ||| || || |||||| ||
Db        329 TATATATCTATATATATGTATATATATATATGCATGTAGATATATCTATATATATGTATC 270

Qy        688 CATATATATACGTATATATATGTATATATATATGCTCCA 726
              || ||||||   |||||| | || |       | |||||
Db        269 TATTTATATATACATATATGTCTAGAGGATCCTCCTCCA 231
```

```
RESULT 92
CC075205/c
LOCUS       CC075205               556 bp    DNA     linear   GSS 16-APR-2003
DEFINITION  CSU-K33r.31C19.T7 CSU-K33r Aedes aegypti genomic clone
            CSU-K33r.31C19, genomic survey sequence.
ACCESSION   CC075205
VERSION     CC075205.1  GI:29916666
KEYWORDS    GSS.
SOURCE      Aedes aegypti (yellow fever mosquito)
  ORGANISM  Aedes aegypti
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Aedes;
            Stegomyia.
REFERENCE   1  (bases 1 to 556)
  AUTHORS   Loftus,B., Shetty,J., Severson,D., Brown,S. and Knudson,D.
  TITLE     End sequencing of Aedes aegypti BACs
  JOURNAL   Unpublished (2003)
COMMENT     Other_GSSs: CSU-K33r.31C19.SP6
            Contact: Brendan Loftus
            Department of Eukaryotic Genomics
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-3543
            Fax: 301-838-0208
            Email: enta@tigr.org
            Library was provided by Susan Brown and Dennis Knudson at Colorado
            State University.
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .556
                     /organism="Aedes aegypti"
                     /mol_type="genomic DNA"
                     /strain="Rexville"
                     /db_xref="taxon:7159"
                     /clone="CSU-K33r.31C19"
                     /clone_lib="CSU-K33r"
                     /note="Vector: pBeloBAC11; Site_1: HindIII"
ORIGIN

  Query Match             5.5%;  Score 273.4;  DB 8;  Length 556;
  Best Local Similarity   71.7%;  Pred. No. 8.9e-37;
  Matches  372;  Conservative    0;  Mismatches  146;  Indels    1;  Gaps    1;

Qy        230 ATATATATATATATATATATATATACACACACACACACATATATATATACACATATATATAC 289
              ||||||||||||||||||||||| | | | | | ||||||||||| | |||||||||
Db        556 ATATATATATATATATATATATATATATATATATATCTATATATATATATATATATATAT 497

Qy        290 ACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATAT 349
              | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | ||||
Db        496 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 437

Qy        350 ATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATAC 409
              ||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||||
Db        436 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 377

Qy        410 ACACATATATACACATATATACACATATATACACATATACACATATATACACATATATAC 469
              | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||||
Db        376 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 317

Qy        470 ATATATACACATATATATAATATACACACATATATATACACATATATACACACATATATA 529
              ||||||| | |||||||| ||||| | | ||||||||| | ||||||| | | |||||||
Db        316 ATATATATATATATATAT-ATATATATATATATATATATATATATATATATATATATATA 258

Qy        530 CACATATATACACATATATATACACATATATACACATATATACATATATACACATATATA 589
              | |||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||||||
Db        257 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 198

Qy        590 TACATATATACACATATATACATATATACACATATATACATATATACACACATATATACA 649
              || |||||||| | ||||||| |||||    | |||||  |    ||  ||  | | |  |
Db        197 TATATATATATATATATATATATATAAGTATATATAAATTAATCAACCTAGAAAAGTTTC 138

Qy        650 CATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATATG 709
              | |  | |         | |   | | |  || |||   |||  || ||  ||   ||
Db        137 AACAGCTGTCGTGGTGAAAAATAAAAAAAGTAGATAATAATAACTACACTAATCCCTAAT 78

Qy        710 TATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTT 748
              ||| ||  ||||   |        ||||  | | |||||
Db         77 TATCTACCTATGTAGCTTGTCCTTTAAGGTGAAACAGTT 39


RESULT 93
BX146163/c
LOCUS       BX146163               648 bp    DNA     linear   GSS 28-JAN-2003
DEFINITION  Danio rerio genomic clone DKEY-104K20, genomic survey sequence.
ACCESSION   BX146163
VERSION     BX146163.1  GI:27977521
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 648)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the SP6 end of BAC 104K20. 104K20
            is part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .648
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-104K20"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             5.5%;  Score 272.2;  DB 9;  Length 648;
  Best Local Similarity   70.4%;  Pred. No. 1.4e-36;
  Matches  392;  Conservative    0;  Mismatches  163;  Indels    2;  Gaps    2;

Qy        238 ATATATATATATATACACACACACACATATATATATACACATATATATACACATATATAT 297
              |||||||||||||| | | | | | | |||||||||| | ||||||||| | ||||||
Db        648 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 589

Qy        298 ACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACAC 357
              | ||||||| | ||||||| | ||||||||| | ||| |||| || | ||||||||| |
Db        588 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 529

Qy        358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417
              |||| | || | || | ||   ||||||| | | ||||||| | ||||||| | | ||||
Db        528 ATATATGTATATATGTGTATGTATATATATATATATATATATATATATATATATATATAT 469

Qy        418 ATACACATATATACACATATATACACATATACACATATATACACATATATACATATATAC 477
              ||| | ||||||| | ||||||| | ||||| | ||||||| | ||||||| ||| ||||
Db        468 ATATATATATATATATATATATATATATATATATATATATATATATATATATAAATACATAC 409

Qy        478 ACATATATATAATATACACACATATATATACACATATATACACACATATATACACATATA 537
              | |||||||| |||||||||| | | | ||||| | ||||||| | | ||||| | |
Db        408 ATATATATAT-ATATACACACACACACACACACACACACACACACACACACACACACA 350

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATACATATA 597
              |||||| | ||||||| ||||||| | |||||||||||||||| | ||||||||| |||||
Db        349 CACACACACATATACATATATATATATATATATACATATATATATATATATATATATA 290

Qy        598 TACACATATATACATATATACACATATATACATATATACACACATATATACACATACATA 657
              || | ||||||| ||||||| | ||||||| ||||||||| |||||| ||  |||   ||
Db        289 TATATATATATATATATATATATATATATATATATATATATACATACATATTTAATCATTTTTA 230
```

```
Qy        658 TACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATA 717
              |    |  || | ||  |       ||||  |  |    |  |  || |||| | | ||| ||
Db        229 T-TTTATTCAAATATTTTATATAATATTTATTTTATTTCATTTAAATATTTTTTTATTTA 171

Qy        718 TATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGA 777
               ||  |  |   |   |||    | | | || ||   |   |   ||| |   | | | |
Db        170 AATATTTAATTTTGTTTAAATATTTGTATTTCATATGCTAGGTAAATATATATATATATA 111

Qy        778 GTTTGACAGCAGTGTAT 794
                |  | |    | |||
Db        110 TATATATATATATATAT 94


RESULT 94
CW020046/c
LOCUS       CW020046                892 bp    DNA     linear   GSS 23-SEP-2004
DEFINITION  ZMMBLc0013O08.f ZMMBLc Zea mays genomic clone ZMMBLc0013O08 5',
            genomic survey sequence.
ACCESSION   CW020046
VERSION     CW020046.1  GI:52613214
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 892)
  AUTHORS   Wing,R., Luo,M., Soderlund,C. and Haller,K.
  TITLE     ZMMBL sequences
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            Plate: 0013  row: O  column: 08
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .892
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /cultivar="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBLc0013O08"
                     /tissue_type="immature ears"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="ZMMBLc"
                     /note="Vector: pAGIBAC1; Site_1: Hpa II; Site_2: Hpa II"
ORIGIN

  Query Match             5.5%;  Score 271;  DB 9;  Length 892;
  Best Local Similarity   72.7%;  Pred. No. 2.2e-36;
  Matches  363;  Conservative    0;  Mismatches  135;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | |||||||||||||||||||||||||| | | | | | | ||||||||||| | |||||||
Db        790 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 731

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | |||||||||                    |
Db        730 TATATATATATATATATATATATATATATATATATATATAATATATATATATATATTA 671

Qy        347 TATATATACACATATGTATACACATATATACACATATA-TACACACATATATACACATAT 405
              |||||||| | |||| |||| | ||||||| | ||||| || | | ||||||| | ||||
Db        670 TATATATATATATATATATATATATATATATATATATACTATATATATATATATATATAT 611

Qy        406 ATACACACATATATACACATATATACACATATATACACATATACACATATATACACATAT 465
              ||| | | ||||||| | ||||||| | ||||||| | ||||  | ||||||| | ||||
Db        610 ATATATATATATATATATATATATATATATATATATATATATTTATATATATATATAT 551
```

```
Qy        466 ATACATATATACACATATATATAATATACACACATATATATACACATATATACACACATA 525
              ||| ||||||| | |||||||||         | |||    || |    |||    | ||
Db        550 ATATATATATATATATATATATATATATATATAATATAATATTATAATATTATTATTATATT 491

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
               ||    |||    |    |||||    ||||||| | ||||||| ||||||| | |||
Db        490 ATTAATATTATTATTATTATTATATTATTATATATATATATATATATATATATATATATA 431

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              ||||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db        430 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 371

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| |||||||||  | |||||||||  ||||||
Db        370 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 311

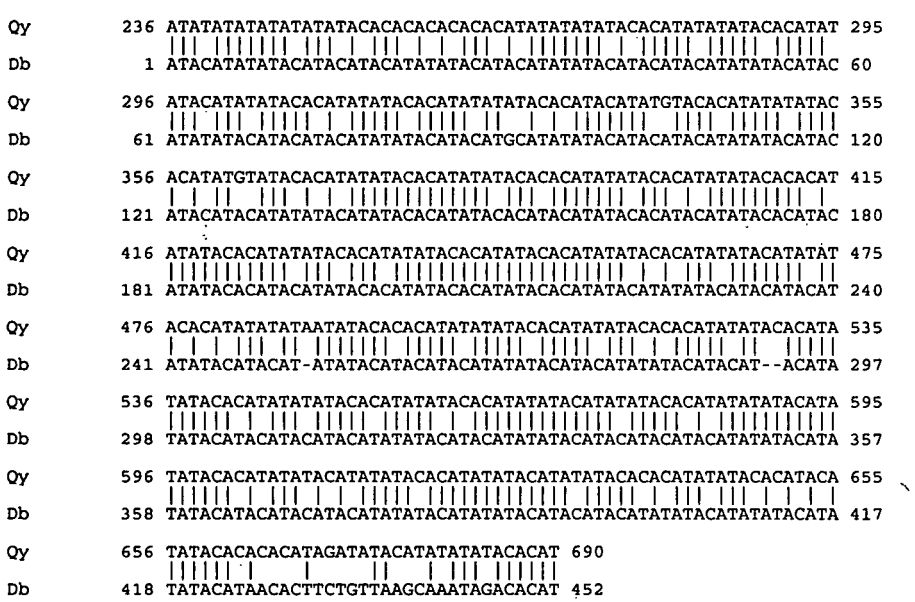
Qy        706 TATGTATATATATATGCTC 724
              ||| ||||||||||||| |
Db        310 TATATATATATATATGCAC 292


RESULT 95
AZ318983
LOCUS       AZ318983                647 bp    DNA     linear   GSS 29-SEP-2000
DEFINITION  1M0038D20F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0038D20 F, genomic survey sequence.
ACCESSION   AZ318983
VERSION     AZ318983.1  GI:10369306
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 647)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,
            Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von
            Niederhausern,A. and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0038  row: D  column: 20
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 647.
FEATURES             Location/Qualifiers
     source          1. .647
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0038D20"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
```

adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

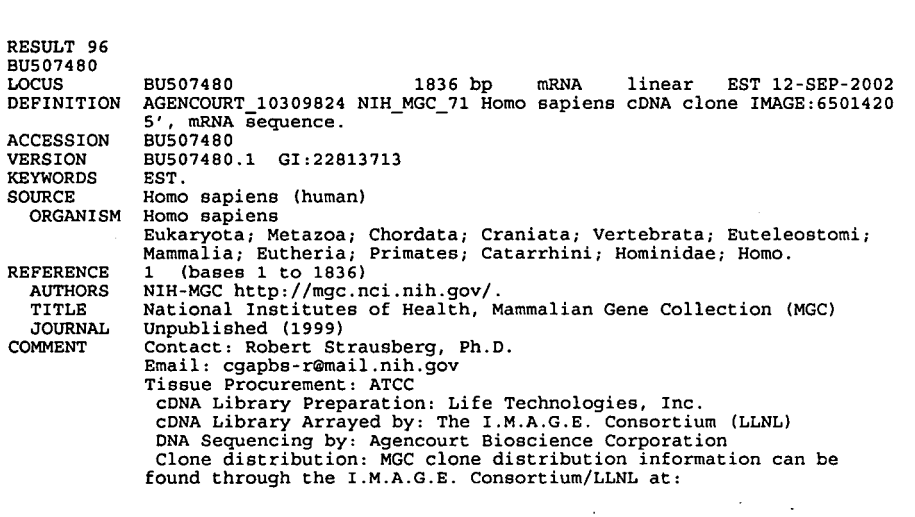
ORIGIN

Query Match 5.4%; Score 270.6; DB 8; Length 647;
Best Local Similarity 77.6%; Pred. No. 2.6e-36;
Matches 353; Conservative 0; Mismatches 99; Indels 3; Gaps 2;



RESULT 96
BU507480
LOCUS BU507480 1836 bp mRNA linear EST 12-SEP-2002
DEFINITION AGENCOURT_10309824 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:6501420 5', mRNA sequence.
ACCESSION BU507480
VERSION BU507480.1 GI:22813713
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1836)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM14055 row: n column: 13
High quality sequence start: 74
High quality sequence stop: 447.
FEATURES Location/Qualifiers
source 1. .1836
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6501420"
/tissue_type="leiomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_71"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2.1 kb. "
ORIGIN

Query Match 5.4%; Score 270.6; DB 5; Length 1836;
Best Local Similarity 72.5%; Pred. No. 2.3e-36;
Matches 359; Conservative 0; Mismatches 135; Indels 1; Gaps 1;



RESULT 97
BH329576/c
LOCUS BH329576 458 bp DNA linear GSS 03-DEC-2001
DEFINITION CH230-106A4.TV CHORI-230 Segment 1 Rattus norvegicus genomic clone CH230-106A4, genomic survey sequence.
ACCESSION BH329576
VERSION BH329576.1 GI:17260290
KEYWORDS GSS.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 458)

```
  AUTHORS   Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K.,
            Shvartsbeyn,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D.,
            Riggs,F., de Jong,P. and Fraser,C.M.
  TITLE     Rat BAC End Sequences from Library CHORI-230 EcoRI segment
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: CH230-106A4.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the rat BAC library CHORI-230
            (http://www.chori.org/bacpac/rat230.htm). For BAC library
            availability, please contact Pieter de Jong (pdejong@mail.cho.org).
            Clones may be purchased from BACPAC Resources
            (http://www.chori.org/bacpac/or ering_information.htm). BAC end
            page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
            Plate: 106  row: A  column: 4
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .458
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /strain="BN/SsNHsd/MCW"
                     /db_xref="taxon:10116"
                     /clone="CH230-106A4"
                     /sex="Female"
                     /cell_type="Brain"
                     /clone_lib="CHORI-230 Segment 1"
                     /note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;
                     CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
                     Pieter de Jong"
ORIGIN

  Query Match             5.4%;  Score 269.8;  DB 8;  Length 458;
  Best Local Similarity   77.2%;  Pred. No. 3.8e-36;
  Matches  353;  Conservative    0;  Mismatches  102;  Indels    2;  Gaps    2;

Qy        230 ATATATATATATATATATATATACACACACACACACATATATATATACACATATATATAC 289
              ||||||||||||||||||||||| | | | |   |||||||||| | |||||||||
Db        458 ATATATATATATATATATATATATATATATATAGGGATATATATATATATATATATATAT 399

Qy        290 ACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATAT 349
              | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | ||||
Db        398 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 339

Qy        350 ATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATAC 409
              ||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||||
Db        338 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 279

Qy        410 ACACATATATACACATATATACACATATATACACATATACACATATATACACATATATAC 469
              | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||||
Db        278 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 219

Qy        470 ATATATACACATATATATAATATACACACATATATATACACATATATACACACATATATA 529
              ||||||| | |||||||| ||||| | | ||||||||| | ||||||| | | |||||||
Db        218 ATATATATATATATATAT-ATATATATATATATATATATATATATATATATATATATATA 160

Qy        530 CACATATATACACATATATATACACATATATACACATATATACATATATACACATATATA 589
              | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||||||
Db        159 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 100

Qy        590 TACATATATACACATATATACATATATACACATATATACATATAT-ACACACATATATAC 648
              || ||||||| | ||||||| ||||||| | |||||||  | | | ||||    | | ||
Db         99 TATATATATATATATATATATATATATATATATATATATTTCTCTCACACTATCACAGAC 40

Qy        649 ACATACATATACACACACATAGATATACATATATATA 685
              ||| | | | | | | || |||  | | | | | | |
Db         39 ACAGAGAGACAGAGAGACGTAGTGACAGACACAGAGA 3
```

```
RESULT 98
CW518121/c
LOCUS       CW518121                663 bp    DNA     linear   GSS 08-OCT-2004
DEFINITION  OP__Ba0025C13.f OP__Ba Oryza punctata genomic clone OP__Ba0025C13
            5', genomic survey sequence.
ACCESSION   CW518121
VERSION     CW518121.1  GI:53992343
KEYWORDS    GSS.
SOURCE      Oryza punctata
  ORGANISM  Oryza punctata
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 663)
  AUTHORS   SanMiguel,P., Westerman,R., Kim,H., Yu,Y., Wissotski,M., Yost,D.,
            Stum,D., Rao,K., Luo,M., Jetty,R., Kudrna,D., Muller,C.,
            Hatfield,J., Soderlund,C., Wing,R. and Jackson,S.A.
  TITLE     OMAP Project - Purdue University
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Scott A. Jackson
            Jackson Laboratory
            Purdue University
            915 W. State St., West Lafayette, IN 47907, USA
            Tel: 7654963621
            Fax: 7654967255
            Email: sjackson@purdue.edu
            Basecalling by phred version 0.020425.c. This sequence was derived
            from the raw sequence read by clipping with lucy version 1.19s.
            Bases 50-712 of the raw sequence (length 1364) were retained after
            clipping.
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Insert Length: 161000   Std Error: 0.00
            Plate: 0025  row: C  column: 13
            Seq primer: TAA TAC GAC TCA CTA TAG GG
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .663
                     /organism="Oryza punctata"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4537"
                     /clone="OP__Ba0025C13"
                     /tissue_type="young leaves"
                     /lab_host="DH10B-T1 phage resistant"
                     /clone_lib="OP__Ba"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

  Query Match             5.4%;  Score 268.8;  DB 9;  Length 663;
  Best Local Similarity   77.9%;  Pred. No. 5.4e-36;
  Matches  349;  Conservative    0;  Mismatches   97;  Indels    2;  Gaps    2;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              |||||||||||||||||||||||||| | | | | | |||||||||| | ||||||||
Db        660 TTTATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 601

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| || | ||   |
Db        600 TATATATATATATATATATATATATATATATATATATATATATATATATATTTATA-TTA 542

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              ||||| ||   |||| |||  | || | || | ||||||| | | ||||||| | |||||
Db        541 TATATCTATCTATATTTATTTATATGTGTATATATATATATATATATATATATATATATA 482

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db        481 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 422

Qy        467 TACATATATACACATATATATAATATACACACATATATATACACATATATACACACATAT 526
              || ||||||| | |||||||| ||||| | | ||||||||| | ||||||| | | ||||
Db        421 TATATATATATATATATATAT-ATATATATATATATATATATATATATATATATATATAT 363
```

```
Qy        527 ATACACATATATACACATATATATACACATATATACACATATATACATATATACACACATAT 586
              ||| | ||||||| | |||||||||| | ||||||| | ||||||| ||||||| | ||||
Db        362 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 303

Qy        587 ATATACATATATACACATATATACATATATACACATATATACATATATACACACATATAT 646
              ||||| ||||||| | |||||||| ||||||| | ||||||| ||||||| | | ||||||
Db        302 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 243

Qy        647 ACACATACATATACACACACATAGATAT 674
              | | ||| ||||       ||| || ||
Db        242 ATATATATATATCTTTTTGCATTGACAT 215


RESULT 99
CW019737
LOCUS       CW019737                600 bp    DNA     linear   GSS 23-SEP-2004
DEFINITION  ZMMBLc0013H15.r ZMMBLc Zea mays genomic clone ZMMBLc0013H15 3',
            genomic survey sequence.
ACCESSION   CW019737
VERSION     CW019737.1  GI:52612602
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 600)
  AUTHORS   Wing,R., Luo,M., Soderlund,C. and Haller,K.
  TITLE     ZMMBL sequences
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            Plate: 0013  row: H  column: 15
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .600
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /cultivar="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBLc0013H15"
                     /tissue_type="immature ears"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="ZMMBLc"
                     /note="Vector: pAGIBAC1; Site_1: Hpa II; Site_2: Hpa II"
ORIGIN

  Query Match             5.4%;  Score 267.2;  DB 9;  Length 600;
  Best Local Similarity   78.8%;  Pred. No. 1e-35;
  Matches  331;  Conservative    0;  Mismatches   88;  Indels    1;  Gaps    1;

Qy        228 TTATATATATATATATATATATACACACACACACACATATATATATACACATATAT 287
              || | ||  | | | |      | ||||||||||||||||||||||||| | ||||||||
Db        182 TTCTCTAATTTTTTTTGCTCGAACACACACACACACACATATATATATATATATATAT 241

Qy        288 ACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACAT 347
              | | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | ||
Db        242 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 301

Qy        348 ATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATAT 407
              ||||||| | |||| |||| | ||||||| | ||||||| | | |||||||| | ||||||
Db        302 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 361

Qy        408 ACACACATATATACACATATATACACATATATACACATATACACATATATACACATATAT 467
              | | | ||||||| | ||||||| | ||||||| | ||||| | |||||||   |||| |
Db        362 ATATATATATATATATATATATATATATATATATATATATATATATATATATGTATATGT 421

Qy        468 ACATATATACACATATATATAATATACACACATATATATACACATATATACACACATATA 527
              | ||||||| | |||||||| ||||| | | ||||||||| | ||||||| | | |||||
Db        422 ATATATATATATATATATAT-ATATATATATATATATATATATATATATATATATATA 480

Qy        528 TACACATATATACACATATATATACACATATATACACATATATACATATATACACATATA 587
              || | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||||
Db        481 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 540

Qy        588 TATACATATATACACATATATACATATATACACATATATACATATATACACACATATATA 647
              |||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||||
Db        541 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 600


RESULT 100
CC075205
LOCUS       CC075205                556 bp    DNA     linear   GSS 16-APR-2003
DEFINITION  CSU-K33r.31C19.T7 CSU-K33r Aedes aegypti genomic clone
            CSU-K33r.31C19, genomic survey sequence.
ACCESSION   CC075205
VERSION     CC075205.1  GI:29916666
KEYWORDS    GSS.
SOURCE      Aedes aegypti (yellow fever mosquito)
  ORGANISM  Aedes aegypti
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Aedes;
            Stegomyia.
REFERENCE   1  (bases 1 to 556)
  AUTHORS   Loftus,B., Shetty,J., Severson,D., Brown,S. and Knudson,D.
  TITLE     End sequencing of Aedes aegypti BACs
  JOURNAL   Unpublished (2003)
COMMENT     Other_GSSs: CSU-K33r.31C19.SP6
            Contact: Brendan Loftus
            Department of Eukaryotic Genomics
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-3543
            Fax: 301-838-0208
            Email: enta@tigr.org
            Library was provided by Susan Brown and Dennis Knudson at Colorado
            State University.
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .556
                     /organism="Aedes aegypti"
                     /mol_type="genomic DNA"
                     /strain="Rexville"
                     /db_xref="taxon:7159"
                     /clone="CSU-K33r.31C19"
                     /clone_lib="CSU-K33r"
                     /note="Vector: pBeloBAC11; Site_1: HindIII"
ORIGIN

  Query Match             5.4%;  Score 267;  DB 8;  Length 556;
  Best Local Similarity   71.4%;  Pred. No. 1.1e-35;
  Matches  365;  Conservative    0;  Mismatches  145;  Indels    1;  Gaps    1;

Qy        186 CAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATAT 245
              || |   |  || | |  |  | | | | || | ||  |  | ||  |||    ||| |
Db         47 CACCTTAAAGGACAAGCTACATAGGTAGATAATTAGGGATTAGTGTAGTTATTATTATCT 106

Qy        246 ATATATACACACACACACATATATATACACATATATATACACATATATACATATAT 305
              |  | |           |||    | |  | |  | |  | | ||      ||  |  |||||
Db        107 ACTTTTTTTATTTTCACCACGACAGCTGTTGAAACTTTTCTAGGTTGATTAATTTATAT 166

Qy        306 ACACATATATACACATATATATACACATACATATGTACACATATATATACACATATGTAT 365
              | || |||||| | |||||||| | ||| |||| || | |||||||| | |||| |||
Db        167 ATACTTATATATATATATATATATATATATATATATATATATATATATATATATATAT 226

Qy        366 ACACATATATACACATATATACACACATATATACACATATATACACACATATATACACAT 425
              | | ||||||| | ||||||| | | ||||||| | ||||||| | | ||||||| | ||
Db        227 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 286
```

```
Qy        426 ATATACACATATATACACATATACACATATATACACATATATACATATATACACATATAT 485
              ||||| | ||||||| | ||||| | ||||||| | ||||||| ||||||| | ||||||
Db        287 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 346

Qy        486 ATAATATACACACATATATATACACATATATACACACATATATACACATATATACACATA 545
              || ||||| | | |||||||||| | ||||||| | | ||||||| | ||||||| | |||
Db        347 AT-ATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 405

Qy        546 TATATACACATATATACACATATATACATATATACACATATATATACATATATACACATA 605
              |||||| | ||||||| | ||||||| ||||||| | ||||||||| ||||||| | |||
Db        406 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 465

Qy        606 TATACATATATACACATATATACATATATACACACATATATACACATACATATACACACA 665
              |||| ||||||| | ||||||| ||||||| | | ||||||| | ||| ||||| | | |
Db        466 TATATATATATATATATATATATATATATATATATATATATATATATATATATAGATATA 525

Qy        666 CATAGATATACATATATATACACATATATAT 696
               ||| ||||| |||||||||| | ||||||||
Db        526 TATATATATATATATATATATATATATATAT 556
```

Search completed: April 22, 2005, 04:39:04
Job time : 9960.71 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 20, 2005, 22:49:00 ; Search time 1619.07 Seconds
(without alignments)
18171.562 Million cell updates/sec

Title: US-09-939-209A-3_COPY_1_4970
Perfect score: 4970
Sequence: 1 agttcaagaccagcctgagc.....gtggctgaatgtctgccccc 4970

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_16Dec04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	4970	100.0	20300	6	ABK47337	Abk47337 Genomic n
c	2	356.6	7.2	121600	6	ABT10748	Abt10748 Human bre
c	3	356.6	7.2	121600	12	ADQ18307	Adq18307 Human sof
c	4	327.4	6.6	188971	12	ADL08108	Adl08108 Human gen
	5	323.6	6.5	163382	13	ABD32659	Abd32659 Human can
c	6	322.2	6.5	26997	4	AAS46748	Aas46748 Tumour su
c	7	311.2	6.3	26997	4	AAS46747	Aas46747 Tumour su
	8	308	6.2	1990	10	ADC87572	Adc87572 Human GPC
	9	296	6.0	15418	3	AAA63785	Aaa63785 Nucleotid
	10	296	6.0	15418	6	AAL38601	Aal38601 Human TER
	11	296	6.0	15418	6	ABS54997	Abs54997 Lambda cl
	12	296	6.0	15418	6	AAL38595	Aal38595 DNA of pl
	13	296	6.0	15418	10	ADC21253	Adc21253 Lambda cl
c	14	294.8	5.9	1486	10	ADC87552	Adc87552 Human GPC
	15	293.6	5.9	27067	10	AAD50021	Aad50021 Human sec
c	16	291.6	5.9	1672	10	ADC87570	Adc87570 Human GPC
	17	288	5.8	30393	4	AAK67239	Aak67239 Human imm
c	18	286.6	5.8	44147	6	ABK84481	Abk84481 Human cDN
c	19	286.6	5.8	44147	10	ADD14691	Add14691 Human src
	20	280.8	5.6	32206	4	AAK89374	Aak89374 Human dig
	21	270.4	5.4	20420	4	AAK73165	Aak73165 Human imm
	22	270.4	5.4	20420	6	ABK69933	Abk69933 Human sec
c	23	268.6	5.4	110000	13	ABD32806_4	Continuation (5 of
c	24	268.6	5.4	191127	10	ADH13788	Adh13788 Prostate
	25	267.6	5.4	110000	13	ABD32806_4	Continuation (5 of
	26	267.4	5.4	19965	4	AAK73166	Aak73166 Human imm
	27	267.4	5.4	19965	6	ABK69932	Abk69932 Human sec
c	28	266.4	5.4	6971	6	ABL33237	Abl33237 Human imm
c	29	263.6	5.3	2522	11	ADM01858	Adm01858 Human cDN
c	30	263	5.3	117962	8	AAD54480	Aad54480 Human CIP
	31	263	5.3	165199	6	ABK83460	Abk83460 Human cDN
c	32	263	5.3	227968	6	ABK83497	Abk83497 Human cDN
c	33	263	5.3	227968	12	ADQ18538	Adq18538 Human sof
c	34	261.4	5.3	12069	6	ABK39930	Abk39930 Human che
c	35	260.6	5.2	20420	4	AAK73165	Aak73165 Human imm
c	36	260.6	5.2	20420	6	ABK69933	Abk69933 Human sec
	37	258.8	5.2	1489	10	ADC87554	Adc87554 Human GPC
c	38	258.8	5.2	2761	12	ADO15781	Ado15781 4 synthes
	39	258.8	5.2	201766	10	ADL13771	Adl13771 Osteoarth
c	40	258.6	5.2	163382	13	ABD32659	Abd32659 Human can
	41	258.4	5.2	82938	6	ABV72623	Abv72623 Human tra
c	42	258	5.2	19965	4	AAK73166	Aak73166 Human imm
c	43	258	5.2	19965	6	ABK69932	Abk69932 Human sec
	44	257.6	5.2	12069	6	ABK39930	Abk39930 Human che
	45	256.4	5.2	1486	10	ADC87552	Adc87552 Human GPC

ALIGNMENTS

RESULT 1
ABK47337
ID ABK47337 standard; DNA; 20300 BP.
XX
AC ABK47337;
XX
DT 18-JUN-2002 (first entry)
XX
DE Genomic nucleotide sequence encoding human RGS-4 protein.
XX
KW RGS-4; schizophrenia; human; regulator of G protein signalling 4;
KW neuroleptic; polymorphism; schizotypy; schizoid; chromosome 1q21-22;
KW gene; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT variation replace(4121,T)
FT /*tag= a
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace(4123,A)
FT /*tag= b
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace(4368,C)
FT /*tag= c
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace(4621,C)
FT /*tag= d
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace(4790,T)
FT /*tag= e
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace(4816,T)
FT /*tag= f
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace(4970,T)
FT /*tag= g
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace(5055,G)
FT /*tag= h
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace(5295,A)
FT /*tag= i

```
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(5695,A)
FT                   /*tag= j
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(7375,T)
FT                   /*tag= k
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(7759,A)
FT                   /*tag= l
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(8596,A)
FT                   /*tag= m
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(9602. .9610,CA)
FT                   /*tag= n
FT                   /note= "Deletion polymorphism"
FT   variation       replace(9892,A)
FT                   /*tag= o
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(9963,A)
FT                   /*tag= p
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(10132,A)
FT                   /*tag= q
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(11056,C)
FT                   /*tag= r
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(11091,T)
FT                   /*tag= s
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(11106,A)
FT                   /*tag= t
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(11774,T)
FT                   /*tag= u
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(12143,A)
FT                   /*tag= v
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(12145,T)
FT                   /*tag= w
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(14367,G)
FT                   /*tag= x
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(17027. .17029,GA)
FT                   /*tag= y
FT                   /note= "Deletion polymorphism"
FT   variation       replace(17630,T)
FT                   /*tag= z
FT                   /standard_name= "Single-nucleotide polymorphism"
XX
PN   WO200216653-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US026622.
XX
PR   24-AUG-2000; 2000US-0228021P.
XX
PA   (UYPI-) UNIV PITTSBURGH.
XX
PI   Levitt PR,  Mirnics K,  Kodavali VC,  Nimgaonkar VL;
XX
DR   WPI; 2002-292070/33.
XX
PT   Diagnosing, assessing susceptibility and treating schizophrenia, involves
PT   observing regulator of G-protein signaling 4, RGS4 levels in a subject.
XX
PS   Claim 1; Page 20-33; 112pp; English.
XX
CC   This invention relates to a novel method for diagnosing schizophrenia or
CC   determining susceptibility to schizophrenia in a human. The method
CC   comprises obtaining from a patient a DNA sample and detecting variations
CC   in the regulator of G-protein signalling 4 (RGS4) gene. Alternatively,
CC   the method involves measuring RGS4 mRNA or protein levels in a tissue
CC   sample from the patient and determining if there is a reduced level. The
CC   method of the invention is useful for diagnosing and determining
CC   susceptibility to schizophrenia. The invention also comprises a method
CC   that is useful for treating schizophrenia which includes a prophylactic
CC   treatment. The method of genotyping polymorphic variants in the RGS-4
CC   gene is applied to diagnosing pathologies of the schizophrenic spectrum,
CC   such as in particular schizotypy, schizoid individuals, etc. This method
CC   offers the possibility of diagnosing schizophrenia by a biological test
CC   and no longer exclusively by clinical evaluations. The present sequence
CC   represents the genomic DNA encoding the human regulator of G-protein
CC   signalling 4 (RGS4) protein used in the method of the invention. The gene
CC   for the RGS4 protein is located on human chromosome 1q21-22
XX
SQ   Sequence 20300 BP; 6157 A; 4102 C; 3775 G; 6266 T; 0 U; 0 Other;

  Query Match            100.0%;  Score 4970;  DB 6;  Length 20300;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 4970;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAATTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAATTA 60

Qy         61 GACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAAT 120

Qy        121 CACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAGC 180

Qy        181 CCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATA 240

Qy        241 TATATATATATACACACACACACATATATATATACACATATATATACACATATATACA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TATATATATATACACACACACACATATATATATACACATATATATACACATATATACA 300

Qy        301 TATATACACATATATACACATATATATACACATACATATGTACACATATATATACACATA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TATATACACATATATACACATATATATACACATACATATGTACACATATATATACACATA 360

Qy        361 TGTATACACATATATACACATATATACACACATATATACACATATATACACACATATATA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGTATACACATATATACACATATATACACACATATATACACATATATACACACATATATA 420

Qy        421 CACATATATACACATATATACACATATACACATATATACACATATATACATATATACACA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CACATATATACACATATATACACATATACACATATATACACATATATACATATATACACA 480

Qy        481 TATATATAATATACACACATATATATACACATATATACACACATATATACACATATATAC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TATATATAATATACACACATATATATACACATATATACACACATATATACACATATATAC 540

Qy        541 ACATATATATACACATATATACACATATATACATATATACACATATATATACATATATAC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ACATATATATACACATATATACACATATATACATATATACACATATATATACATATATAC 600

Qy        601 ACATATATACATATATACACATATATACATATATACACACATATATACACATACATATAC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ACATATATACATATATACACATATATACATATATACACACATATATACACATACATATAC 660

Qy        661 ACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATAT 720

Qy        721 GCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        721 GCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTT 780
Qy        781 TGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATAC 840
Db        781 TGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATAC 840
Qy        841 ATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATATATATA 900
Db        841 ATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATATATATA 900
Qy        901 TAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTATAGAAGAATGG 960
Db        901 TAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTATAGAAGAATGG 960
Qy        961 CTAAGATGCCAAAGCTCAGTGTATGTGGCAGAGGCATGGTGTAGGGTGTGTCCAGGTTCA 1020
Db        961 CTAAGATGCCAAAGCTCAGTGTATGTGGCAGAGGCATGGTGTAGGGTGTGTCCAGGTTCA 1020
Qy       1021 TATATTGCATTAAGTGTGAGAACACCCTGGAGTATGAACCAAGAAAATGCAAAAGCCAGA 1080
Db       1021 TATATTGCATTAAGTGTGAGAACACCCTGGAGTATGAACCAAGAAAATGCAAAAGCCAGA 1080
Qy       1081 AGTGATGGAGGAAATGAGACACAATAATGAAGATATTGAGAGGAGGGTGTGGGCCTAGAG 1140
Db       1081 AGTGATGGAGGAAATGAGACACAATAATGAAGATATTGAGAGGAGGGTGTGGGCCTAGAG 1140
Qy       1141 TGAAGCTTTTCGTGCCAGTACTTCTTTTGAAGGCCCAGTTCTCTTCTCTCTCGGGGGCTC 1200
Db       1141 TGAAGCTTTTCGTGCCAGTACTTCTTTTGAAGGCCCAGTTCTCTTCTCTCTCGGGGGCTC 1200
Qy       1201 CTTCATCTCTCATAGAGTCCACAGCTTTTAAGGGCCAACACTTGAGGTCAGCCTGGCTCT 1260
Db       1201 CTTCATCTCTCATAGAGTCCACAGCTTTTAAGGGCCAACACTTGAGGTCAGCCTGGCTCT 1260
Qy       1261 CTCATTTGAGCTGGATAGAACATTTTAGAGCACCATCTATTCTTCAAGAGGAAGTTTAAA 1320
Db       1261 CTCATTTGAGCTGGATAGAACATTTTAGAGCACCATCTATTCTTCAAGAGGAAGTTTAAA 1320
Qy       1321 AATAAAAGAACCTTGAAGAGGAAAAAATGTAGACATTCAATCTAACCTTTTCATTTTACT 1380
Db       1321 AATAAAAGAACCTTGAAGAGGAAAAAATGTAGACATTCAATCTAACCTTTTCATTTTACT 1380
Qy       1381 AGCCAAAGCTAAATAGAATGCAGGTTACCTGTTTTTCAGCCAGGCACCATCATTTCCTAA 1440
Db       1381 AGCCAAAGCTAAATAGAATGCAGGTTACCTGTTTTTCAGCCAGGCACCATCATTTCCTAA 1440
Qy       1441 TTGTTATAAAATTTATTATTATTGTTGTTATTATTATTATTTGCCATAAGAAGTTTCCCA 1500
Db       1441 TTGTTATAAAATTTATTATTATTGTTGTTATTATTATTATTTGCCATAAGAAGTTTCCCA 1500
Qy       1501 TATCCTTTTAGTATAACAAAAACACAATTCACAAGCATTATAAAACCCATGGTGTCTAAC 1560
Db       1501 TATCCTTTTAGTATAACAAAAACACAATTCACAAGCATTATAAAACCCATGGTGTCTAAC 1560
Qy       1561 TATTAAAAAAAATTAAGTGGAACACACTTGTCCCAGCTACTGGGGAGGCTGAGGAGGGAGG 1620
Db       1561 TATTAAAAAAAATTAAGTGGAACACACTTGTCCCAGCTACTGGGGAGGCTGAGGAGGGAGG 1620
Qy       1621 ATCACGTGATCCCAGGGGGTCAAGGTTATGGAGAGCTATGATTGTGCCACTGCACTCCAG 1680
Db       1621 ATCACGTGATCCCAGGGGGTCAAGGTTATGGAGAGCTATGATTGTGCCACTGCACTCCAG 1680
Qy       1681 CCTGGGTGACAGGGAAAGACCCTGTCTCTAAAATTTTTTTTAAAAAAACTAAACTGGTTT 1740
Db       1681 CCTGGGTGACAGGGAAAGACCCTGTCTCTAAAATTTTTTTTAAAAAAACTAAACTGGTTT 1740
Qy       1741 TATTACAGAGATTCTGGAGACAGCTACACATAAAAGGGTGGTATGCCTCATATTAGCTAC 1800
Db       1741 TATTACAGAGATTCTGGAGACAGCTACACATAAAAGGGTGGTATGCCTCATATTAGCTAC 1800
Qy       1801 CCAGGGAGGTGGAATGCCAACTTAGGTGGTGTCACCACTATTAAAAATGCCCCAAAGCAA 1860
Db       1801 CCAGGGAGGTGGAATGCCAACTTAGGTGGTGTCACCACTATTAAAAATGCCCCAAAGCAA 1860
```

```
Qy       1861 TCAAAACTGAGAACTTCCTGGGAGCTTAGCATTGTGCAAAAGCAGCACAAAACACTTAAA 1920
Db       1861 TCAAAACTGAGAACTTCCTGGGAGCTTAGCATTGTGCAAAAGCAGCACAAAACACTTAAA 1920
Qy       1921 CAATTCACAGTTGTGTTGGAATGGGAAGGCCTGGAAATATAAACCAAAGAGTATATTGTC 1980
Db       1921 CAATTCACAGTTGTGTTGGAATGGGAAGGCCTGGAAATATAAACCAAAGAGTATATTGTC 1980
Qy       1981 TAAATTGATAGAGATTACAATTGCCTGAAAGAAAAAGTTGACTTTTAACTAGAATGTTCA 2040
Db       1981 TAAATTGATAGAGATTACAATTGCCTGAAAGAAAAAGTTGACTTTTAACTAGAATGTTCA 2040
Qy       2041 GAGTAGGTTTACAGAAGAAGCTCTTAAACTGGGCTCCAGTGGATTTGTCAATGCTTTGGA 2100
Db       2041 GAGTAGGTTTACAGAAGAAGCTCTTAAACTGGGCTCCAGTGGATTTGTCAATGCTTTGGA 2100
Qy       2101 AGCTGGTGGGGTGGGAGGGTTGGAGGGGGCATAAAAAGTCATGTTGGTATGCTCTGCTCA 2160
Db       2101 AGCTGGTGGGGTGGGAGGGTTGGAGGGGGCATAAAAAGTCATGTTGGTATGCTCTGCTCA 2160
Qy       2161 AGTCTCCATTCTGTTTCCTTTTCCTCTTTTCAATGTCATGTCCCATTATTTCATTATGGG 2220
Db       2161 AGTCTCCATTCTGTTTCCTTTTCCTCTTTTCAATGTCATGTCCCATTATTTCATTATGGG 2220
Qy       2221 CTTCCCTTTATCCAGGATCAATATGCCACCTCTTGGTTGTCTTTTACCTACTTCTCCACC 2280
Db       2221 CTTCCCTTTATCCAGGATCAATATGCCACCTCTTGGTTGTCTTTTACCTACTTCTCCACC 2280
Qy       2281 TCACTATGGAATCGTCCTTGGGTAGCTCCTGTGCTTGGGAACCTGCACGGGCACTTTTCT 2340
Db       2281 TCACTATGGAATCGTCCTTGGGTAGCTCCTGTGCTTGGGAACCTGCACGGGCACTTTTCT 2340
Qy       2341 GATGTCTTGATTCCAGCTTTACTCCTAAAACTTAAATGCTGAGGGGCCAACACCATGGCA 2400
Db       2341 GATGTCTTGATTCCAGCTTTACTCCTAAAACTTAAATGCTGAGGGGCCAACACCATGGCA 2400
Qy       2401 GTGGTAGGGATGGGAATGGGGGTCTTGTAACACACTACATAAACTACACGAAATAAACTA 2460
Db       2401 GTGGTAGGGATGGGAATGGGGGTCTTGTAACACACTACATAAACTACACGAAATAAACTA 2460
Qy       2461 CATGAAACTCAACATGTTTGCAAGACTCAGTTCACATCCATGAGGAGCTCATGCTTCTCC 2520
Db       2461 CATGAAACTCAACATGTTTGCAAGACTCAGTTCACATCCATGAGGAGCTCATGCTTCTCC 2520
Qy       2521 CTCCTGCTCCCCTAGCACACATGATTATCTCTATTTGGAAATGTTTGGCATTTTTGGTGA 2580
Db       2521 CTCCTGCTCCCCTAGCACACATGATTATCTCTATTTGGAAATGTTTGGCATTTTTGGTGA 2580
Qy       2581 AGTGAATGGTTCAATAACTTTCTCCACCATCAGAACAAAAGCTCTTTAAGGTTAGGGATG 2640
Db       2581 AGTGAATGGTTCAATAACTTTCTCCACCATCAGAACAAAAGCTCTTTAAGGTTAGGGATG 2640
Qy       2641 GGATCATACACACTTCCCTTGTCCAAGTCCCCATCACCCCTTATCTAGACAATTGCTACA 2700
Db       2641 GGATCATACACACTTCCCTTGTCCAAGTCCCCATCACCCCTTATCTAGACAATTGCTACA 2700
Qy       2701 GTTTCCTACACACTCTTCTAACCTCTTGCAGTCTATTTTCATAAAACAGCTAGAGAACTT 2760
Db       2701 GTTTCCTACACACTCTTCTAACCTCTTGCAGTCTATTTTCATAAAACAGCTAGAGAACTT 2760
Qy       2761 TGAGATGTAAGTCAAAAAATAGAACATGTCGCTCTTTCCCATTGTTTTTGAAATAAAGTT 2820
Db       2761 TGAGATGTAAGTCAAAAAATAGAACATGTCGCTCTTTCCCATTGTTTTTGAAATAAAGTT 2820
Qy       2821 CAACCCCCTTACCAGGGTCAACAAGGCCCTGCAATGATTTGGTCCTGTTAAAAATTCTTT 2880
Db       2821 CAACCCCCTTACCAGGGTCAACAAGGCCCTGCAATGATTTGGTCCTGTTAAAAATTCTTT 2880
Qy       2881 AGCCTTAACTCATGCTGTTCTTCCTTACACTCACTGCATTCTAGCCATTGAGGTTTCTAT 2940
Db       2881 AGCCTTAACTCATGCTGTTCTTCCTTACACTCACTGCATTCTAGCCATTGAGGTTTCTAT 2940
```

```
Qy       2941 GCATCAAACTTTTTTTGGTCCCAGCACTGTGCACATCCTTCTGGGTAGAATGCCCCTTGA 3000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2941 GCATCAAACTTTTTTTGGTCCCAGCACTGTGCACATCCTTCTGGGTAGAATGCCCCTTGA 3000

Qy       3001 TTTGTATAATTAGCACCTCCTTCATCATTTAGGTCTTAGTATAACTACTACCTTCTTAGA 3060
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3001 TTTGTATAATTAGCACCTCCTTCATCATTTAGGTCTTAGTATAACTACTACCTTCTTAGA 3060

Qy       3061 GAAGCTCTGCTTCTTCATCCTATAAAAAAGTAAAATTCCTTACCCTGTTATTTTTTAAGT 3120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3061 GAAGCTCTGCTTCTTCATCCTATAAAAAAGTAAAATTCCTTACCCTGTTATTTTTTAAGT 3120

Qy       3121 CATCCGTGTTTCATTCTGTTAAAGTTCTTATCACAATTTATCATTATTTTATTTACAGTC 3180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3121 CATCCGTGTTTCATTCTGTTAAAGTTCTTATCACAATTTATCATTATTTTATTTACAGTC 3180

Qy       3181 ATGTGCCACATAACAATGTTTCAGTCAGGGATAGAACACAAATGTATCTGGCCCCATAAT 3240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3181 ATGTGCCACATAACAATGTTTCAGTCAGGGATAGAACACAAATGTATCTGGCCCCATAAT 3240

Qy       3241 ATTATAAGCTGAGAAATTTCTATTAACTAGTGATATCGCAGCCATCATAAGTGTAATGCA 3300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3241 ATTATAAGCTGAGAAATTTCTATTAACTAGTGATATCGCAGCCATCATAAGTGTAATGCA 3300

Qy       3301 GGACATTACCTTTTCTATGTTTAGATATGTTAGATACACAAATATATTTCATTGTGTTAT 3360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3301 GGACATTACCTTTTCTATGTTTAGATATGTTAGATACACAAATATATTTCATTGTGTTAT 3360

Qy       3361 AATTTCCTACAGTATTCAGTACAGTAACATGCTGTACAGGTTTGTAACCTAGGAGTAATA 3420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3361 AATTTCCTACAGTATTCAGTACAGTAACATGCTGTACAGGTTTGTAACCTAGGAGTAATA 3420

Qy       3421 GGCTATACCATACAGCTTAGGTGTGTAGTAGGCTATAACCATCTAGGTTTGTGTAAGTAC 3480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3421 GGCTATACCATACAGCTTAGGTGTGTAGTAGGCTATAACCATCTAGGTTTGTGTAAGTAC 3480

Qy       3481 ATTCTATGATATTCCCACAATGATGAAATCACCTAACTACACATTTCTCAGAATGTTTCA 3540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3481 ATTCTATGATATTCCCACAATGATGAAATCACCTAACTACACATTTCTCAGAATGTTTCA 3540

Qy       3541 CTGTTGTGAAGTGACCCATGACTATATTTTCCTATATACTTGATATTTTTGTGCATCTGC 3600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3541 CTGTTGTGAAGTGACCCATGACTATATTTTCCTATATACTTGATATTTTTGTGCATCTGC 3600

Qy       3601 CCATGAGAATGTAGTGTAAGATCAAAGGATGCAAGAATGGGTTCTATCCAGTATAGTACC 3660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3601 CCATGAGAATGTAGTGTAAGATCAAAGGATGCAAGAATGGGTTCTATCCAGTATAGTACC 3660

Qy       3661 CACTACACTGGTGGATGTCAATATGTATTTGTTAGATTAATATCTCAAGAATGAGCACCT 3720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3661 CACTACACTGGTGGATGTCAATATGTATTTGTTAGATTAATATCTCAAGAATGAGCACCT 3720

Qy       3721 TTCTCAGACACATAAAAGATGCTCAATATAAAAGTTTGTTGAACTGAACGTTATTGGCAA 3780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3721 TTCTCAGACACATAAAAGATGCTCAATATAAAAGTTTGTTGAACTGAACGTTATTGGCAA 3780

Qy       3781 ATGTAACATGATCGGATTTAAAGAGGAGCGAAACAGAGGTCTGGCTCAAACACCATACTT 3840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3781 ATGTAACATGATCGGATTTAAAGAGGAGCGAAACAGAGGTCTGGCTCAAACACCATACTT 3840

Qy       3841 CTAGAGTGCATAAGAGGTAGCAGTTGATTACCACTGGCGACAGGAGAAAAAAGAGCTTGA 3900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3841 CTAGAGTGCATAAGAGGTAGCAGTTGATTACCACTGGCGACAGGAGAAAAAAGAGCTTGA 3900

Qy       3901 CCGCAGGGTACTGTGAAGACATTTCAGGTTGATGGCACAGAACAGGGGAAATACATAAAT 3960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3901 CCGCAGGGTACTGTGAAGACATTTCAGGTTGATGGCACAGAACAGGGGAAATACATAAAT 3960

Qy       3961 GTGTGGGAATATTCAGTGGTCTGGGATGACTACATAGTAGAATATAATGAAGAAAAGAGT 4020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3961 GTGTGGGAATATTCAGTGGTCTGGGATGACTACATAGTAGAATATAATGAAGAAAAGAGT 4020

Qy       4021 GGAAGGGAAAGATGAAAAGTTGGAATGGGGATGAATTATGAAAGTACCAGAATGTTATGC 4080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4021 GGAAGGGAAAGATGAAAAGTTGGAATGGGGATGAATTATGAAAGTACCAGAATGTTATGC 4080

Qy       4081 TAAGGAATCTAGATTTTAAAATGTGAGGGCAAATTGAAGTCCTGGGCACGTTACAAAACT 4140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4081 TAAGGAATCTAGATTTTAAAATGTGAGGGCAAATTGAAGTCCTGGGCACGTTACAAAACT 4140

Qy       4141 AGAGGTCATAAAGTTTACCCTAATTTACCAAGATTTCCTAGAGGATCTATAATTGGAATC 4200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4141 AGAGGTCATAAAGTTTACCCTAATTTACCAAGATTTCCTAGAGGATCTATAATTGGAATC 4200

Qy       4201 CAGATCTGCCTCTCTGTAAAGTTCAAGCACTTTCCATGACACCATACTGTTTCTTTCCAC 4260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4201 CAGATCTGCCTCTCTGTAAAGTTCAAGCACTTTCCATGACACCATACTGTTTCTTTCCAC 4260

Qy       4261 CTGCACAATGCAAATGAACTCTTATGAAACTGCTGTTTCTATCCTGGGCTAAATGTTGCA 4320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4261 CTGCACAATGCAAATGAACTCTTATGAAACTGCTGTTTCTATCCTGGGCTAAATGTTGCA 4320

Qy       4321 GAAAAAAGATTTAATCTTTGGGATAAGGCTATTTTGGGTTTTCTCCTACTTCTTGGGAAA 4380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4321 GAAAAAAGATTTAATCTTTGGGATAAGGCTATTTTGGGTTTTCTCCTACTTCTTGGGAAA 4380

Qy       4381 CAAGGTTTTCTTCCCCTGGCTAATTAAGTGTGGTATTGTTCTTCCAGGGAAATCAGTGAT 4440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4381 CAAGGTTTTCTTCCCCTGGCTAATTAAGTGTGGTATTGTTCTTCCAGGGAAATCAGTGAT 4440

Qy       4441 GCATCACCTGCTGCTATCAAATGTCAGGGTTGGAGTTCCTGATTTATTGCATGTGCCCAC 4500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4441 GCATCACCTGCTGCTATCAAATGTCAGGGTTGGAGTTCCTGATTTATTGCATGTGCCCAC 4500

Qy       4501 AAAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCT 4560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4501 AAAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCT 4560

Qy       4561 GTTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAA 4620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4561 GTTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAA 4620

Qy       4621 ATCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGG 4680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4621 ATCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGG 4680

Qy       4681 AGGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCT 4740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4681 AGGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCT 4740

Qy       4741 GAGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAACATCTGGTAGA 4800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4741 GAGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAACATCTGGTAGA 4800

Qy       4801 ACTAAAGAACGCAACGCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCC 4860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4801 ACTAAAGAACGCAACGCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCC 4860

Qy       4861 TTGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCT 4920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4861 TTGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCT 4920

Qy       4921 AGAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCC 4970
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4921 AGAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCC 4970
```

RESULT 2
ABT10748/c
ID ABT10748 standard; cDNA; 121600 BP.
XX
AC ABT10748;
XX
DT 04-DEC-2002 (first entry)
XX
DE Human breast cancer associated coding sequence SEQ ID NO: 882.

```
XX
KW   Human; breast specific gene; breast cancer; differential expression;
KW   cytostatic; gene therapy; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200259271-A2.
XX
PD   01-AUG-2002.
XX
PF   25-JAN-2002; 2002WO-US002176.
XX
PR   25-JAN-2001; 2001US-0263757P.
PR   25-APR-2001; 2001US-0286090P.
PR   23-MAY-2001; 2001US-0292517P.
XX
PA   (GENE-) GENE LOGIC INC.
XX
PI   Orr MS,  Nation M,  Diggans JC,  Zeng W;
XX
DR   WPI; 2002-674803/72.
XX
PT   Diagnosing breast cancer in a patient comprises detecting the level of
PT   gene expression in cell or tissue samples, where a differential gene
PT   expression is indicative of breast cancer.
XX
PS   Claim 1; SEQ ID NO 882; 260pp + Sequence Listing; English.
XX
CC   The present invention relates to methods of diagnosing breast cancer in a
CC   patient, which comprise detecting the level of expression in a tissue
CC   sample of two or more genes selected from those shown in ABT09867-
CC   ABT11112, where a differential expression of the genes indicates breast
CC   cancer. The methods are useful in diagnosing, treating, detecting the
CC   progression, and in monitoring treatment of breast cancer in patients.
CC   The methods are also useful as a screening tool for agents that modulate
CC   the onset or progression of breast cancer. The breast cancer genes may be
CC   used as diagnostic markers for the prediction or identification of the
CC   malignant state of breast tissue, for confirming the type and progression
CC   of cancer, and for drug screening and assays. The present sequence is a
CC   coding sequence of the invention. Note: The sequence data for this patent
CC   did not form part of the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub.published_pct_sequences
XX
SQ   Sequence 121600 BP; 38372 A; 24652 C; 24048 G; 34528 T; 0 U; 0 Other;

  Query Match             7.2%;  Score 356.6;  DB 6;  Length 121600;
  Best Local Similarity   74.7%;  Pred. No. 2.3e-52;
  Matches  543;  Conservative    0;  Mismatches  169;  Indels   15;  Gaps    7;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAA--- 57
              || || |||||| ||||   ||||  |||||||||| |||||||||||||| |||||
Db      44776 AGATCGAGACCATCCTGGCTAACACAGTGAAACCCCGTCTCTACTAAAAATCCAAAAAAA 44717

Qy         58 --TTAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGG 115
                |||| | ||| |||||      | |||||| ||||||||||  |||||| ||||||||
Db      44716 AGTTAGCCGGGCGTGGTGGCGGGCACCTGTAGTCCCAGCTACTCTGGAGGCTGAGGCAGG 44657

Qy        116 AGAATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCAC 174
              |||||  | |||||  ||  || || |||| |||| | ||| | ||||   |||| |||||
Db      44656 AGAATGGCGTGAACTTG--GGAGGCGGAGCTTGCAGTGAGCCGAGATAGCGCCACTGCAC 44599

Qy        175 TCCAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATA 234
              |||||  |  ||| || ||||||| |   ||||||| |||||||| |       | ||||||
Db      44598 TCCAGTCTGGGCGAC-AGAGCGAGACTCCATCTCAAAAAAAAAATA----TATATATATA 44544

Qy        235 TATATATATATATATACACACACACACACATATATATATACACATATATATACACATA 294
              || ||||||||| ||||||| | | | | | ||||||| ||| | | ||||||| | |||
Db      44543 TAGATATATATAGATATACAGATATATAGATATATATAGATATATAGATATATATAGATA 44484

Qy        295 TATACATATATACACATATATACACATATATATACACATACATATGTACACATATATATA 354
              |||| ||||||| | ||||||| | ||||||| | | ||| |||| || | ||||||| |
Db      44483 TATAGATATATATAGATATATAGATATATATAGATATATAGATATATATAGATATATAGA 44424

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              | ||||  ||| | | ||||| | | ||||| | | ||||| | | ||||| | | | |
Db      44423 TATATATAGATATATAGATATATATAGATATATAGATATATAGATATATATAGATATATA 44364

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              ||||| | | ||||| | ||||||| | ||||| | ||||||| | ||||||| |||||
Db      44363 GATATATATAGATATATAGATATATATAGATATATAGATATATAGATATATATAGATATA 44304

Qy        475 TACACATATATATAATATACACACATATATATACACATATATACACACATATATACACAT 534
              || |  |||| || ||||| | | ||||| ||| | ||| ||| | | ||||||| | ||
Db      44303 TAGATGTATAGAT-ATATATAGATATATAGATATATATAGATATATAGATATATAGATAT 44245

Qy        535 ATATACACATATATATACACATATATACACATATATACATATATACACATATATATACAT 594
              ||||| | ||||| ||| | ||| ||| | ||||| | ||||| | | ||| |||||
Db      44244 ATATAGATATATAGATATATATAGATATATATATAGATATATAGATATATAGATATATAG 44185

Qy        595 ATATACACATATATACATATATACA-CATATATACATATATACACACATATATACACATA 653
              ||||| | ||||||| ||||||| |  ||||||| ||||||| | | ||||| | | |||
Db      44184 ATATATAGATATATAGATATATAGATGATATATAGATATATATAGATATATAGATATATA 44125

Qy        654 CATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATA 713
              ||||| | | | ||||||||| ||| ||||| | ||||||| |  |||| |||| ||||
Db      44124 GATATATAGATATATAGATATATATAGATATATAGATATATAGATATATAGATATATATA 44065

Qy        714 TATATAT 720
              |||||||
Db      44064 GATATAT 44058


RESULT 3
ADQ18307/c
ID   ADQ18307 standard; DNA; 121600 BP.
XX
AC   ADQ18307;
XX
DT   26-AUG-2004  (first entry)
XX
DE   Human soft tissue sarcoma-upregulated DNA - SEQ ID 1125.
XX
KW   soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO2004048938-A2.
XX
PD   10-JUN-2004.
XX
PF   26-NOV-2003; 2003WO-US038193.
XX
PR   26-NOV-2002; 2002US-0429739P.
XX
PA   (PROT-) PROTEIN DESIGN LABS INC.
XX
PI   Aziz N,  Ginsburg WM,  Zlotnik A;
XX
DR   WPI; 2004-441208/41.
XX
PT   Early detection of soft tissue sarcoma comprises determining expression
PT   of a gene in a first soft tissue sample and a normal soft tissue sample
PT   and comparing the gene expression, also useful in treating soft tissue
PT   sarcoma.
XX
PS   Example 2; SEQ ID NO 1125; 210pp; English.
XX
CC   The invention relates to a novel method for detecting soft tissue sarcoma
CC   which comprises obtaining a first soft tissue sample from an individual
CC   and a normal soft tissue sample from the same or different individual,
CC   determining the expression of a gene in both samples and comparing the
CC   expression of the gene in both soft tissue samples, where a higher level
CC   of protein expression in the first soft tissue sample indicates the
CC   presence of soft tissue sarcoma. The method of the invention has
```

```
CC   cytostatic applications and may be useful for detecting soft tissue
CC   sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC   acid sequences may be useful in diagnostic and screening applications.
CC   The current sequence is that of a human soft tissue sarcoma-upregulated
CC   DNA of the invention. The current sequence is not shown within the
CC   specification per se but was submitted in CD format by the inventor.
XX
SQ   Sequence 121600 BP; 38372 A; 24652 C; 24048 G; 34528 T; 0 U; 0 Other;

  Query Match             7.2%;  Score 356.6;  DB 12;  Length 121600;
  Best Local Similarity   74.7%;  Pred. No. 2.3e-52;
  Matches  543;  Conservative    0;  Mismatches  169;  Indels   15;  Gaps    7;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAA--- 57
              || || |||||| ||||    ||||  |||||||||| ||||||||||||||| |||||
Db      44776 AGATCGAGACCATCCTGGCTAACACAGTGAAACCCCGTCTCTACTAAAAATCCAAAAAAA 44717

Qy         58 --TTAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGG 115
                |||| | ||| |||||     | |||||| |||||||||||  |||||| ||||||||
Db      44716 AGTTAGCCGGGCGTGGTGGCGGGCACCTGTAGTCCCAGCTACTCTGGAGGCTGAGGCAGG 44657

Qy        116 AGAATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCAC 174
              |||||  | ||||| ||  || || |||| |||| | ||| | ||||   |||| |||||
Db      44656 AGAATGGCGTGAACTTG--GGAGGCGGAGCTTGCAGTGAGCCGAGATAGCGCCACTGCAC 44599

Qy        175 TCCAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATA 234
              ||||| |  ||| || ||||||| |   ||||||| |||||||| |       | ||||||
Db      44598 TCCAGTCTGGGCGAC-AGAGCGAGACTCCATCTCAAAAAAAAAATA----TATATATATA 44544

Qy        235 TATATATATATATATATACACACACACACACATATATATATACACATATATATACACATA 294
              || ||||||||| |||||||| | | | | | ||||||| ||| | | |||||||| | |||
Db      44543 TAGATATATATAGATATACAGATATATAGATATATATAGATATATAGATATATATAGATA 44484

Qy        295 TATACATATATACACATATATACACATATATATACACATACATATGTACACATATATATA 354
              |||| ||||||| | ||||||| | ||||||| | ||| |||| || | ||||||| |
Db      44483 TATAGATATATATAGATATATAGATATATATAGATATATAGATATATATAGATATATAGA 44424

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              | ||||  ||| | | ||||| | | ||||| | | ||||| | | ||||| | | | |
Db      44423 TATATATAGATATATAGATATATATAGATATATAGATATATAGATATATATAGATATATA 44364

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              |||||  | | ||||| | ||||||| | ||||| | ||||||| | ||||||| |||||
Db      44363 GATATATATAGATATATAGATATATATAGATATATAGATATATAGATATATATAGATATA 44304

Qy        475 TACACATATATATAATATACACACATATATATACACATATATACACACATATATACACAT 534
              || |  |||| || ||||| | | ||||| ||| | ||| ||| | | ||||||| | ||
Db      44303 TAGATGTATAGAT-ATATATAGATATATAGATATATATAGATATATAGATATATAGATAT 44245

Qy        535 ATATACACATATATATACACATATATACACATATATACATATATACACATATATATACAT 594
              |||||  | ||||| ||| | ||| ||| | ||||| | ||||| | | ||| ||||| |
Db      44244 ATATAGATATATAGATATATATAGATATATATATAGATATATAGATATATAGATATATAG 44185

Qy        595 ATATACACATATATACATATATACA-CATATATACATATATACACACATATATACACATA 653
              |||||  | ||||||| ||||||| |  ||||||| ||||||| | | ||||| | | |||
Db      44184 ATATATAGATATATAGATATATAGATGATATATAGATATATATAGATATATAGATATATA 44125

Qy        654 CATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATA 713
              |||||  | | | ||||||||| ||| ||||| | ||||||| |  |||| |||| ||||
Db      44124 GATATATAGATATATAGATATATATAGATATATAGATATATAGATATATAGATATATATA 44065

Qy        714 TATATAT 720
              |||||||
Db      44064 GATATAT 44058


RESULT 4
ADL08108/c
ID   ADL08108 standard; DNA; 188971 BP.
XX
AC   ADL08108;
XX
DT   20-MAY-2004  (first entry)
XX
DE   Human gene associated with low HDL-C APOA1.
XX
KW   Human; ds; SNP; single nucleotide polymorphism;
KW   high density lipoprotein-C; HDL-C; vascular disease; metabolic disease;
KW   coronary artery disease; gene.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       replace(123408,T)
FT                   /*tag=  a
FT                   /standard_name= "Single nucleotide polymorphism"
XX
PN   US2004043389-A1.
XX
PD   04-MAR-2004.
XX
PF   04-SEP-2002; 2002US-00235192.
XX
PR   04-SEP-2002; 2002US-00235192.
XX
PA   (VITI-) VITIVITY INC.
XX
PI   Mccarthy J;
XX
DR   WPI; 2004-214170/20.
XX
PT   Determining whether a subject has, or is at risk of developing, an
PT   abnormally low high density lipoprotein-C (HDL-C) level comprises
PT   detecting an allelic variant of a polymorphic region from any of a set of
PT   27 genes.
XX
PS   Disclosure; SEQ ID NO 27; 37pp; English.
XX
CC   The invention relates to determining whether a subject has, or is at risk
CC   of developing, an abnormally low high density lipoprotein-C (HDL-C) level
CC   comprises determining whether the subject has an allelic variant of a
CC   polymorphic region from any of 27 genes (alleles listed in Table 5 of the
CC   specification). Also included are determining whether a male subject has,
CC   or is at risk of developing, an abnormally low HDL-C level, comprising
CC   determining whether the male subject has an allelic variant of a
CC   polymorphic region listed in Table 5 which is associated with abnormally
CC   low HDL-C levels in males, and determining whether a female subject has,
CC   or is at risk of developing, an abnormally low HDL-C level, comprising
CC   determining whether the female subject has an allelic variant of a
CC   polymorphic region listed in Table 5 which is associated with abnormally
CC   low HDL-C levels in females. The allelic variant in determining whether a
CC   subject has, or is at risk of developing, an abnormally low HDL-C level
CC   is APOA 1 CC, CD14 1 CT, COL5A2 1 GG, EDNRB 1 AG or AA, FABP3 1 CT, GBE1
CC   1 AG or GG, LIPC 5 AA, MTHFR 1 CC, VWF 2 GG, or their complements. The
CC   allelic variant in determining whether a male subject has, or is at risk
CC   of developing, an abnormally low HDL-C level, LRP1 3 CC or CT, PAI2 4 GG,
CC   or PPARG 1 CG, or their complements. The allelic variants are also COL5A2
CC   1 GG, CD14 1 CT or CC, and FABP3 1 CT, in combination, or their
CC   complements. The methods are useful for diagnosing (a predisposition to)
CC   abnormally low levels of low high density lipoprotein-C (HDL-C) in a
CC   subject, The methods are useful in diagnosing (a predisposition to) or
CC   prognosticating diseases and disorders associated with abnormal lipid
CC   levels such as vascular and metabolic diseases, e.g., coronary artery
CC   disease. The present sequence is a human gene containing a SNP (single
CC   nucleotide polymorphism associated with low high density lipoprotein-C
CC   (HDL-C) levels.
XX
SQ   Sequence 188971 BP; 53091 A; 44581 C; 42653 G; 48646 T; 0 U; 0 Other;

  Query Match             6.6%;  Score 327.4;  DB 12;  Length 188971;
  Best Local Similarity   67.6%;  Pred. No. 2.7e-47;
  Matches  601;  Conservative    0;  Mismatches  261;  Indels   27;  Gaps    9;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC--AAAAT 58
              || ||||||||| |||| |||||||||||||| || ||||||||||||||||  |||||
Db     104585 AGATCAAGACCATCCTGGCCAACATGGTGAAACTCCGTCTCTACTAAAAATACAAAAAAT 104526
```

```
Qy         59 TAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGA 118
Db     104525 TAGCTGGGCGTGGTGGTGGGCTCCTGTAATCCCAGCTACTTGGGAGGCTGAAGCAGGAGA 104466

Qy        119 ATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTG---CAC 174
Db     104465 ATCGCTTGAACC--CAGGAGGCGGAGGTTGCAGTGAGCCAAGATCGCGCCATTGCACCAC 104408

Qy        175 TCCAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATA 234
Db     104407 TCCAGCCTGGGCAAAAAGAGAGAAACTCTGTCTCAAAAAAAAAAAAAAAAAAAAAAAAAA 104348

Qy        235 TATATATATATATATATACACACACACACACATATATATATACACATATATATACACATA 294
Db     104347 AATATATATATATATATATATATATACACACACATATATATTCATATATA-------A 104295

Qy        295 TATACATATATACACATATATACACATATATATACACATACATATGTACACATATATATA 354
Db     104294 TTCATATATATATTCATATATA---ATTCATATATATATTCATATATATTCATATATATA 104238

Qy        355 CACATATGT-ATACACATATATACACATATATACACACATATATACACATATATACACAC 413
Db     104237 TTCATATATAATTCATATATATATTCATATATATTCATATATATATTCATATATATATTC 104178

Qy        414 ATATATACACATATATACACATATATACACATATACACATATATACACATATATACATAT 473
Db     104177 ATATATA-----ATTCATATATATATTCATATATATTCATATATATATTCATATATATAT 104123

Qy        474 ATACACATATATATAATATACACACATATATATACACATATATACACACATATATACACA 533
Db     104122 TCATATATAATTCATATATATATTCATATATATTCATATATATATTCATATATATTCATA 104063

Qy        534 TATATACACATATATATACACATATATACACATATATACATATATACACATATATATACA 593
Db     104062 TATATATTCATATATATATTCATATATATTCATATATATATTCATATATATATTCATATA 104003

Qy        594 TATATACACATATATACATATATACACATATATACATATATACACACATATATACACATA 653
Db     104002 TATATTCATATATATATTCATATATATATTCATATATATATTCATATATATTCATATATA 103943

Qy        654 CATATACACACACATAGATATACA---TATATATACACATATATATACGTATATATATGT 710
Db     103942 TATTCATATATATATTCATATATAATTCATATATATATTCATATATATATTCATATATAA 103883

Qy        711 ATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGA 770
Db     103882 ATTCATATATAATTCATATATATATTCATATAGATATTCATATATAATTCATATAGATAT 103823

Qy        771 AAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTA 830
Db     103822 TCATAGATATTCATATATATTCATATAGATATTCATATATTCATATATATTCATATATTC 103763

Qy        831 GGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGA 879
Db     103762 ATGTATATTCATATAGATATTCATATATAGATATTCATAGATATTCATA 103714
```

RESULT 5
ABD32659
ID ABD32659 standard; DNA; 163382 BP.
XX
AC ABD32659;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human cancer-associated genomic DNA HD13-065.
XX
KW Human; ds; cancer-associated protein; gene; cytostatic; cancer;
KW leukaemia; lymphoma; CAP.
XX
OS Homo sapiens.
XX
PN WO2004074320-A2.
XX
PD 02-SEP-2004.
XX
PF 17-FEB-2004; 2004WO-US004730.
XX
PR 14-FEB-2003; 2003US-00367094.
PR 14-MAR-2003; 2003US-00388838.
PR 15-APR-2003; 2003US-00417375.
PR 13-JUN-2003; 2003US-00461862.
PR 15-SEP-2003; 2003US-00663431.
PR 15-DEC-2003; 2003US-00737318.
XX
PA (SAGR-) SAGRES DISCOVERY INC.
XX
PI Morris DW, Morris DW, Malandro MS;
XX
DR WPI; 2004-652914/63.
XX
PT New isolated cancer-associated polynucleotides and polypeptides useful
PT for diagnosing, preventing or treating cancers, especially lymphoma and
PT leukemia, or in screening for agents that modulate cancer.
XX
PS claim 16; seqid 199; 310pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising at least 10
CC contiguous nucleotides of any of the 233 polynucleotide sequences given
CC in the specification, or its complement. The nucleic acids encode cancer-
CC associated proteins. Also included are an expression vector comprising
CC the isolated nucleic acid cited above, a host cell comprising the above
CC recombinant nucleic acid or expression vector, a microarray for detecting
CC a cancer-associated (CA) nucleic acid comprising at least one probe
CC comprising at least 10 contiguous nucleotides of any of the above-
CC mentioned nucleotide sequences, an isolated polypeptide (encoded within
CC an open reading frame of a CA sequence selected from any of the 95
CC polynucleotide sequences as mentioned in the specification, or its
CC complement), an isolated antibody, (or its antigen binding fragment) that
CC binds to the above polypeptide, a hybridoma that produces the above
CC monoclonal antibody, a pharmaceutical composition comprising the above
CC antibody and a pharmaceutical excipient, a kit for detecting cancer
CC cells(comprising the antibody cited above, methods for diagnosing cancer
CC or for detecting the presence or absence of cancer cells in an
CC individual, a method for inhibiting growth of cancer cells in an
CC individual, a method for delivering a therapeutic agent to cancer cells
CC in an individual, an electronic library comprising the above
CC polynucleotide or polypeptide (or their fragments), methods of screening
CC for anticancer activity or for a bioactive agent capable of modulating
CC the activity of a CA protein (CAP), methods for detecting cancer
CC associated with expression of a polypeptide in a test cell sample, a
CC method for treating cancers and a method for inhibiting the expression of
CC CA gene in a cell. The composition and methods are useful for detecting,
CC diagnosing, preventing and treating cancers, especially lymphoma and
CC leukaemia. These may also be used in screening for agents that modulate
CC cancer. The present sequence is a human CAP genomic sequence. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 163382 BP; 50073 A; 25805 C; 27837 G; 53184 T; 0 U; 6483 Other;

Query Match 6.5%; Score 323.6; DB 13; Length 163382;
Best Local Similarity 68.2%; Pred. No. 1.2e-46;
Matches 495; Conservative 0; Mismatches 224; Indels 7; Gaps 3;

```
Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
Db     144405 AGTTCAAGACCAGCCTGGCCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAAATT 144464

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
Db     144465 AGCCAGGCGTGGTGGCACGTGCCTGTAATCCCAGCTACCCAGGAGGCTGAGGCAGGAGAA 144524

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
```

```
Db     144525 TTGCTTGAAGC-CAAGAGGCAGAAGTTGCAGTGAGCCGAGATCATGCCATTGCACTCCAG 144583

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239

Db     144584 CCTGGGGGACAAGAACGAGACTTCATGTCAAAAAAAAAAAATACACAGTCTTCCTTGGCT 144643

Qy        240 ATATATATATATACACACACACACACATATATATATACACATATATATACACATATATAC 299

Db     144644 TAGCTGTATGGGAAATATATATATATATTTATTTATATATATATTTATATATATTTATTT 144703

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACAT 359

Db     144704 ATATATATTTATTTATATATATTTATTTATATATATTTATTTATATATTTATTTATATTT 144763

Qy        360 ATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATAT 419

Db     144764 ATTTTTATATATTTATTTATATATATTTGTATATATTTATTTATATATTTATTTATATAT 144823

Qy        420 ACACATATATACACATATATACACATATACACATATATACACATATATACATA-----TA 474

Db     144824 TTATTTATATATTTATATATATTAGTTCATATATATTTATATATTTATATATATTAGTTA 144883

Qy        475 TACACATATATATAATATACACACATATATATACACATATATACACACATATATACACAT 534

Db     144884 TATATATTTATATATTAGTTATATATATTTATATATATTTTTATATATTTATTTATATAT 144943

Qy        535 ATATACACATATATATACACATATATACACATATATACATATATACACATATATATACAT 594

Db     144944 ATGTATATATTAACATATTTATATATATTCATATATTTATATATAAATATATATATTAAT 145003

Qy        595 ATATACACATATATACATATATACACATATATACATATATACACACATATATACACATAC 654

Db     145004 ATATTTATATATATTTATATATTAATATATGTTTATATATTTATACATATTTATGTATAT 145063

Qy        655 ATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATAT 714

Db     145064 TAATATATATATTTATATATATTTATATATGTATTTATATATATTTATATATTTATATTT 145123

Qy        715 ATATAT 720

Db     145124 ATTTAT 145129
```

RESULT 6
AAS46748/c
ID AAS46748 standard; DNA; 26997 BP.
XX
AC AAS46748;
XX
DT 18-DEC-2001 (first entry)
XX
DE Tumour suppressor gene derived chemically modified sequence #472.
XX
KW Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer;
KW tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW cytosine methylation; ds.
XX
OS Homo sapiens.
XX
PN WO200168912-A2.
XX
PD 20-SEP-2001.
XX
PF 15-MAR-2001; 2001WO-EP002955.
XX
PR 15-MAR-2000; 2000DE-01013847.
PR 06-APR-2000; 2000DE-01019058.
PR 07-APR-2000; 2000DE-01019173.
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2001-602752/68.
XX
PT Fragments of chemically modified genes associated with tumor suppressor
PT genes and oncogenes, useful in designing primers and probes for analyzing
PT diseases associated with cytosine methylation state e.g. cancer.
XX
PS Claim 1; SEQ ID NO 472; 27pp; English.
XX
CC The invention relates to a nucleic acid comprising a sequence of 18
CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC bisulphite, of genes associated with tumour suppression and oncogenes
CC having a sequence taken from 536 (actually 533 since numbers 408, 458 and
CC 500 are missing from the sequence listing) sequences (Ss) and sequences
CC complementary to (Ss). The nucleic acid may be a peptide nucleic acid-
CC oligomer (PNA) of at least 9 nucleotides and may form part of a set of
CC probes for detecting the cytosine methylation state and/or single
CC nucleotide polymorphisms and also to be used in an array for analysing
CC diseases associated with CpG dinucleotides e.g. cancers and tumours. The
CC probes can also be used in a method for ascertaining genetic and/or
CC epigenetic parameters for the diagnosis and/or therapy of existing
CC diseases or the predisposition to specific diseases, by analysing
CC cytosine methylations. The parameters may be compared to another set of
CC genetic and/or epigenetic parameters, the differences serving as basis
CC for diagnosis and/or prognosis events which are disadvantageous to
CC patients. The present sequence is one of the 533 genomic sequences
CC derived from tumour suppressor genes and oncogenes. Sequences with even
CC numbered Seq ID numbers are the complementary sequence of the
CC corresponding odd numbered sequence (e.g. ID 2 and ID1, ID 536 and ID
CC 535, except for those whose partner sequence is missing). Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 26997 BP; 7650 A; 354 C; 5924 G; 13069 T; 0 U; 0 Other;

Query Match 6.5%; Score 322.2; DB 4; Length 26997;
Best Local Similarity 66.0%; Pred. No. 1.9e-46;
Matches 498; Conservative 0; Mismatches 253; Indels 4; Gaps 2;

```
Qy        207 TCAGAAAAAAAAAAAGGCATTTTATATATATATATATATATACACACACACACACA 266

Db       6221 TAAATAAAAAAAATATAATATATATATATATATATATATATATACACACATATATATA 6162

Qy        267 TATATATATACACATATATATACACATATATACATATATACACATATATACACATATATA 326

Db       6161 AACATATATACACACACATATACATATATACACACATATATACATATATACACATATATA 6102

Qy        327 TACACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATA 386

Db       6101 TATATATATACATATATATACATATATATATACACATATATATATATACATATATATA 6042

Qy        387 CACACATATATACACATATATACACACATATATACACATATATACACATATATACACATA 446

Db       6041 CATATACATATACACATATATATATATATACATATATATACATATACATATACACATATA 5982

Qy        447 TACACATATATACACATATATACATATATACACATATATATA-ATATACACACATATATA 505

Db       5981 TATATATATACATATACATACATATATATACACATATATATATATATACATATACATACA 5922

Qy        506 TACACATATATACACACATATATACACATATATACACATATATATACACATATATACACA 565

Db       5921 TATATATACACATATATATATATACATATATACACGTATATACATATATACACACATA 5862

Qy        566 TATATACATATATACACATATATATACATATATACACATATATACATATATACACA---T 622

Db       5861 TTTATATATACGTATATACATATATATATATATACATATATATATATATATACATATT 5802

Qy        623 ATATACATATATACACACATATATACACATACATATACACACACATAGATATACATATAT 682

Db       5801 ATATATATATATATATACATATATAAATATACATATATATATATATATACATATATACAT 5742
```

```
Qy        683 ATACACATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTA 742
              ||| | ||||||| |  |||||||||| ||||  |||||          || |      ||
Db       5741 ATATATATATATAAATATATATATATATATACGTATATATATATACACACACACACCATA 5682

Qy        743 GCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAA 802
              |    ||   | |      | | |  |||| |   |||   |  | |  | |   | |
Db       5681 AAATACTATACACCCATAAAAAAAAAAAAAAATAACATTTACAACAACTTAAATAAAATTA 5622

Qy        803 GGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTAT 862
              |    | |   | | | ||  | |     | | | |   |  | | ||| |  ||
Db       5621 AAAACCATTATTCTAACCGAAATAACTCAAAAATAAAAAACCAAATATTATATATTCTCA 5562

Qy        863 CTGTAAGGTAGTTAAGAAGGTAACACTATATATATATATAGTGAAAGCAGTGTAAACCTA 922
              ||   |  ||   |  ||  ||  |  | | | | | |||   | |  |   |||| |
Db       5561 CTTATAACTAAAAACTAAACTATAAAAACACAAAAACATAAAAATAATACAATAAACTTT 5502

Qy        923 AAGGATGGGCCAAGGATTTAAATGTTATAGAAGAA 957
              ||  |      ||  |   ||  || | || ||
Db       5501 AAAAACTCTAAAAAAAACATAAAAATAAATAAAAA 5467
```

RESULT 7
AAS46747/c
ID AAS46747 standard; DNA; 26997 BP.
XX
AC AAS46747;
XX
DT 18-DEC-2001 (first entry)
XX
DE Tumour suppressor gene derived chemically modified sequence #471.
XX
KW Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer;
KW tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW cytosine methylation; ds.
XX
OS Homo sapiens.
XX
PN WO200168912-A2.
XX
PD 20-SEP-2001.
XX
PF 15-MAR-2001; 2001WO-EP002955.
XX
PR 15-MAR-2000; 2000DE-01013847.
PR 06-APR-2000; 2000DE-01019058.
PR 07-APR-2000; 2000DE-01019173.
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2001-602752/68.
XX
PT Fragments of chemically modified genes associated with tumor suppressor
PT genes and oncogenes, useful in designing primers and probes for analyzing
PT diseases associated with cytosine methylation state e.g. cancer.
XX
PS Claim 1; SEQ ID NO 471; 27pp; English.
XX
CC The invention relates to a nucleic acid comprising a sequence of 18
CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC bisulphite, of genes associated with tumour suppression and oncogenes
CC having a sequence taken from 536 (actually 533 since numbers 408, 458 and
CC 500 are missing from the sequence listing) sequences (Ss) and sequences
CC complementary to (Ss). The nucleic acid may be a peptide nucleic acid-
CC oligomer (PNA) of at least 9 nucleotides and may form part of a set of
CC probes for detecting the cytosine methylation state and/or single
CC nucleotide polymorphisms and also to be used in an array for analysing
CC diseases associated with CpG dinucleotides e.g. cancers and tumours. The
CC probes can also be used in a method for ascertaining genetic and/or
CC epigenetic parameters for the diagnosis and/or therapy of existing
CC diseases or the predisposition to specific diseases, by analysing
CC cytosine methylations. The parameters may be compared to another set of
CC genetic and/or epigenetic parameters, the differences serving as basis
CC for diagnosis and/or prognosis events which are disadvantageous to
CC patients. The present sequence is one of the 533 genomic sequences
CC derived from tumour suppressor genes and oncogenes. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 26997 BP; 7233 A; 354 C; 6190 G; 13220 T; 0 U; 0 Other;

```
  Query Match             6.3%;  Score 311.2;  DB 4;  Length 26997;
  Best Local Similarity   76.9%;  Pred. No. 1.6e-44;
  Matches  409;  Conservative    0;  Mismatches  113;  Indels   10;  Gaps    2;

Qy        203 CATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATATACACACACACA 262
              | | ||| ||| | | ||        | ||||||||||||||||| | | || || | | |
Db      21339 CTTTTCATAAATACATAATATTCCATAATATATATATATATATACACACACGCATATATA 21280

Qy        263 CACATATATATATACACATATATATACACATATATACATATATACACATATATACACATA 322
              |||||||    | ||| ||| ||||| | |||||||||||||| ||| ||||||| |  ||
Db      21279 CACATATACCCACACATATACATATATATATATATACATATACACATATATATATATCTA 21220

Qy        323 TATATACACATACATATGT-----ACACATATATATACACATATGTATACACATATATAC 377
              |||||||||  | |||  |     | | ||| | ||| ||||||  |||| ||| ||||||
Db      21219 CATATACACACATATACATATATAATATATACACATATACATATATATATACACATATAC 21160

Qy        378 ACATATATACACACATATATACACATATATACACACATATATACACATATATACACATAT 437
              |||||||  | || ||  |||||  |||||||||  | ||||||||||||  ||||  ||| |||
Db      21159 ACATATACATACGCACATATAAACATATATATATACATATATACACGTATACACATATAC 21100

Qy        438 ATACACATATACACATATATACACATATATACATATATACACATATATATA-----ATAT 492
              |||  ||| |||  | |||||||||  |||||  |||||||  |  |  |||  |||||      |||
Db      21099 ATATACACATATATATATATACATATATACACATATACATATATACATATATATATATAC 21040

Qy        493 ACACACATATATATACACATATATACACACATATATACACATATATACACATATATATAC 552
              | | | |  | ||||||| |||||  | |  | |||||||  ||||||||||||||||  |||
Db      21039 ATATATACACATATACATATATACATATATATATATATACATATATACACATATACATAT 20980

Qy        553 ACATATATACACATATATACATATATACACATATATATACATATATACACATATATACAT 612
              | | |||||  | ||||||||||||||||||||||||  |||  |||  |||  | |||||||  |
Db      20979 ATACATATATATATATATACATATATACACATATACATATATACATATATATATATAC 20920

Qy        613 ATATACACATATATACATATATACACACATATATACACATACATATACACACACATAGAT 672
              |||||  |||  ||||||||||||||  | |||||||  | |||  |||||  | |  | |||  ||
Db      20919 ATATATACACATATACATATATACATATATATATATATATATATATATATATATACAT 20860

Qy        673 ATACATATATATACACATATATATACGTATATATATGTATATATATATGCTC 724
              |||  ||||||||||| ||||||  ||  ||||||||| ||||||||||  | |
Db      20859 ATATATATATATACATATATATCTATATATATATATATATATATATATACAC 20808
```

RESULT 8
ADC87572
ID ADC87572 standard; DNA; 1990 BP.
XX
AC ADC87572;
XX
DT 01-JAN-2004 (first entry)
XX
DE Human GPCR gene SEQ ID NO:2025.
XX
KW ds; gene; human; GPCR;
KW guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1270724-A2.
XX
PD 02-JAN-2003.
XX

```
PF   18-JUN-2002; 2002EP-00013517.
XX
PR   18-JUN-2001; 2001JP-00246789.
XX
PA   (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA   (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI   Suwa M,  Asai K,  Akiyama Y,  Aburatani H;
XX
DR   WPI; 2003-315783/31.
DR   P-PSDB; ADC87573.
XX
PT   New polynucleotide, useful for preparing a composition for treating a
PT   patient in need of increased or suppressed activity or expression of the
PT   guanosine triphosphate-binding protein coupled receptor.
XX
PS   Claim 1; SEQ ID NO 2025; 28pp; English.
XX
CC   The invention relates to a novel polynucleotide encoding a guanosine
CC   triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC   the invention may have a use in gene therapy. The polynucleotide and
CC   polypeptide are useful for preparing a composition for treating a patient
CC   in need of increased or suppressed activity or expression of the
CC   guanosine triphosphate-binding protein coupled receptor. The
CC   polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC   invention.
XX
SQ   Sequence 1990 BP; 833 A; 366 C; 180 G; 611 T; 0 U; 0 Other;

  Query Match             6.2%;  Score 308;  DB 10;  Length 1990;
  Best Local Similarity   80.2%;  Pred. No. 4.8e-44;
  Matches  401;  Conservative    0;  Mismatches   90;  Indels    9;  Gaps    3;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | || | | |||| |||||||| | | | | ||||||| || ||||||||||||||||||
Db        826 TATACACACATATGTATATATACACATATATACACACATATGTATATATACACATATATA 885

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              |||||||| || ||||| ||| ||||| |||||||| |||| | | |||||| ||||||
Db        886 CACACATATGTATATATACACATATATACACACATATGTATATATACACATATATACACA 945

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||| || | |||   ||| | | |  || |  |||||| ||||||||||||||||| |
Db        946 CATATGTATATATACACATATATACACGTATATGTATATATACACATATATACACATACA 1005

Qy        407 TACACA----CATATATACACATATATACACATATATACACATATACACATATATACACA 462
              ||||||     ||||| ||||||||||||||||| ||||||||||  ||||||||||
Db       1006 TACACATATGTATATACACACATATATACACATACATACACATATACGTATATATACACA 1065

Qy        463 TATATACATATATACACATATATATAATATACACACATATATATACACATATATACACAC 522
              |||| ||||| |||||||||||  | ||||| |||||||||| ||| ||   || | ||
Db       1066 TATACACATACATACACATATACGT-ATATATACACATATATACACATATGCGTATATAC 1124

Qy        523 ATATATACACATATATACACATATATATACACATATATACACATATATACATATATACAC 582
              ||||||||||||||||||||    ||||||  ||||||||||||||||||| ||||||
Db       1125 ATATATACACATATATACAC----ATATACGTATATATACACATATATACACATATACGT 1180

Qy        583 ATATATATACATATATACACATATATACATATATACACATATATACATATATACACACAT 642
              ||||||| |||||||||||||||||||||| |||||   ||||| ||||||||||||| ||
Db       1181 ATATATACACATATATACACATATATACACATATGTGTATATACACATATATACACATAT 1240

Qy        643 ATATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTAT 702
              |||| | | ||||||| ||||| |||  |||| ||| ||||| |||  ||| ||  |||
Db       1241 GTATATATACACATATATACACATATATGTATATATACATATATACACGTATGTATATAT 1300

Qy        703 ATATATGTATATATATATGC 722
              | | || ||| ||||||| |
Db       1301 ACACATATATGTATATATAC 1320


RESULT 9
AAA63785
ID   AAA63785 standard; DNA; 15418 BP.
XX
AC   AAA63785;
XX
DT   04-DEC-2000  (first entry)
XX
DE   Nucleotide sequence of the human TERT promoter and gene from pGRN142.
XX
KW   Telomerase reverse transcriptase; TERT; promoter; telomerase complex;
KW   telomere length; hyperplasic disease; cancer; oncolytic virus;
KW   cis-acting transcriptional control sequence; viral replication;
KW   cell proliferation; aging; immunological disorder; infertility; ss.
XX
OS   Homo sapiens.
XX
PN   WO200046355-A2.
XX
PD   10-AUG-2000.
XX
PF   04-FEB-2000; 2000WO-US003104.
XX
PR   04-FEB-1999;   99US-00244438.
XX
PA   (GERO-) GERON CORP.
XX
PI   Morin GB,  Lichtsteiner S,  Vasserot A,  Adams R,  Cardoza LM;
PI   Lebkowski JS;
XX
DR   WPI; 2000-532898/48.
XX
PT   New polynucleotides comprising cis-acting transcriptional control
PT   sequences, e.g. promoter sequence, of telomerase reverse transcriptase
PT   genes, useful in the treatment of cancer.
XX
PS   Claim 9; Page 58-61; 63pp; English.
XX
CC   The present sequence represents the human telomerase reverse
CC   transcriptase (TERT) promoter and gene. TERT is part of the telomerase
CC   complex responsible for maintaining telomere length and increasing
CC   replicative capacity of progenitor cells. Telomerase activity is turned
CC   off in mature differentiated cells, but is turned back on again in
CC   hyperplasic diseases, including many cancers. The polynucleotide
CC   comprises cis-acting transcriptional control sequences, e.g. promoter
CC   sequence. These promoter sequences are used to produce oncolytic viruses,
CC   in which a toxin or genetic element essential for viral replication is
CC   placed under the control of a TERT promoter. As a result, the virus
CC   replicates preferentially in cells expressing TERT, and selectively lyses
CC   cancer cells. The oncolytic viruses are useful for treating cancer in
CC   humans or animals. The TERT promoter sequences are useful in the
CC   treatment of cancer and other diseases of cell proliferation such as
CC   degenerative and aging processes and diseases of aging, immunological
CC   disorders, or infertility
XX
SQ   Sequence 15418 BP; 4518 A; 3797 C; 3765 G; 3338 T; 0 U; 0 Other;

  Query Match             6.0%;  Score 296;  DB 3;  Length 15418;
  Best Local Similarity   69.2%;  Pred. No. 6.5e-42;
  Matches  479;  Conservative    0;  Mismatches  200;  Indels   13;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||| |||||||||||  ||| || | |||||||||||| ||||||||| || ||||||
Db        646 AGTTCGAGACCAGCCTGGCCAATATAGCGAAACCCCATCTGTACTAAAAACACAAAAATT 705

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   ||  |||||| ||| |||||||||||||| ||||  ||||||  ||||||||||
Db        706 AGCTGGGTGTGGTGACACATGCCTGTAATCCCAGGTACTCAGGAGGCTAAGGCAGGAGAA 765

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              | |||||||  |  ||  | || |||     |   ||||||    ||| ||||||||||
Db        766 TTGCTTGAACTGGGAGGCAGAGGTGGT-----GAGCCAAGATTGCACCAGTGCACTCCAG 820

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  ||  || |||| ||||   ||||||| ||| ||| |||   | |  ||||| |||
Db        821 CCTTGGTGAC-AGAGTGAAACTCCATCTCAAAAACAAACAAACAAAATACATATACATAA 879
```

```
Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              ||||||||  | | | | | ||| | ||| ||||||| ||| ||||||| |  |||||||
Db        880 ATATATATGCACATATATATACATATATAAATATATATACACATATATAAATCTATATAC 939

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACAT-----ATATATA 354
              ||||||||| ||||| ||| ||||| ||  | |||||||| || ||||     ||||||
Db        940 ATATATACATATATACACATATATAAATCTATATACATATATATACATATATAATATATT 999

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              |||||| || | | ||| ||| | | ||||||| | ||| ||||| ||||| | | |||
Db       1000 TACATATATAAATATATACATATATAAATATACATATATAAATACATATATAAATATACA 1059

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              |||||| | ||| ||| | | ||||||| ||||| | ||||| | | ||| ||| | |||
Db       1060 TATATAAATATACATATATAAATATACATATATAAATATATACATATATAAATATACATA 1119

Qy        475 TACACATATATATA-ATATACACACATATATATACACATATATACACACATATATACACA 533
              || | ||||||||| ||||| | | ||||| ||| |||  |||| ||| |||||||||  |
Db       1120 TATAAATATATATACATATATAAATATATAAATATACAAGTATATACAAATATATACATA 1179

Qy        534 TATATACACATATATATACACATATATACACATATATACATATATACACATATATATACA 593
              |||| |   |||||  || | | ||||| | ||| ||| ||| | | || | |   |
Db       1180 TATAAATGTATATACGTATATACATATATATATAAATATATAAAAAAACTTTTGGCTGGG 1239

Qy        594 TATATACACATATATACATATATACACATATATACATATATACACACATATATACACATA 653
              |  |   || || |      | | | |  | |     ||  | || ||| | | |||||
Db       1240 CACCTTTCCAAATCTCATGGCACATATAAGTCTCATGGTAACCTCAAATAAAAAAACATA 1299

Qy        654 CATATACACACACATAGATATACATATATATA 685
              |          || | ||| | |   ||| |
Db       1300 TAACAGATACACCAAAAATAAAAACCAATAAA 1331
```

RESULT 10
AAL38601
ID AAL38601 standard; DNA; 15418 BP.
XX
AC AAL38601;
XX
DT 16-AUG-2002 (first entry)
XX
DE Human TERT promoter and upstream sequence.
XX
KW Cytostatic; glycosyltransferase; tumour; cell-surface carbohydrate;
KW tissue specific; transcriptional control element; cancer cell;
KW gene therapy; human TERT; telomerase reverse transcriptase; ds.
XX
OS Homo sapiens.
XX
PN WO200242468-A2.
XX
PD 30-MAY-2002.
XX
PF 26-NOV-2001; 2001WO-US044306.
XX
PR 27-NOV-2000; 2000US-0253395P.
XX
PA (GERO-) GERON CORP.
XX
PI Schiff MJ;
XX
DR WPI; 2002-479954/51.
XX
PT New polynucleotide encoding glycosyltransferase enzymes including histo
PT blood group transferase useful for treating conditions associated with
PT hyperproliferation, such as cancers and other neoplasias.
XX
PS Claim 9; Page 17-20; 49pp; English.
XX
CC The invention relates to a polynucleotide comprising an encoding sequence
CC for a glycosyltransferase, under control of a heterologous tumour
CC specific or tissue specific transcriptional control element, where
CC expression of the polynucleotide in a human cell causes the cell to
CC express a cell-surface carbohydrate determinant to which some or all
CC humans have a naturally occurring antibody. The polynucleotide of the
CC invention is useful for killing cancer cells and for preparing a
CC medicament for the treatment of cancer. The polynucleotide sequence of
CC the invention can be used to treat disorders by gene therapy. This
CC polynucleotide sequence represents the human TERT (telomerase reverse
CC transcriptase) promoter and upstream sequence related to the invention
XX
SQ Sequence 15418 BP; 4518 A; 3797 C; 3765 G; 3338 T; 0 U; 0 Other;

Query Match 6.0%; Score 296; DB 6; Length 15418;
Best Local Similarity 69.2%; Pred. No. 6.5e-42;
Matches 479; Conservative 0; Mismatches 200; Indels 13; Gaps 5;

```
Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||| ||||||||||  ||| || | ||||||||||| |||||||| || ||||||
Db        646 AGTTCGAGACCAGCCTGGCCAATATAGCGAAACCCCATCTGTACTAAAAACACAAAAATT 705

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   ||  |||||| ||| |||||||||||||| ||||  ||||||  ||||||||||
Db        706 AGCTGGGTGTGGTGACACATGCCTGTAATCCCAGGTACTCAGGAGGCTAAGGCAGGAGAA 765

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              |  |||||||  |  ||  | || |||     |   ||||||    ||| ||||||||||
Db        766 TTGCTTGAACTGGGAGGCAGAGGTGGT-----GAGCCAAGATTGCACCAGTGCACTCCAG 820

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  ||  || |||| ||||   ||||||| ||| ||| |||   | |  ||||| |||
Db        821 CCTTGGTGAC-AGAGTGAAACTCCATCTCAAAAACAAACAAACAAAATACATATACATAA 879

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              ||||||||  | | | | | ||| | ||| ||||||| ||| ||||||| |  |||||||
Db        880 ATATATATGCACATATATATACATATATAAATATATATACACATATATAAATCTATATAC 939

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACAT-----ATATATA 354
              ||||||||| ||||| ||| ||||| ||  | |||||||| || ||||     ||||||
Db        940 ATATATACATATATACACATATATAAATCTATATACATATATATACATATATAATATATT 999

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              |||||| || | | ||| ||| | | ||||||| | ||| ||||| ||||| | | |||
Db       1000 TACATATATAAATATATACATATATAAATATACATATATAAATACATATATAAATATACA 1059

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              |||||| | ||| ||| | | ||||||| ||||| | ||||| | | ||| ||| | |||
Db       1060 TATATAAATATACATATATAAATATACATATATAAATATATACATATATAAATATACATA 1119

Qy        475 TACACATATATATA-ATATACACACATATATATACACATATATACACACATATATACACA 533
              || | ||||||||| ||||| | | ||||| ||| |||  |||| ||| |||||||||  |
Db       1120 TATAAATATATATACATATATAAATATATAAATATACAAGTATATACAAATATATACATA 1179

Qy        534 TATATACACATATATATACACATATATACACATATATACATATATACACATATATATACA 593
              |||| |   |||||  || | | ||||| | ||| ||| ||| | | || | |   |
Db       1180 TATAAATGTATATACGTATATACATATATATATAAATATATAAAAAAACTTTTGGCTGGG 1239

Qy        594 TATATACACATATATACATATATACACATATATACATATATACACACATATATACACATA 653
              |  |   || || |      | | | |  | |     ||  | || ||| | | |||||
Db       1240 CACCTTTCCAAATCTCATGGCACATATAAGTCTCATGGTAACCTCAAATAAAAAAACATA 1299

Qy        654 CATATACACACACATAGATATACATATATATA 685
              |          || | ||| | |   ||| |
Db       1300 TAACAGATACACCAAAAATAAAAACCAATAAA 1331
```

RESULT 11
ABS54997
ID ABS54997 standard; DNA; 15418 BP.
XX
AC ABS54997;
XX
DT 10-DEC-2002 (first entry)
XX
DE Lambda clone containing human TERT genomic insert.

```
XX
KW   Telomerase reverse transcriptase; TERT; replication-conditional virus;
KW   adenovirus replication gene; cancer cell; lung; pancreatic cancer;
KW   medulloblastoma; cervical carcinoma; fibrosarcoma; osteosarcoma;
KW   cytolysis; replication defective adenovirus vector; congenital defect;
KW   proinflammatory; antiinflammatory; heterologous effector gene;
KW   cancer therapy; cytostatic; gene therapy; lambda clone; human; ds.
XX
OS   Bacteriophage lambda.
OS   Homo sapiens.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    1. .43
FT                   /*tag=  a
FT                   /note= "From Lambda clone"
FT   misc_feature    44. .15375
FT                   /*tag=  b
FT                   /note= "Human telomerase reverse transcriptase (hTERT)
FT                   sequence including the hTERT promoter and upstream
FT                   sequence"
FT   misc_feature    15376. .15418
FT                   /*tag=  c
FT                   /note= "From Lambda clone"
XX
PN   WO200253760-A2.
XX
PD   11-JUL-2002.
XX
PF   17-DEC-2001; 2001WO-US048785.
XX
PR   18-DEC-2000; 2000US-0256418P.
XX
PA   (GERO-) GERON CORP.
XX
PI   Irving JM,  Lebkowski JS;
XX
DR   WPI; 2002-723123/78.
XX
PT   Novel replication-conditional virus useful for cytolysis of target cells
PT   e.g. cancer cells and preparing a medicament for treating cancer,
PT   comprises heterologous replication element in an adenovirus-based
PT   construct.
XX
PS   Claim 11; Page 26-29; 32pp; English.
XX
CC   The present invention relates to a new replication-conditional virus with
CC   a genome comprising adenovirus replication genes and one or more
CC   heterologous gene(s) that functionally replaces one or more adenovirus
CC   gene(s) required for replication or assembly of the virus. The invention
CC   is useful for killing a cancer cell (such as lung, pancreatic cancer,
CC   medulloblastoma, cervical carcinoma, fibrosarcoma or osteosarcoma),
CC   killing a cell expressing TERT (telomerase reverse transcriptase), and in
CC   preparing a medicament for treating cancer and a condition associated
CC   with increased expression of TERT in affected cells, in a subject. The
CC   invention is also useful for cytolysis of specific target cells. The
CC   invention is further useful for producing replication defective
CC   adenovirus vector which is useful for transient expression of a
CC   heterologous therapeutic gene to correct a congenital defect, introducing
CC   proinflammatory or antiinflammatory activity, enhancing telomerase
CC   function, and delivering heterologous effector genes that induce killing
CC   of the transduced cells. The invention is more safe for use in cancer
CC   therapy. The present nucleic acid sequence represents the human TERT
CC   sequence contained within a lambda clone sequence of the invention
XX
SQ   Sequence 15418 BP; 4518 A; 3797 C; 3765 G; 3338 T; 0 U; 0 Other;

  Query Match             6.0%;  Score 296;  DB 6;  Length 15418;
  Best Local Similarity   69.2%;  Pred. No. 6.5e-42;
  Matches  479;  Conservative    0;  Mismatches  200;  Indels   13;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||| ||||||||||  ||| || | ||||||||||| |||||||| || ||||||
Db        646 AGTTCGAGACCAGCCTGGCCAATATAGCGAAACCCCATCTGTACTAAAAACACAAAAATT 705

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   ||  |||||| ||| |||||||||||||| ||||  ||||||  ||||||||||
Db        706 AGCTGGGTGTGGTGACACATGCCTGTAATCCCAGGTACTCAGGAGGCTAAGGCAGGAGAA 765

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              | |||||||  |  ||  | || |||     |   ||||||    ||| ||||||||||
Db        766 TTGCTTGAACTGGGAGGCAGAGGTGGT-----GAGCCAAGATTGCACCAGTGCACTCCAG 820

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  ||  || |||| ||||   ||||||| ||| ||| |||   | |  ||||| |||
Db        821 CCTTGGTGAC-AGAGTGAAACTCCATCTCAAAAACAAACAAACAAAATACATATACATAA 879

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              ||||||||  | | | | | ||| | ||| ||||||| ||| ||||||| |  |||||||
Db        880 ATATATATGCACATATATATACATATATAAATATATATACACATATATAAATCTATATAC 939

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACAT-----ATATATA 354
              |||||||| ||||| ||| ||||| ||  | |||||||| || ||||     ||||||
Db        940 ATATATACATATATACACATATATAAATCTATATACATATATATACATATATAATATATT 999

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              |||||| || | | ||| ||| | | ||||||| | ||| |||||  ||||| | | |||
Db       1000 TACATATATAAATATATACATATATAAATATACATATATAAATACATATATAAATATACA 1059

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              |||||| | ||| ||| | | ||||||| ||||| | ||||| | | ||| ||| | |||
Db       1060 TATATAAATATACATATATAAATATACATATATAAATATATACATATATAAATATACATA 1119

Qy        475 TACACATATATATA-ATATACACACATATATATACACATATATACACACATATATACACA 533
              || | |||||||||| ||||| | | ||||| ||| |||  |||| ||| |||||||| |
Db       1120 TATAAATATATATACATATATAAATATATAAATATACAAGTATATACAAATATATACATA 1179

Qy        534 TATATACACATATATATACACATATATACACATATATACATATATACACATATATATACA 593
              |||| |   ||||| || | | ||||| | ||| ||| ||| | | || | |   |
Db       1180 TATAAATGTATATACGTATATACATATATATATAAATATATAAAAAAACTTTTGGCTGGG 1239

Qy        594 TATATACACATATATACATATATACACATATATACATATATACACACATATATACACATA 653
              |  |   || || |     | | | |  | |     ||  | || ||| | | |||||
Db       1240 CACCTTTCCAAATCTCATGGCACATATAAGTCTCATGGTAACCTCAAATAAAAAAACATA 1299

Qy        654 CATATACACACACATAGATATACATATATATA 685
              |          || | ||| | |   ||| |
Db       1300 TAACAGATACACCAAAAATAAAAACCAATAAA 1331


RESULT 12
AAL38595
ID   AAL38595 standard; DNA; 15418 BP.
XX
AC   AAL38595;
XX
DT   16-AUG-2002  (first entry)
XX
DE   DNA of plasmid pGRN144 containing human hTERT gene.
XX
KW   Anticonvulsant; cerebroprotective; vasotropic; nootropic; stem cell;
KW   neuroprotective; antiparkinsonian; antiinflammatory; undifferentiated;
KW   cardiant; transcriptional control element; human embryonic; tissue graft;
KW   regenerative medicine; tissue reconstitution; neural progenitor cell;
KW   nerve system; epilepsy; stroke; ischaemia; Huntington's disease;
KW   Parkinson's disease; multiple sclerosis; leukodystrophy; neuritis;
KW   hepatocyte; liver damage; cardiomyocyte; cryoinjury; cardiac muscle;
KW   pGRN144; human hTERT gene; chimeric; ds.
XX
OS   Homo sapiens.
OS   Unidentified.
OS   Chimeric.
XX
PN   WO200242445-A2.
XX
PD   30-MAY-2002.
```

```
XX
PF   26-NOV-2001; 2001WO-US044309.
XX
PR   27-NOV-2000; 2000US-0253357P.
PR   27-NOV-2000; 2000US-0253443P.
PR   13-FEB-2001; 2001US-00783203.
XX
PA   (GERO-) GERON CORP.
XX
PI   Gold JD,  Lebrowski JS;
XX
DR   WPI; 2002-479952/51.
XX
PT   Depleting a cell e.g., human embryonic stem cell population of
PT   undifferentiated stem cells (UC) for use in regenerative medicine
PT   comprises genetically altering UC in a population to express nucleic acid
PT   encoding a lethal product.
XX
PS   Example 9; Page 62-66; 67pp; English.
XX
CC   The invention relates to a system for depleting a cell population of
CC   undifferentiated stem cells, by introducing nucleic acids of structure P-
CC   X, where X is a nucleic acid encoding product that is lethal to a cell in
CC   which it is expressed; or renders a cell in which it is expressed
CC   susceptible to the lethal effect of an external agent, and P is a
CC   transcriptional control element causing X to be preferentially expressed
CC   in the undifferentiated stem cells. The system is used for depleting a
CC   cell population (preferably, human embryonic stem cells) of
CC   undifferentiated stem cells. A population of differentiated cells is
CC   useful in regenerative medicine, and for preparing antibodies and cDNA
CC   libraries that are specific for a differentiated phenotype. The cell
CC   populations are also useful for drug screening and therapeutic
CC   applications. The differentiated cells are useful for tissue
CC   reconstitution or regeneration in a human patient in need of treatment.
CC   The cells are administered in a manner that permits to graft to the
CC   intended tissue side and reconstitute or regenerate the functionally
CC   deficient area. The neural progenitor cells are useful for treating acute
CC   or chronic damage to the nerve system e.g. epilepsy, stroke, ischaemia,
CC   Huntington's disease, Parkinson's disease, multiple sclerosis,
CC   leukodystrophies, neuritis, etc. The hepatocytes and hepatocyte
CC   precursors are useful for assessing animal models for ability to repair
CC   liver damage. The cardiomyocyte population is useful for assessing animal
CC   models for cryoinjury, regenerating cardiac muscle and to treat
CC   insufficient cardiac function. This polynucleotide sequence represents
CC   the DNA of plasmid pGRN144 containing the human hTERT gene relating to
CC   the invention
XX
SQ   Sequence 15418 BP; 4519 A; 3797 C; 3764 G; 3338 T; 0 U; 0 Other;

  Query Match             6.0%;  Score 296;  DB 6;  Length 15418;
  Best Local Similarity   69.2%;  Pred. No. 6.5e-42;
  Matches  479;  Conservative    0;  Mismatches  200;  Indels   13;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||| ||||||||||  ||| || | ||||||||||||| |||||||| || ||||||
Db        646 AGTTCGAGACCAGCCTGGCCAATATAGCGAAACCCCATCTGTACTAAAAACACAAAAATT 705

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   ||  |||||| ||| ||||||||||||| ||||  ||||||  ||||||||||
Db        706 AGCTGGGTGTGGTGACACATGCCTGTAATCCCAGGTACTCAGGAGGCTAAGGCAGGAGAA 765

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              |  |||||||  |  ||  | || |||      |   ||||||    ||| ||||||||||
Db        766 TTGCTTGAACTGGGAGGCAGAGGTGGT-----GAGCCAAGATTGCACCAGTGCACTCCAG 820

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  ||  || |||| ||||   ||||||| ||| ||| |||   | |  ||||| |||
Db        821 CCTTGGTGAC-AGAGTGAAACTCCATCTCAAAAACAAACAAACAAAATACATATACATAA 879

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              ||||||||  | | | | ||| | ||| ||||||| ||| ||||||| |  |||||||
Db        880 ATATATATGCACATATATATACATATATAAATATATATACACATATATAAATCTATATAC 939

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACAT-----ATATATA 354
              |||||||| ||||| ||| ||||| ||  | |||||||| || ||||     ||||||
Db        940 ATATATACATATATACACATATATAAATCTATATACATATATATACATATATAATATATT 999

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              |||||| || | | ||| ||| | | ||||||| | ||| ||||| ||||| | | |||
Db       1000 TACATATATAAATATATACATATATAAATATACATATATAAATACATATATAAATATACA 1059

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              |||||| | ||| ||| | | ||||||| ||||| | ||||| | | ||| ||| | |||
Db       1060 TATATAAATATACATATATAAATATACATATATAAATATATACATATATAAATATACATA 1119

Qy        475 TACACATATATATA-ATATACACACATATATATACACATATATACACACATATATACACA 533
              || | |||||||||  ||||| | | ||||| ||| |||  |||| ||| |||||||||  |
Db       1120 TATAAATATATATACATATATAAATATATAAATATACAAGTATATACAAATATATACATA 1179

Qy        534 TATATACACATATATATACACATATATACACATATATACATATATACACATATATATACA 593
              |||| |   |||||  || | | ||||| | ||| ||| ||| | | || | |   |
Db       1180 TATAAATGTATATACGTATATACATATATATATAAATATATAAAAAAACTTTTGGCTGGG 1239

Qy        594 TATATACACATATATACATATATACACATATATACATATATACACACATATATACACATA 653
              | |   || || |      | | | |  | |     ||  | || ||| | | |||||
Db       1240 CACCTTTCCAAATCTCATGGCACATATAAGTCTCATGGTAACCTCAAATAAAAAAACATA 1299

Qy        654 CATATACACACACATAGATATACATATATATA 685
              |          || | ||| | |   ||| |
Db       1300 TAACAGATACACCAAAAATAAAAACCAATAAA 1331


RESULT 13
ADC21253
ID   ADC21253 standard; DNA; 15418 BP.
XX
AC   ADC21253;
XX
DT   18-DEC-2003  (first entry)
XX
DE   Lambda clone (lambdaGphi5) containing human TERT DNA.
XX
KW   Viral vector; heterologous control element; gene expression;
KW   human telomerase reverse transcriptase; hTERT; tumour specific gene;
KW   cell death; transcriptional control element;
KW   human telomerase RNA component; hTR; cancer cell; liver cancer;
KW   prostate cancer; muscle cancer; neural cell cancer; lung cancer;
KW   pancreatic cancer; medulloblastoma; cervical carcinoma; fibrosarcoma;
KW   osteosarcoma; lambda clone; lambdaGphi5; human; ds.
XX
OS   Synthetic.
OS   Homo sapiens.
OS   Bacteriophage lambda.
XX
PN   US2003099616-A1.
XX
PD   29-MAY-2003.
XX
PF   25-JUL-2002; 2002US-00206447.
XX
PR   25-JUL-2001; 2001US-0308029P.
XX
PA   (IRVI/) IRVING J M.
PA   (KARP/) KARPF D B.
PA   (SCHI/) SCHIFF J M.
XX
PI   Irving JM,  Karpf DB,  Schiff JM;
XX
DR   WPI; 2003-730140/69.
XX
PT   New dual specificity vectors driven by the telomerase promoter, useful
PT   for killing or slowing the growth of tumor cells, or for treating cancer,
PT   e.g. liver cancer, prostate cancer, lung cancer, or pancreatic cancer.
XX
PS   Example 1; Page 13-20; 25pp; English.
XX
```

```
CC   The present invention relates to a viral vector comprising first and
CC   second genes controlled by heterologous control elements, where the first
CC   gene is preferentially expressed in cells expressing human telomerase
CC   reverse transcriptase (hTERT), and another gene under the control of a
CC   heterologous transcriptional control element for a tissue or tumour
CC   specific gene other than TERT, and where transduction of the vector into
CC   a mammalian cell expressing TERT causes the death of the cell or its
CC   progeny. In particular, the second gene may be under the control of a
CC   transcriptional control element for a tissue specific gene selected from
CC   albumin, alpha-fetoprotein, prostate-specific antigen, mitochondrial
CC   creatine kinase, myelin basic protein, glial fibrillary acidic protein,
CC   and neuron-specific enolase. The second gene may be under the control of
CC   a transcriptional control element for a human telomerase RNA component
CC   (hTR). The vector of the invention is useful for killing a cancer cell,
CC   or treating a subject for a condition associated with increased
CC   expression of TERT in affected cells. The vector is also useful in the
CC   preparation of a medicament for treatment of a condition associated with
CC   increased expression of TERT, particularly cancer, and especially liver
CC   cancer, prostate cancer, muscle cancer, neural cell cancer, lung cancer,
CC   pancreatic cancer, medulloblastoma, cervical carcinoma, fibrosarcoma, and
CC   osteosarcoma. The present sequence represents a lambda clone
CC   (lambdaGphi5) containing human TERT DNA.
XX
SQ   Sequence 15418 BP; 4518 A; 3797 C; 3765 G; 3338 T; 0 U; 0 Other;

  Query Match             6.0%;  Score 296;  DB 10;  Length 15418;
  Best Local Similarity   69.2%;  Pred. No. 6.5e-42;
  Matches  479;  Conservative    0;  Mismatches  200;  Indels   13;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||| |||||||||||  ||| || | ||||||||||||| ||||||||| || ||||||
Db        646 AGTTCGAGACCAGCCTGGCCAATATAGCGAAACCCCATCTGTACTAAAAACACAAAAATT 705

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   ||  |||||| ||| |||||||||||||| ||||  ||||||  ||||||||||
Db        706 AGCTGGGTGTGGTGACACATGCCTGTAATCCCAGGTACTCAGGAGGCTAAGGCAGGAGAA 765

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              | |||||||  |  ||  | || |||     |   ||||||    ||| ||||||||||
Db        766 TTGCTTGAACTGGGAGGCAGAGGTGGT-----GAGCCAAGATTGCACCAGTGCACTCCAG 820

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  ||  || |||| ||||   ||||||| ||| ||| |||   | |  ||||| |||
Db        821 CCTTGGTGAC-AGAGTGAAACTCCATCTCAAAAACAAACAAACAAAATACATATACATAA 879

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              ||||||||  | | | | | ||| | ||| |||||||| ||| |||||||  |  |||||||
Db        880 ATATATATGCACATATATATACATATATAAATATATATACACATATATAAATCTATATAC 939

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACAT-----ATATATA 354
              |||||||||  ||||| ||| ||||| ||  | |||||||| || ||||      ||||||
Db        940 ATATATACATATATACACATATATAAATCTATATACATATATATACATATATAATATATT 999

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              |||||| || | | ||| ||| | | |||||| | ||| ||||| ||||| | | |||
Db       1000 TACATATATAAATATATACATATATAAATATACATATATAAATACATATATAAATATACA 1059

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              |||||| | ||| ||| | | |||||||| ||||| | ||||| | | ||| ||| | |||
Db       1060 TATATAAATATACATATATAAATATACATATATAAATATATACATATATAAATATACATA 1119

Qy        475 TACACATATATATA-ATATACACACATATATATACACATATATACACACATATATACACA 533
              || | |||||||||| ||||| | | ||||| ||| |||  |||| ||| |||||||||  |
Db       1120 TATAAATATATATACATATATAAATATATAAATATACAAGTATATACAAATATATACATA 1179

Qy        534 TATATACACATATATATACACATATATACACATATATACATATATACACATATATATACA 593
              |||| |   |||||  || | | ||||| | ||| ||| ||| | | || | |   |
Db       1180 TATAAATGTATATACGTATATACATATATATATAAATATATAAAAAAACTTTTGGCTGGG 1239

Qy        594 TATATACACATATATACATATATACACATATATACATATATACACACATATATACACATA 653
              | |   || || |      | | | |  | |      ||  | || ||| | | |||||
Db       1240 CACCTTTCCAAATCTCATGGCACATATAAGTCTCATGGTAACCTCAAATAAAAAAACATA 1299

Qy        654 CATATACACACACATAGATATACATATATATA 685
              |           || | ||| | |   ||| |
Db       1300 TAACAGATACACCAAAAATAAAAACCAATAAA 1331


RESULT 14
ADC87552/c
ID   ADC87552 standard; DNA; 1486 BP.
XX
AC   ADC87552;
XX
DT   01-JAN-2004  (first entry)
XX
DE   Human GPCR gene SEQ ID NO:2005.
XX
KW   ds; gene; human; GPCR;
KW   guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1270724-A2.
XX
PD   02-JAN-2003.
XX
PF   18-JUN-2002; 2002EP-00013517.
XX
PR   18-JUN-2001; 2001JP-00246789.
XX
PA   (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA   (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI   Suwa M,  Asai K,  Akiyama Y,  Aburatani H;
XX
DR   WPI; 2003-315783/31.
DR   P-PSDB; ADC87553.
XX
PT   New polynucleotide, useful for preparing a composition for treating a
PT   patient in need of increased or suppressed activity or expression of the
PT   guanosine triphosphate-binding protein coupled receptor.
XX
PS   Claim 1; SEQ ID NO 2005; 28pp; English.
XX
CC   The invention relates to a novel polynucleotide encoding a guanosine
CC   triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC   the invention may have a use in gene therapy. The polynucleotide and
CC   polypeptide are useful for preparing a composition for treating a patient
CC   in need of increased or suppressed activity or expression of the
CC   guanosine triphosphate-binding protein coupled receptor. The
CC   polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC   invention.
XX
SQ   Sequence 1486 BP; 517 A; 149 C; 242 G; 578 T; 0 U; 0 Other;

  Query Match             5.9%;  Score 294.8;  DB 10;  Length 1486;
  Best Local Similarity   74.9%;  Pred. No. 9.2e-42;
  Matches  397;  Conservative    0;  Mismatches  127;  Indels    6;  Gaps    2;

Qy        197 AAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACA 256
              | || |   | || | || | | |      || || | | ||  | ||||||| |||||||
Db        999 ATATATACACACATATAATATATACATATTTATACACACATGCACATATATAAATACACA 940

Qy        257 CACACACACATATATATATACACATATATATACACATATATACATATATACACATATATA 316
              | | ||| ||||| | | | | || |||||| ||||||||||||||||||||||||| ||
Db        939 TATATACATATATACACACATATATGTATATATACATATATACATATATACACATATGTA 880

Qy        317 CACATATATATACACATACATATGTACACATATATATACACATATGTATACACATATATA 376
              | ||| ||||| ||||| |||   | |  |||||||||||||||||||| | | | |||
Db        879 TATATACATATATACATATATACACATATGTATATATACACATATGTATATATACACATA 820

Qy        377 CACATATATACACACATATATACACATATATACACACATATATACACATATATACACATA 436
              ||||||| || | | | | |||||||||| || | | | | |||||||||| || | |||
Db        819 CACATATGTATATATACACATACACATATGTATATATACACATACACATATGTATATATA 760
```

```
Qy        437 TATACACATATACACATATATACACATATATACATATATACACATATATATA-----ATA 491
              | | ||| |||   |||||||||||||| ||  |||||||||||||||| ||      | |
Db        759 CACATACACATATGTATATATACACATACATGTATATATACACATATATGTATATATACA 700

Qy        492 TACACACATATATATACACATATATACACACATATATACACATATATACACATATATATA 551
              | |||||| |||| ||| | | ||| |   |||||||||||||| | | | ||| | |||
Db        699 CATACACATGTATACACATACACATATATGTATATATACACATACACATATATACACATA 640

Qy        552 CACATATATACACATATATACATATATACACATATATATACATATATACACATATATACA 611
              ||| ||||||||||||||| ||||||||||||||   |||||||||||||| |||
Db        639 TACACATATACACATATATACACATATACACATATATAAGTATATATACACATATACACA 580

Qy        612 TATATACACATATATACATATATACACACATATATACACATACATAT-ACACACACATAG 670
              ||||   | ||||| |||||||||||||| |||  || | ||||| || | ||||| |||
Db        579 TATACGTATATATACACATATATACACATATACGTATATATACACATAATACACATATAC 520

Qy        671 ATATACATATATATACACATATATATACGTATATATATGTATATATATAT 720
              |||| ||| | ||| | | | |||||||||||||||| ||||||| | ||
Db        519 GTATAAATACACATATATACACATATACGTATATATACGTATATACACAT 470
```

RESULT 15
AAD50021
ID AAD50021 standard; DNA; 27067 BP.
XX
AC AAD50021;
XX
DT 24-MAR-2003 (first entry)
XX
DE Human secreted protein encoding gene.
XX
KW Human; secreted protein; cancer; gene therapy; cytostatic; gene; ds.
XX
OS Homo sapiens.
XX

FH	Key	Location/Qualifiers
FT	variation	replace(921,T)
FT		/*tag= a
FT		/standard_name= "Single nucleotide polymorphism"
FT	variation	replace(1022,T)
FT		/*tag= ba
FT		/standard_name= "Single nucleotide polymorphism"
FT	variation	replace(1781,T)
FT		/*tag= b
FT		/standard_name= "Single nucleotide polymorphism"
FT	variation	replace(1850,A)
FT		/*tag= c
FT		/standard_name= "Single nucleotide polymorphism"
FT	variation	replace(1882,T)
FT		/*tag= bb
FT		/standard_name= "Single nucleotide polymorphism"
FT	variation	replace(1951,A)
FT		/*tag= bc
FT		/standard_name= "Single nucleotide polymorphism"
FT	CDS	2031. .25941
FT		/*tag= d
FT		/product= "Human secreted protein"
FT	exon	2031. .2094
FT		/*tag= e
FT	intron	2095. .5569
FT		/*tag= f
FT	exon	2132. .2195
FT		/*tag= bd
FT	intron	2196. .5670
FT		/*tag= be
FT	variation	replace(2839,G)
FT		/*tag= g
FT		/standard_name= "Single nucleotide polymorphism"
FT	variation	replace(2940,G)
FT		/*tag= bf
FT		/standard_name= "Single nucleotide polymorphism"
FT	variation	replace(3730,A)
FT		/*tag= h
FT		/standard_name= "Single nucleotide polymorphism"
FT	variation	replace(3831,A)
FT		/*tag= bg
FT		/standard_name= "Single nucleotide polymorphism"
FT	exon	5570. .5762
FT		/*tag= i
FT	exon	5671. .5863
FT		/*tag= bh
FT	intron	5763. .7571
FT		/*tag= j
FT	intron	5864. .7672
FT		/*tag= bi
FT	variation	replace(6631,A)
FT		/*tag= k
FT		/standard_name= "Single nucleotide polymorphism"
FT	variation	replace(6732,A)
FT		/*tag= bj
FT		/standard_name= "Single nucleotide polymorphism"
FT	variation	replace(6945,A)
FT		/*tag= l
FT		/standard_name= "Single nucleotide polymorphism"
FT	variation	replace(6952,T)
FT		/*tag= m
FT		/standard_name= "Single nucleotide polymorphism"
FT	variation	replace(7457,A)
FT		/*tag= n
FT		/standard_name= "Single nucleotide polymorphism"
FT	variation	replace(7558,A)
FT		/*tag= bk
FT		/standard_name= "Single nucleotide polymorphism"
FT	exon	7572. .7741
FT		/*tag= o
FT	exon	7673. .7761
FT		/*tag= bl
FT	intron	7742. .10000
FT		/*tag= p
FT	intron	7762. .9149
FT		/*tag= bm
FT	variation	replace(7830,A)
FT		/*tag= q
FT		/standard_name= "Single nucleotide polymorphism"
FT	variation	replace(7931,A)
FT		/*tag= bn
FT		/standard_name= "Single nucleotide polymorphism"
FT	variation	replace(8089,C)
FT		/*tag= r
FT		/standard_name= "Single nucleotide polymorphism"
FT	variation	replace(8190,C)
FT		/*tag= bo
FT		/standard_name= "Single nucleotide polymorphism"
FT	variation	replace(8551,T)
FT		/*tag= s
FT		/standard_name= "Single nucleotide polymorphism"
FT	variation	replace(8652,T)
FT		/*tag= bp
FT		/standard_name= "Single nucleotide polymorphism"
FT	exon	9150. .9302
FT		/*tag= bq
FT	variation	replace(9269,C)
FT		/*tag= t
FT		/standard_name= "Single nucleotide polymorphism"
FT	intron	9303. .10101
FT		/*tag= br
FT	variation	replace(9362,T)
FT		/*tag= u
FT		/standard_name= "Single nucleotide polymorphism"
FT	variation	replace(9370,C)
FT		/*tag= bs
FT		/standard_name= "Single nucleotide polymorphism"
FT	variation	replace(9463,T)
FT		/*tag= bt
FT		/standard_name= "Single nucleotide polymorphism"
FT	variation	replace(9782,T)

```
FT                   /*tag=  v
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(9883,T)
FT                   /*tag=  bu
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            10001. .10173
FT                   /*tag=  w
FT   exon            10102. .10274
FT                   /*tag=  bv
FT   intron          10174. .10298
FT                   /*tag=  x
FT   intron          10275. .10399
FT                   /*tag=  bw
FT   exon            10299. .10485
FT                   /*tag=  y
FT   exon            10400. .10586
FT                   /*tag=  bx
FT   intron          10486. .12027
FT                   /*tag=  z
FT   intron          10587. .12128
FT                   /*tag=  by
FT   variation       replace(11493,W)
FT                   /*tag=  aa
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(11594,W)
FT                   /*tag=  bz
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            12028. .12193
FT                   /*tag=  ab
FT   exon            12129. .12294
FT                   /*tag=  ca
FT   intron          12194. .25821
FT                   /*tag=  ac
FT   variation       replace(12260,G)
FT                   /*tag=  ad
FT                   /standard_name= "Single nucleotide polymorphism"
FT   intron          12295. .25922
FT                   /*tag=  cb
FT   variation       replace(12361,G)
FT                   /*tag=  cc
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(13086,C)
FT                   /*tag=  ae
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(13183,C)
FT                   /*tag=  af
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(13187,C)
FT                   /*tag=  cd
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(13284,C)
FT                   /*tag=  ce
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(21240,G)
FT                   /*tag=  ag
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(21341,G)
FT                   /*tag=  cf
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(21695,G)
FT                   /*tag=  ah
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(21796,G)
FT                   /*tag=  cg
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(22058,T)
FT                   /*tag=  ai
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(22159,T)
FT                   /*tag=  ch
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(22233,A)
FT                   /*tag=  aj
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(22244. .22245,A)
FT                   /*tag=  ak
FT                   /standard_name= "Single nucleotide polymorphism"

  Query Match             5.9%;  Score 293.6;  DB 10;  Length 27067;
  Best Local Similarity   75.8%;  Pred. No. 1.7e-41;
  Matches  391;  Conservative    0;  Mismatches  119;  Indels    6;  Gaps    2;

Qy        230 ATATATATATATATATATATATACACACACACACATATAT-ATATACACATATATATA 288
              || ||||  |  | |  |  ||| |     | ||| |||||| | ||| | ||||| | |
Db      26509 ATGTATAACTTAAAAATTTCATATATGTGTATACATATATATGAAATATATATATACAGA 26568

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              |||| ||||| || |||||| ||||||||||||||||||  | | | ||||  |||||||
Db      26569 CACACATATATATGTATACATATATATACACATATATATGTATACATATATACACACATA 26628

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              ||| || |||||| |||||||| ||||||||||||||||||||| ||| | | ||| | |
Db      26629 TATGTATACATATATATACACACATATACACATATATACACACACATACATATATACACA 26688

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              |||| ||||| |||||||||||||| ||  |||||||||| |  |||| | | ||| | |
Db      26689 CACATATATACACACATATATACACACATGCACACATATATATGTATACATATATACACA 26748

Qy        469 CATATATACACATATATATA-----ATATACACACATATATATACACATATATACACACA 523
              ||| |||||  ||||||| |     ||||||||||||||||||||||| ||||||||||
Db      26749 CATGTATACGTATATATACACACATATATACACACATATATATACACACATATACACACA 26808

Qy        524 TATATACACATATATACACATATATATACACATATATACACATATATACATATATACACA 583
              || | |||||||||||||| |||||||||| ||||| ||| | ||| |||| || |||
Db      26809 TACACACACATATATACACACATATATACACACATATATACACACATATATATGTATACA 26868

Qy        584 TATATATACATATATACACATATATACATATATACACATATATACATATATACACACATA 643
              |||||||||| | ||| | | | ||||| | ||||| ||||| ||||||| ||| | | |
Db      26869 TATATATACACACATATATACACATACACACATACATATATACACATATACACATATACA 26928

Qy        644 TATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATA 703
              | ||| ||||| | |   | ||||| ||||||| | || || |||||| ||||| | |
Db      26929 CACACATATACACACATGTATACATATATATACACACATGTATACATATGTATACACACA 26988

Qy        704 TATATGTATATATATATGCTCCAGAGTTCATAAGAG 739
              |||| | |||| ||||   |     | |||| | |
Db      26989 CATATATGTATACATATATACACACATACATATGTG 27024


RESULT 16
ADC87570/c
ID   ADC87570 standard; DNA; 1672 BP.
XX
AC   ADC87570;
XX
DT   01-JAN-2004  (first entry)
XX
DE   Human GPCR gene SEQ ID NO:2023.
XX
KW   ds; gene; human; GPCR;
KW   guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1270724-A2.
XX
PD   02-JAN-2003.
XX
PF   18-JUN-2002; 2002EP-00013517.
XX
PR   18-JUN-2001; 2001JP-00246789.
XX
PA   (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA   (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI   Suwa M,  Asai K,  Akiyama Y,  Aburatani H;
```

```
XX
DR   WPI; 2003-315783/31.
DR   P-PSDB; ADC87571.
XX
PT   New polynucleotide, useful for preparing a composition for treating a
PT   patient in need of increased or suppressed activity or expression of the
PT   guanosine triphosphate-binding protein coupled receptor.
XX
PS   Claim 1; SEQ ID NO 2023; 28pp; English.
XX
CC   The invention relates to a novel polynucleotide encoding a guanosine
CC   triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC   the invention may have a use in gene therapy. The polynucleotide and
CC   polypeptide are useful for preparing a composition for treating a patient
CC   in need of increased or suppressed activity or expression of the
CC   guanosine triphosphate-binding protein coupled receptor. The
CC   polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC   invention.
XX
SQ   Sequence 1672 BP; 499 A; 171 C; 300 G; 602 T; 0 U; 100 Other;

  Query Match             5.9%;  Score 291.6;  DB 10;  Length 1672;
  Best Local Similarity   75.5%;  Pred. No. 3.3e-41;
  Matches  376;  Conservative    0;  Mismatches  119;  Indels    3;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||||| | | ||||||||| | ||||||||||||||||||
Db       1390 TATATATATATATATATATATATATATATATACACACACACACATATATACACATATATA 1331

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              |||||| | | | ||||| ||| ||||| ||||||||||||||||||| |||    | | |
Db       1330 TACACACACACATATATACACATATATACACACATATATATACACATATATACACATATA 1271

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||| | | | ||||  | | | |||| ||||||||||||||||||| |||| || ||||
Db       1270 TATACACATATATATACACATATATATGTACACATATATACACACATGTATATACGTATA 1211

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              |    |||||||||||||| | ||| |  ||| |  | ||||||||||||||  | |||||
Db       1210 TGTGTACATATATACACACACATATATGTATGTGTATATATACACATATATGTATATATA 1151

Qy        467 TACATATATACACATATATATA---ATATACACACATATATATACACATATATACACACA 523
              |||| | | | |||| ||||||    |||| | ||| | ||| | | | || |||| |
Db       1150 TACACACACAGACATGTATATATGTGTATATATACACACATAGATATATGTATATATGTG 1091

Qy        524 TATATACACATATATACACATATATATACACATATATACACATATATACATATATACACA 583
              |||||||||| | ||| |  | |||||||||| |||||    |||||||| | | | | |
Db       1090 TATATACACACACATATATGTGTATATACACACATATATGTGTATATACACACACATATA 1031

Qy        584 TATATATACATATATACACATATATACATATATACACATATATACATATATACACACATA 643
              | | |||| | | | ||| |||| |  ||||| ||||| |||||  | |||| ||||| |
Db       1030 TGTGTATATACACACACATATATGTGTATATACACACACATATATGTGTATATACACACA 971

Qy        644 TATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATA 703
              ||| |  |  |||||||||||||  |||||  | ||||||||||||||| || |||
Db        970 CATATATGTGTATATACACACACATGTATATATGTGTATATACACATATATGTATATATG 911

Qy        704 TATATGTATATATATATG 721
              | ||| || | |||||||
Db        910 TGTATATACACATATATG 893


RESULT 17
AAK67239
ID   AAK67239 standard; DNA; 30393 BP.
XX
AC   AAK67239;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:22051.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001354.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
```

PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.

XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS Disclosure; SEQ ID NO 22051; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention
XX
SQ Sequence 30393 BP; 8271 A; 7315 C; 8076 G; 6731 T; 0 U; 0 Other;

Query Match 5.8%; Score 288; DB 4; Length 30393;
Best Local Similarity 76.1%; Pred. No. 1.6e-40;
Matches 382; Conservative 0; Mismatches 115; Indels 5; Gaps 2;

```
Qy      227 TTTATATATATATATATATATATACACACACACACATATATATATACACATATATA 286
            |||||||| | ||||| ||||||| | | |||||||||| | | | | ||| |||||||
Db    14755 TATATATATACACATATACATATATATATAGACACACACACACACACACATATATATA 14814

Qy      287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
            ||||| | ||||||||||||||| ||||||| ||||| | ||| |||||||| |||| |
Db    14815 CACACACACATACATATATACACACATATACATATATACACACATATACATATATACATA 14874

Qy      347 TATA----TATACACATATGTATACACATATATACACATATATACACACATATATACACA 402
            ||||    | |||| ||||    | | | ||||| ||| ||||||| | ||||| |||||
Db    14875 TATACACGTGTACATATATACGCATATACATATATACACATATACATATATATACACACA 14934

Qy      403 TATATACACACATATATACACATATATACACATATATACACATATACACATATATACACA 462
            |||| | | | ||||| | | | ||  |||||||||| |||||||| ||||||||||
Db    14935 TATACATATATACATATATATACACATGTACATATATACGCATATACATATATATACACA 14994

Qy      463 TATATACATATATACACATATATATA-ATATACACACATATATATACACATATATACACA 521
            |||| | |||||||||||  |||| | ||||||| | ||| | ||  |||||||||| ||
Db    14995 TATACATATATATACACACCTATACATATATACATATATACACATGTACATATATACGCA 15054

Qy      522 CATATATACACATATATACACATATATATACACATATATACACATATATACATATATACA 581
            ||| ||| | | | | | | | || |||||| ||||| ||||||||| |  |||| | |
Db    15055 TATACATATATACACACATATACATGTATACATATATACACACATATACATGTATACATA 15114

Qy      582 CATATATATACATATATACACATATATACATATATACACATATATACATATATACACACA 641
            ||| | ||  | ||||| ||  ||||||||||||||||||| | ||||| ||||| |
Db    15115 TATACACATGTACATATATACGCATATACATATATACACATATACATATATACACACATA 15174

Qy      642 TATATACACATACATATATACACACACATAGATATACATATATATACACATATATATACGTA 701
            ||||| ||||| ||| |||   ||||| |||  |||||| ||| | | | |||||| ||
Db    15175 CATATATACATATATACACATGTACATATATACGCATATACATATATACACATATACATA 15234

Qy      702 TATATATGTATATATATATGCT 723
            |||| |  | || |||||| ||
Db    15235 TATACACATGTACATATATACT 15256
```

```
RESULT 18
ABK84481/c
ID   ABK84481 standard; cDNA; 44147 BP.
XX
AC   ABK84481;
XX
DT   14-AUG-2002  (first entry)
XX
DE   Human cDNA differentially expressed in granulocytic cells #1052.
XX
KW   Human; ss; granulocytic cell; DNA chip; bacterial infection;
KW   viral infection; parasitic infection; protozoal infection;
KW   fungal infection; sterile inflammatory disease; psoriasis;
KW   rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KW   cardiac reperfusion injury; renal reperfusion injury; ARDS;
KW   adult respiratory distress syndrome; inflammatory bowel disease;
KW   Crohn's disease; ulcerative colitis; periodontal disease;
KW   granulocyte activation; chronic inflammation; allergy.
XX
OS   Homo sapiens.
XX
PN   WO200228999-A2.
XX
PD   11-APR-2002.
XX
PF   03-OCT-2001; 2001WO-US030821.
XX
PR   03-OCT-2000; 2000US-0237189P.
XX
PA   (GENE-) GENE LOGIC INC.
XX
PI   Beazer-Barclay Y,  Weissman SM,  Yamaga S,  Vockley J;
XX
DR   WPI; 2002-435328/46.
XX
PT   Detecting granulocyte activation by detecting differential expression of
PT   genes associated with granulocyte activation, which serves as diagnostic
PT   markers that is useful for monitoring disease states and drug toxicity.
XX
PS   Claim 1; SEQ ID NO 1052; 114pp; English.
XX
CC   The invention relates to detecting (M1) granulocyte (GC) activation
CC   (GCA), by detecting the level of expression of gene(s) (Gs) identified by
CC   DNA chip analysis as given in the specification, and comparing the
CC   expression level to an expression level in an unactivated GC, where
CC   differential expression of Gs is indicative of GCA. Also included are
CC   modulating (M2) GA by contacting GC with an agent that alters the
CC   expression of at least one gene in Gs; (2) screening (M3) for an agent
CC   capable of modulating GCA or an inflammation (especially chronic) in a
CC   tissue, an allergic response in a subject, exposure of a subject to a
CC   pathogen or sterile inflammatory disease using the gene expression
CC   profile; (3) detecting (M4) an inflammation (especially chronic) in a
CC   tissue, an allergic response in a subject, exposure of a subject to a
CC   pathogen or sterile inflammatory disease, by detecting the level of
CC   expression in a sample of the tissue of gene(s) from Gs, where the level
CC   of expression of the gene is indicative of inflammation; (4) treating
CC   (M5) an inflammation (especially chronic) or in a tissue, an allergic
CC   response in a subject, exposure of a subject to a pathogen or sterile
CC   inflammatory disease, by contacting a tissue having inflammation with an
CC   agent that modulates the expression of gene(s) from Gs in the tissue. M1
CC   is useful for detecting GCA; M2 is useful for modulating GA; M3 is useful
CC   for screening an agent capable of modulating GCA preferably in an
CC   inflammation in a tissue; M4 is useful for detecting an inflammation
CC   (especially chronic) in a tissue, an allergic response in a subject,
CC   exposure of a subject to a pathogen or sterile inflammatory disease (e.g.
CC   psoriasis, rheumatoid arthritis, glomerulonephritis, asthma, thrombosis,
CC   cardiac reperfusion injury, renal reperfusion injury, ARDS, adult
CC   respiratory distress syndrome, inflammatory bowel disease, Crohn's
CC   disease, ulcerative colitis, periodontal disease; also bacterial
CC   infection, viral infection, parasitic infection, protozoal infection,
CC   fungal infection and M5 is useful for treating one of the above
CC   conditions. The present sequence represents a gene differentially
CC   expressed in granulocytes. Note: The sequence data for this patent did
CC   not form part of the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 44147 BP; 9813 A; 11709 C; 11789 G; 10833 T; 0 U; 3 Other;

  Query Match             5.8%;  Score 286.6;  DB 6;  Length 44147;
  Best Local Similarity   72.9%;  Pred. No. 2.9e-40;
  Matches  453;  Conservative    0;  Mismatches  154;  Indels   14;  Gaps    6;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC---AAAA 57
              ||||  |||||||||||| |||||||| |||||||| |||||| |||||||||    ||||
Db      10942 AGTTTGAGACCAGCCTGACCAACATGGAGAAACCCCGTCTCTATTAAAAATACAAAAAAA 10883

Qy         58 TTAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAG 117
              |||| ||||| |||||  ||| |||||||||||||||||||| | |||| || |||||||
Db      10882 TTAGCCAGGCGTGGTGGCACATGCCTGTAATCCCAGCTACTTGGAAGGCTGAAGCAGGAG 10823

Qy        118 AATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTC 176
              || | ||||||||  | || || ||||||| ||| ||| | ||||||||||||||||||
Db      10822 AAGCGCTTGAACC--CGGGAGGCGGAGGTTTCGGTGAGCCCAGATCATGCCATTGCACTG 10765

Qy        177 CAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATA 236
              |||||  |||||||||||| ||||       || |   || ||| |   |||| |||||||
Db      10764 CAGCCTGGGCAACAAGAGCAAAAT------TCCGTTTAATAAATAATTATTTAATATATA 10711

Qy        237 TATATATATATATATACACACACACACATATATATATACACATATATATACACATATA 296
              ||||||| | | | || | | | | |||| ||||||||||  ||||||||||||   ||
Db      10710 TATATATGTGTGTGTATATATATATACACGTATATATATACGTATATATATACACGCGTA 10651

Qy        297 TACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACA 356
              || |||||   | ||||||||||   |||||| || || |||| ||| | |||||||| |
Db      10650 TATATATACGTATATATATACACGCGTATATATACGTATATATATACGCGTATATATATA 10591

Qy        357 CATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATA 416
              | ||| |||| |||  || | | |  |||| | | |||  || | ||||| ||  |||||
Db      10590 CGTATATATATACACGTACATATACGTATATATATATACGTATATATATACACGTACATA 10531

Qy        417 TATACACATATATACACATATATACACATATACACATATATACACATATATACATATATA 476
              ||      ||||||| ||| ||||||  ||||| | | | ||| | |  |||| ||| | |
Db      10530 TACGTGTATATATATACACATATACGTATATATATACACATATATACGTATATATACACA 10471

Qy        477 CACATATATATAATATACACACATATATATACACATATATACACACATATATACACATAT 536
              | ||||| |  |||| | |||| |||||||| ||||||| |    ||   |  || |||
Db      10470 TATATATACA-CATATGTATACATCTATATACATATATATAAAATAATTATTTTACGTAT 10412

Qy        537 ATACACATATATATACACATATATA-CACATATATACATATATACACATATATATACATA 595
              |||         |  | |||  ||  | |  ||||| | |||| | |||||||||| |||
Db      10411 ATAATTTACGTAAATAATATACGTATAATAATTATACGTGTATATATATATATATATATA 10352

Qy        596 TATACACATATATACATATAT 616
              | ||||| ||||| ||||||
Db      10351 CACACACACATATATATATAT 10331


RESULT 19
ADD14691/c
ID   ADD14691 standard; cDNA; 44147 BP.
XX
AC   ADD14691;
XX
DT   01-JAN-2004  (first entry)
XX
DE   Human src biomarker polynucleotide SEQ ID NO:85.
XX
KW   predictor set; protein tyrosine kinase activity modulator;
KW   protein tyrosine kinase pathway; protein tyrosine kinase; cytostatic;
KW   gene therapy; drug sensitivity; genetic profile; cancer; human; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO2003062395-A2.
XX
```

```
PD   31-JUL-2003.
XX
PF   17-JAN-2003; 2003WO-US001981.
XX
PR   18-JAN-2002; 2002US-0350061P.
XX
PA   (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX
PI   Huang F,  Fairchild CR,  Lee FY,  Shaw P;
XX
DR   WPI; 2003-636735/60.
DR   P-PSDB; ADD14095.
XX
PT   New polynucleotides and polypeptides for predicting the activity of
PT   compounds that interact with protein tyrosine kinases and/or protein
PT   tyrosine kinase pathways.
XX
PS   Claim 2; SEQ ID NO 85; 139pp; English.
XX
CC   The present invention describes a predictor set comprising a plurality of
CC   polynucleotides or polypeptides whose expression pattern is predictive of
CC   the response of cells to treatment with a compound that modulates protein
CC   tyrosine kinase activity or members of the protein tyrosine kinase
CC   pathway. Also described: (1) predicting whether a compound is capable of
CC   modulating the activity of cells, comprising obtaining a sample of cells,
CC   determining whether the cells express a plurality of markers, and
CC   correlating the expression of the markers to the compound's ability to
CC   modulate the activity of the cells; (2) a plurality of cell lines for
CC   identifying polynucleotides and polypeptides whose expression levels
CC   correlate with compound sensitivity or resistance of cells associated
CC   with a disease state; and (3) identifying polynucleotides and
CC   polypeptides that predict compound sensitivity or resistance of cells
CC   associated with a disease state, comprising subjecting the plurality of
CC   cell lines to one or more compounds, analysing the expression pattern of
CC   a microarray of polynucleotides or polypeptides, and selecting
CC   polynucleotides or polypeptides that predict the sensitivity or
CC   resistance of cells associated with a disease state by using the
CC   expression pattern of the microarray. The polynucleotides and
CC   polypeptides have cytostatic activities, and can be used in gene therapy.
CC   The polynucleotides and polypeptides are useful in predicting the
CC   activity of compounds that interact with protein tyrosine kinases and/or
CC   protein tyrosine kinase pathways. These may be used in determining drug
CC   sensitivity in patients to allow the development of individualized
CC   genetic profiles which aid in treating diseases and disorders (e.g.
CC   cancer) based on patient response at a molecular level. The present
CC   sequence is used in the exemplification of the present invention.
XX
SQ   Sequence 44147 BP; 9813 A; 11709 C; 11789 G; 10833 T; 0 U; 3 Other;

  Query Match             5.8%;  Score 286.6;  DB 10;  Length 44147;
  Best Local Similarity   72.9%;  Pred. No. 2.9e-40;
  Matches  453;  Conservative    0;  Mismatches  154;  Indels   14;  Gaps    6;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC---AAAA 57
              ||||  |||||||||||| |||||||| |||||||| |||||| |||||||||   ||||
Db      10942 AGTTTGAGACCAGCCTGACCAACATGGAGAAACCCCGTCTCTATTAAAAATACAAAAAAA 10883

Qy         58 TTAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAG 117
              |||| ||||| |||||  ||| ||||||||||||||||||| | |||| || |||||||
Db      10882 TTAGCCAGGCGTGGTGGCACATGCCTGTAATCCCAGCTACTTGGAAGGCTGAAGCAGGAG 10823

Qy        118 AATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTC 176
              || | ||||||||  | || || ||||||| ||| ||| | ||||||||||||||||||
Db      10822 AAGCGCTTGAACC--CGGGAGGCGGAGGTTTCGGTGAGCCCAGATCATGCCATTGCACTG 10765

Qy        177 CAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATA 236
              |||||  |||||||||||| ||||      || |   || ||| |   |||| |||||||
Db      10764 CAGCCTGGGCAACAAGAGCAAAAT------TCCGTTTAATAAATAATTATTTAATATATA 10711

Qy        237 TATATATATATATATACACACACACACATATATATATACACATATATATACACATATA 296
              ||||||| | | | || | | | | |||| ||||||||||  ||||||||||    ||
Db      10710 TATATATGTGTGTGTATATATATATACACGTATATATATACGTATATATATACACGCGTA 10651

Qy        297 TACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACA 356
              || |||||   | ||||||||||   |||||| || || |||| ||| | ||||||||
Db      10650 TATATATACGTATATATATACACGCGTATATATACGTATATATATACGCGTATATATATA 10591

Qy        357 CATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATA 416
              | ||| |||| ||| || | | |  |||| | | |||  || | ||||| ||  |||||
Db      10590 CGTATATATATACACGTACATATACGTATATATATATACGTATATATATACACGTACATA 10531

Qy        417 TATACACATATATACACATATATACACATATACACATATATACACATATATACATATATA 476
              ||     ||||||| ||| ||||||  ||||| | | | ||| | |  |||| ||| | |
Db      10530 TACGTGTATATATATACACATATACGTATATATATACACATATATACGTATATATACACA 10471

Qy        477 CACATATATATAATATACACACATATATATACACATATATACACACATATATACACATAT 536
              | ||||| |  ||||  | |||| |||||||| |||||||  |    ||    |  || |||
Db      10470 TATATATACA-CATATGTATACATCTATATACATATATATAAAATAATTATTTTACGTAT 10412

Qy        537 ATACACATATATATACACATATATA-CACATATATACATATATACACATATATATACATA 595
              |||          |   | |||  ||  | |  ||||| | |||| | |||||||||| |||
Db      10411 ATAATTTACGTAAATAATATACGTATAATAATTATACGTGTATATATATATATATATATA 10352

Qy        596 TATACACATATATACATATAT 616
              | ||||| ||||| ||||||
Db      10351 CACACACACATATATATATAT 10331


RESULT 20
AAK89374
ID   AAK89374 standard; DNA; 32206 BP.
XX
AC   AAK89374;
XX
DT   05-NOV-2001  (first entry)
XX
DE   Human digestive system antigen genomic sequence SEQ ID NO: 2950.
XX
KW   Human; digestive system antigen; gene therapy; cancer; appendicitis;
KW   ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
KW   digestive system disorder; Meckel's diverticulum; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155314-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001324.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
```

PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-502630/55.
XX
PT Polynucleotides encoding digestive system antigens, useful for
PT diagnosing, treating, preventing and/or prognozing disorders of the
PT digestive system, particularly cancer and cancer metastases.
XX
PS Disclosure; SEQ ID NO 2950; 986pp; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human digestive system antigens. These can be used in the
CC diagnosis, treatment and prevention of digestive system disorders,
CC including cancer, Meckel's diverticulum, bacterial or parasitic
CC infections, appendicitis, Hirschsprung's disease, chronic colitis or
CC ulcerative colitis. The present sequence is a genomic DNA fragment
CC encoding a digestive system antigen of the invention
XX
SQ Sequence 32206 BP; 9503 A; 5480 C; 7017 G; 10206 T; 0 U; 0 Other;

Query Match 5.6%; Score 280.8; DB 4; Length 32206;
Best Local Similarity 74.4%; Pred. No. 2.9e-39;
Matches 369; Conservative 0; Mismatches 122; Indels 5; Gaps 1;

```
Qy      229 TATATATATATATATATATATACACACACACACACATATATATATACACATATATATA 288
            ||||| |||||||| | ||| | | | | ||| | |  |||||||| ||| | | |||||
Db    21803 TATATGTATATATACACATACACATATATACATATATGTATATATACACATACACATATA 21862

Qy      289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
            ||||||||  |||||||||||||| | | |  ||| |||||| | | ||||  ||| ||
Db    21863 TACATATATGTATATATACACATACACATATGTATGTATACATACATATATACACATATG 21922

Qy      349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
```

```
              |||||| |||||  |||| ||| || |  | ||||| |  || ||||||||||| | | ||
Db      21923 TATATATACATACATATATACACATGTGTATATATACATGCAGATATATACACGTGTGTA 21982

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              | ||| ||||  | ||| || |||||||| |||| | || | ||| | | ||| ||||
Db      21983 TATACACATATGTATATACATGCACATATACACACGTGTATATATACACATACACGTATA 22042

Qy        469 CATATATACACATATATATA-----ATATACACACATATATATACACATATATACACACA 523
              ||| ||| ||| |||||||     | ||| |||||| |||||| ||||| | | | |||
Db      22043 TATACATATACACATATATACATACACATATACACATGTATATATACATACACATATACA 22102

Qy        524 TATATACACATATATACACATATATATACACATATATACACATATATACATATATACACA 583
              || || | ||| ||||| |||||  |||||  |||||| ||||| ||| ||| |||||
Db      22103 CATGTATATATACATACATATATACATACATGTATATATACATACATATATACATACATG 22162

Qy        584 TATATATACATATATACACATATATACATATATACACATATATACATATATACACACATA 643
              |||||||||||| ||| | | ||| |  |||||| ||||| ||||||||||||||  |||
Db      22163 TATATATACATACATATATACATACATGTATATATACATACATACATATATACACGTATA 22222

Qy        644 TATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATA 703
              |||||| | | ||||||||||||||| ||| | | |||||| || | || ||||  || |
Db      22223 TATACATACATATATACACACACATATATACATACATATATGCATACATGTATATATACA 22282

Qy        704 TATATGTATATATATA 719
              || |  |||||| |||
Db      22283 TACAGATATATACATA 22298


RESULT 21
AAK73165
ID   AAK73165 standard; DNA; 20420 BP.
XX
AC   AAK73165;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:27977.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001354.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0235484P.
PR   27-SEP-2000; 2000US-0235834P.
PR   27-SEP-2000; 2000US-0235836P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   13-OCT-2000; 2000US-0239937P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241221P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   20-OCT-2000; 2000US-0241826P.
PR   01-NOV-2000; 2000US-0244617P.
PR   08-NOV-2000; 2000US-0246474P.
PR   08-NOV-2000; 2000US-0246475P.
PR   08-NOV-2000; 2000US-0246476P.
PR   08-NOV-2000; 2000US-0246477P.
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246526P.
```

PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS Disclosure; SEQ ID NO 27977; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention
XX
SQ Sequence 20420 BP; 4834 A; 5403 C; 5630 G; 4553 T; 0 U; 0 Other;

Query Match 5.4%; Score 270.4; DB 4; Length 20420;
Best Local Similarity 61.4%; Pred. No. 1.8e-37;

Matches 467; Conservative 0; Mismatches 291; Indels 2; Gaps 2;

```
Qy       144 GGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAGCCCAGGCAACAAGAGCGAAATGTC 203
Db      4055 GCTGGAGTGCAGTAGCATGATCACAGCTCACTGCAGCCATATATATGTATATATATATAT 4114

Qy       204 ATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATATACACACACACAC 263
Db      4115 ATATGAAAAAATATATATATATATATATATATGAAAAAATATATATAAATATATATATGA 4174

Qy       264 ACATATATATATACACATATATATACACATATATACATATATACACAT-ATATACACATA 322
Db      4175 AAATATATATATAAATATATATATGAAAATATATATAAATATATATATAATATATATAAA 4234

Qy       323 TATATACACATACATATGTACACATATATATACACATATGTATACACATATATACACATA 382
Db      4235 TATATATATATATAAATATATATATGAAAATATATATAAATATATATATAAATATATATA 4294

Qy       383 TATACACACATATATACACATATATACACACATATATACACATATATACACATATATACA 442
Db      4295 TGAAAATATATATATGAAAATATATATGAAAATATATATATAAATATATATATATGAAAA 4354

Qy       443 CATATACACATATATACACATATATACATATATACACATATATATAATATACACACATAT 502
Db      4355 TATATATATAAATATATATATGAAAATATATATATAAATATATAT-ATATGAAAATATAT 4413

Qy       503 ATATACACATATATACACACATATATACACATATATACACATATATATACACATATATAC 562
Db      4414 ATATAAATATATATATATGAAAATATATATATAAATATATATATATGAAAATATATATAT 4473

Qy       563 ACATATATACATATATACACATATATATACATATATACACATATATACATATATACACAT 622
Db      4474 AAATATATATATATGAAAATATATATATAAATATATATATATGAAAATATATATATAAAT 4533

Qy       623 ATATACATATATACACACATATATACACATACATATACACACACATAGATATACATATAT 682
Db      4534 ATATATATATGAAAATATATATATAAATATATATATATGAAAATATATATATAAATATAT 4593

Qy       683 ATACACATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTA 742
Db      4594 ATATGAAAATATATATATAAATATATATATGAAAATATATATATAAATATATATATATGA 4653

Qy       743 GCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAA 802
Db      4654 AAATATATATATAAATATATATATATGAAAATATATATAAATATATATATGAAAATATAT 4713

Qy       803 GGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTAT 862
Db      4714 ATATATATATATATGAAAATATATATATATGAAAATATATATATAAATATATATATGAAA 4773

Qy       863 CTGTAAGGTAGTTAAGAAGGTAACACTATATATATATATA 902
Db      4774 ATAATATATATAAATATATATAAATATATATAAAAATATA 4813
```

RESULT 22
ABK69933
ID ABK69933 standard; DNA; 20420 BP.
XX
AC ABK69933;
XX
DT 15-JUL-2002 (first entry)
XX
DE Human secreted protein gene 68 genomic DNA fragment #21.
XX
KW Human; ds; secreted protein; gene therapy; immunosuppressive;
KW antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;
KW vasotropic; cerebroprotective; nootropic; neuroprotective; antibacterial;
KW virucide; fungicide; opthalmalogical; autoimmune disease; neoplasm;
KW rheumatoid arthritis; hyperproliferative disorder; cardiac arrest;
KW cardiovascular disorder; cerebrovascular disorder; cerebral ischaemia;
KW angiogenesis; nervous system disorder; Alzheimer's disease; infection;
KW ocular disorder; corneal infection; wound healing; skin aging;

KW epithelial cell proliferation; food additive.
XX
OS Homo sapiens.
XX
PN WO200226931-A2.
XX
PD 04-APR-2002.
XX
PF 24-SEP-2001; 2001WO-US029871.
XX
PR 25-SEP-2000; 2000US-0234925P.
PR 12-JAN-2001; 2001WO-US000911.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Komatsoulis G, Duan DR, Rosen CA, Moore PA, Shi Y;
PI Lafleur DW, Olsen H, Brewer LA, Florence KA, Young PE, Soppet DR;
PI Endress GA, Mucenski M, Ebner R;
XX
DR WPI; 2002-362489/39.
XX
PT Novel 71 isolated secreted polypeptides and polynucleotides encoding the
PT polypeptides, useful for treating Huntington's disease, sepsis,
PT meningitis, thrombocytopenia, hemolytic anemia, rheumatoid arthritis,
PT asthma.
XX
PS Example 2; Page 1440-1445; 1478pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (or its
CC fragment, homologue complement or allelic variant) encoding a human
CC secreted protein (and its fragment, domain, epitope, variant, secreted
CC form and species variant). Also included are a recombinant vector
CC comprising the nucleic acid, a recombinant host cell comprising the
CC vector, an antibody against the secreted protein, a recombinant host cell
CC that expresses the secreted protein and a method of identifying a binding
CC partner of the secreted protein. The nucleic acid and protein are used to
CC prevent, diagnose, treat or ameliorate a medical condition in e.g.
CC humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep for
CC example autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC angiogenesis, nervous system disorders e.g. Alzheimer's disease,
CC infections caused by bacteria, viruses and fungi and ocular disorders
CC e.g. corneal infection. Many other diseases and disorders are listed in
CC the specification. The polypeptides can also be used to aid wound healing
CC an epithelial cell proliferation, to prevent skin aging due to sunburn,
CC to maintain organs before transplantation, for supporting cell culture of
CC primary tissues, to regenerate tissues and in chemotaxis. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities. The present sequence
CC represents a ds DNA fragment of the gene for a novel human secreted
CC protein of the invention
XX
SQ Sequence 20420 BP; 4834 A; 5403 C; 5630 G; 4553 T; 0 U; 0 Other;

Query Match 5.4%; Score 270.4; DB 6; Length 20420;
Best Local Similarity 61.4%; Pred. No. 1.8e-37;
Matches 467; Conservative 0; Mismatches 291; Indels 2; Gaps 2;

```
Qy       144 GGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAGCCCAGGCAACAAGAGCGAAATGTC 203
Db      4055 GCTGGAGTGCAGTAGCATGATCACAGCTCACTGCAGCCATATATATGTATATATATATAT 4114

Qy       204 ATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATATACACACACACAC 263
Db      4115 ATATGAAAAAATATATATATATATATATATATGAAAAAATATATATAAATATATATATGA 4174

Qy       264 ACATATATATATACACATATATATACACATATATACATATATACACAT-ATATACACATA 322
Db      4175 AAATATATATATAAATATATATATGAAAATATATATAAATATATATATAATATATATAAA 4234

Qy       323 TATATACACATACATATGTACACATATATATACACATATGTATACACATATATACACATA 382
```

```
Db      4235 TATATATATATATAAATATATATATGAAAATATATATAAATATATATATAAATATATATA 4294

Qy       383 TATACACACATATATACACATATATACACACATATATACACATATATACACATATATACA 442
Db      4295 TGAAAATATATATATGAAAATATATATGAAAATATATATATAAATATATATATATGAAAA 4354

Qy       443 CATATACACATATATACACATATATACATATATACACATATATATAATATACACACATAT 502
Db      4355 TATATATATAAATATATATATGAAAATATATATATAAATATATAT-ATATGAAAATATAT 4413

Qy       503 ATATACACATATATACACACATATATACACATATATACACATATATATACACATATATAC 562
Db      4414 ATATAAATATATATATATGAAAATATATATATAAATATATATATATGAAAATATATATAT 4473

Qy       563 ACATATATACATATATACACATATATATACATATATACACATATATACATATATACACAT 622
Db      4474 AAATATATATATATGAAAATATATATATAAATATATATATATGAAAATATATATATAAAT 4533

Qy       623 ATATACATATATACACACATATATACACATACATATACACACACATAGATATACATATAT 682
Db      4534 ATATATATATGAAAATATATATATAAATATATATATATGAAAATATATATATAAATATAT 4593

Qy       683 ATACACATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTA 742
Db      4594 ATATGAAAATATATATATAAATATATATGAAAATATATATATAAATATATATATATGA 4653

Qy       743 GCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAA 802
Db      4654 AAATATATATATAAATATATATATATGAAAATATATATAAATATATATATGAAAATATAT 4713

Qy       803 GGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTAT 862
Db      4714 ATATATATATATATGAAAATATATATATATGAAAATATATATATAAATATATATATGAAA 4773

Qy       863 CTGTAAGGTAGTTAAGAAGGTAACACTATATATATATATA 902
Db      4774 ATAATATATATAAATATATATAAATATATATAAAAATATA 4813
```

RESULT 23
ABD32806_4/c
Continuation (5 of 7) of ABD32806 from base 400001 (Human cancer-associated genomic DNA F
WP Sequence split into 7 fragments LOCUS ABD32806 Accession Abd32806

WP	Fragment Name	Begin	End
WP	ABD32806_0	1	110000
WP	ABD32806_1	100001	210000
WP	ABD32806_2	200001	310000
WP	ABD32806_3	300001	410000
WP	ABD32806_4	400001	510000
WP	ABD32806_5	500001	610000
WP	ABD32806_6	600001	653458

Query Match 5.4%; Score 268.6; DB 13; Length 110000;
Best Local Similarity 69.2%; Pred. No. 4e-37;
Matches 441; Conservative 0; Mismatches 184; Indels 12; Gaps 5;

```
Qy         1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC--AAAAT 58
Db     95587 AGTTTGAGACCAGCCTGGCCAACATGGTAAAACCCCATCTCTATTTAAAAAAACAAAAAT 95528

Qy        59 TAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGA 118
Db     95527 TAACCAGGTGTGATGGTACACATCTGTACTCCCGGCTACTCAGGAGGCTGAGGCACGAGA 95468

Qy       119 ATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCA-AGATCATGCCATTGCACTCC 177
Db     95467 ATTGTTTGAACCTG--GAAAGTGGAGGTTGTAGTGAGCAGAGATCGCAACACTGTACTCC 95410

Qy       178 AGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATAT 237
Db     95409 AGCCTGGGTGACAGAGCAAGACTGTC------TGAAAAAAAAAATCAACAAAATAAACAA 95356

Qy       238 ATATATATATATATACACACACACACACATATATATATACACATATATATACACATATAT 297
```

```
Db      95355 AGAGGTAATGACTGACATATATATATATACATATATATATTCCCATATATATACACATAT 95296

Qy        298 ACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACAC 357
              || |||||| || |||||||| |||||| ||  |||| | | || || | |||| ||
Db      95295 ACGTATATATACGTATATACACGTATATACACGTATACGTGTATATACGTGTATACACGT 95236

Qy        358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417
              |||  |   | || | || | |||  || | |||| |||| || |||||| ||| ||||
Db      95235 GTATACACGTATATGTGTATATATACGTATATACATGTATATACGTATATATACATATAT 95176

Qy        418 ATACACATATATACACATATATACACATATACACATATATACACATATATACATATATAC 477
              | | |   ||||||  |||||||| |||||| ||  ||| |  |||||| |  | |||
Db      95175 ACATATTCATATACGTATATATACGTATATACATATTCATATATGTATATATACGTGTAC 95116

Qy        478 ACATATATATAATATACACACATATATATACACATATATACACACATATATACACATATA 537
              | |||||| | ||||| ||| |||||  ||||| ||||||   | |||||| | | |||
Db      95115 ATATATATGT-ATATATACATATATACGTACACGTATATATGTATATATACATATATACG 95057

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATACATATA 597
              ||||| |||||||  | ||||| | | |||  ||||  |||| |  |||||||| | | |
Db      95056 TACACGTATATATGTATATATACATATATACGTACACGTATATATGTATATATATACACA 94997

Qy        598 TACACATATATACATATATACACATATATACATATAT 634
              ||| || | |   |||||| | ||||||| || | |
Db      94996 CACATATGTGTGTGTATATATATATATATATATTTTT 94960


RESULT 24
ADH13788/c
ID   ADH13788 standard; DNA; 191127 BP.
XX
AC   ADH13788;
XX
DT   11-MAR-2004  (first entry)
XX
DE   Prostate leucine zipper regulatory element/PrLZ gene SEQ ID NO:2.
XX
KW   prostate leucine zipper; PrLZ; prostate; cytostatic; cancer; dysplasia;
KW   ds; human; gene.
XX
OS   Chimeric.
OS   Unidentified.
OS   Homo sapiens.
XX
PN   WO2003087311-A2.
XX
PD   23-OCT-2003.
XX
PF   07-APR-2003; 2003WO-US010536.
XX
PR   05-APR-2002; 2002US-0370557P.
XX
PA   (UYEM-) UNIV EMORY.
XX
PI   Wang R,  Chung LWK;
XX
DR   WPI; 2003-845313/78.
DR   P-PSDB; ADH13786.
XX
PT   New isolated prostate leucine zipper (PrLZ) nucleic acid molecule, useful
PT   in diagnosing, prognosticating and/or treating cancers or high grade
PT   dysplasias affecting PrLZ-positive tissues, such as those in the
PT   prostate.
XX
PS   Claim 23; SEQ ID NO 2; 103pp; English.
XX
CC   The invention relates to a novel isolated nucleic acid molecule that
CC   regulats the expression of prostate-expressed genes, specifically the
CC   prostate leucine zipper (PrLZ) gene. A polypeptide of the invention has
CC   cytostatic activity, and may act as a leucine zipper antagonist. The
CC   agent or vector that inhibits the expression or activity of PrLZ is
CC   useful for the treatment of cancer or high grade dysplasia, and for the
CC   preparation of a medicament for the treatment of cancer or high grade
CC   dysplasia, where the cancer or dysplasia affects prostate tissue, breast
CC   tissue, renal tissue, tissue within the gastrointestinal tract, or any
CC   other PrLZ-positive tissue. The methods and compositions of the present
CC   invention are also useful for the diagnosis and prognosis of such
CC   cancers. The present sequence represents a genomic sequence that
CC   regulates the expression of PrLZ, an additional 5' flanking sequence, the
CC   PrLZ gene, and 3' flanking sequence.
XX
SQ   Sequence 191127 BP; 51194 A; 37977 C; 39502 G; 62454 T; 0 U; 0 Other;

  Query Match             5.4%;  Score 268.6;  DB 10;  Length 191127;
  Best Local Similarity   64.4%;  Pred. No. 4.1e-37;
  Matches  467;  Conservative    0;  Mismatches  249;  Indels    9;  Gaps    4;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAATTA 60
              || || |||||| |||| ||||||||||||||||           ||||     || ||||
Db     122784 AGATCGAGACCATCCTGGCCAACATGGTGAAACCCGTCTCAACTAAAAACACAAAGATTA 122725

Qy         61 GACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAAT 120
              |   ||| |||||      | |||||| ||||| |||||   |||||| ||||||||||||||
Db     122724 GCTGGGCGTGGTGGCGTGCACCTGTAGTCCCAACTACTCATGAGGCTGAGGCAGGAGAAT 122665

Qy        121 CACTTGAACCTGCTGGGGGGTGGAGGTTGCGGGGAGC-AAGATCATGCCATTGCACTCCAG 179
              | | |||||    | || ||  |||||||| | ||||  ||||   ||||   ||||||||
Db     122664 CTCCTGAAC--TCGGGAGGCAGAGGTTGCAGTGAGCTGAGATTGCGCCACCACACTCCAG 122607

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  |  |  |||| ||    ||     | |||| | | | |   | | | | | | |
Db     122606 CCTGGCGACAGAGAGAGA---CTCCATCTAAAAAACATATATGTGTGTGTGTGTGTGTGT 122550

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              | | | | | ||       | | |||| || | ||| | ||   | || | | || |||
Db     122549 GTGTGTGTGTGTATGTGTGTATATACATGTACACATAGATATGCGTGTATATACATGTAC 122490

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACAT 359
              | ||| | |    | || | || || ||| | | ||  ||  | ||| || |||||||
Db     122489 ACATAGATATGCGTGTATATACATGTACACATAGATATGCGTGTATATACATGTACACAT 122430

Qy        360 ATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATAT 419
              |  ||| |  |||||||  || | | | ||   |  | ||| ||  ||||||| ||
Db     122429 AGATATGCGTGTATATACATGTACACATAGATATGCGTGTATATACATGTACACATAGAT 122370

Qy        420 ACACATATATACACATATATACACATATACACATATATACACATATATACATATATACAC 479
              |  | | |||| |||| || ||| | |||  | | |||| |||| || ||||| |||  |
Db     122369 ATGCGTGTATATACATGTACACATAGATATGCGTGTATATACATGTACACATAGATATGC 122310

Qy        480 ATATATATA---ATATACACACATATATATACACATATATACACACATATATACACATAT 536
              | ||||||     || ||| | ||||    |  | ||| ||  |||||| |||  | | |
Db     122309 GTGTATATACATGTACACATAGATATGCGTGTATATACATGTACACATAGATATGCGTGT 122250

Qy        537 ATACACATATATATACACATATATACACATATATACATATATACACATATATATACATAT 596
              ||| |||| || | | || | |  | ||| ||  | | ||| | || | | || |||
Db     122249 ATATACATGTACACAGATATGTGTGTATATACATGTACACATAGATATGTGTGTATATAC 122190

Qy        597 ATACACATATATACATATATACACATATATACATATATACACACATATATACACATACAT 656
              | ||| ||||| |  |||| |||| ||||| | ||||| | ||| | | | | ||||||
Db     122189 ACACATATATACATCTATACGCACACATATATACATATATATACACACACATATATACAT 122130

Qy        657 ATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATAT 716
              ||| ||||||| | ||||| ||| | ||||||| | | | || |||||||||| ||||||
Db     122129 ATATACACACACACATATATATACACATACACACACACACATATATATATATATATAA 122070

Qy        717 ATATG 721
              |  ||
Db     122069 ACTTG 122065


RESULT 25
ABD32806_4
Continuation (5 of 7) of ABD32806 from base 400001 (Human cancer-associated genomic DNA F
WP Sequence split into 7 fragments  LOCUS ABD32806 Accession Abd32806
WP     Fragment Name      Begin     End
```

```
WP      ABD32806_0              1    110000
WP      ABD32806_1         100001    210000
WP      ABD32806_2         200001    310000
WP      ABD32806_3         300001    410000
WP      ABD32806_4         400001    510000
WP      ABD32806_5         500001    610000
WP      ABD32806_6         600001    653458

  Query Match             5.4%;  Score 267.6;  DB 13;  Length 110000;
  Best Local Similarity   69.9%;  Pred. No. 6e-37;
  Matches  432;  Conservative    0;  Mismatches  179;  Indels    7;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||| |||||||||||  |||||||| ||||||||||||||||| ||||| || ||||||
Db      94745 AGTTCGAGACCAGCCTGGCCAACATGGCGAAACCCCATCTCTACAAAAAACACAAAAATT 94804

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              || ||||| |||||  | |  ||| || |||||| ||||| |||||| ||||| ||||||
Db      94805 AGCCAGGCGTGGTGGCATATACCTATAGTCCCAG-TACTTGGGAGGCTGAGGCGGGAGAA 94863

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              ||||||||||     || |       |   |||   | ||||||||||| || |||||||
Db      94864 TCACTTGAACCCAGGAGGTGAAAGATTGCAGGGAGCCGAGATCATGCCACTGAACTCCAG 94923

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  ||  || ||||| | |      ||||| ||||||||||      | ||||||||||
Db      94924 CCTGGGTGAC-AGAGCCAGACTCTGTCTCAAAAAAAAAAAAA---TATATATATATATAT 94979

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              ||| | | | |||          | | ||||| ||||| || | ||  || | || ||||
Db      94980 ATACACACACATATGTGTGTGTATATATATACATATATACGTGTACGTATATATGTATAT 95039

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACAT 359
              ||| ||| | |  | |||  |||||| || | |||||||| |||   ||  |||| |  |
Db      95040 ATACATATATACGTGTACGTATATATGTATATATACATATATACGTGTACGTATATATGT 95099

Qy        360 ATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATAT 419
              || |||||| ||||||  |||  |||| | ||||||||  | || |||| ||  ||||||
Db      95100 ATATATACATATATATGTACACGTATATATACATATATGAATATGTATATACGTATATAT 95159

Qy        420 ACACATATATACACATATATACACATATACACATATATACACATATATACATATATACAC 479
              ||  ||||  | |  |||||||   ||||| || ||||||||  |||||||| ||||||| ||
Db      95160 ACGTATATGAATATGTATATATGTATATATACGTATATACATGTATATACGTATATATAC 95219

Qy        480 ATATATA-TAATATACACACATATATATACACATATATACACACATATATACACATATAT 538
              | |||||    |||||||   |||| |   | ||| |   | || | |||| || | |||
Db      95220 ACATATACGTGTATACACGTGTATACACGTATATACACGTATACGTGTATATACGTGTAT 95279

Qy        539 ACACATATATATACACATATATACACATATATACATATATACACATATATATACATATAT 598
              | || |||||||||  |||| |  | ||||||    ||||| | || |||||| ||||||
Db      95280 ATACGTATATATACGTATATGTGTATATATATGGGAATATATATATGTATATATATATAT 95339

Qy        599 ACACATATATACATATAT 616
                    |   |||  | | |
Db      95340 GTCAGTCATTACCTCTTT 95357


RESULT 26
AAK73166
ID   AAK73166 standard; DNA; 19965 BP.
XX
AC   AAK73166;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:27978.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001354.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0235484P.
PR   27-SEP-2000; 2000US-0235834P.
PR   27-SEP-2000; 2000US-0235836P.
PR   29-SEP-2000; 2000US-0236327P.
```

PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS Disclosure; SEQ ID NO 27978; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention
XX
SQ Sequence 19965 BP; 4699 A; 5296 C; 5505 G; 4465 T; 0 U; 0 Other;

Query Match 5.4%; Score 267.4; DB 4; Length 19965;
Best Local Similarity 63.9%; Pred. No. 5.9e-37;
Matches 453; Conservative 0; Mismatches 251; Indels 5; Gaps 3;

Qy 215 AAAAAAAGGCATTTTATATATATATATATATATATACACACACACACACATATATATA 274
Db 4148 AAAAAATATATATAAATATATATATGAAAATATATATATAAATATATATATGAAAATATA 4207

Qy 275 TACACATATATATACACATATATACATATATACACATATATACACATATATATACACATA 334
Db 4208 TATAAATATATATATGAAAATATATATATGAAAATATATATAAATATATATATGAAAATA 4267

Qy 335 CATATGTACACATATATATACACATATGTATACACATATATACACATATATACACACATA 394
Db 4268 TATATGAAAATATATATATGAAAATATATATGAAAATATATATATAAATATATATATATG 4327

Qy 395 TATACACATATATACACACATATATACACATATATACACATATATACACATATACACATA 454
Db 4328 AAAATATATATATAAATATATATATGAAAATATATATATAAATATATATATGAAAATATA 4387

Qy 455 TATACACATATATACATATATACACATATATAT-AATATACACACATATATATACACATA 513
Db 4388 TATATAAATATATATATATGAAAATATATATATAAATATATATATATGAAAATATATATA 4447

Qy 514 TATACACACATATATACACATATATACACATATATATACACATATATACACATATATACA 573
Db 4448 TAAATATATATATATGAAAATATATATATAAATATATATATATGAAAATATATATATAAA 4507

Qy 574 TATATACACATATATATACATATATACACATATATACATATATACACATATATACATATA 633
Db 4508 TATATATATGAAAATATATATATAAATATATATATATGAAAATATATATATAAATATATA 4567

Qy 634 TACACACATATATACACATACATATACACACACATAGATATACATATATATACACATATA 693
Db 4568 TATATGAAAATATATATAAATATATATATGAATATATATATATATATATGAAAATATATA 4627

Qy 694 TATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTA 753
Db 4628 TATATGAAAATATATATATAAATATATATATGAAAATATATATAAATATATATATGAATA 4687

Qy 754 CCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAG 813
Db 4688 TATATATATATATATGAAAATATATATATATGAAAATATATATATAAATATATATATGAA 4747

Qy 814 GTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAG 873

```
Db       4748 AATAAT---ATATATAAATATATATAAATATATATAAAAATATAAATATATATAA-AAAT 4803

Qy        874 TTAAGAAGGTAACACTATATATATATATAGTGAAAGCAGTGTAAACCTA 922
              ||  ||  ||    |||| | |||||||   |||    | || |  ||
Db       4804 ATATAAATATATATATATAAACATATATATATAAATATATATATATATA 4852


RESULT 27
ABK69932
ID   ABK69932 standard; DNA; 19965 BP.
XX
AC   ABK69932;
XX
DT   15-JUL-2002  (first entry)
XX
DE   Human secreted protein gene 68 genomic DNA fragment #20.
XX
KW   Human; ds; secreted protein; gene therapy; immunosuppressive;
KW   antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;
KW   vasotropic; cerebroprotective; nootropic; neuroprotective; antibacterial;
KW   virucide; fungicide; opthalmalogical; autoimmune disease; neoplasm;
KW   rheumatoid arthritis; hyperproliferative disorder; cardiac arrest;
KW   cardiovascular disorder; cerebrovascular disorder; cerebral ischaemia;
KW   angiogenesis; nervous system disorder; Alzheimer's disease; infection;
KW   ocular disorder; corneal infection; wound healing; skin aging;
KW   epithelial cell proliferation; food additive.
XX
OS   Homo sapiens.
XX
PN   WO200226931-A2.
XX
PD   04-APR-2002.
XX
PF   24-SEP-2001; 2001WO-US029871.
XX
PR   25-SEP-2000; 2000US-0234925P.
PR   12-JAN-2001; 2001WO-US000911.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Komatsoulis G,  Duan DR,  Rosen CA,  Moore PA,  Shi Y;
PI   Lafleur DW,  Olsen H,  Brewer LA,  Florence KA,  Young PE,  Soppet DR;
PI   Endress GA,  Mucenski M,  Ebner R;
XX
DR   WPI; 2002-362489/39.
XX
PT   Novel 71 isolated secreted polypeptides and polynucleotides encoding the
PT   polypeptides, useful for treating Huntington's disease, sepsis,
PT   meningitis, thrombocytopenia, hemolytic anemia, rheumatoid arthritis,
PT   asthma.
XX
PS   Example 2; Page 1434-1440; 1478pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule (or its
CC   fragment, homologue complement or allelic variant) encoding a human
CC   secreted protein (and its fragment, domain, epitope, variant, secreted
CC   form and species variant). Also included are a recombinant vector
CC   comprising the nucleic acid, a recombinant host cell comprising the
CC   vector, an antibody against the secreted protein, a recombinant host cell
CC   that expresses the secreted protein and a method of identifying a binding
CC   partner of the secreted protein. The nucleic acid and protein are used to
CC   prevent, diagnose, treat or ameliorate a medical condition in e.g.
CC   humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep for
CC   example autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
CC   disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
CC   e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC   angiogenesis, nervous system disorders e.g. Alzheimer's disease,
CC   infections caused by bacteria, viruses and fungi and ocular disorders
CC   e.g. corneal infection. Many other diseases and disorders are listed in
CC   the specification. The polypeptides can also be used to aid wound healing
CC   an epithelial cell proliferation, to prevent skin aging due to sunburn,
CC   to maintain organs before transplantation, for supporting cell culture of
CC   primary tissues, to regenerate tissues and in chemotaxis. The
CC   polypeptides can also be used as a food additive or preservative to
CC   increase or decrease storage capabilities. The present sequence
CC   represents a ds DNA fragment of the gene for a novel human secreted
CC   protein of the invention
XX
SQ   Sequence 19965 BP; 4699 A; 5296 C; 5505 G; 4465 T; 0 U; 0 Other;

  Query Match             5.4%;  Score 267.4;  DB 6;  Length 19965;
  Best Local Similarity   63.9%;  Pred. No. 5.9e-37;
  Matches  453;  Conservative    0;  Mismatches  251;  Indels    5;  Gaps    3;

Qy        215 AAAAAAAGGCATTTTATATATATATATATATATACACACACACACATATATATA 274
              ||||||      |  |||||||||||  | |||||||||| | | | | | |   | |||||
Db       4148 AAAAAATATATATAAATATATATATGAAAATATATATATAAATATATATATGAAAATATA 4207

Qy        275 TACACATATATATACACATATATACATATATACACATATATACACATATATATACACATA 334
              || | |||||||||    | |||||| ||||  | | |||||||| | ||||||||  | |||
Db       4208 TATAAATATATATATGAAAATATATATATGAAAATATATATAAATATATATATGAAAATA 4267

Qy        335 CATATGTACACATATATATACACATATGTATACACATATATACACATATATACACACATA 394
              |||||  | | ||||||||  | |||| |||  | |||||||  | | |||||  | | ||
Db       4268 TATATGAAAATATATATATGAAAATATATATGAAAATATATATATAAATATATATATATG 4327

Qy        395 TATACACATATATACACACATATATACACATATATACACATATATACACATATACACATA 454
              | | | ||||||| | | ||||||  | |||||||  | | |||||  | ||  | | |||
Db       4328 AAAATATATATATAAATATATATATGAAAATATATATATAAATATATATATGAAAATATA 4387

Qy        455 TATACACATATATACATATATACACATATATAT-AATATACACACATATATATACACATA 513
              ||||  | |||||||  ||||   | | ||||||||  ||||||  | | ||  | |||  | |||
Db       4388 TATATAAATATATATATATGAAAATATATATATAAATATATATATATGAAAATATATATA 4447

Qy        514 TATACACACATATATACACATATATACACATATATATACACATATATACACATATATACA 573
              || | | | ||||||  | |||||||  | | |||||||  | ||  | | |  |||||||  |
Db       4448 TAAATATATATATATGAAAATATATATATAAATATATATATATGAAAATATATATATAAA 4507

Qy        574 TATATACACATATATATACATATATACACATATATACATATATACACATATATACATATA 633
              ||||||  |   | |||||  |||||  | | |||||||   | |||  | |||||  | |||||
Db       4508 TATATATATGAAAATATATATATAAATATATATATATGAAAATATATATATAAATATATA 4567

Qy        634 TACACACATATATACACATACATATACACACACATAGATATACATATATATACACATATA 693
              ||  |   | |||||  | |  | |||||  |   | |||  |||||  ||||||  | |  |||||
Db       4568 TATATGAAAATATATATAAATATATATATGAATATATATATATATATATGAAAATATATA 4627

Qy        694 TATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTA 753
              ||||  | |  ||||||  |||| ||||||      | | |  |||  |       |  ||| ||
Db       4628 TATATGAAAATATATATATAAATATATATATGAAAATATATATAAATATATATATGAATA 4687

Qy        754 CCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAG 813
                  |   ||| | ||||| | |  |  |    | | |||    .| |   | || | |
Db       4688 TATATATATATATATGAAAATATATATATATGAAAATATATATATAAATATATATATGAA 4747

Qy        814 GTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAG 873
              | ||   | |||  |    | ||| ||| || |  ||  | |  ||| | |||   |
Db       4748 AATAAT---ATATATAAATATATATAAATATATATAAAAATATAAATATATATAA-AAAT 4803

Qy        874 TTAAGAAGGTAACACTATATATATATATAGTGAAAGCAGTGTAAACCTA 922
              ||  ||  ||    |||| | |||||||   |||    | || |  ||
Db       4804 ATATAAATATATATATATAAACATATATATATAAATATATATATATATA 4852


RESULT 28
ABL33237/c
ID   ABL33237 standard; DNA; 6971 BP.
XX
AC   ABL33237;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 1210.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
```

```
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP007537.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful for
PT   diagnosis and treatment of diseases associated with abnormal cytosine
PT   methylation.
XX
PS   Claim 1; SEQ ID NO 1210; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention
XX
SQ   Sequence 6971 BP; 1665 A; 242 C; 1831 G; 3232 T; 0 U; 1 Other;

  Query Match             5.4%;  Score 266.4;  DB 6;  Length 6971;
  Best Local Similarity   71.1%;  Pred. No. 8.3e-37;
  Matches  367;  Conservative    0;  Mismatches  146;  Indels    3;  Gaps    1;

Qy        212 AAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACACATATAT 271
              | | | | |    |  | ||||||||  |||||||||| |||||| | | | | ||||||
Db       6472 ATACATATATACCTATATATATATACCTATATATATAAATACACATATATATATATATAT 6413

Qy        272 ATATACACATATATATACACATATATACATATATACACATATATACACATATATATACAC 331
              |||   | ||||||||| |||  |||| ||||||||||  |||| |||||||||   |
Db       6412 ATACCAATATATATATATACACGTATATATATATACACACTTATATACATATATACCTAT 6353

Qy        332 ATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACACAC 391
              ||| |||  |||||||||||||| | |||  |||| | ||| |||||||| |||| | |
Db       6352 ATATATAAATACACATATATATATATACCTATATATATAAATACACATTTATATATAT 6293

Qy        392 ATATATACACATATATACACACATATATACACATATATACACATATATACACATATACAC 451
              | ||| |||  |||||| | | |    || | | ||| |||  |||||| | ||||||
Db       6292 AAATACACATTTATATATATATACGCCTATATAAATACACATTTATATATATATATACCT 6233

Qy        452 ATATATACACATATATACATATATACACATATATATAATATACACACATATATATACACA 511
              ||||||| || |||||| ||| ||||  |||||||||| | |||   |||||||| | |
Db       6232 ATATATATACCTATATATATAAATACCTATATATATAAAACACATTTATATATATATATA 6173

Qy        512 TATATACACACATATATACACATATATACACATATATATACACATATATACACATATATA 571
                |||| | | ||| | |   ||||||| || | ||||||  | |  |||| | | | |||
Db       6172 CCTATATATAAATACACATTTATATATATACCTCTATATATATACCTATATATAAAAATA 6113

Qy        572 CATATATACACATATATATACATATATACACATATATACATATATACACATATATACATA 631
              || || || | |||    || ||||| | | | | ||  |||||||| | |||||| |||
Db       6112 CACATTTATATATACGCCTATATATAAAAATACACATTTATATATACGCCTATATATATA 6053

Qy        632 TATACACACATATATACACATACATATACACACACATAGATATACATATATAT---ACAC 688
```

```
              ||||||| ||||||  | ||||   || | | | | |.|  || ||||||||   ||||
Db       6052 AATACACATTTATATATATATACGCCTATATATATAAATACCTATATATATATAAAACAC 5993

Qy        689 ATATATATACGTATATATATGTATATATATATGCTC 724
              || ||||||  ||||||||  |||||||| || | |
Db       5992 ATTTATATATATATATATACCTATATATAAATACAC 5957


RESULT 29
ADM01858/c
ID   ADM01858 standard; cDNA; 2522 BP.
XX
AC   ADM01858;
XX
DT   20-MAY-2004  (first entry)
XX
DE   Human cDNA of the invention SEQ ID NO:543.
XX
KW   ss; gene; human; gene therapy; diagnostic marker; pharmaceutical.
XX
OS   Homo sapiens.
XX
PN   EP1347046-A1.
XX
PD   24-SEP-2003.
XX
PF   12-APR-2002; 2002EP-00008400.
XX
PR   22-MAR-2002; 2002JP-00137785.
XX
PA   (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI   Isogai T,  Sugiyama T,  Otsuki T,  Wakamatsu A,  Sato H,  Ishii S;
PI   Yamamoto J,  Isono Y,  Hio Y,  Otsuka K,  Nagai K,  Irie R,  Tamechika I;
PI   Seki N,  Yoshikawa T,  Otsuka M,  Nagahari K,  Masuho Y;
XX
DR   WPI; 2003-723558/69.
DR   P-PSDB; ADM04301.
XX
PT   New polynucleotides and polypeptides are useful in gene therapy, for
PT   developing a diagnostic marker or medicines for regulating their
PT   expression and activity, or as a target of gene therapy.
XX
PS   Claim 1; SEQ ID NO 543; 305pp; English.
XX
CC   The invention relates to a novel human polynucleotide and the encoded
CC   polypeptide. A polynucleotide of the invention may have a use in gene
CC   therapy. An oligonucleotide of the invention ADM06202-ADM06773 is useful
CC   as a primer for synthesizing the polynucleotide or as a probe for
CC   detecting the polynucleotide. The polynucleotides ADM01316-ADM03758 are
CC   useful in gene therapy, for developing a diagnostic marker or medicines
CC   for regulating their expression and activity, or as a target of gene
CC   therapy. The proteins ADM03759-ADM06201 encoded by the polynucleotides
CC   are useful as pharmaceutical agents. The present sequence represents a
CC   cDNA sequence of the invention.
XX
SQ   Sequence 2522 BP; 789 A; 399 C; 547 G; 787 T; 0 U; 0 Other;

  Query Match             5.3%;  Score 263.6;  DB 11;  Length 2522;
  Best Local Similarity   68.9%;  Pred. No. 2.4e-36;
  Matches  376;  Conservative    0;  Mismatches  169;  Indels    1;  Gaps    1;

Qy        236 ATATATATATATATATACACACACACACACATATATATATACACATATATATACACATAT 295
              ||   |  | | ||||||  | |||||||  |  |||||||     |||||||||||||  |
Db       2217 ATTCAACTGTGTATATATATATACACACATACGTATATATGTGTATATATATACACACGT 2158

Qy        296 ATACATATATACACATATATACACATATATATACACATACATATGTACACATATATATAC 355
              ||| ||| | | | || |||||| | | |   | | |||||  || ||   ||| ||| ||
Db       2157 ATATATACACATATATGTATACATACACACGCATATATACACATATATGTATACATACAC 2098

Qy        356 ACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACAT 415
              || ||| ||||||||||||||  | | ||| |||| ||||||||||||||||||  | ||||
Db       2097 ACGTATATATACACATATATGTATACATACACACGCATATATACACATATATGTATACAT 2038
```

```
Qy        416 ATATACACATATATACACATATATACACATATACACATATATACACATATATACATATAT 475
              | | ||  |||||||||||||||| || | |||||||   ||| | | | ||| ||  ||
Db       2037 ACACACGTATATATACACATATATGCATACATACACACGCATATATACACATATATGCAT 1978

Qy        476 ACACATATATATAATATACACACATATATATACACATATATACACACATATATACACATA 535
              ||| | ||| . | ||||| |||||||||| || ||||| ||||  | | ||| ||| |||
Db       1977 ACATACATACGT-ATATATACACATATATGTATACATACATACGTATACATACACATATA 1919

Qy        536 TATACACATATATATACACATATATACACATATATACATATATACACATATATATACATA 595
              | || ||||| ||   || ||||| ||| |||| || | ||| | || ||||||||||| |
Db       1918 TGTATACATACATGCGCATATATACACATATATGTATACATACATACGTATATATACACA 1859

Qy        596 TATACACATATATACATATATACACATATATACATATATACACACATATATACACATACA 655
              ||||   ||| |||||||  || ||||| | | || |||||| |||||  || | |||||
Db       1858 TATATGTATACATACATACGTATACATACACATATGTATACATACATACGTATATATACA 1799

Qy        656 TATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATA 715
              |||  |   | | | || | ||||||||||||||||| ||   ||| ||  |||||||
Db       1798 CATATATGTATACATACATGCGTATATATACACATATATGTATACATACATGCGTATATA 1739

Qy        716 TATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGA 775
              || |       |  | ||||      | ||  |   ||    || | ||| |   | | |
Db       1738 TACACATATATGCATACATACATATGAACACATATATATGTGTGTGTATATATACATATA 1679

Qy        776 GAGTTT 781
                |||
Db       1678 AGATTT 1673
```

RESULT 30
AAD54480/c
ID AAD54480 standard; DNA; 117962 BP.
XX
AC AAD54480;
XX
DT 26-JUN-2003 (first entry)
XX
DE Human CIP DNA #1.
XX
KW Human; p53 pathway; chloride cotransporter interactor protein; CIP;
KW angiogenic disorder; cell proliferation disorder; apoptotic disorder;
KW breast cancer; gene therapy; ds.
XX
OS Homo sapiens.
XX
PN WO200299055-A2.
XX
PD 12-DEC-2002.
XX
PF 03-JUN-2002; 2002WO-US017473.
XX
PR 05-JUN-2001; 2001US-0296076P.
PR 10-OCT-2001; 2001US-0328605P.
PR 15-FEB-2002; 2002US-0357253P.
XX
PA (EXEL-) EXELIXIS INC.
XX
PI Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP;
XX
DR WPI; 2003-175140/17.
XX
PT Identifying p53 pathway modulators for treating or diagnosing disorders
PT with defective p53 function e.g. breast cancer, by providing an assay
PT system having a purified cotransporter interactor protein (CIP)
PT polypeptide or nucleic acid.
XX
PS Disclosure; Page 38-101; 123pp; English.
XX
CC The invention relates to a method of Identifying p53 pathway modulating
CC agent. The method involves contacting a test agent with an assay system
CC comprising a purified cation Cl- cotransporter interactor protein (CIP)
CC polypeptide or polynucleotide, or their functionally active fragment or
CC derivative. The method is useful for identifying modulators of the p53
CC pathway particularly for identifying agents for treating disorders (e.g.
CC breast cancer) associated with defective p53 function. Modulators of the
CC invention are useful as targets for novel therapeutics. CIP sequences are
CC useful as modifiers of the p53 pathway, and as therapeutic targets for
CC disorders associated with defective p53 function e.g. angiogenic,
CC apoptotic or cell proliferation disorders. The invention is useful in
CC gene therapy. The present sequence is human CIP DNA
XX
SQ Sequence 117962 BP; 27840 A; 32096 C; 30624 G; 27402 T; 0 U; 0 Other;

Query Match 5.3%; Score 263; DB 8; Length 117962;
Best Local Similarity 71.6%; Pred. No. 3.8e-36;
Matches 373; Conservative 0; Mismatches 145; Indels 3; Gaps 2;

```
Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||  |||||||||||  ||||||||||||||||||||||||||||||||||| ||||||
Db     107030 AGTTTGAGACCAGCCTGGCCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAAATT 106971

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              || ||||||||||  ||| |||||||||||||||||| | |||||| || | |||||||
Db     106970 AGCCAGGCATGGTGGCACATGCCTGTAATCCCAGCTACCTGGGAGGCTGAAGTAGGAGAA 106911

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              || || ||||| |  || ||  |||| ||| | |||| |                | | |   ||
Db     106910 TCTCTGGAACCAG--GGAGGCAGAGGCTGCAGTGAGCCAAGATCATTCCTGGGTGTCAGA 106853

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              || |  ||      |       |  | || ||||||||||    |    | | | | |
Db     106852 GAGAGACTCCATCCCTGCCCGTCCCCCGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 106793

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              | | | ||||||| | | | | | | |||||||||||| | ||||| | ||||| | |||
Db     106792 AAAAAAATATATATATATATATATATATATATATATATATATACACACACACACATAT 106733

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACAT 359
              ||||| | | ||||| |||||||||||||| ||| |||  || | ||| ||||| | ||
Db     106732 ATATACATATATATACACACATATATATACATATATATACATATATATACATATATAT 106673

Qy        360 ATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATAT 419
              | ||||| | ||| ||| | ||||| ||| | ||||| ||| |||||||||| | ||||||
Db     106672 ACGTATATATATACATATATATATACACATATATATACACATATATATACATATATATAT 106613

Qy        420 ACACATATATACACATATATACACATATACACATATATACACATATATACATATATACAC 479
              |||||||||| ||| ||||| | | ||| | ||||| ||| ||||||||||||||| |
Db     106612 ACACATATATATACACATATATATACATATATATATACACATATATATACATATATATAT 106553

Qy        480 ATATATATAATATACACACATATATATACACATATATACAC 520
              || |||||  |||  | |      || |    || || |||
Db     106552 ATGTATATTTTATTGAGATCACATTAAATGTGTACATTCAC 106512
```

RESULT 31
ABK83460
ID ABK83460 standard; cDNA; 165199 BP.
XX
AC ABK83460;
XX
DT 14-AUG-2002 (first entry)
XX
DE Human cDNA differentially expressed in granulocytic cells #31.
XX
KW Human; ss; granulocytic cell; DNA chip; bacterial infection;
KW viral infection; parasitic infection; protozoal infection;
KW fungal infection; sterile inflammatory disease; psoriasis;
KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KW cardiac reperfusion injury; renal reperfusion injury; ARDS;
KW adult respiratory distress syndrome; inflammatory bowel disease;
KW Crohn's disease; ulcerative colitis; periodontal disease;
KW granulocyte activation; chronic inflammation; allergy.
XX
OS Homo sapiens.
XX

PN WO200228999-A2.
XX
PD 11-APR-2002.
XX
PF 03-OCT-2001; 2001WO-US030821.
XX
PR 03-OCT-2000; 2000US-0237189P.
XX
PA (GENE-) GENE LOGIC INC.
XX
PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
XX
DR WPI; 2002-435328/46.
XX
PT Detecting granulocyte activation by detecting differential expression of
PT genes associated with granulocyte activation, which serves as diagnostic
PT markers that is useful for monitoring disease states and drug toxicity.
XX
PS Claim 1; SEQ ID NO 31; 114pp; English.
XX
CC The invention relates to detecting (M1) granulocyte (GC) activation
CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
CC DNA chip analysis as given in the specification, and comparing the
CC expression level to an expression level in an unactivated GC, where
CC differential expression of Gs is indicative of GCA. Also included are
CC modulating (M2) GA by contacting GC with an agent that alters the
CC expression of at least one gene in Gs; (2) screening (M3) for an agent
CC capable of modulating GCA or an inflammation (especially chronic) in a
CC tissue, an allergic response in a subject, exposure of a subject to a
CC pathogen or sterile inflammatory disease using the gene expression
CC profile; (3) detecting (M4) an inflammation (especially chronic) in a
CC tissue, an allergic response in a subject, exposure of a subject to a
CC pathogen or sterile inflammatory disease, by detecting the level of
CC expression in a sample of the tissue of gene(s) from Gs, where the level
CC of expression of the gene is indicative of inflammation; (4) treating
CC (M5) an inflammation (especially chronic) or in a tissue, an allergic
CC response in a subject, exposure of a subject to a pathogen or sterile
CC inflammatory disease, by contacting a tissue having inflammation with an
CC agent that modulates the expression of gene(s) from Gs in the tissue. M1
CC is useful for detecting GCA; M2 is useful for modulating GA; M3 is useful
CC for screening an agent capable of modulating GCA preferably in an
CC inflammation in a tissue; M4 is useful for detecting an inflammation
CC (especially chronic) in a tissue, an allergic response in a subject,
CC exposure of a subject to a pathogen or sterile inflammatory disease (e.g.
CC psoriasis, rheumatoid arthritis, glomerulonephritis, asthma, thrombosis,
CC cardiac reperfusion injury, renal reperfusion injury, ARDS, adult
CC respiratory distress syndrome, inflammatory bowel disease, Crohn's
CC disease, ulcerative colitis, periodontal disease; also bacterial
CC infection, viral infection, parasitic infection, protozoal infection,
CC fungal infection and M5 is useful for treating one of the above
CC conditions. The present sequence represents a gene differentially
CC expressed in granulocytes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 165199 BP; 48510 A; 33223 C; 34406 G; 49060 T; 0 U; 0 Other;

Query Match 5.3%; Score 263; DB 6; Length 165199;
Best Local Similarity 69.3%; Pred. No. 3.8e-36;
Matches 472; Conservative 0; Mismatches 200; Indels 9; Gaps 8;

```
Qy        46 AAAAATACAAAATTAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGG 105
             | ||| | | |    | |||||||| | ||| ||||||||||||||||   |||||
Db    138669 AGAAAGAAAGATGGATGGATGCATGGTGGTGCACACCTGTAATCCCAGCTACACAGGAGG 138728

Qy       106 CCGAGGCAGGAGAATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGA-GCAAGATCAT 164
             | | ||||  | ||||||||||||||  || ||| ||| | ||   ||  ||||||  |
Db    138729 CTGGGGCATTAAAATCACTTGAACCTG--GGAGGTAGAGTTAGCAATGACACAAGATTGT 138786

Qy       165 GCCATTGCACTCCAGCCCAGGCAACAAGAGCGAAATGTC-ATCTCAGAAAAAAAAAAAGG 223
             ||   ||||||||||||  ||   ||  |    | ||||  |||| | |    |   |
Db    138787 GCTGCTGCACTCCAGCCTGGGTGGCAGAATGAGACTGTCTCTCTCTGTATCTAGCTATCT 138846

Qy       224 CATTTTATATATATATATATATATATACACACACA-CACACATATATATATACACATA 282
                 | ||||||||||  ||||||  | ||  | |  | ||  | ||||  |||||  | |||
Db    138847 ATCCATCCATATATATATCCATATATGTCCATATATATCATATATATCCATATATAGATA 138906

Qy       283 TATATACACATATATACATATATACACATATATACACATATATATACACATACATATGT- 341
             ||||||   |||||||  ||||||  | ||   |||  |  ||||  ||||  |  ||  ||||  |
Db    138907 TATATATCTATATATATCTATATATA-ATCCATATATATATCTATATATATCTATATATA 138965

Qy       342 ACACATATATATACACATATGTATACACATATATACACATATATACACACATATATACAC 401
             |  ||||||||||  |||||  |||  ||  |||||  |  |  ||||||  ||  |  ||||   |  |
Db    138966 ATCCATATATATATCCATATATATCCATATATAGATATATATATCCATATATATCCATAT 139025

Qy       402 ATATATACACACATATATACACATATATACACATATATACA-CATATACACATATATACA 460
             |||  |||  |  |  ||  |||  |  ||||  ||  |  ||  ||||  |   ||   |  |  ||||||  ||
Db    139026 ATAGATATATATATCCATATATATATCTATATATCTATATATAATCAATATATATATCCA 139085

Qy       461 CATATATACATATATACACATATATA-TAATATACACACATATATATACACATATATACA 519
             |||||  |  ||||||  ||||||||  |  |||||  |  ||||||||||  ||||||||  |
Db    139086 TATATAGATGTATATATCCATATATAGTTATATATATCCATATATATATCCATATATAGA 139145

Qy       520 CACATATATACACATATATACACATATATATACACATATATACACATATATACATATATA 579
             |  ||||||  ||  |||||||  ||||||||  |  |  ||||||  ||  |||||||   |||||
Db    139146 TATATATATCCATATATATATCCATATATAGATATATATATCCATATATATATCCATATA 139205

Qy       580 CACATATATATACATATATACACATATATACATATATACACATATATACATATATACACA 639
             |  |||||||||   |||||  |  ||  |||  |||  |||  |  ||||  |  |||||||  ||
Db    139206 TAGATATATATATCCATATATATATCCATATATAGATATATATATCCATATATATATCCA 139265

Qy       640 CATATATACACATACATATACACACACATAGATATACATATATATACACATATATATACG 699
             |||||  |  |  |||   |||  |  |  |   |  |||||   |||||||  |  ||||||||  |
Db    139266 TATATAGATATATATCCATATATATATCCATATATATCCATATATAGATATATATATCCA 139325

Qy       700 TATATATATGTATATATATAT 720
             |||||||||  |||||||  ||
Db    139326 TATATATATCCATATATAGAT 139346
```

RESULT 32
ABK83497/c
ID ABK83497 standard; cDNA; 227968 BP.
XX
AC ABK83497;
XX
DT 14-AUG-2002 (first entry)
XX
DE Human cDNA differentially expressed in granulocytic cells #68.
XX
KW Human; ss; granulocytic cell; DNA chip; bacterial infection;
KW viral infection; parasitic infection; protozoal infection;
KW fungal infection; sterile inflammatory disease; psoriasis;
KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KW cardiac reperfusion injury; renal reperfusion injury; ARDS;
KW adult respiratory distress syndrome; inflammatory bowel disease;
KW Crohn's disease; ulcerative colitis; periodontal disease;
KW granulocyte activation; chronic inflammation; allergy.
XX
OS Homo sapiens.
XX
PN WO200228999-A2.
XX
PD 11-APR-2002.
XX
PF 03-OCT-2001; 2001WO-US030821.
XX
PR 03-OCT-2000; 2000US-0237189P.
XX
PA (GENE-) GENE LOGIC INC.
XX
PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
XX
DR WPI; 2002-435328/46.
XX

```
PT   Detecting granulocyte activation by detecting differential expression of
PT   genes associated with granulocyte activation, which serves as diagnostic
PT   markers that is useful for monitoring disease states and drug toxicity.
XX
PS   Claim 1; SEQ ID NO 68; 114pp; English.
XX
CC   The invention relates to detecting (M1) granulocyte (GC) activation
CC   (GCA), by detecting the level of expression of gene(s) (Gs) identified by
CC   DNA chip analysis as given in the specification, and comparing the
CC   expression level to an expression level in an unactivated GC, where
CC   differential expression of Gs is indicative of GCA. Also included are
CC   modulating (M2) GA by contacting GC with an agent that alters the
CC   expression of at least one gene in Gs; (2) screening (M3) for an agent
CC   capable of modulating GCA or an inflammation (especially chronic) in a
CC   tissue, an allergic response in a subject, exposure of a subject to a
CC   pathogen or sterile inflammatory disease using the gene expression
CC   profile; (3) detecting (M4) an inflammation (especially chronic) in a
CC   tissue, an allergic response in a subject, exposure of a subject to a
CC   pathogen or sterile inflammatory disease, by detecting the level of
CC   expression in a sample of the tissue of gene(s) from Gs, where the level
CC   of expression of the gene is indicative of inflammation; (4) treating
CC   (M5) an inflammation (especially chronic) or in a tissue, an allergic
CC   response in a subject, exposure of a subject to a pathogen or sterile
CC   inflammatory disease, by contacting a tissue having inflammation with an
CC   agent that modulates the expression of gene(s) from Gs in the tissue. M1
CC   is useful for detecting GCA; M2 is useful for modulating GA; M3 is useful
CC   for screening an agent capable of modulating GCA preferably in an
CC   inflammation in a tissue; M4 is useful for detecting an inflammation
CC   (especially chronic) in a tissue, an allergic response in a subject,
CC   exposure of a subject to a pathogen or sterile inflammatory disease (e.g.
CC   psoriasis, rheumatoid arthritis, glomerulonephritis, asthma, thrombosis,
CC   cardiac reperfusion injury, renal reperfusion injury, ARDS, adult
CC   respiratory distress syndrome, inflammatory bowel disease, Crohn's
CC   disease, ulcerative colitis, periodontal disease; also bacterial
CC   infection, viral infection, parasitic infection, protozoal infection,
CC   fungal infection and M5 is useful for treating one of the above
CC   conditions. The present sequence represents a gene differentially
CC   expressed in granulocytes. Note: The sequence data for this patent did
CC   not form part of the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 227968 BP; 53185 A; 60440 C; 61985 G; 52358 T; 0 U; 0 Other;

  Query Match             5.3%;  Score 263;  DB 6;  Length 227968;
  Best Local Similarity   71.6%;  Pred. No. 3.9e-36;
  Matches  373;  Conservative    0;  Mismatches  145;  Indels    3;  Gaps    2;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||  ||||||||||  ||||||||||||||||||||||||||||||||| ||||||
Db       5103 AGTTTGAGACCAGCCTGGCCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAAATT 5044

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              || ||||||||||  ||| |||||||||||||||||| | |||||| || | |||||||
Db       5043 AGCCAGGCATGGTGGCACATGCCTGTAATCCCAGCTACCTGGGAGGCTGAAGTAGGAGAA 4984

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              || || ||||| |  || ||  |||| ||| | |||| |       | | |   ||
Db       4983 TCTCTGGAACCAG--GGAGGCAGAGGCTGCAGTGAGCCAAGATCATTCCTGGGTGTCAGA 4926

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
                || |  ||      |       |  | || ||||||||||    |    | | | | |
Db       4925 GAGAGACTCCATCCCTGCCCGTCCCCCGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4866

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              | | | ||||||| | | | | | | ||||||||||| | ||||| | ||||| | |||
Db       4865 AAAAAAATATATATATATATATATATATATATATATATATATACACACACACATAT 4806

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACAT 359
              ||||| | | ||||| ||||||||||||||| ||| |||  || | ||| ||||| | ||
Db       4805 ATATACATATATATACACACATATATATACATATATATACATATATATACATATATATAT 4746

Qy        360 ATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATAT 419
```

```
              | ||||| | ||| ||| | ||||| ||| | ||||| ||| ||||||||| | ||||||
Db       4745 ACGTATATATATACATATATATATACACATATATATACACATATATATACATATATAT 4686

Qy        420 ACACATATATACACATATATACACATATACACATATATACACATATATACATATATACAC 479
              |||||||||||| ||| ||||| | | ||| | ||||| ||| ||||||||||||||| | |
Db       4685 ACACATATATATACACATATATATACATATATATATACACATATATATACATATATAT 4626

Qy        480 ATATATATAATATACACACATATATATACACATATATACAC 520
              || |||||  |||  | |       || |    || || |||
Db       4625 ATGTATATTTTATTGAGATCACATTAAATGTGTACATTCAC 4585


RESULT 33
ADQ18538/c
ID   ADQ18538 standard; DNA; 227968 BP.
XX
AC   ADQ18538;
XX
DT   26-AUG-2004  (first entry)
XX
DE   Human soft tissue sarcoma-upregulated DNA - SEQ ID 1357.
XX
KW   soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO2004048938-A2.
XX
PD   10-JUN-2004.
XX
PF   26-NOV-2003; 2003WO-US038193.
XX
PR   26-NOV-2002; 2002US-0429739P.
XX
PA   (PROT-) PROTEIN DESIGN LABS INC.
XX
PI   Aziz N,  Ginsburg WM,  Zlotnik A;
XX
DR   WPI; 2004-441208/41.
XX
PT   Early detection of soft tissue sarcoma comprises determining expression
PT   of a gene in a first soft tissue sample and a normal soft tissue sample
PT   and comparing the gene expression, also useful in treating soft tissue
PT   sarcoma.
XX
PS   Example 2; SEQ ID NO 1357; 210pp; English.
XX
CC   The invention relates to a novel method for detecting soft tissue sarcoma
CC   which comprises obtaining a first soft tissue sample from an individual
CC   and a normal soft tissue sample from the same or different individual,
CC   determining the expression of a gene in both samples and comparing the
CC   expression of the gene in both soft tissue samples, where a higher level
CC   of protein expression in the first soft tissue sample indicates the
CC   presence of soft tissue sarcoma. The method of the invention has
CC   cytostatic applications and may be useful for detecting soft tissue
CC   sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC   acid sequences may be useful in diagnostic and screening applications.
CC   The current sequence is that of a human soft tissue sarcoma-upregulated
CC   DNA of the invention. The current sequence is not shown within the
CC   specification per se but was submitted in CD format by the inventor.
XX
SQ   Sequence 227968 BP; 53185 A; 60440 C; 61985 G; 52358 T; 0 U; 0 Other;

  Query Match             5.3%;  Score 263;  DB 12;  Length 227968;
  Best Local Similarity   71.6%;  Pred. No. 3.9e-36;
  Matches  373;  Conservative    0;  Mismatches  145;  Indels    3;  Gaps    2;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||  ||||||||||  ||||||||||||||||||||||||||||||||| ||||||
Db       5103 AGTTTGAGACCAGCCTGGCCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAAATT 5044

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
```

```
Db       5043 AGCCAGGCATGGTGGCACATGCCTGTAATCCCAGCTACCTGGGAGGCTGAAGTAGGAGAA 4984

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179

Db       4983 TCTCTGGAACCAG--GGAGGCAGAGGCTGCAGTGAGCCAAGATCATTCCTGGGTGTCAGA 4926

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239

Db       4925 GAGAGACTCCATCCCTGCCCGTCCCCCGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4866

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299

Db       4865 AAAAAAATATATATATATATATATATATATATATATATATACACACACACACATAT 4806

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACAT 359

Db       4805 ATATACATATATATACACACATATATATACATATATATACATATATATACATATATATAT 4746

Qy        360 ATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATAT 419

Db       4745 ACGTATATATATACATATATATATACACATATATATACACATATATATACATATATATAT 4686

Qy        420 ACACATATATACACATATATACACATATACACATATATACACATATATACATATATACAC 479

Db       4685 ACACATATATATACACATATATATACATATATATATACACATATATATACATATATAT 4626

Qy        480 ATATATATAATATACACACATATATATACACATATATACAC 520

Db       4625 ATGTATATTTTATTGAGATCACATTAAATGTGTACATTCAC 4585


RESULT 34
ABK39930/c
ID   ABK39930 standard; DNA; 12069 BP.
XX
AC   ABK39930;
XX
DT   21-MAY-2002  (first entry)
XX
DE   Human chemically pretreated gene sequence #6 strand 1.
XX
KW   Human; ds; bisulphite treatment; CpG; DNA methylation; cancer; tumour;
KW   cytostatic; ALDH6; CYP11A; CYP11B1; CYP3A3; DPYD; EPHX2; OCLN; TXNRD1;
KW   UGT8; MRP; pharmacogenomics; SNP; single nucleotide polymorphism.
XX
OS   Homo sapiens.
XX
PN   WO200202806-A2.
XX
PD   10-JAN-2002.
XX
PF   29-JUN-2001; 2001WO-EP007470.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-154757/20.
XX
PT   New nucleic acid, oligonucleotides and peptide nucleic acid-oligomers,
PT   useful for detecting cytosine methylation state of genes associated with
PT   pharmacogenomics and for therapy of diseases e.g. cancer.
XX
PS   Claim 1; SEQ ID NO 11; 24pp; English.
XX
CC   The invention relates to a nucleic acid comprising a sequence at least 18
CC   bases in length of a segment of the chemically pretreated DNA of genes
CC   associated with pharmacogenomics according to one of the sequences of the
CC   genes ALDH6 (NM_000693), CYP11A (NM_000781), CYP11B1 (NM_000497), CYP3A3
CC   (NM_000776 and NM_017460), DPYD (NM_000110), EPHX2 (NM_001979), OCLN
CC   (NM_002538), TXNRD1 (NM_003330), UGT8 (NM_003360), MRP (NM_004996,
CC   NM_019900, NM_019901, NM_019902, NM_019862, NM_019898, NM_019899) and
CC   their complementary sequences, or a sequence (S1) chosen from 87
CC   sequences and their complements. The chemical pretreatment is bisulphite
CC   treatment to convert cytosines (but not methyl-cytosines) into uracils.
CC   Also included are an oligomer (II) in particular an oligonucleotide or a
CC   peptide nucleic acid (PNA)-oligomer, comprising in each case at least one
CC   base sequence having a length of 9 nucleotides which hybridises to or is
CC   identical to a chemically pretreated DNA of genes associated with
CC   pharmacogenomics and their complements, arranged in an array for
CC   analysing diseases associated with the methylation state (CpG) and/or
CC   detecting SNPs (single nucleotide polymorphisms) of the 87 sequences. The
CC   oligomers may also be used as PCR primers. The set of 87 nucleic acids
CC   and their complements is useful for diagnosis and therapy of solid
CC   tumours and cancer. The present sequence represents one the 87 DNA
CC   sequences or its complement. Note: The sequence data for this patent did
CC   not form part of the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 12069 BP; 3210 A; 185 C; 2849 G; 5823 T; 0 U; 2 Other;

  Query Match             5.3%;  Score 261.4;  DB 6;  Length 12069;
  Best Local Similarity   70.5%;  Pred. No. 6.3e-36;
  Matches  363;  Conservative    0;  Mismatches  151;  Indels    1;  Gaps    1;

Qy        204 ATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACAC 263

Db       9846 ATATCTATATATATATAAATATATCTATATATATATAAATATCTATATATATATAAATAT 9787

Qy        264 ACATATATATATACACATATATATACACATATATACATATATACACATATATACACATAT 323

Db       9786 CTATATACGTATAAATATCTATATATATATAAATATCTATATCCACGTATATCTATATAT 9727

Qy        324 ATATACACATACATATGTACACATATATATACACATATGTATACACATATATACACATAT 383

Db       9726 ATATAAATATCTATATATATATAAATATATATATCCACGTATATCTATATATATATAAAT 9667

Qy        384 ATACACACATATATACACATATATACACACATATATACACATATATACACATATATACAC 443

Db       9666 ATCTATATATATATAAATATATATATCCACGTATATCTATATATATATAAATATCTACGT 9607

Qy        444 ATATACACATATATACACATATATACATATATACACATATATAATATACACACATATA 503

Db       9606 ATAAATATATATATCCACGTATATCTATATATATATAAATATCT-ACGTATAAATATATA 9548

Qy        504 TATACACATATATACACACATATATACACATATATACACATATATATACACATATATACA 563

Db       9547 TATCCACGTATATCTATATATATATAAATATCTATATATATATAAATATATATATCCACG 9488

Qy        564 CATATATACATATATACACATATATATACATATATACACATATATACATATATACACATA 623

Db       9487 TATATCTATATATATATAAATATCTATATATATATAAATATCTCTATATATATAAATATA 9428

Qy        624 TATACATATATACACACATATATACACATACATATACACACACATAGATATACATATATA 683

Db       9427 TATATCCACATATATCTATATATATATAAATATCTATATATATATAAATATCTATATATA 9368

Qy        684 TACACATATATATACGTATATATATGTATATATAT 718

Db       9367 TATATAAATATCTATATATATATAAATATCTATAT 9333


RESULT 35
AAK73165/c
ID   AAK73165 standard; DNA; 20420 BP.
XX
AC   AAK73165;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:27977.
```

XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001354.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.

PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS Disclosure; SEQ ID NO 27977; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention
XX
SQ Sequence 20420 BP; 4834 A; 5403 C; 5630 G; 4553 T; 0 U; 0 Other;

Query Match 5.2%; Score 260.6; DB 4; Length 20420;
Best Local Similarity 71.7%; Pred. No. 8.9e-36;
Matches 355; Conservative 0; Mismatches 139; Indels 1; Gaps 1;

```
Qy       227 TTTATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
            |||  ||||||||||||||||||||    || | | |  |||||||||   ||||||||
Db      4732 TTTTCATATATATATATATATATTTTCATATATATATTTATATATATTTTCATATATA 4673

Qy       287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
            || |  |||||| |||| | || ||||||| |  |||||||| |  | ||||| || |
Db      4672 TATATTTATATATATATTTTCATATATATATATTTATATATATATTTTCATATATATATT 4613

Qy       347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
            ||||||||  |  |   |||| | || |||| | |||| | ||  | |||||||   |||||
Db      4612 TATATATATATTTTCATATATATATTTATATATATATTTTCATATATATATATTTATATA 4553

Qy       407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
            || |     ||||| | |||| || | ||||||   ||||||  | |||| || | |||||
Db      4552 TATATTTTCATATATATATATTTATATATATATTTTCATATATATATATTTATATATATA 4493

Qy       467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
            |    ||||| | |||| |||| |||||    ||||||||| |  |||||| | |  |
Db      4492 TTTTCATATATATATATTTATATATATATTTTCATATATATATATTTATATATATATTTT 4433

Qy       526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              ||| | ||||||  | ||||||||   |||||||| | || |||| ||||||   ||||
Db      4432 CATATATATATATTTATATATATATTTTCATATATATATATTTATATATATATTTTCATA 4373

Qy       586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
            ||||||  ||||||  | || |  | |||||||  |  ||||||  |||| | || | ||||
Db      4372 TATATATTTATATATATATTTTCATATATATATATTTATATATATATTTTCATATATATT 4313

Qy       646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
            | || ||| ||||   || | ||| || || ||||||||  | ||||||  | | ||||||
Db      4312 TTCATATATATATTTTCATATATATATTTATATATATATTTATATATATTTTCATATATA 4253

Qy       706 TATGTATATATATAT 720
            ||| ||||||||||||
Db      4252 TATTTATATATATAT 4238
```

RESULT 36
ABK69933/c
ID ABK69933 standard; DNA; 20420 BP.
XX
AC ABK69933;
XX
DT 15-JUL-2002 (first entry)
XX
DE Human secreted protein gene 68 genomic DNA fragment #21.
XX
KW Human; ds; secreted protein; gene therapy; immunosuppressive;
KW antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;
KW vasotropic; cerebroprotective; nootropic; neuroprotective; antibacterial;
KW virucide; fungicide; opthalmalogical; autoimmune disease; neoplasm;
KW rheumatoid arthritis; hyperproliferative disorder; cardiac arrest;
KW cardiovascular disorder; cerebrovascular disorder; cerebral ischaemia;
KW angiogenesis; nervous system disorder; Alzheimer's disease; infection;
KW ocular disorder; corneal infection; wound healing; skin aging;
KW epithelial cell proliferation; food additive.
XX
OS Homo sapiens.
XX
PN WO200226931-A2.
XX
PD 04-APR-2002.
XX
PF 24-SEP-2001; 2001WO-US029871.
XX
PR 25-SEP-2000; 2000US-0234925P.
PR 12-JAN-2001; 2001WO-US000911.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Komatsoulis G, Duan DR, Rosen CA, Moore PA, Shi Y;
PI Lafleur DW, Olsen H, Brewer LA, Florence KA, Young PE, Soppet DR;
PI Endress GA, Mucenski M, Ebner R;
XX
DR WPI; 2002-362489/39.
XX
PT Novel 71 isolated secreted polypeptides and polynucleotides encoding the
PT polypeptides, useful for treating Huntington's disease, sepsis,
PT meningitis, thrombocytopenia, hemolytic anemia, rheumatoid arthritis,
PT asthma.
XX
PS Example 2; Page 1440-1445; 1478pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (or its
CC fragment, homologue complement or allelic variant) encoding a human
CC secreted protein (and its fragment, domain, epitope, variant, secreted
CC form and species variant). Also included are a recombinant vector
CC comprising the nucleic acid, a recombinant host cell comprising the
CC vector, an antibody against the secreted protein, a recombinant host cell
CC that expresses the secreted protein and a method of identifying a binding
CC partner of the secreted protein. The nucleic acid and protein are used to
CC prevent, diagnose, treat or ameliorate a medical condition in e.g.
CC humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep for
CC example autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC angiogenesis, nervous system disorders e.g. Alzheimer's disease,
CC infections caused by bacteria, viruses and fungi and ocular disorders
CC e.g. corneal infection. Many other diseases and disorders are listed in
CC the specification. The polypeptides can also be used to aid wound healing
CC an epithelial cell proliferation, to prevent skin aging due to sunburn,
CC to maintain organs before transplantation, for supporting cell culture of
CC primary tissues, to regenerate tissues and in chemotaxis. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities. The present sequence
CC represents a ds DNA fragment of the gene for a novel human secreted
CC protein of the invention
XX

```
SQ   Sequence 20420 BP; 4834 A; 5403 C; 5630 G; 4553 T; 0 U; 0 Other;

  Query Match             5.2%;  Score 260.6;  DB 6;  Length 20420;
  Best Local Similarity   71.7%;  Pred. No. 8.9e-36;
  Matches  355;  Conservative    0;  Mismatches  139;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              |||  ||||||||||||||||||||    || | | | |  |||||||||    ||||||||
Db       4732 TTTTCATATATATATATATATATTTTCATATATATATTTATATATATTTTCATATATA 4673

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || |  |||||| |||| | || |||||||  |  ||||||||  |  | |||||  || |
Db       4672 TATATTTATATATATATTTTCATATATATATATTTATATATATATTTTCATATATATATT 4613

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              ||||||||  |  |   |||| | || |||| | |||| | || | |||||||   |||||
Db       4612 TATATATATATTTTCATATATATATTTATATATATATTTTCATATATATATATTTATATA 4553

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || |     ||||| | |||| || | ||||||   ||||||  | |||| || | |||||
Db       4552 TATATTTTCATATATATATATTTATATATATATTTTCATATATATATATTTATATATATA 4493

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              |    |||||  | |||| |||| |||||     |||||||||  |  ||||||  | |  |
Db       4492 TTTTCATATATATATATTTATATATATATTTTCATATATATATATTTATATATATATTTT 4433

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
               ||| | ||||||  | |||||||   |||||||| | || |||| ||||||   ||||
Db       4432 CATATATATATATTTATATATATATTTTCATATATATATATTTATATATATATTTTCATA 4373

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              ||||||  ||||||  | || |  | |||||||  |  |||||| |||| | || | ||||
Db       4372 TATATATTTATATATATATTTTCATATATATATATTTATATATATATTTTCATATATATT 4313

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              | || ||| ||||   || | ||| || || ||||||||  | ||||||  | | ||||||
Db       4312 TTCATATATATATTTTCATATATATATTTATATATATATTTATATATATTTTCATATATA 4253

Qy        706 TATGTATATATATAT 720
              ||| ||||||||||||
Db       4252 TATTTATATATATAT 4238


RESULT 37
ADC87554
ID   ADC87554 standard; DNA; 1489 BP.
XX
AC   ADC87554;
XX
DT   01-JAN-2004  (first entry)
XX
DE   Human GPCR gene SEQ ID NO:2007.
XX
KW   ds; gene; human; GPCR;
KW   guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1270724-A2.
XX
PD   02-JAN-2003.
XX
PF   18-JUN-2002; 2002EP-00013517.
XX
PR   18-JUN-2001; 2001JP-00246789.
XX
PA   (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA   (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI   Suwa M,  Asai K,  Akiyama Y,  Aburatani H;
XX
DR   WPI; 2003-315783/31.
DR   P-PSDB; ADC87555.
XX
PT   New polynucleotide, useful for preparing a composition for treating a
PT   patient in need of increased or suppressed activity or expression of the
PT   guanosine triphosphate-binding protein coupled receptor.
XX
PS   Claim 1; SEQ ID NO 2007; 28pp; English.
XX
CC   The invention relates to a novel polynucleotide encoding a guanosine
CC   triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC   the invention may have a use in gene therapy. The polynucleotide and
CC   polypeptide are useful for preparing a composition for treating a patient
CC   in need of increased or suppressed activity or expression of the
CC   guanosine triphosphate-binding protein coupled receptor. The
CC   polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC   invention.
XX
SQ   Sequence 1489 BP; 553 A; 208 C; 160 G; 568 T; 0 U; 0 Other;

  Query Match             5.2%;  Score 258.8;  DB 10;  Length 1489;
  Best Local Similarity   72.9%;  Pred. No. 1.6e-35;
  Matches  360;  Conservative    0;  Mismatches  132;  Indels    2;  Gaps    2;

Qy        229 TATATATATATATATATATATATACACACACACACACATATATATATACACATATATATA 288
              | |||||||||  ||| ||||||||  ||| | | ||||||||  ||||||  || ||   |||
Db        519 TGTATATATATGTATGTATATATATACATATATACACATATGTATATATACGTACC-ATA 577

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              |  |||||   ||||||  || ||||||  ||||  ||||||  ||  ||  ||||  ||| |||
Db        578 TATGTATATGTGTATATATACGTATATATACATGTATATATACGTATATATACACATATA 637

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              |||||  ||||  | ||||  ||  ||||||  ||  ||||  | |  | |||  |   | |||||
Db        638 CATATATACATTTATATATACGTATATATACGTATACATATATATACACGTATATATACA 697

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              |||  |||||  ||    ||||||  ||  ||||  |  |  |||  ||    ||||||  ||  ||||  |
Db        698 TACATATATACACGTGTATATATACGTATACATATATACACGTGTATATATACGTATACA 757

Qy        469 CATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATATA 527
              |||||||||  |  ||||||  |   ||  |||  |||||  |      ||||||||   |  |  |||
Db        758 TATATATACACGTGTATATATACGTATACATATATACACGTGTATATATACGTATACATA 817

Qy        528 TACACATATATACACATATATATACACATATATACACATATATACATATATACACACATATA 587
              ||  |||   |  ||  |  |||   ||||||  |||||  ||    ||||||  |   ||||||  |||||
Db        818 TATACACGTGTATATATACGTATACATATATACACGTGTATATATACGTATACATATATA 877

Qy        588 TATACATATATACACATATATACATATATACACATATATACATATATACACACATATATA 647
              |    ||||||||  |  |||||  |  |   |   |  |||  |||   |||||||||   ||||||
Db        878 CACGTGTATATACATACATATATACACGTGTATATACATATGTATATACACGTGTATATA 937

Qy        648 CACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATA 707
              |  |||  ||||||  |  ||  ||  ||||||||  |||||||||  ||||||||||  ||||||  .
Db        938 TGTATATGTATACATATACGTATATATACATGTATATACACGTATATATACGTGTATATA 997

Qy        708 TGTATATATATATG 721
              ||||||  |||||  |
Db        998 TGTATACATATACG 1011


RESULT 38
ADO15781/c
ID   ADO15781 standard; DNA; 2761 BP.
XX
AC   ADO15781;
XX
DT   29-JUL-2004  (first entry)
XX
DE   4 synthesis-period of neuroblastoma related DNA, SEQ ID 43.
XX
KW   Human; 4 synthesis-period; neuroblastoma; stage 4S; ds.
XX
OS   Homo sapiens.
XX
```

PN WO2004039975-A1.
XX
PD 13-MAY-2004.
XX
PF 30-OCT-2003; 2003WO-JP013932.
XX
PR 30-OCT-2002; 2002JP-00316586.
XX
PA (HISM) HISAMITSU PHARM CO LTD.
PA (CHIB-) CHIBA PREFECTURE.
XX
PI Nakagawara A, Ohira M;
XX
DR WPI; 2004-390323/36.
XX
PT Novel nucleic acid obtained from 4 synthesis-period of neuroblastoma
PT cells useful for prognosing and determining progress stage of
PT neuroblastomas.
XX
PS Claim 1; SEQ ID NO 43; 455pp; Japanese.
XX
CC The present invention relates to human nucleic acid sequences (I;
CC ADO15739-ADO15912) obtained from 4 synthesis-period (stage 4S) of
CC neuroblastoma cell. (I) is useful for prognosing and determining the
CC progress stage of 4 synthesis-period of neuroblastoma.
XX
SQ Sequence 2761 BP; 847 A; 475 C; 598 G; 841 T; 0 U; 0 Other;

Query Match 5.2%; Score 258.8; DB 12; Length 2761;
Best Local Similarity 71.0%; Pred. No. 1.6e-35;
Matches 375; Conservative 0; Mismatches 142; Indels 11; Gaps 2;

```
Qy      229 TATATATATATATATATATATATACACACACACACACATATATATATACACATATATATA 288
            | | ||||||||||| ||||||||  | | |     |||||||| |||| ||||||||
Db     2444 TGTGTATATATATATACATATATACGTATATATGTGTATATATATACACACGTATATATA 2385

Qy      289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
            |||||||||  ||| |||||||  |||| | | | |||||   |||||||   || ||||
Db     2384 CACATATATGTATACATACACACGTATATATACACATATATGTATACATACACACGCATA 2325

Qy      349 TATATACACATATGTATACA--------CATATATACACATATATACACACATATATACA 400
            |||| ||| ||||||||||         ||||||||||||||||  | ||||| | ||
Db     2324 TATACACATATATGTATACATACACACGTATATATACACATATATGTATACATACACACG 2265

Qy      401 CATATATACACACATATATACACATATATACACATATATACACATATACACATATATACA 460
            |||||||||| |||| || ||||| ||||  ||||||||||||||    ||| |||||
Db     2264 TATATATACACATATATGTATACATACATACGTATATATACACATATATGTATACATACA 2205

Qy      461 CATATATACATATATACACATATATATAATATACACACATATATATACACATATATACAC 520
              | |||| ||| | | | || |||   | ||||| || ||||||||||||||||  |
Db     2204 TACGTATATATACACATATATGTAT---ACATACATACGTATATATACACATATATGTAT 2148

Qy      521 ACATATATACACATATATACACATATATATACACATATATACACATATATACATATATAC 580
            ||||| ||||  ||||||||||||||| || ||||| ||||  ||||||||| |||||
Db     2147 ACATACATACGTATATATACACATATATGTATACATACATACGTATATATACACATATAT 2088

Qy      581 ACATATATATACATATATACACATATATACATATATACACATATATACATATATACACAC 640
              ||| |||  |    |||| | | | ||| || |||||| | ||    ||||||||||
Db     2087 GTATACATACATGCGTATATATACACATATATGTATACATACATGCGTATATATACACAT 2028

Qy      641 ATATATACACATACATATACACACACATAGATATACATATATATACACATATATATACGT 700
            |||| || ||||| ||||  ||||| ||| || |   | |||| | | |||||      |
Db     2027 ATATGTATACATATATATGAACACATATATATGTGTGTGTATACATATATATAAGATTTT 1968

Qy      701 ATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTT 748
              |    |  ||    ||| |   |||| ||  || | ||| | ||
Db     1967 TCTTGACCATGAATCACCATGTTTGGGAGTACAGTAGGGTTAGGAATT 1920
```

RESULT 39
ADL13771
ID ADL13771 standard; DNA; 201766 BP.
XX
AC ADL13771;
XX
DT 06-MAY-2004 (first entry)
XX
DE Osteoarthritis-associated polymorphic nucleotide #303.
XX
KW ds; gene; osteopathic; antiinflammatory; antiarthritic; gene therapy;
KW joint space narrowing; osteophyte development; joint pain;
KW osteoarthritis; SNP; single nucleotide polymorphism.
XX
OS Homo sapiens.
XX
PN WO2003054166-A2.
XX
PD 03-JUL-2003.
XX
PF 19-DEC-2002; 2002WO-US041225.
XX
PR 20-DEC-2001; 2001US-0342603P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Jones KA, Schafer A;
XX
DR WPI; 2003-559141/52.
XX
PT Determining susceptibility of an individual to joint space narrowing,
PT osteophyte development and/or joint pain comprises identifying whether
PT the individual has at least one polymorphism in a polynucleotide encoding
PT a protein.
XX
PS Disclosure; SEQ ID NO 303; 297pp; English.
XX
CC The invention relates to a method of determining susceptibility of an
CC individual to joint space narrowing and/or osteophyte development and/or
CC joint pain comprising identifying whether the individual has at least one
CC polymorphism in a polynucleotide encoding at least one of the protein
CC listed in the specification. The methods, composition and agent are
CC useful for modulating the susceptibility of an individual to joint space
CC narrowing and/or osteophyte development and/or joint pain that is
CC associated with a disease, preferably osteoarthritis. The cell line and
CC the non-human animal are useful for screening for an agent for diagnosing
CC an individual having susceptibility to joint space narrowing and/or
CC osteophyte development and/or joint pain. This sequence corresponds to
CC the polynucleotide encoding a protein listed in the specification. (Note:
CC The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences).
XX
SQ Sequence 201766 BP; 57593 A; 35575 C; 36821 G; 54688 T; 0 U; 17089 Other;

Query Match 5.2%; Score 258.8; DB 10; Length 201766;
Best Local Similarity 72.9%; Pred. No. 2.1e-35;
Matches 360; Conservative 0; Mismatches 132; Indels 2; Gaps 2;

```
Qy      229 TATATATATATATATATATATATACACACACACACACATATATATATACACATATATATA 288
            | |||||||||| ||| |||||||| ||| | | |||||||| |||||| || ||   |||
Db   188638 TGTATATATATGTATGTATATATATACATATATACACATATGTATATATACGTACC-ATA 188696

Qy      289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
            |  |||||   |||||| || |||||| |||| |||||| || || ||||  ||| |||
Db   188697 TATGTATATGTGTATATATACGTATATATACATGTATATATACGTATATATACACATATA 188756

Qy      349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
            ||||| |||| | |||| || |||||| || |||| | | | ||| |   | ||||| |
Db   188757 CATATATACATTTATATATACGTATATATACGTATACATATATATACACGTATATATACA 188816

Qy      409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
            ||| ||||| ||   |||||| || |||| | | ||| ||   |||||| || |||| |
Db   188817 TACATATATACACGTGTATATATACGTATACATATATACACGTGTATATATACGTATACA 188876

Qy      469 CATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATATA 527
            |||||||||| | |||||| |  || ||| ||||| |     ||||||||  | | |||
```

```
Db     188877 TATATATACACGTGTATATATACGTATACATATATACACGTGTATATATACGTATACATA 188936

Qy        528 TACACATATATACACATATATATACACATATATACACATATATACATATATACACATATA 587
              || |||  | || | |||  |||||| ||||| ||   ||||||  |  |||||| |||||
Db     188937 TATACACGTGTATATATACGTATACATATATACACGTGTATATATACGTATACATATATA 188996

Qy        588 TATACATATATACACATATATACATATATACACATATATACATATATACACACATATATA 647
               |    |||||||| | ||||| | |   |  | ||| |||  |||||||||   ||||||
Db     188997 CACGTGTATATACATACATATATACACGTGTATATACATATGTATATACACGTGTATATA 189056

Qy        648 CACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATA 707
                 |||  |||||| | || || |||||||| ||||||||| |||||||||| ||||||
Db     189057 TGTATATGTATACATATACGTATATATACATGTATATACACGTATATATACGTGTATATA 189116

Qy        708 TGTATATATATATG 721
              |||||| ||||| |
Db     189117 TGTATACATATACG 189130


RESULT 40
ABD32659/c
ID   ABD32659 standard; DNA; 163382 BP.
XX
AC   ABD32659;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Human cancer-associated genomic DNA HD13-065.
XX
KW   Human; ds; cancer-associated protein; gene; cytostatic; cancer;
KW   leukaemia; lymphoma; CAP.
XX
OS   Homo sapiens.
XX
PN   WO2004074320-A2.
XX
PD   02-SEP-2004.
XX
PF   17-FEB-2004; 2004WO-US004730.
XX
PR   14-FEB-2003; 2003US-00367094.
PR   14-MAR-2003; 2003US-00388838.
PR   15-APR-2003; 2003US-00417375.
PR   13-JUN-2003; 2003US-00461862.
PR   15-SEP-2003; 2003US-00663431.
PR   15-DEC-2003; 2003US-00737318.
XX
PA   (SAGR-) SAGRES DISCOVERY INC.
XX
PI   Morris DW,  Morris DW,  Malandro MS;
XX
DR   WPI; 2004-652914/63.
XX
PT   New isolated cancer-associated polynucleotides and polypeptides useful
PT   for diagnosing, preventing or treating cancers, especially lymphoma and
PT   leukemia, or in screening for agents that modulate cancer.
XX
PS   claim 16; seqid 199; 310pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising at least 10
CC   contiguous nucleotides of any of the 233 polynucleotide sequences given
CC   in the specification, or its complement. The nucleic acids encode cancer-
CC   associated proteins. Also included are an expression vector comprising
CC   the isolated nucleic acid cited above, a host cell comprising the above
CC   recombinant nucleic acid or expression vector, a microarray for detecting
CC   a cancer-associated (CA) nucleic acid comprising at least one probe
CC   comprising at least 10 contiguous nucleotides of any of the above-
CC   mentioned nucleotide sequences, an isolated polypeptide (encoded within
CC   an open reading frame of a CA sequence selected from any of the 95
CC   polynucleotide sequences as mentioned in the specification, or its
CC   complement), an isolated antibody, (or its antigen binding fragment) that
CC   binds to the above polypeptide, a hybridoma that produces the above
CC   monoclonal antibody, a pharmaceutical composition comprising the above
CC   antibody and a pharmaceutical excipient, a kit for detecting cancer
CC   cells(comprising the antibody cited above, methods for diagnosing cancer
CC   or for detecting the presence or absence of cancer cells in an
CC   individual, a method for inhibiting growth of cancer cells in an
CC   individual, a method for delivering a therapeutic agent to cancer cells
CC   in an individual, an electronic library comprising the above
CC   polynucleotide or polypeptide (or their fragments), methods of screening
CC   for anticancer activity or for a bioactive agent capable of modulating
CC   the activity of a CA protein (CAP), methods for detecting cancer
CC   associated with expression of a polypeptide in a test cell sample, a
CC   method for treating cancers and a method for inhibiting the expression of
CC   CA gene in a cell. The composition and methods are useful for detecting,
CC   diagnosing, preventing and treating cancers, especially lymphoma and
CC   leukaemia. These may also be used in screening for agents that modulate
CC   cancer. The present sequence is a human CAP genomic sequence. Note: The
CC   sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 163382 BP; 50073 A; 25805 C; 27837 G; 53184 T; 0 U; 6483 Other;

  Query Match             5.2%;  Score 258.6;  DB 13;  Length 163382;
  Best Local Similarity   70.0%;  Pred. No. 2.2e-35;
  Matches  362;  Conservative    0;  Mismatches  154;  Indels    1;  Gaps    1;

Qy        204 ATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATATACACACACACAC 263
              || | |  | | | |||         |||| ||||||| |||| |||| | ||| | | |
Db     145185 ATATAAATATATATAAATAAATAAATATAAATATATAAATATGTATAAATACATATATAA 145126

Qy        264 ACATATATATATACACATATATATACACATATATACATATATACACATATATACACATAT 323
              | | ||||| ||| | | |||||||| | ||| ||| ||| ||| | ||| ||| | ||||
Db     145125 ATAAATATAAATATATAAATATATATAAATACATATATAAATATATATAAATATATATAT 145066

Qy        324 ATATACACATACATATGTACACATATATATACACATATGTATACACATATATACACATAT 383
               ||| ||||| |||||||| | ||||||| ||| ||||  ||| | | ||||| | | ||
Db     145065 TAATATACATAAATATGTATAAATATATAAACATATATTAATATATAAATATATATAAAT 145006

Qy        384 ATACACACATATATACACATATATACACACATATATACACATATATACACATATATACAC 443
              |||   | |||||||   ||||||| | | ||  ||| | ||| |||     ||||| ||
Db     145005 ATATTAATATATATATTTATATATAAATATATGAATATATATAAATATGTTAATATATAC 144946

Qy        444 ATATACACATATATACACATATATACATATATACACATATATATAATATACACACATATA 503
              ||||| | | ||| | | ||| | | ||||||| | |||||||    ||| | | |||||
Db     144945 ATATATATAAATAAATATATAAAAATATATATAAATATATATAACTAATATATAAATATA 144886

Qy        504 TATACACATATATACACACATATATACACATATATACACATATATATACACATATATACA 563
              |||| ||  ||||| | | ||||||| | ||||||  ||  |||||||  | |||||||  |
Db     144885 TATA-ACTAATATATATAAATATATAAATATATATGAACTAATATATATAAATATATAAA 144827

Qy        564 CATATATACATATATACACATATATATACATATATACACATATATACATATATACACATA 623
              | ||||| | ||| | | ||| ||| | ||||| | | |||||||  ||| ||| | | |
Db     144826 TAAATATATAAATAAATATATAAATAAATATATACAAATATATATAAATAAATATATAAA 144767

Qy        624 TATACATATATACACACATATATACACATACATATACACACACATAGATATACATATATA 683
              ||| ||||| | | | ||||| | | ||| ||||| | | | ||| ||| | | |||||
Db     144766 AATAAATATAAATAAATATATAAATAAATATATATAAATAAATATATATAAATAAATATA 144707

Qy        684 TACACATATATATACGTATATATATGTATATATATAT 720
              || | ||| |||||  || |||||| |||| ||| ||
Db     144706 TATAAATAAATATATATAAATATATATATAAATAAAT 144670


RESULT 41
ABV72623
ID   ABV72623 standard; DNA; 82938 BP.
XX
AC   ABV72623;
XX
DT   27-NOV-2002  (first entry)
XX
DE   Human transporter protein encoding gene.
XX
KW   Human; transporter protein; hTP; gamma-aminobutyric-acid; GABA;
```

```
KW   gene therapy; protein therapy; vaccine; gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1120. .81932
FT                   /*tag= a
FT                   /product= "Transporter protein"
FT                   /note= "Contains 8 introns"
FT   exon            1120. .1223
FT                   /*tag= b
FT                   /number= 1
FT   intron          1224. .27733
FT                   /*tag= c
FT                   /number= 1
FT   variation       replace(2419,C)
FT                   /*tag= d
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(3191,A)
FT                   /*tag= e
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(4195,T)
FT                   /*tag= f
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(5910,C)
FT                   /*tag= g
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(7119,T)
FT                   /*tag= h
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(7529,G)
FT                   /*tag= h
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(8855,C)
FT                   /*tag= i
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(8956,A)
FT                   /*tag= j
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(10220,G)
FT                   /*tag= k
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(11628,C)
FT                   /*tag= l
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(11670,A)
FT                   /*tag= m
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(11671,G)
FT                   /*tag= n
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(11931,C)
FT                   /*tag= o
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(15984,G)
FT                   /*tag= p
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(17735,A)
FT                   /*tag= q
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(17910,A)
FT                   /*tag= r
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(18130,C)
FT                   /*tag= s
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19293. .19294,TWA)
FT                   /*tag= t
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19294. .19295,ATT)
FT                   /*tag= u
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19296. .19297,CYT)
FT                   /*tag= v
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19303. .19304,TYC)
FT                   /*tag= w
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19304. .19305,CYC)
FT                   /*tag= x
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19307. .19308,TMC)
FT                   /*tag= y
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19308. .19309,CCC)
FT                   /*tag= z
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19313. .19314,TMA)
FT                   /*tag= aa
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19314. .19315,AWT)
FT                   /*tag= ab
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19315. .19316,TCC)
FT                   /*tag= ac
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19316. .19317,CYT)
FT                   /*tag= ad
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19317. .19318,TWA)
FT                   /*tag= ae
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19318. .19319,ATT)
FT                   /*tag= af
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19860,G)
FT                   /*tag= ag
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(20843,G)
FT                   /*tag= ah
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(21144,T)
FT                   /*tag= ai
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(21371,G)
FT                   /*tag= aj
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(22061. .22062,A)
FT                   /*tag= ak
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(22321,T)
FT                   /*tag= al
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(22470. .22471,A)
FT                   /*tag= am
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(22472. .22473,A)
FT                   /*tag= an
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(24990,C)
FT                   /*tag= ao
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(25284,A)
FT                   /*tag= ap
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(25934,G)
FT                   /*tag= aq
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(27509. .27510,TAT)
FT                   /*tag= ar
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(27513. .27514,TTT)
FT                   /*tag= as
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            27734. .27882
FT                   /*tag= at
```

```
FT                   /number= 2
FT   intron          27883. .41045
FT                   /*tag=  au
FT                   /number= 2
FT   variation       replace(28608,G)
FT                   /*tag=  av
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(28627,C)
FT                   /*tag=  aw
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(29709,T)
FT                   /*tag=  ax
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(29788,T)
FT                   /*tag=  ay
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(29878,T)
FT                   /*tag=  az
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(30294,T)
FT                   /*tag=  ba
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(30896. .30897,T)
FT                   /*tag=  bb
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(31815,G)
FT                   /*tag=  bc
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(31999,TTT)
FT                   /*tag=  bd
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(32026,T)
FT                   /*tag=  be
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(32146,A)
FT                   /*tag=  bf
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(32184,S)
FT                   /*tag=  bg
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(32214. .32215,T)
FT                   /*tag=  bh
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(32215,C)
FT                   /*tag=  bi

  Query Match             5.2%;  Score 258.4;  DB 6;  Length 82938;
  Best Local Similarity   70.5%;  Pred. No. 2.3e-35;
  Matches  361;  Conservative    0;  Mismatches  146;  Indels    5;  Gaps    1;

Qy        214 AAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACACATATATAT 273
              | | | | |    | |||| | | | |||| |||| |||||||||   | |||||| |||
Db       6946 ATACATATGTATATATATACACACACATATGTATACATACACACATGTATACATATGTAT 7005

Qy        274 ATACACATATATATACACATATATACATATATACACATATATACACATATATATACACAT 333
              |||   ||||| |||||||| ||||||||| || | || |||| | ||| | ||  | |
Db       7006 ATATGTATATAGATACACATGTATACATATGTATATATGTATATAGATACACATGTATAC 7065

Qy        334 ACATATGTACACATATATATACACATATGTATACACATATATACACATATATACACACAT 393
              | || | || |  |||||| | ||| ||||||||| || |||| |  |||||| | |||
Db       7066 ATATGTATATATGTATATAGATACACATGTATACATATGTATATATGTATATAGATACAC 7125

Qy        394 ATATACACATATATACACACATATATACACATATATACACATATATACACATATACACAT 453
              || || |||||| || | |  |||||| | | | ||  | | ||||  | ||||  | ||
Db       7126 ATGTATACATATGTATATATGTATATAGATACACATGTATACATATGTATATATGTATAT 7185

Qy        454 ATATACACATATATACATATATACACATATATATA-----ATATACACACATATATATAC 508
              | |||||||| ||||||||| || | || ||||||      | ||   | |||||| ||||
Db       7186 AGATACACATGTATACATATGTATATATGTATATAGATACACATGTATACATATGTATAT 7245

Qy        509 ACATATATACACACATATATACACATATATACACATATATATACACATATATACACATAT 568
              |  |||||| | ||| || || |||||| || | || ||||||||  ||| || | || |
Db       7246 ATGTATATAGATACACATGTATACATATGTATATATGTATATACATGTATGTATATATGT 7305

Qy        569 ATACATATATACACATATATATACATATATACACATATATACATATATACACATATATAC 628
              |||  ||||||||  ||| |||| | ||||  ||||||  |||| || | | || ||
Db       7306 ATATGTATATACATGTATGTATATGTGTATATGTATATATGTATATGTATATACATGTAT 7365

Qy        629 ATATATACACACATATATACACATACATATACACACACATAGATATACATATATATACAC 688
              ||||||  |||  ||||||||  |  ||||| | |||  || |||  |||||  ||| |
Db       7366 GTATATATACAGCTATATACATGTGTATATATATACAGCTATATACGTATATACATATAT 7425

Qy        689 ATATATATACGTATATATATGTATATATATAT 720
              || |||||| ||||| |||||| |||||||||
Db       7426 ATGTATATATGTATACATATGTGTATATATAT 7457


RESULT 42
AAK73166/c
ID   AAK73166 standard; DNA; 19965 BP.
XX
AC   AAK73166;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:27978.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001354.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
```

PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS Disclosure; SEQ ID NO 27978; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention
XX
SQ Sequence 19965 BP; 4699 A; 5296 C; 5505 G; 4465 T; 0 U; 0 Other;

Query Match 5.2%; Score 258; DB 4; Length 19965;
Best Local Similarity 64.7%; Pred. No. 2.5e-35;
Matches 431; Conservative 0; Mismatches 230; Indels 5; Gaps 3;

Qy 59 TAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGA 118
Db 4999 TGGCCAGGAGCGGTGGCTCACGCCTATAATCCCAGCACGTTGGGAGGCCGAGGCAGGGAG 4940

Qy 119 ATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCA 178
Db 4939 GATCGCTTGAGAT-CAAGAGCTCCAGACCAGCCTAGGCAACACGGTGAAATTCCATCTCT 4881

Qy 179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATA 238
Db 4880 ACAAAAAAAATACAAAAAAAAAAAAAAAATATATATATATATATTTATATATATATGTTTA 4821

```
Qy        239 TATATATATATATACACACACACACATATATATATACACATATATATACACATATATA 298
              ||||||||||| || | |      | | |||| |||||    |||||||| | ||||||
Db       4820 TATATATATATTTATATATTTTTATATATATTTATATTTTATATATATTTATATATATT 4761

Qy        299 CATATATA-CACATATATACACATATATATACACATACATATGTACACATATATATACAC 357
              ||||||| |  |  ||| | |||| |||| | ||  |    || | |||||| | ||
Db       4760 TATATATATTATTTTCATATATATATTTATATATATATTTTCATATATATATATTTTCAT 4701

Qy        358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417
              |||| |||| | ||||  | | ||||||  | | |||| | || |||||||   | ||||
Db       4700 ATATATATATATATATTCATATATATATTTATATATATTTTCATATATATATTTATATAT 4641

Qy        418 ATACACATATATACACATATATACACATATACACATATATACACATATATACATATATAC 477
              |||     ||||| | |||| | || ||||| | |||||||  |||||||| || ||||
Db       4640 ATATTTTCATATATATATATTTTCATATATATATATATATTCATATATATATTTATAT 4581

Qy        478 ACATATATATA---ATATACACACATATATATACACATATATACACACATATATACACAT 534
              | || |  |||   |||||   | ||||||||  ||||||||  | |  |||||| | ||
Db       4580 ATATTTTCATATATATATATTTATATATATATTTTCATATATATATATTTATATATATAT 4521

Qy        535 ATATACACATATATATACACATATATACACATATATACATATATACACATATATATACAT 594
              |  | | |||||| || | ||||||  ||||||||  |||| || | |||||| | |||
Db       4520 TTTCATATATATATTTATATATATATTTTCATATATATATATTTATATATATATTTTCAT 4461

Qy        595 ATATACACATATATACATATATACACATATATACATATATACACACATATATACACATAC 654
              ||||| | || |||| ||||||   |||||||| |||| || | | |||| | || |||
Db       4460 ATATATATATTTATATATATATTTTCATATATATATATTTATATATATATTTTCATATAT 4401

Qy        655 ATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATAT 714
              |||||   | | ||| || | ||||||||||   |||||||| |  |||||| ||| |
Db       4400 ATATATTTATATATATATTTTCATATATATATTTATATATATATTTTCATATATATATTT 4341

Qy        715 ATATAT 720
              ||||||
Db       4340 ATATAT 4335
```

```
RESULT 43
ABK69932/c
ID   ABK69932 standard; DNA; 19965 BP.
XX
AC   ABK69932;
XX
DT   15-JUL-2002  (first entry)
XX
DE   Human secreted protein gene 68 genomic DNA fragment #20.
XX
KW   Human; ds; secreted protein; gene therapy; immunosuppressive;
KW   antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;
KW   vasotropic; cerebroprotective; nootropic; neuroprotective; antibacterial;
KW   virucide; fungicide; opthalmalogical; autoimmune disease; neoplasm;
KW   rheumatoid arthritis; hyperproliferative disorder; cardiac arrest;
KW   cardiovascular disorder; cerebrovascular disorder; cerebral ischaemia;
KW   angiogenesis; nervous system disorder; Alzheimer's disease; infection;
KW   ocular disorder; corneal infection; wound healing; skin aging;
KW   epithelial cell proliferation; food additive.
XX
OS   Homo sapiens.
XX
PN   WO200226931-A2.
XX
PD   04-APR-2002.
XX
PF   24-SEP-2001; 2001WO-US029871.
XX
PR   25-SEP-2000; 2000US-0234925P.
PR   12-JAN-2001; 2001WO-US000911.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Komatsoulis G,  Duan DR,  Rosen CA,  Moore PA,  Shi Y;
PI   Lafleur DW,  Olsen H,  Brewer LA,  Florence KA,  Young PE,  Soppet DR;
PI   Endress GA,  Mucenski M,  Ebner R;
XX
DR   WPI; 2002-362489/39.
XX
PT   Novel 71 isolated secreted polypeptides and polynucleotides encoding the
PT   polypeptides, useful for treating Huntington's disease, sepsis,
PT   meningitis, thrombocytopenia, hemolytic anemia, rheumatoid arthritis,
PT   asthma.
XX
PS   Example 2; Page 1434-1440; 1478pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule (or its
CC   fragment, homologue complement or allelic variant) encoding a human
CC   secreted protein (and its fragment, domain, epitope, variant, secreted
CC   form and species variant). Also included are a recombinant vector
CC   comprising the nucleic acid, a recombinant host cell comprising the
CC   vector, an antibody against the secreted protein, a recombinant host cell
CC   that expresses the secreted protein and a method of identifying a binding
CC   partner of the secreted protein. The nucleic acid and protein are used to
CC   prevent, diagnose, treat or ameliorate a medical condition in e.g.
CC   humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep for
CC   example autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
CC   disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
CC   e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC   angiogenesis, nervous system disorders e.g. Alzheimer's disease,
CC   infections caused by bacteria, viruses and fungi and ocular disorders
CC   e.g. corneal infection. Many other diseases and disorders are listed in
CC   the specification. The polypeptides can also be used to aid wound healing
CC   an epithelial cell proliferation, to prevent skin aging due to sunburn,
CC   to maintain organs before transplantation, for supporting cell culture of
CC   primary tissues, to regenerate tissues and in chemotaxis. The
CC   polypeptides can also be used as a food additive or preservative to
CC   increase or decrease storage capabilities. The present sequence
CC   represents a ds DNA fragment of the gene for a novel human secreted
CC   protein of the invention
XX
SQ   Sequence 19965 BP; 4699 A; 5296 C; 5505 G; 4465 T; 0 U; 0 Other;

  Query Match             5.2%;  Score 258;  DB 6;  Length 19965;
  Best Local Similarity   64.7%;  Pred. No. 2.5e-35;
  Matches  431;  Conservative    0;  Mismatches  230;  Indels    5;  Gaps    3;
```

```
Qy         59 TAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGA 118
              | | ||||   ||||   ||||||| ||||||||||   || ||||||||||||||||
Db       4999 TGGCCAGGAGCGGTGGCTCACGCCTATAATCCCAGCACGTTGGGAGGCCGAGGCAGGGAG 4940

Qy        119 ATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCA 178
                |  | | |  | | |  ||           |||| |   ||  ||| ||   |
Db       4939 GATCGCTTGAGAT-CAAGAGCTCCAGACCAGCCTAGGCAACACGGTGAAATTCCATCTCT 4881

Qy        179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATA 238
              | |   || |  |   |||    | |    | | | | |      | ||||| | ||
Db       4880 ACAAAAAAAATACAAAAAAAAAAAAAAAATATATATATATATTTATATATATATGTTTA 4821

Qy        239 TATATATATATATACACACACACACATATATATATACACATATATATACACATATATA 298
              |||||||||| || | |      | | |||| |||||    |||||||| | ||||||
Db       4820 TATATATATATTTATATATTTTTATATATATTTATATTTTATATATATTTATATATATT 4761

Qy        299 CATATATA-CACATATATACACATATATATACACATACATATGTACACATATATATACAC 357
              ||||||| |  |  ||| | |||| |||| | ||  |    || | |||||| | ||
Db       4760 TATATATATTATTTTCATATATATATTTATATATATATTTTCATATATATATATTTTCAT 4701

Qy        358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417
              |||| |||| | ||||  | | ||||||  | | |||| | || |||||||   | ||||
Db       4700 ATATATATATATATATTCATATATATATTTATATATATTTTCATATATATATTTATATAT 4641

Qy        418 ATACACATATATACACATATATACACATATACACATATATACACATATATACATATATAC 477
              |||     ||||| | |||| | || ||||| | |||||||  |||||||| || ||||
Db       4640 ATATTTTCATATATATATATTTTCATATATATATATATATTCATATATATATTTATAT 4581

Qy        478 ACATATATATA---ATATACACACATATATATACACATATATACACACATATATACACAT 534
              | || |  |||   |||||   | ||||||||  ||||||||  | |  |||||| | ||
Db       4580 ATATTTTCATATATATATATTTATATATATATTTTCATATATATATATTTATATATATAT 4521
```

```
Qy        535 ATATACACATATATATACACATATATACACATATATACATATATACACATATATATACAT 594
              |  | | |||||| || | ||||||   |||||||| |||| || | |||||| | |||
Db       4520 TTTCATATATATATTTATATATATATTTTCATATATATATTTATATATATATTTTCAT 4461

Qy        595 ATATACACATATATACATATATACACATATATACATATATACACACATATATACACATAC 654
              ||||| | || |||| ||||||   |||||||| |||| || | | |||| | || |||
Db       4460 ATATATATATTTATATATATATTTTCATATATATATATTTATATATATATTTTCATATAT 4401

Qy        655 ATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATAT 714
              |||||   | | ||| || | ||||||||||   |||||||| |  |||||| ||| |
Db       4400 ATATATTTATATATATATTTTCATATATATATTTATATATATATTTTCATATATATATTT 4341

Qy        715 ATATAT 720
              ||||||
Db       4340 ATATAT 4335
```

RESULT 44
ABK39930
ID ABK39930 standard; DNA; 12069 BP.
XX
AC ABK39930;
XX
DT 21-MAY-2002 (first entry)
XX
DE Human chemically pretreated gene sequence #6 strand 1.
XX
KW Human; ds; bisulphite treatment; CpG; DNA methylation; cancer; tumour;
KW cytostatic; ALDH6; CYP11A; CYP11B1; CYP3A3; DPYD; EPHX2; OCLN; TXNRD1;
KW UGT8; MRP; pharmacogenomics; SNP; single nucleotide polymorphism.
XX
OS Homo sapiens.
XX
PN WO200202806-A2.
XX
PD 10-JAN-2002.
XX
PF 29-JUN-2001; 2001WO-EP007470.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-154757/20.
XX
PT New nucleic acid, oligonucleotides and peptide nucleic acid-oligomers,
PT useful for detecting cytosine methylation state of genes associated with
PT pharmacogenomics and for therapy of diseases e.g. cancer.
XX
PS Claim 1; SEQ ID NO 11; 24pp; English.
XX
CC The invention relates to a nucleic acid comprising a sequence at least 18
CC bases in length of a segment of the chemically pretreated DNA of genes
CC associated with pharmacogenomics according to one of the sequences of the
CC genes ALDH6 (NM_000693), CYP11A (NM_000781), CYP11B1 (NM_000497), CYP3A3
CC (NM_000776 and NM_017460), DPYD (NM_000110), EPHX2 (NM_001979), OCLN
CC (NM_002538), TXNRD1 (NM_003330), UGT8 (NM_003360), MRP (NM_004996,
CC NM_019900, NM_019901, NM_019902, NM_019862, NM_019898, NM_019899) and
CC their complementary sequences, or a sequence (S1) chosen from 87
CC sequences and their complements. The chemical pretreatment is bisulphite
CC treatment to convert cytosines (but not methyl-cytosines) into uracils.
CC Also included are an oligomer (II) in particular an oligonucleotide or a
CC peptide nucleic acid (PNA)-oligomer, comprising in each case at least one
CC base sequence having a length of 9 nucleotides which hybridises to or is
CC identical to a chemically pretreated DNA of genes associated with
CC pharmacogenomics and their complements, arranged in an array for
CC analysing diseases associated with the methylation state (CpG) and/or
CC detecting SNPs (single nucleotide polymorphisms) of the 87 sequences. The
CC oligomers may also be used as PCR primers. The set of 87 nucleic acids
CC and their complements is useful for diagnosis and therapy of solid
CC tumours and cancer. The present sequence represents one the 87 DNA
CC sequences or its complement. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 12069 BP; 3210 A; 185 C; 2849 G; 5823 T; 0 U; 2 Other;

Query Match 5.2%; Score 257.6; DB 6; Length 12069;
Best Local Similarity 62.9%; Pred. No. 2.8e-35;
Matches 433; Conservative 0; Mismatches 249; Indels 6; Gaps 2;

```
Qy        226 TTTTATATATATATATATATATATATACACACACACACACATATATATATACACATATAT 285
              || |||||||||| |||| |||||||| | | | | ||     |||||||||   ||||||
Db       9239 TTATATATATATAGATATTTATATATATATAGATATACGTGGATATATATATTTATATAT 9298

Qy        286 ATACACATATATACATATATACACATATATACACATATATATACACATACATATGTACAC 345
              ||| | |||| || ||||||| | |||||     |||||  |||   ||| |||| || |
Db       9299 ATATAGATATTTATATATATATAGATATACGTGGATATAGATATTTATATATATATAGAT 9358

Qy        346 ATATATATACACATATGTATACACATATATACACATATATACACACATATATACACATAT 405
              || ||||||  | ||||  |||   |||||||  | | ||||  | | |||||||  | ||||
Db       9359 ATTTATATATATATATAGATATTTATATATATATAGATATTTATATATATATAGATATAT 9418

Qy        406 ATACACACATATATACACATATATACACATATATACACATATACACATATATACACATAT 465
              | | | ||||||  | |||||||  | |||| || | |||||  | |||| || | ||||
Db       9419 GTGGATATATATATTTATATATATAGAGATATTTATATATATATAGATATTTATATATAT 9478

Qy        466 ATACATATATACACATATATATA-----ATATACACACATATATATACACATATATACAC 520
              ||| |||||      |||||||||     |||||  | | |||| |||| | |||||  | |
Db       9479 ATAGATATACGTGGATATATATATTTATATATATAGATATTTATATATATATAGATAT 9538

Qy        521 ACATATATACACATATATACACATATATATACACATATATACACATATATACATATATAC 580
              ||  |  ||| | |||| || || || ||||  | |||||||  | |||||      |||||
Db       9539 ACGTGGATATATATATTTATACGTAGATATTTATATATATATAGATATACGTGGATATAT 9598

Qy        581 ACATATATATACATATATACACATATATACATATATACACATATATACATATATACACAC 640
              | || ||||   | ||||  | |||||| | ||||||    ||||| |||| || | |
Db       9599 ATATTTATACGTAGATATTTATATATATATAGATATACGTGGATATATATATTTATATAT 9658

Qy        641 ATATATACACATACATATACACACACATAGATATACATATATATACACATATATATACGT 700
              ||||| | |  || ||||| | | | |||   |  | ||||||||  | |||||||||  |
Db       9659 ATATAGATATTTATATATATATAGATATACGTGGATATATATATTTATATATATATAGAT 9718

Qy        701 ATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGG 760
              || ||||| |||||| |||| |        |  |||           |  |   ||    |
Db       9719 ATTTATATATATATAGATATACGTGGATATAGATATTTATATATATATAGATATTTATAC 9778

Qy        761 GGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATG 820
              | ||| ||  |   | |  |  | ||  | |||    | |  || || |   |
Db       9779 GTATATAGATATTTATATATATATAGATATTTATATATATATAGATATATTTATATATAT 9838

Qy        821 GCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAA 880
              | |||     ||| ||| |||  | | | ||| |   ||| | ||      |  | |
Db       9839 ATAGATATATATGGATATATATATTTATAT-ATATATAATATTTATATATAGATATATAT 9897

Qy        881 GGTAACACTATATATATATATAGTGAAA 908
              ||  | | ||| ||||||||||  || |
Db       9898 GGATATAGTATTTATATATATATAGATA 9925
```

RESULT 45
ADC87552
ID ADC87552 standard; DNA; 1486 BP.
XX
AC ADC87552;
XX
DT 01-JAN-2004 (first entry)
XX
DE Human GPCR gene SEQ ID NO:2005.
XX
KW ds; gene; human; GPCR;
KW guanosine triphosphate-binding protein coupled receptor; gene therapy.

```
XX
OS   Homo sapiens.
XX
PN   EP1270724-A2.
XX
PD   02-JAN-2003.
XX
PF   18-JUN-2002; 2002EP-00013517.
XX
PR   18-JUN-2001; 2001JP-00246789.
XX
PA   (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA   (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI   Suwa M,  Asai K,  Akiyama Y,  Aburatani H;
XX
DR   WPI; 2003-315783/31.
DR   P-PSDB; ADC87553.
XX
PT   New polynucleotide, useful for preparing a composition for treating a
PT   patient in need of increased or suppressed activity or expression of the
PT   guanosine triphosphate-binding protein coupled receptor.
XX
PS   Claim 1; SEQ ID NO 2005; 28pp; English.
XX
CC   The invention relates to a novel polynucleotide encoding a guanosine
CC   triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC   the invention may have a use in gene therapy. The polynucleotide and
CC   polypeptide are useful for preparing a composition for treating a patient
CC   in need of increased or suppressed activity or expression of the
CC   guanosine triphosphate-binding protein coupled receptor. The
CC   polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC   invention.
XX
SQ   Sequence 1486 BP; 517 A; 149 C; 242 G; 578 T; 0 U; 0 Other;

  Query Match             5.2%;  Score 256.4;  DB 10;  Length 1486;
  Best Local Similarity   70.2%;  Pred. No. 4.1e-35;
  Matches  344;  Conservative    0;  Mismatches  146;  Indels    0;  Gaps    0;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | |||||||| || ||| | |||||| ||| |      |  | |||||||||| || | ||
Db        718 TGTATATATACATGTATGTGTATATATACATATGTGTATGTGTATATATACATATGTGTA 777

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              |    |||||||||||| |  |  | |||| | | |||| |  |  |  |||| |||| |
Db        778 TGTGTATATATACATATGTGTATGTGTATATATACATATGTGTATGTGTATATATACATA 837

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              | | |||| | | || | |  | |||| || | || |||| | ||||||  |  | ||||
Db        838 TGTGTATATATACATATGTGTATATATGTATATATGTATATATACATATGTGTATATATG 897

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | |  |||||| ||||||||      ||||||   |||||    ||| || | || |
Db        898 TATATATGTATATATACATATATGTGTGTATATATGTATATATGTGTATTTATATATGTG 957

Qy        467 TACATATATACACATATATATAATATACACACATATATATACACATATATACACACATAT 526
              |  | | || | |||| ||||   || |    | |||||| ||||||||  |   ||||
Db        958 CATGTGTGTATAAATATGTATATATTATATGTGTGTATATATACATATATGTATGTATAT 1017

Qy        527 ATACACATATATACACATATATATACACATATATACACATATATACATATATACACACATAT 586
              |||||  |||||  | |||||||||  | |  ||| | ||| |||   |||||| ||| ||||
Db       1018 ATACATATATACATACATATATACATATATACACACATATACGCACATATAGACATATAT 1077

Qy        587 ATATACATATATACACATATATACATATATACACATATATACATATATACACACATATAT 646
              || || |||||  |||||||||||||||||||||   |||||||| | ||| | |||||||
Db       1078 ATGTATATATACACACATATATACATATATATGTATATATACACACATATATACATATAT 1137

Qy        647 ACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATAT 706
              |   ||| ||| ||||| | ||| |||||  ||||||||||||||||||   ||||||
Db       1138 ATGTATATATACACACATATATACATATATGTATATATACACATATATATGTATATATAC 1197

Qy        707 ATGTATATAT 716
```

```
              | |||||||
Db       1198 ACATATATAT 1207


RESULT 46
AAQ39248/c
ID   AAQ39248 standard; DNA; 870 BP.
XX
AC   AAQ39248;
XX
DT   25-MAR-2003  (revised)
DT   10-AUG-1993  (first entry)
XX
DE   Sequence of cosmid clone 4F7 isolated from a chromosome 11q-specific
DE   cosmid library genomic polymorphic locus (GPL).
XX
KW   Genomic polymorphic locus; restriction site; chromosome 11 q13; ss.
XX
OS   Synthetic.
XX
PN   WO9307166-A1.
XX
PD   15-APR-1993.
XX
PF   07-OCT-1992;   92WO-US008539.
XX
PR   09-OCT-1991;   91US-00773099.
XX
PA   (SALK ) SALK INST BIOLOGICAL STUDIES.
XX
PI   Evans GA,  Selleri L,  Eubanks JH;
XX
DR   WPI; 1993-134379/16.
XX
PT   Oligo-nucleotide(s) useful as primers - for identifying individuals on
PT   basis of presence of polymorphic locus, allows specific detection.
XX
PS   Disclosure; Page 33-34; 51pp; English.
XX
CC   Two cosmids, 4F7 and 7H12, were identified by screening cosmid libraries
CC   with a sequence from the 2nd transmembrane domain of a rat D2 dopamine
CC   recptor cDNA. The cosmids had potential structural homology to G-protein
CC   receptor gene family as well as several loci on chromosome 11. A highly
CC   repetitive element was also detected in the region of of overlap of the 2
CC   cosmids and was subcloned into Bluescript and its DNA sequence
CC   determined. It revealed a highly degenerate repetative hexamer motif
CC   spanning 438 bp consisting of 73 repeats. (Updated on 25-MAR-2003 to
CC   correct PN field.)
XX
SQ   Sequence 870 BP; 185 A; 127 C; 247 G; 311 T; 0 U; 0 Other;

  Query Match             5.2%;  Score 256;  DB 2;  Length 870;
  Best Local Similarity   73.9%;  Pred. No. 4.7e-35;
  Matches  340;  Conservative    0;  Mismatches  115;  Indels    5;  Gaps    1;

Qy        238 ATATATATATATATACACACACACACATATATATATACACATATATATACACATATAT 297
              ||| ||| | ||||||| ||||| ||||||||| ||| ||| ||| ||||||| | | ||
Db        810 ATACATACACATATACATACACATACACATATACATACACATATACATATACATACAT 751

Qy        298 ACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACAC 357
              | | ||||||| ||| | | ||||||| ||| ||||||| || |||| ||| ||||||||
Db        750 ATACATATACATATACACATACATATACATATACATACACATATACATATACATATACAC 691

Qy        358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417
              |||  |||||| ||| |||||||||||| | | ||||||| ||| | ||||||| | | ||
Db        690 ATACATATACATATACATACACATATACATATACATATACACATACATATACATATACAT 631

Qy        418 ATACACATATATACACATATATACACATATACACATATATACACATATATACATATATAC 477
              ||||| | | ||| ||| ||| ||| | ||| ||||||| | ||| ||| ||||||| |
Db        630 ATACATACACATATACACATACACATACATATACATATACATACACATACACATATACAT 571

Qy        478 ACATATATATAATATACACACATATATATACACATATATACACACATATATACACATATA 537
              | | |      ||| | | ||||||| |||||||||| | | ||||| | | ||| |||
```

```
Db        570 ATACA-----CATACATATACATATACATACACATATACATATACATACACATACACATA 516

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATACATATA 597
              |||| | | ||| ||| | | |||||||||| | ||| | |||||| || |  ||||  ||
Db        515 TACATACACATACACATACACATACACATACACACACACATACACGTACACGTACACGTA 456

Qy        598 TACACATATATACATATATACACATATATACATATATACACACATATATACACATACATA 657
              ||  | |  ||||  |||||| ||| ||| | ||||||| | | ||||||| |||||||
Db        455 CACGTACACGTACACGTATACATATACATATACATATACATATACATATACATATACATA 396

Qy        658 TACACACACATAGATATACATATATATACACATATATATA 697
              |||| | ||||| | | | | | ||| |   | |   |||
Db        395 TACATATACATACACACATACACATACAAGAAGAAGGATA 356


RESULT 47
ADC87554/c
ID   ADC87554 standard; DNA; 1489 BP.
XX
AC   ADC87554;
XX
DT   01-JAN-2004  (first entry)
XX
DE   Human GPCR gene SEQ ID NO:2007.
XX
KW   ds; gene; human; GPCR;
KW   guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1270724-A2.
XX
PD   02-JAN-2003.
XX
PF   18-JUN-2002; 2002EP-00013517.
XX
PR   18-JUN-2001; 2001JP-00246789.
XX
PA   (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA   (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI   Suwa M,  Asai K,  Akiyama Y,  Aburatani H;
XX
DR   WPI; 2003-315783/31.
DR   P-PSDB; ADC87555.
XX
PT   New polynucleotide, useful for preparing a composition for treating a
PT   patient in need of increased or suppressed activity or expression of the
PT   guanosine triphosphate-binding protein coupled receptor.
XX
PS   Claim 1; SEQ ID NO 2007; 28pp; English.
XX
CC   The invention relates to a novel polynucleotide encoding a guanosine
CC   triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC   the invention may have a use in gene therapy. The polynucleotide and
CC   polypeptide are useful for preparing a composition for treating a patient
CC   in need of increased or suppressed activity or expression of the
CC   guanosine triphosphate-binding protein coupled receptor. The
CC   polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC   invention.
XX
SQ   Sequence 1489 BP; 553 A; 208 C; 160 G; 568 T; 0 U; 0 Other;

  Query Match             5.1%;  Score 255.4;  DB 10;  Length 1489;
  Best Local Similarity   63.8%;  Pred. No. 6.1e-35;
  Matches  438;  Conservative    0;  Mismatches  241;  Indels    8;  Gaps    3;

Qy        236 ATATATATATATATATACACACACACACATATATATATACACATATATATACACATAT 295
              | || | |||||||||||| | | ||  | |||| | |||| | | ||| |||   ||||
Db       1197 ACATGTGTATATATATACATATATACGTATATATGTGTATATATACATACATAGGTATAT 1138

Qy        296 ATACATATATACACATATATACACATATATATACACATACATATGTACACATATATATAC 355
              | | ||||| ||| | ||||| || |||||||  | |||  ||| || |||| ||||||

Db       1137 ACATATATACACATACATATATACGTATATATGTATATATGTATATATACATGTATATAT 1078

Qy        356 ACATATG-TATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              ||||| | || |||||||||||| |||||| ||  |||| || ||||||||||||  |
Db       1077 GTATATGTTGTATACATATATACACGTATATATACGTATATGTATACATATATACACGTA 1018

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              |||||||  |||| || |||||||||||| |||| | |  | || | | || || |||||
Db       1017 TATATACGTATATGTATACATATATACACGTATATATACGTGTATATACATGTATATATA 958

Qy        475 TACACATATATATAATATACACACATATATATACACATATATACACACATATATACACAT 534
              | || |     ||| | | ||||||||| ||   |||||||| |  ||||||||| |
Db        957 CGTATATGT-----ATACATATACATATATACACGTGTATATACATATGTATATACACGT 903

Qy        535 ATATACACATATATATACACATATATACACATATATACATATATACACATATATATACAT 594
              |||| |  ||| |||| |||  | || | || ||||| |||||| |||  | |||| ||
Db        902 GTATATATGTATGTATATACACGTGTATATATGTATACGTATATATACACGTGTATATAT 843

Qy        595 ATATACACATATATACATATATACACATATATACATATATACACACATATATACACATAC 654
              |||||  |||||||||  | || | || ||||| |||||| |||| | |||| |  ||
Db        842 GTATACGTATATATACACGTGTATATATGTATACGTATATATACACGTGTATATATGTAT 783

Qy        655 ATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATAT 714
              | || | | |||    |||| |||| |||||  ||||||| |||| ||||||||||||
Db        782 ACGTATATATACACGTGTATATATATGTATACGTATATATACACGTGTATATATGTATAC 723

Qy        715 ATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAG 774
              |||||    |  | ||  ||| |         | |   |      |     || | | | |
Db        722 GTATATATACACGTGTATATATGTATGTATATATACGTGTATATATATGTATACGTATAT 663

Qy        775 AGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGG 834
              | |  | | |  ||||    |     | | |     |    | |||    |
Db        662 ATACGTATATATAAATGTAT--ATATGTATATGTGTATATATACGTATATATACATGTAT 605

Qy        835 AAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATA 894
              | |||| || || |  |   | || ||| ||  | |||  ||   | |     |||||
Db        604 ATATACGTATATATACACATATACATATATGGTACGTATATATACATATGTGTATATATG 545

Qy        895 TATATATAGTGAAAGCAGTGTAAACCT 921
              ||||||||   | |    | || || |
Db        544 TATATATATACATACATATATATACAT 518


RESULT 48
ABD33148/c
ID   ABD33148 standard; DNA; 104399 BP.
XX
AC   ABD33148;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Murine cancer-associated (CA) gene MD07-019.
XX
KW   Mouse; cancer-associated protein; CAP; cancer-associated gene; CA; gene;
KW   ds; cancer; cytostatic.
XX
OS   Mus musculus.
XX
PN   WO2004058146-A2.
XX
PD   15-JUL-2004.
XX
PF   15-DEC-2003; 2003WO-US040081.
XX
PR   17-DEC-2002; 2002US-00322281.
XX
PA   (SAGR-) SAGRES DISCOVERY INC.
XX
PI   Morris DW,  Malandro MS;
XX
DR   WPI; 2004-499109/47.
XX
PT   Novel human cancer associated protein encoded within open reading frame
```

```
PT   of cancer associated gene, useful as targets for diagnosing cancer.
XX
PS   Disclosure; SEQ ID NO 123; 182pp; English.
XX
CC   The invention relates to cancer-associated proteins (CAP) and the cancer-
CC   associated (CA) nucleic acids encoding them. The invention also relates
CC   to a method for treating cancers involving administering to a patient an
CC   inhibitor of CAP, and a method of screening for anticancer activity in a
CC   potential drug involving providing a cell that expresses a CA gene,
CC   contacting a tissue sample derived from a cancer cell with an anticancer
CC   drug candidate and monitoring the effect of the anticancer drug candidate
CC   on expression of the CA gene. The CAP proteins are useful for detecting
CC   cancer associated with expression of a CAP protein in a test cell sample
CC   and for screening for a bioactive agent capable of modulating the
CC   activity of a CAP protein. The CA nucleic acids are useful for diagnosing
CC   cancer, involving determining the expression of a CA nucleic acid in a
CC   tissue. This sequence represents a murine CA gene of the invention. Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 104399 BP; 25540 A; 20083 C; 20918 G; 29473 T; 0 U; 8385 Other;

  Query Match             5.1%;  Score 255.4;  DB 13;  Length 104399;
  Best Local Similarity   73.3%;  Pred. No. 7.7e-35;
  Matches  367;  Conservative    0;  Mismatches  131;  Indels    3;  Gaps    3;

Qy        230 ATATATATATATATATATATATACACACACACACACATATATATATACACATATATATAC 289
              ||| ||| ||| ||||||| |||| ||| | | ||||| ||| |||| | ||||||||
Db      13721 ATACATACATACATATATACATACGTACATATATACATACATACATACGTACATATATAC 13662

Qy        290 ACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATAT 349
              | |  || | ||||| | ||||| ||||| ||||| ||||  | | ||||  | |||||
Db      13661 ATACGTACATATATACATACATACATACATATATACATACGTACATATATACATACATAC 13602

Qy        350 ATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATAC 409
              |||   |||||| || | || || ||| | | ||| | ||| | ||||| |||||  |||
Db      13601 ATACGTACATATATACATACGTACATATATACATACATACATACATATATACATACGTAC 13542

Qy        410 ACACATATATACACATATATACACATATATACACATATACACATATATACACATATATAC 469
              | | ||| ||||| | |||    |||||||||| | ||||| | ||||| ||||| ||||
Db      13541 ATATATACATACATACATACGTACATATATACATACATACATACATATATACATACATAC 13482

Qy        470 ATATATACACATATATATAATATACACACATATATATACACATATATACACACATATATA 529
              |||||||| | ||| ||||| | | | | ||  ||| | | ||| | ||| ||||| |
Db      13481 ATATATACATACATACATAATACATATATACATTCATATATACATA-ATACATATATACA 13423

Qy        530 CACATATATACACATATATAT-ACACATATATACACATAT-ATACATATATACACATATA 587
              |||||||||| | ||| || | |||||||||| | || ||||||| || || |||||
Db      13422 TACATATATACATACATACATAATACATATATACATACATAATACATACATGCATATATA 13363

Qy        588 TATACATATATACACATATATACATATATACACATATATACATATATACACACATATATA 647
              || |||| ||| | | ||| ||| | ||||| | ||| | | ||||| ||| |  || |
Db      13362 CATGCATACATATATACATACACACACATACATACATACATATATACATACGTACA 13303

Qy        648 CACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATA 707
              | ||||||| | ||| || || ||||| | ||| ||||| ||||| |||| || |||||
Db      13302 TATATACATACATACATACGTACATATATACATACATACATATATACATACATACATATA 13243

Qy        708 TGTATATATATATGCTCCAGA 728
              |  ||| ||| ||  | || |
Db      13242 TACATACATACATAATACATA 13222


RESULT 49
ADC86154/c
ID   ADC86154 standard; DNA; 84510 BP.
XX
AC   ADC86154;
XX
DT   01-JAN-2004  (first entry)
XX
DE   Human GPCR gene SEQ ID NO:607.
XX
KW   ds; gene; human; GPCR;
KW   guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1270724-A2.
XX
PD   02-JAN-2003.
XX
PF   18-JUN-2002; 2002EP-00013517.
XX
PR   18-JUN-2001; 2001JP-00246789.
XX
PA   (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA   (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI   Suwa M,  Asai K,  Akiyama Y,  Aburatani H;
XX
DR   WPI; 2003-315783/31.
DR   P-PSDB; ADC86155.
XX
PT   New polynucleotide, useful for preparing a composition for treating a
PT   patient in need of increased or suppressed activity or expression of the
PT   guanosine triphosphate-binding protein coupled receptor.
XX
PS   Claim 1; SEQ ID NO 607; 28pp; English.
XX
CC   The invention relates to a novel polynucleotide encoding a guanosine
CC   triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC   the invention may have a use in gene therapy. The polynucleotide and
CC   polypeptide are useful for preparing a composition for treating a patient
CC   in need of increased or suppressed activity or expression of the
CC   guanosine triphosphate-binding protein coupled receptor. The
CC   polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC   invention.
XX
SQ   Sequence 84510 BP; 21253 A; 19989 C; 19504 G; 23764 T; 0 U; 0 Other;

  Query Match             5.1%;  Score 255;  DB 10;  Length 84510;
  Best Local Similarity   72.6%;  Pred. No. 8.9e-35;
  Matches  358;  Conservative    0;  Mismatches  130;  Indels    5;  Gaps    2;

Qy        232 ATATATATATATATATATATACACACACACACATATATATATACACATATATATACAC 291
              |||||||| ||||   |||| | | ||||| || |||||||||| ||  ||||| ||
Db      80292 ATATATATACATATTCTTATATATATATACACATATGTATATATACATATGAATATATAC 80233

Qy        292 ATATATACATATATACACATATATACACATATATATACACATACATATGTACACATATAT 351
              | ||||  |||||||| ||  ||| | | | ||||||| ||  |||| ||||||||||
Db      80232 ACATATGTATATATACATATGAATATATACACATATACATATGAATATATACACATATAC 80173

Qy        352 ATACACATATGTATACACATATATACACATATATACACACATATATACACATATATACAC 411
              |||   |||| || | | ||| |||   |||||||||| |||||   |   ||||| ||
Db      80172 ATATGAATATATACATATATACATATGAATATATACACATATATACGTATGAATATATAC 80113

Qy        412 ACATATATAC----ACATATATACACATATATACACATATACACATATATACACATATAT 467
              ||||||||||      ||||||| ||| ||||| || |||  | ||||||||||||||||
Db      80112 ACATATATACGTATGAATATATATACACATATATACGTATGAATATATATACACATATAT 80053

Qy        468 ACATATATACACATATATATAATATACACACATATATATACACATATATACACACATATA 527
              || |||  | | ||||| |  ||||||  |  |||||||  || ||||||  |   |||
Db      80052 ACGTATGAATATATATACA-CATATACGTATGAATATATATACCCATATACGTATGAATA 79994

Qy        528 TACACATATATACACATATATATACACATATATACACATATATACATATATACACATATA 587
              |||||||||||| || ||  |||| | ||| | | | |||  ||| |||||| | | | |
Db      79993 TACACATATATATACGTACGTATATATATACACATATATACGTACGTATATATATACACA 79934

Qy        588 TATACATATATACACATATATACATATATACACATATATACATATATACACACATATATA 647
              |||| |  ||   | |||||||||| ||||| || ||   | |||||||||||| |||||
Db      79933 TATATACGTACGTATATATATACACATATATACGTACGAATATATATACACATATATACG 79874

Qy        648 CACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATA 707
              ||  | ||||| ||||| ||| |  |||  ||||||| ||| ||||||||||||   |||
```

```
Db      79873 TACGAATATATATACACATATATACGTACGAATATATATACACATATATACGTACGAATA 79814

Qy        708 TGTATATATATAT 720
              | |||| | ||||
Db      79813 TATATACACATAT 79801


RESULT 50
ADL13771/c
ID   ADL13771 standard; DNA; 201766 BP.
XX
AC   ADL13771;
XX
DT   06-MAY-2004  (first entry)
XX
DE   Osteoarthritis-associated polymorphic nucleotide #303.
XX
KW   ds; gene; osteopathic; antiinflammatory; antiarthritic; gene therapy;
KW   joint space narrowing; osteophyte development; joint pain;
KW   osteoarthritis; SNP; single nucleotide polymorphism.
XX
OS   Homo sapiens.
XX
PN   WO2003054166-A2.
XX
PD   03-JUL-2003.
XX
PF   19-DEC-2002; 2002WO-US041225.
XX
PR   20-DEC-2001; 2001US-0342603P.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Jones KA,  Schafer A;
XX
DR   WPI; 2003-559141/52.
XX
PT   Determining susceptibility of an individual to joint space narrowing,
PT   osteophyte development and/or joint pain comprises identifying whether
PT   the individual has at least one polymorphism in a polynucleotide encoding
PT   a protein.
XX
PS   Disclosure; SEQ ID NO 303; 297pp; English.
XX
CC   The invention relates to a method of determining susceptibility of an
CC   individual to joint space narrowing and/or osteophyte development and/or
CC   joint pain comprising identifying whether the individual has at least one
CC   polymorphism in a polynucleotide encoding at least one of the protein
CC   listed in the specification. The methods, composition and agent are
CC   useful for modulating the susceptibility of an individual to joint space
CC   narrowing and/or osteophyte development and/or joint pain that is
CC   associated with a disease, preferably osteoarthritis. The cell line and
CC   the non-human animal are useful for screening for an agent for diagnosing
CC   an individual having susceptibility to joint space narrowing and/or
CC   osteophyte development and/or joint pain. This sequence corresponds to
CC   the polynucleotide encoding a protein listed in the specification. (Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification but was obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences).
XX
SQ   Sequence 201766 BP; 57593 A; 35575 C; 36821 G; 54688 T; 0 U; 17089 Other;

  Query Match             5.1%;  Score 255;  DB 10;  Length 201766;
  Best Local Similarity   63.6%;  Pred. No. 9.4e-35;
  Matches  437;  Conservative    0;  Mismatches  242;  Indels    8;  Gaps    3;

Qy        236 ATATATATATATATATACACACACACACACATATATATATACACATATATATACACATAT 295
              | || | ||||||||||||| | | ||  | |||| | |||| | | ||| |||   ||||
Db     189316 ACATGTGTATATATATACATATATACGTATATATGTGTATATATACATACATAGGTATAT 189257

Qy        296 ATACATATATACACATATATACACATATATATACACATACATATGTACACATATATATAC 355
              | | ||||| ||| | ||||| || |||||||  | |||  ||| || |||| ||||||
Db     189256 ACATATATACACATACATATATACGTATATATGTATATATGTATATATACATGTATATAT 189197

Qy        356 ACATATGTA-TACACATATATACACATATATACACACATATATACACATATATACACACA 414
              |||||   || ||||||||||||| |||||| ||  |||| || ||||||||||||  |
Db     189196 GTATATGNNGTATACATATATACACGTATATATACGTATATGTATACATATATACACGTA 189137

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              |||||||  |||| || |||||||||||| |||| | |  | || | | || || |||||
Db     189136 TATATACGTATATGTATACATATATACACGTATATATACGTGTATATACATGTATATATA 189077

Qy        475 TACACATATATATAATATACACACATATATATACACATATATACACACATATATACACAT 534
              | || |     ||| | | ||||||||| ||   |||||||| |  ||||||||  |
Db     189076 CGTATATGT-----ATACATATACATATATACACGTGTATATACATATGTATATACACGT 189022

Qy        535 ATATACACATATATATACACATATATACACATATATACATATATACACATATATATACAT 594
              |||| |  ||| |||| |||  | || | || ||||| |||||| |||  | |||| ||
Db     189021 GTATATATGTATGTATATACACGTGTATATATGTATACGTATATATACACGTGTATATAT 188962

Qy        595 ATATACACATATATACATATATACACATATATACATATATACACACATATATACACATAC 654
              |||||  |||||||||  | || | || ||||| |||||| |||| | |||| |  ||
Db     188961 GTATACGTATATATACACGTGTATATATGTATACGTATATATACACGTGTATATATGTAT 188902

Qy        655 ATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATAT 714
              |  || | | |||    |||| |||| |||||  ||||||| |||| ||||||||||||
Db     188901 ACGTATATATACACGTGTATATATATGTATACGTATATATACACGTGTATATATGTATAC 188842

Qy        715 ATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAG 774
              |||||   | | ||  ||| |       | |   |     |     || | | | |
Db     188841 GTATATATACACGTGTATATATGTATGTATATATACGTGTATATATATGTATACGTATAT 188782

Qy        775 AGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGG 834
              | |  | | | | |||||     |     | | |     |       | |||   |
Db     188781 ATACGTATATATAAATGTAT--ATATGTATATGTGTATATATACGTATATATACATGTAT 188724

Qy        835 AAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATA 894
              | |||| || || |  |   | || ||| ||  | |||  ||   |  |    |||||
Db     188723 ATATACGTATATATACACATATACATATATGGTACGTATATATACATATGTGTATATATG 188664

Qy        895 TATATATAGTGAAAGCAGTGTAAACCT 921
              ||||||||   | |    | || || |
Db     188663 TATATATATACATACATATATATACAT 188637


RESULT 51
ADC87564
ID   ADC87564 standard; DNA; 1114 BP.
XX
AC   ADC87564;
XX
DT   01-JAN-2004  (first entry)
XX
DE   Human GPCR gene SEQ ID NO:2017.
XX
KW   ds; gene; human; GPCR;
KW   guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1270724-A2.
XX
PD   02-JAN-2003.
XX
PF   18-JUN-2002; 2002EP-00013517.
XX
PR   18-JUN-2001; 2001JP-00246789.
XX
PA   (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA   (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI   Suwa M,  Asai K,  Akiyama Y,  Aburatani H;
XX
DR   WPI; 2003-315783/31.
DR   P-PSDB; ADC87565.
XX
```

```
PT   New polynucleotide, useful for preparing a composition for treating a
PT   patient in need of increased or suppressed activity or expression of the
PT   guanosine triphosphate-binding protein coupled receptor.
XX
PS   Claim 1; SEQ ID NO 2017; 28pp; English.
XX
CC   The invention relates to a novel polynucleotide encoding a guanosine
CC   triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC   the invention may have a use in gene therapy. The polynucleotide and
CC   polypeptide are useful for preparing a composition for treating a patient
CC   in need of increased or suppressed activity or expression of the
CC   guanosine triphosphate-binding protein coupled receptor. The
CC   polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC   invention.
XX
SQ   Sequence 1114 BP; 398 A; 107 C; 167 G; 442 T; 0 U; 0 Other;

  Query Match             5.1%;  Score 254.2;  DB 10;  Length 1114;
  Best Local Similarity   64.0%;  Pred. No. 9.7e-35;
  Matches  399;  Conservative    0;  Mismatches  223;  Indels    1;  Gaps    1;

Qy        229 TATATATATATATATATATATATACACACACACACACATATATATATACACATATATATA 288
              ||||| ||||| | ||| |||||||| | | | ||| |||| |||||     || ||||||
Db        212 TATATGTATATTTGTATGTATATACAAATATATACATATATGTATATGTGTATGTATATA 271

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              || ||||||||||||||    || |||| ||| |||||||| |||  | | |  | |||
Db        272 CAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATATA 331

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
               | ||| | | || |||    || |||| ||| ||||| ||| |||| |  |  ||||||
Db        332 CAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATATA 391

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              || | ||||| | | || | ||   ||||| | | ||||||| |||| |  |  ||||||
Db        392 CAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATATA 451

Qy        469 CATATATACACATATATATAATATACACACATATATATACACATATATACACACATATAT 528
              || ||||| |||||||| |  |||  | | | | ||||| ||||||||     |  |||||
Db        452 CAAATATATACATATATGT-GTATGTATATACAAATATATACATATATGTGTATGTATAT 510

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACATATAT 588
              ||| |||||||| |||| | ||   ||||| | | ||||| | |||| |  |  |||||
Db        511 ACAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATAT 570

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
              | | |||||||||| |||| |  || |||| ||| ||||| | |||||     || ||||
Db        571 ACAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATAT 630

Qy        649 ACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATAT 708
              ||| | ||||||| | |  |  || || ||| | ||| | | ||||||  |||| |||||
Db        631 ACAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATAT 690

Qy        709 GTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAG 768
              | |||||||      | ||   ||      |  |  |   ||    ||     || | |
Db        691 ACAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATAT 750

Qy        769 GAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAG 828
               ||| | |     | ||  || | |    | ||  | ||      ||       |||
Db        751 ACAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATAT 810

Qy        829 TAGGGGAAATACATAAATGTGTA 851
                  | ||||||| |||||||
Db        811 ACAAATATATACATATATGTGTA 833


RESULT 52
ACC84012
ID   ACC84012 standard; DNA; 2200 BP.
XX
AC   ACC84012;
XX
DT   22-SEP-2003  (first entry)
XX
DE   Pacific white shrimp skeletal muscle actin gene promoter.
XX
KW   Shrimp; actin; promoter; transgenic; disease resistance;
KW   stress tolerance; cold tolerance; ds.
XX
OS   Litopenaeus vannamei.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    878. .893
FT                   /*tag=  a
FT                   /note= "CATA-rich region"
FT   misc_feature    1071. .1078
FT                   /*tag=  a
FT                   /note= "CATA-rich region"
FT   misc_feature    1538. .1549
FT                   /*tag=  a
FT                   /note= "CACA-rich region"
FT   misc_feature    1554. .1567
FT                   /*tag=  a
FT                   /note= "CACA-rich region"
FT   misc_feature    1570. .1585
FT                   /*tag=  a
FT                   /note= "CACA-rich region"
XX
PN   WO2003048325-A2.
XX
PD   12-JUN-2003.
XX
PF   04-DEC-2002; 2002WO-US038523.
XX
PR   04-DEC-2001; 2001US-0336603P.
XX
PA   (UYHA-) UNIV HAWAII.
XX
PI   Sun PS,  Arakaki KL,  Sun SSM;
XX
DR   WPI; 2003-513755/48.
XX
PT   New isolated Beta-lactin nucleic acid promoter molecule from shrimp,
PT   useful for imparting resistance against a virus and increasing stress
PT   tolerance in an animal and for regulating growth of an animal.
XX
PS   Claim 52; Page 14-15; 97pp; English.
XX
CC   The present sequence is that of the promoter region of the Pacific white
CC   shrimp skeletal muscle actin gene. This promoter is suitable for inducing
CC   expression of a protein encoded by a nucleic acid molecular operably
CC   associated with the promoter molecule in an expression vector. It
CC   contains promoter-associated TATA and CAAT boxes approximately 500 bp
CC   upstream from the translation start site. A unique characteristic of the
CC   promoter are CACA-rich and CATA-rich regions located upstream of the TATA
CC   and CAAT boxes. The invention relates to actin and beta-actin promoter
CC   molecules from shrimp, expression cassettes including the promoter,
CC   expression vectors, host cells and transgenic animals transduced with the
CC   promoter. Methods for imparting resistance to a pathogen, for regulating
CC   growth, for improving stress tolerance, and for increasing cold tolerance
CC   in an animal, especially a shrimp, involve transforming the animal with a
CC   nucleic acid construct including the isolated actin and beta-actin
CC   promoter molecules of the invention
XX
SQ   Sequence 2200 BP; 837 A; 372 C; 263 G; 728 T; 0 U; 0 Other;

  Query Match             5.1%;  Score 254;  DB 9;  Length 2200;
  Best Local Similarity   71.7%;  Pred. No. 1.1e-34;
  Matches  360;  Conservative    0;  Mismatches  140;  Indels    2;  Gaps    2;

Qy        234 ATATATATATATATATATACACACACACACATATATATATACACATATATATACACAT 293
              | || ||||||||||||||    | |    | | |||| ||| || | | ||| ||| |||
Db        647 ACATTTATATATATATATGTATATGTATATATATGTATTTATATACATAAATATACAC 706

Qy        294 ATATACATATATACACATATATACACATATATATACACATACATATGTACACATATATAT 353
              |||| ||| |||| | |||| || | ||| ||||| |  || |||| | || |||||||||
```

```
Db        707 ATATTCATGTATATATATATGTATATATACATATATATTTATATATATGCATATATATAT 766

Qy        354 ACACATATGTATACACATATATACACATATATACACACATATATACACATATATACACAC 413
              | |||| |||| | | |||| || ||  ||| | |  |||||| | | |||||||| |
Db        767 GTATATATATATATATACATATGCATATTCATATATATGTATATATATACATATACATAT 826

Qy        414 ATATATACACATATATACACATATATACACATATACACATATATACACATATATACATAT 473
              |||| || | ||||  | | ||||||  | ||  | | |||||||| | ||| |||||||
Db        827 ATATGTATATATATGCATAAATATATGTATATGAATATATATATATATACATACATAC 886

Qy        474 ATACACATATATATAATATACACACATATATATACACATATATACACACATATATACACA 533
              ||||| |||||||| ||||| | | ||||||||| | |||||| | | ||| | |||||
Db        887 ATACATATATATAT-ATATATATATATATATATATATATATATATATATATACACACACA 945

Qy        534 TATATACACATATATATACACATATATACACATATATACATATATACACATATATATACA 593
              | || |||  | ||||||||| | | | ||| | | |||||||||   | ||  |||||
Db        946 CACATGCACGCACATATACACACACACAAACACACACACATATATATGTAAATGCATACA 1005

Qy        594 TATATACACATATATACATATATACACATATATACATATATACACACATATATACACATA 653
              ||| |  |||   | ||||||||||||||||||||||| | ||||| | | ||||| ||
Db       1006 TATGTGTACACGGACACATATATACACATATATACATACACACACA-ACAGATACATATG 1064

Qy        654 CATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATA 713
              |||||||| ||| | | |||||  | | | || | | ||||||| ||||||||| ||||||
Db       1065 CATATACATACATACACATATATCTCTCTCTATATACATATATATGTATATATGTGTATA 1124

Qy        714 TATATATGCTCCAGAGTTCATA 735
              ||  ||| |   | |   || |
Db       1125 TACTTATACATAAAAACACACA 1146


RESULT 53
AAV83939/c
ID   AAV83939 standard; DNA; 80595 BP.
XX
AC   AAV83939;
XX
DT   03-MAR-1999  (first entry)
XX
DE   HC-contig derived from normal human chromosome 10q25.2 region.
XX
KW   Yeast artificial chromosome; YAC; probe; eukaryotic chromosome;
KW   neocentromere; replication; extra-chromosomal element; segregation;
KW   cell division; artificial chromosome; gene therapy; mardel(10);
KW   human artificial chromosome; transgenic; chromosome 10; 10q25.2; ss.
XX
OS   Homo sapiens.
XX
PN   WO9851790-A1.
XX
PD   19-NOV-1998.
XX
PF   13-MAY-1998;   98WO-AU000352.
XX
PR   13-MAY-1997;   97AU-00006784.
PR   26-AUG-1997;   97AU-00008791.
XX
PA   (AMRA-) AMRAD OPERATIONS PTY LTD.
XX
PI   Choo K,  Du Sart D,  Cancilla MR;
XX
DR   WPI; 1999-009773/01.
XX
PT   New isolated nucleic acid comprising neocentromere sequences from
PT   eukaryotic chromosome - used to produce replicable, segregating
PT   artificial chromosomes that can carry large amounts of DNA for gene
PT   therapy.
XX
PS   Claim 8; Fig 6; 540pp; English.
XX
CC   The present sequence represents the HC-contig derived from normal human
CC   chromosome 10, 10q25.2 region. This region can be naturally mutated to
CC   produce an unusual chromosomal marker referred to as mardel(10). The
CC   mardel(10) marker is mitotically stable and contains a functional
CC   neocentromere at a location regarded as non-centromeric. This
CC   neocentromere maps to q25.2 on chromosome 10. The specification describes
CC   nucleic acid sequences derived from a eukaryotic chromosome, including a
CC   neocentromere or its functional derivative or hybrid, that are able, in a
CC   compatible cell, of replicating, acting as extra-chromosomal element and
CC   segregating during cell division. The sequences can be used to construct
CC   artificial chromosomes for use in gene therapy comprising a replicable,
CC   segregating nucleic acid that confers a specific phenotype on cells.
CC   Human artificial chromosomes can propagate in human cells and carry large
CC   amounts of DNA (e.g. therapeutic genes), and, being extra-chromosomal,
CC   they are not mutagenic. The artificial chromosomes are also useful for
CC   generation of transgenic plants and animals, in production of proteins
CC   and to make diagnostic reagents, e.g. for expression of cytokines,
CC   receptors and growth factors, or to increase the copy number of a gene in
CC   a cell. The constructs may also be used for functional and structural
CC   analysis of chromosomes
XX
SQ   Sequence 80595 BP; 23183 A; 16613 C; 16824 G; 23975 T; 0 U; 0 Other;

  Query Match             5.1%;  Score 253.6;  DB 2;  Length 80595;
  Best Local Similarity   71.9%;  Pred. No. 1.6e-34;
  Matches  361;  Conservative    0;  Mismatches  134;  Indels    7;  Gaps    2;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | |||| |||| ||| | |||||| ||| | | ||| |  ||| | |||| ||| |||||
Db      16026 TATATACATATTTATGTCTATATACACATATATACATATTTATGTCTATATACACATATA 15967

Qy        287 TACACATATATACATATATACACATATATACACA------TATATATACACATACATATG 340
              |||| || |||   |||||||||||||||||| |      | ||||||||||| |||
Db      15966 TACATATTTATGTCTATATACACATATATACATATTTATGTCTATATACACATATATACA 15907

Qy        341 TACACATATATATACACATATGTATACACATATATACACATATATACACACATATATACA 400
              ||   || | |||| ||| || |||||| || |||     ||||| ||| | ||| |||
Db      15906 TATTTATGTCTATATACACATATATACATATTTATGTCTATATACACATATATACATATT 15847

Qy        401 CATATATACACACATATATACACATATATACACATATATACACATATACACATATATACA 460
              || | || | ||| ||||| |||||| ||    ||||||||||||| |||||| ||
Db      15846 TATGTCTATATACACATATATACATATTTATGTCTATATACACATATATACATATTTATG 15787

Qy        461 CATATATACATATATACACATATATAT-AATATACACACATATATATACACATATATACA 519
              |||||||| ||||| |||||| |||   |||| ||||||||||| | |  ||| | |
Db      15786 TCTATATACACATATATACATATTTATGTCTATATACACATATATACATATTTATGTCTA 15727

Qy        520 CACATATATACACATATATACACATATATATACACATATATACACATATATACATATATA 579
              | | | ||| | | ||||   | | ||||||||||||||||||| || |||   ||||||
Db      15726 TATACACATATATACATATTTATGTCTATATACACATATATACATATTTATGTCTATATA 15667

Qy        580 CACATATATATACATATATACACATATATACATATATACACATATATACATATATACACA 639
              |||||||||| | || |||  | ||||| |||||||||| || |||   ||||||||||
Db      15666 CACATATATACATATTTATGTATATATACACATATATACATATTTATGTCTATATACACA 15607

Qy        640 CATATATACACATACATATACACACACATAGATATACATATATATACACATATATATACG 699
              ||||| | |  ||  |||| |   |  || ||||   | |||||  ||||| |||||
Db      15606 TATATACATATTTATGTATATATTTATGTATATATGTGTGTATATGTACATACATATATA 15547

Qy        700 TATATATATGTATATATATATG 721
              ||| |||  || |||||||||
Db      15546 CATACATAAATACATATATATG 15525


RESULT 54
AAA10225/c
ID   AAA10225 standard; DNA; 106746 BP.
XX
AC   AAA10225;
XX
DT   03-JUL-2000  (first entry)
XX
DE   Human PCTA-1 genomic DNA, SEQ ID NO:1.
XX
KW   Prostate carcinoma antigen-1; PCTA-1; human; galectin-8 homologue;
KW   splice variant; prostate cancer; biallelic marker; polymorphism;
```

```
KW   detection; diagnosis; prognosis; drug screening; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   misc_signal     1. .68647
FT                   /*tag= a
FT                   /note= "5' regulation region"
FT   promoter        66647. .68647
FT                   /*tag= b
FT   exon            68648. .68471
FT                   /*tag= c
FT                   /number
FT   intron          68742. .70646
FT                   /*tag= d
FT                   /number
FT   exon            70647. .70794
FT                   /*tag= e
FT                   /number= 1
FT   intron          70795. .82207
FT                   /*tag= f
FT                   /number= 1
FT   exon            82208. .82296
FT                   /*tag= g
FT                   /number= 2
FT   intron          82297. .83612
FT                   /*tag= h
FT                   /number= 2
FT   exon            83613. .83823
FT                   /*tag= i
FT                   /number= 3
FT   intron          83824. .85297
FT                   /*tag= j
FT                   /number= 3
FT   exon            85298. .85417
FT                   /*tag= k
FT                   /number= 4
FT   intron          85418. .86388
FT                   /*tag= l
FT                   /number= 4
FT   exon            86389. .86445
FT                   /*tag= m
FT                   /number= 5
FT   intron          86446. .87495
FT                   /*tag= n
FT                   /number= 5
FT   exon            87496. .87522
FT                   /*tag= o
FT                   /number= 6
FT   intron          87523. .87649
FT                   /*tag= p
FT                   /number= 6
FT   exon            87650. .87775
FT                   /*tag= q
FT                   /number= 6bis
FT   intron          87776. .88294
FT                   /*tag= r
FT                   /number= 6bis
FT   exon            88295. .88383
FT                   /*tag= s
FT                   /number= 7
FT   intron          88384. .89483
FT                   /*tag= t
FT                   /number= 7
FT   exon            89484. .89649
FT                   /*tag= u
FT                   /number= 8
FT   intron          89650. .92748
FT                   /*tag= v
FT                   /number= 8
FT   exon            92749. .97155
FT                   /*tag= w
FT                   /number= 9
FT   exon            92749. .92883
FT                   /*tag= x
FT                   /number= 9bis
FT   intron          92884. .95820
FT                   /*tag= y
FT                   /number= 9bis
FT   exon            95821. .97155
FT                   /*tag= z
FT                   /number= 9ter
FT   misc_signal     97155. .106746
FT                   /*tag= aa
FT                   /note= "3' regulation region"
XX
PN   WO9964590-A1.
XX
PD   16-DEC-1999.
XX
PF   04-JUN-1999;   99WO-IB001072.
XX
PR   05-JUN-1998;   98US-0088187P.
PR   28-SEP-1998;   98US-0102324P.
XX
PA   (GEST ) GENSET.
XX
PI   Blumenfeld M,  Bougueleret L,  Chumakov I;
XX
DR   WPI; 2000-338635/29.
DR   P-PSDB; AAY87403, AAY87404, AAY87405.
XX
PT   Novel polynucleotide used to develop products for detecting
PT   susceptibility to, diagnosis, prognosis and therapy of prostate cancer.
XX
PS   Claim 1; Page 277-306; 339pp; English.
XX
CC   This sequence represents the genomic DNA sequence encoding human prostate
CC   carcinoma tumour antigen-1 (PCTA-1). At least three PCTA-1 cDNAs
CC   (AAA10226, AAA10227, AAA10228) have been identified all of which comprise
CC   exons 1, 2, 3, 4, 5, 6, 7, 8, and additionally either contain exon 9,
CC   exons 6bis and 9, or exons 9bis and 9ter. PCTA-1 is a human homologue of
CC   rat galectin-8, which can mediate both cell-cell and cell-matrix
CC   interactions. 125 biallelic markers were identified in the genomic
CC   sequence, with 40 biallelic markers being located in exonic regions. Six
CC   of these alter the amino acid sequence of a PCTA-1 protein. 43 markers
CC   were found in the 5' regulatory region, including 16 in the promoter, 39
CC   in intronic sequences, and 3 in the 3' regulatory region. Alleles
CC   containing certain biallelic markers are associated with prostate cancer.
CC   These polymorphisms can be used as the basis of methods for determining
CC   genetic predisposition to prostate cancer, for diagnosis of this cancer
CC   and for the prognosis/detection of an eventual treatment response to
CC   therapeutic agents against prostate cancer. They can be used to design of
CC   drug screening protocols to provide accurate and efficient evaluation of
CC   therapeutic and side-effect potential of new or already existing
CC   medicaments, as well as for screening molecules which modulate or inhibit
CC   the expression of the PCTA-1 gene, preferably those active against
CC   prostate cancer
XX
SQ   Sequence 106746 BP; 29695 A; 23487 C; 23529 G; 30029 T; 0 U; 6 Other;

  Query Match             5.1%;  Score 253;  DB 3;  Length 106746;
  Best Local Similarity   69.4%;  Pred. No. 2e-34;
  Matches  416;  Conservative    0;  Mismatches  175;  Indels    8;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              |||||||||||||||||  ||||||||||||||||| ||||||||||||||  ||||||
Db      38717 AGTTCAAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAAAATT 38658

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   |||||||||| | ||||||||||||||||||||||  |||| | ||||||||||||
Db      38657 AGCTGGGCATGGTGGTGCACGCCTGTAATCCCAGCTACTCAGGAGACTGAGGCAGGAGAA 38598

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCCA 178
              |||||||||||||  || ||||||| ||| || ||| ||| |   | ||| |||||||||
Db      38597 TCACTTGAACCTG--GGAGGTGGAGTTTGTGGTGAGCCAAAACTTTCCCACTGCACTCCA 38540
```

```
Qy        179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCAT-TTTATATATAT 237
              |||      |   | |   | || |   | |  | |      | |   ||    | ||||
Db      38539 GCCAGTATATATATATATATATATACACACGTACACGTGTACACATATACACACATATGC 38480

Qy        238 ATATATATATATATACACACACACACATATATATATACACATATATATACACATATAT 297
              | |  | || ||| ||||||| |||||  | ||| || ||||||| || |  | || ||
Db      38479 ACACGTGTACATACACACACATACACATGCACATACATGCACATATGTACATGCGTACAT 38420

Qy        298 ACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACAC 357
              ||| | ||  |||||||||||| || || | |  ||||    |||||| || || | |
Db      38419 ACACACATGTACATATATACACACATGTACATATGTACACCCATACACACATGTACATAT 38360

Qy        358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417
              ||     | ||| | ||  |||||| ||||| |||||| | |  || |||   |||
Db      38359 GTACACCCATACACACATGTACATATGTACACCCATATACATATGTACATATGTACACCC 38300

Qy        418 ATACACATATATACACATATATACACATATACACATATATACACATATATACATATATAC 477
              |||||||||| |||| || || || |||| |||  || |||   | |   | | | ||
Db      38299 ATACACATATGTACATATGTACACCCATACACATGTACATATGTACACCCATACACATGT 38240

Qy        478 ACATATATATAATATACACACATATATATACACATATATACACACATATATACACATATA 537
              |||||| |   | || |||| || || |||   ||||  | ||||||| |||| |||||
Db      38239 ACATATGT---ACATCCACATATGTACATATGTATATCCATACACATATGTACATATATA 38183

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATACATAT 596
              ||||||||  || | |  || ||| | || | |  |||| ||   |||||||||   ||
Db      38182 CACACATATGTACATATGTACATATATATTTGTGTATATCTATGTATATATATAAGAAT 38124
```

RESULT 55
AAV83940/c
ID AAV83940 standard; DNA; 80240 BP.
XX
AC AAV83940;
XX
DT 03-MAR-1999 (first entry)
XX
DE NC-contig derived from mardel(10) on chromosome 10q25.2.
XX
KW Yeast artificial chromosome; YAC; probe; eukaryotic chromosome;
KW neocentromere; replication; extra-chromosomal element; segregation;
KW cell division; artificial chromosome; gene therapy; mardel(10);
KW human artificial chromosome; transgenic; chromosome 10; 10q25.2; ss.
XX
OS Homo sapiens.
XX
PN WO9851790-A1.
XX
PD 19-NOV-1998.
XX
PF 13-MAY-1998; 98WO-AU000352.
XX
PR 13-MAY-1997; 97AU-00006784.
PR 26-AUG-1997; 97AU-00008791.
XX
PA (AMRA-) AMRAD OPERATIONS PTY LTD.
XX
PI Choo K, Du Sart D, Cancilla MR;
XX
DR WPI; 1999-009773/01.
XX
PT New isolated nucleic acid comprising neocentromere sequences from
PT eukaryotic chromosome - used to produce replicable, segregating
PT artificial chromosomes that can carry large amounts of DNA for gene
PT therapy.
XX
PS Claim 9; Fig 16A; 540pp; English.
XX
CC The present sequence represents the NC-contig derived from a mutated
CC human chromosome 10, 10q25.2 region. The sequence contains an unusual
CC chromosomal marker referred to as mardel(10). The mardel(10) marker is
CC mitotically stable and contains a functional neocentromere at a location
CC regarded as non-centromeric. This neocentromere maps to q25.2 on
CC chromosome 10. The specification describes nucleic acid sequences derived
CC from a eukaryotic chromosome, including a neocentromere or its functional
CC derivative or hybrid, that are able, in a compatible cell, of
CC replicating, acting as extra-chromosomal element and segregating during
CC cell division. The sequences can be used to construct artificial
CC chromosomes for use in gene therapy comprising a replicable, segregating
CC nucleic acid that confers a specific phenotype on cells. Human artificial
CC chromosomes can propagate in human cells and carry large amounts of DNA
CC (e.g. therapeutic genes), and, being extra-chromosomal, they are not
CC mutagenic. The artificial chromosomes are also useful for generation of
CC transgenic plants and animals, in production of proteins and to make
CC diagnostic reagents, e.g. for expression of cytokines, receptors and
CC growth factors, or to increase the copy number of a gene in a cell. The
CC constructs may also be used for functional and structural analysis of
CC chromosomes
XX
SQ Sequence 80240 BP; 23102 A; 16537 C; 16747 G; 23846 T; 0 U; 8 Other;

Query Match 5.1%; Score 252.6; DB 2; Length 80240;
Best Local Similarity 71.9%; Pred. No. 2.3e-34;
Matches 360; Conservative 0; Mismatches 134; Indels 7; Gaps 2;

```
Qy        227 TTTATATATATATATATATATATATACACACACACACATATATATATACACATATATA 286
              | || |||| ||| | |||||| | ||| | | ||| |  ||| | |||| ||| | |||
Db      15787 TATACATATTTATGTCTATATACACACATATATACATATTTATGTCTATATACACACATA 15728

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              |||| || |||  |||||||||||||||||| || ||| |  | ||||| || || |||
Db      15727 TACATATTTATGTCTATATACACATATATACATATTTATGTCTATATACACATATATACA 15668

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              ||| |||    ||||  | | | ||| |||   || | || | ||| ||||| ||||||
Db      15667 TATTTATGTCTATATACACATATATACATATTTATGTCTATATACACATATATACATATT 15608

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              | |    |||||||||||||||||| || |||    |||||||||||||||| || ||
Db      15607 T--ATGTCTATATACACATATATACATATTTATGTCTATATACACATATATACATATTTA 15550

Qy        467 TACATATATACACATATATA-----TAATATACACACATATATATACACATATATACACA 521
              |   ||||||||||||||||     | || |  | | | | ||||| |||||| ||
Db      15549 TGTCTATATACACATATATACATATTTATGTCTATATACACATATATACATATTTATGTC 15490

Qy        522 CATATATACACATATATACACATATATATACACATATATACACATATATACATATATACA 581
              |||||| ||| ||| |||    || | |||| ||| ||||| |||||| ||   | |||| |
Db      15489 TATATACACATATACATATTTATGTCTATATACACATATATACATATTTATGTCTATATA 15430

Qy        582 CATATATATACATATATACACATATATACATATATACACATATATACATATATACACACA 641
              || |||||||||||| ||    |||||||| ||||| |||||| ||  |||||| |||||
Db      15429 CACATATATACATATTTATGTCTATATACACATATATACATATTTATGTATATATACACA 15370

Qy        642 TATATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTA 701
              |||||||| ||  || |  | | ||| | |||||||||| |||  | |||| |||   ||
Db      15369 TATATACATATTTATGTCTATATACACATATATACATATTTATGTATATATTTATGTATA 15310

Qy        702 TATATATGTATATATATATGC 722
              ||| | ||||||| || || |
Db      15309 TATGTGTGTATATGTACATAC 15289
```

RESULT 56
ABV72623/c
ID ABV72623 standard; DNA; 82938 BP.
XX
AC ABV72623;
XX
DT 27-NOV-2002 (first entry)
XX
DE Human transporter protein encoding gene.
XX
KW Human; transporter protein; hTP; gamma-aminobutyric-acid; GABA;
KW gene therapy; protein therapy; vaccine; gene; ds.
XX

```
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1120. .81932
FT                   /*tag= a
FT                   /product= "Transporter protein"
FT                   /note= "Contains 8 introns"
FT   exon            1120. .1223
FT                   /*tag= b
FT                   /number= 1
FT   intron          1224. .27733
FT                   /*tag= c
FT                   /number= 1
FT   variation       replace(2419,C)
FT                   /*tag= d
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(3191,A)
FT                   /*tag= e
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(4195,T)
FT                   /*tag= f
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(5910,C)
FT                   /*tag= g
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(7119,T)
FT                   /*tag= h
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(7529,G)
FT                   /*tag= h
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(8855,C)
FT                   /*tag= i
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(8956,A)
FT                   /*tag= j
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(10220,G)
FT                   /*tag= k
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(11628,C)
FT                   /*tag= l
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(11670,A)
FT                   /*tag= m
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(11671,G)
FT                   /*tag= n
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(11931,C)
FT                   /*tag= o
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(15984,G)
FT                   /*tag= p
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(17735,A)
FT                   /*tag= q
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(17910,A)
FT                   /*tag= r
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(18130,C)
FT                   /*tag= s
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19293. .19294,TWA)
FT                   /*tag= t
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19294. .19295,ATT)
FT                   /*tag= u
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19296. .19297,CYT)
FT                   /*tag= v
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19303. .19304,TYC)
FT                   /*tag= w
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19304. .19305,CYC)
FT                   /*tag= x
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19307. .19308,TMC)
FT                   /*tag= y
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19308. .19309,CCC)
FT                   /*tag= z
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19313. .19314,TMA)
FT                   /*tag= aa
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19314. .19315,AWT)
FT                   /*tag= ab
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19315. .19316,TCC)
FT                   /*tag= ac
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19316. .19317,CYT)
FT                   /*tag= ad
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19317. .19318,TWA)
FT                   /*tag= ae
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19318. .19319,ATT)
FT                   /*tag= af
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19860,G)
FT                   /*tag= ag
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(20843,G)
FT                   /*tag= ah
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(21144,T)
FT                   /*tag= ai
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(21371,G)
FT                   /*tag= aj
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(22061. .22062,A)
FT                   /*tag= ak
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(22321,T)
FT                   /*tag= al
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(22470. .22471,A)
FT                   /*tag= am
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(22472. .22473,A)
FT                   /*tag= an
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(24990,C)
FT                   /*tag= ao
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(25284,A)
FT                   /*tag= ap
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(25934,G)
FT                   /*tag= aq
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(27509. .27510,TAT)
FT                   /*tag= ar
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(27513. .27514,TTT)
FT                   /*tag= as
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            27734. .27882
FT                   /*tag= at
FT                   /number= 2
FT   intron          27883. .41045
```

```
FT                   /*tag=  au
FT                   /number= 2
FT   variation       replace(28608,G)
FT                   /*tag=  av
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(28627,C)
FT                   /*tag=  aw
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(29709,T)
FT                   /*tag=  ax
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(29788,T)
FT                   /*tag=  ay
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(29878,T)
FT                   /*tag=  az
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(30294,T)
FT                   /*tag=  ba
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(30896. .30897,T)
FT                   /*tag=  bb
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(31815,G)
FT                   /*tag=  bc
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(31999,TTT)
FT                   /*tag=  bd
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(32026,T)
FT                   /*tag=  be
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(32146,A)
FT                   /*tag=  bf
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(32184,S)
FT                   /*tag=  bg
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(32214. .32215,T)
FT                   /*tag=  bh
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(32215,C)
FT                   /*tag=  bi

  Query Match             5.1%;  Score 252.6;  DB 6;  Length 82938;
  Best Local Similarity   70.7%;  Pred. No. 2.3e-34;
  Matches  350;  Conservative    0;  Mismatches  144;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | |||| | |||| |||| |||||||||| | | |    | |||  ||||||    ||||||
Db       7454 TATATACACATATGTATACATATATACATATATATGTATATACGTATATAGCTGTATATA 7395

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || ||| || || |||  |  | ||||| | |||| |||||||| |||||||| |||| |
Db       7394 TATACACATGTATATAGCTGTATATATACATACATGTATATACATATACATATATACATA 7335

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              || | ||| |||||  | || | | ||||||| ||||| | |||  |||||||| |||||
Db       7334 TACACATATACATACATGTATATACATATACATATATACATACATGTATATACATATATA 7275

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              | |   ||| ||    || |||| |||||||||||| || || |||| | ||   |||||
Db       7274 CATATGTATACATGTGTATCTATATACATATATACATATGTATACATGTGTATCTATATA 7215

Qy        467 TACATATATACACATATATAT-AATATACACACATATATATACACATATATACACACATA 525
              | ||||| | |  |||| ||   |||  | | | ||||||||| || ||||||    ||
Db       7214 CATATATACATATGTATACATGTGTATCTATATACATATATACATATGTATACATGTGTA 7155

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              | || | | ||||| |||||| ||||||  | |||  | ||| ||| ||| |||   |||
Db       7154 TCTATATACATATATACATATGTATACATGTGTATCTATATACATATATACATATGTATA 7095

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
```

```
              || |  || |||| |||||||||||||| || |||| | ||  |||||||| | ||| |
Db       7094 CATGTGTATCTATATACATATATACATATGTATACATGTGTATCTATATACATATATACA 7035

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              ||   |||||| |  |   | ||| ||||| | |||| || |||| | | || |||||  |
Db       7034 TATGTATACATGTGTATCTATATACATATATACATATGTATACATGTGTGTATGTATACA 6975

Qy        706 TATGTATATATATAT 720
              ||||| | | |||||
Db       6974 TATGTGTGTGTATAT 6960


RESULT 57
ADC87564/c
ID   ADC87564 standard; DNA; 1114 BP.
XX
AC   ADC87564;
XX
DT   01-JAN-2004  (first entry)
XX
DE   Human GPCR gene SEQ ID NO:2017.
XX
KW   ds; gene; human; GPCR;
KW   guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1270724-A2.
XX
PD   02-JAN-2003.
XX
PF   18-JUN-2002; 2002EP-00013517.
XX
PR   18-JUN-2001; 2001JP-00246789.
XX
PA   (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA   (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI   Suwa M,  Asai K,  Akiyama Y,  Aburatani H;
XX
DR   WPI; 2003-315783/31.
DR   P-PSDB; ADC87565.
XX
PT   New polynucleotide, useful for preparing a composition for treating a
PT   patient in need of increased or suppressed activity or expression of the
PT   guanosine triphosphate-binding protein coupled receptor.
XX
PS   Claim 1; SEQ ID NO 2017; 28pp; English.
XX
CC   The invention relates to a novel polynucleotide encoding a guanosine
CC   triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC   the invention may have a use in gene therapy. The polynucleotide and
CC   polypeptide are useful for preparing a composition for treating a patient
CC   in need of increased or suppressed activity or expression of the
CC   guanosine triphosphate-binding protein coupled receptor. The
CC   polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC   invention.
XX
SQ   Sequence 1114 BP; 398 A; 107 C; 167 G; 442 T; 0 U; 0 Other;

  Query Match             5.1%;  Score 251;  DB 10;  Length 1114;
  Best Local Similarity   63.7%;  Pred. No. 3.5e-34;
  Matches  397;  Conservative    0;  Mismatches  225;  Indels    1;  Gaps    1;

Qy        230 ATATATATATATATATATATATACACACACACACATATATATATACACATATATATAC 289
              ||| | |||||| ||||||| |  | | ||| |||||||||| |||| |  | |||||||
Db        864 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 805

Qy        290 ACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATAT 349
              | | | ||| || |||| |  | |||| ||||| | ||| |  || |||| |  | |||
Db        804 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 745

Qy        350 ATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATAC 409
```

```
                ||| ||| ||||||||| |  | |||| ||||| | | | |  ||||||    |||||||
Db        744 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 685

Qy        410 ACACATATATACACATATATACACATATATACACATATACACATATATACACATATATAC 469
              | ||| |||||   ||||||   ||||| | ||| ||| |  ||||||    |||||||
Db        684 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 625

Qy        470 ATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATATAT 528
              ||| | | | || ||||||  | || | ||||| | ||| |  ||||||     |||||
Db        624 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 565

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACATATAT 588
              | ||| |||||   ||||||  |  | ||| ||||||||||||  ||||||    |||||
Db        564 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 505

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
              ||||| |||||   ||||||   ||||||| | | ||| || |||| |    |||||||
Db        504 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 445

Qy        649 ACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATAT 708
              | | |||||||   | | ||   |||| | ||| | | | || ||||||  | ||||||
Db        444 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 385

Qy        709 GTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAG 768
               || | |||||||      |  |||     ||  |   ||    |  | ||| |
Db        384 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 325

Qy        769 GAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAG 828
              | | | | |  | |    |||||    | |   | || |      |  | | |||
Db        324 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 265

Qy        829 TAGGGGAAATACATAAATGTGTA 851
                | ||||||| |||| ||
Db        264 ATACACATATACATATATGTATA 242
```

RESULT 58
ADE95974
ID ADE95974 standard; DNA; 96594 BP.
XX
AC ADE95974;
XX
DT 12-FEB-2004 (first entry)
XX
DE Human SYK gene genomic DNA sequence.
XX
KW cancer diagnosis; cancer treatment; carcinoma; cytostatic; gene therapy;
KW lymphoma; breast cancer; prostate cancer; leukaemia; ds; human; SYK.
XX
OS Homo sapiens.
XX
PN WO2003039484-A2.
XX
PD 15-MAY-2003.
XX
PF 08-NOV-2002; 2002WO-US036071.
XX
PR 08-NOV-2001; 2001US-00052482.
XX
PA (SAGR-) SAGRES DISCOVERY.
XX
PI Morris DW, Engelhard EK;
XX
DR WPI; 2003-441462/41.
XX
PT New carcinoma associated nucleic acids and proteins, useful for screening
PT drug candidates, or for diagnosing and treating carcinomas, e.g.
PT lymphoma, breast cancer, prostate cancer or leukemia.
XX
PS Claim 1; SEQ ID NO 232; 793pp; English.
XX
CC This invention relates to novel recombinant nucleic acids for use in
CC diagnosis and treatment of cancer, especially carcinomas, as well as the
CC use of compositions in screening methods. The compositions of the
CC invention may have cytostatic activity whilst the disclosed sequences may
CC be useful for gene therapy. The carcinoma associated nucleic acids and
CC proteins are useful for diagnosing and treating carcinomas, for example
CC lymphoma, breast cancer, prostate cancer or leukaemia, or for screening
CC drug candidates or bioactive agents capable of binding to, or modulating
CC the activity of, a carcinoma associated protein. The present sequence is
CC the genomic DNA sequence of the human SYK gene which is a carcinoma
CC associated gene of the invention.
XX
SQ Sequence 96594 BP; 27524 A; 20558 C; 21159 G; 26914 T; 0 U; 439 Other;

  Query Match   5.1%; Score 251; DB 10; Length 96594;
  Best Local Similarity   69.5%; Pred. No. 4.4e-34;
  Matches 354; Conservative   0; Mismatches 152; Indels   3; Gaps   1;

```
Qy        229 TATATATATATATATATATATATACACACACACACATATATATATACACATATATATA 288
              || ||| | |||| | |||||||||| ||| | |||||||| | | |||| ||| ||| |
Db      90981 TACATACACATATGTGTATATATACATACATATACACATATGTGTGTACATATACATACA 91040

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              || | | |||  | | |||| ||| ||||||| | ||| ||| ||   ||  |  |||||
Db      91041 CATACACATATGTGTGTACATATACATACACACACATACACATATGTGTACATGTACATA 91100

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              ||| ||| | |  |||    | || |||| | ||| ||| ||| ||||    | || ||
Db      91101 CATACACATACACATATGTGTACATGTACATACATACACATACACATATGTGTACATGTA 91160

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              || ||||| | | ||| ||||    | || |||| |||||||||||||| ||| || | ||
Db      91161 CATACATACACATACACATATGTGTACATGTACATATATACACATATACACATATGTGTA 91220

Qy        469 CATATATACACATATATATAATATACACACATATATATACACATATATACACACATATAT 528
              ||| || | | ||| |   ||||||||| || | || |   | ||||| |||||||||
Db      91221 CATGTACATATATACAC---ATATACACATATGTGTACATGTACATATATACACATATAC 91277

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACATATAT 588
              ||| || | ||||  || |||||||||||||| ||| || | ||||| || | | | ||
Db      91278 ACATATGTGTACATGTACATATACACATATACACATATGTGTACATGTACATATACACAT 91337

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
              ||||| ||||    | ||  ||||||| |||  | || |   | ||| |||||||| |
Db      91338 ATACACATATGTGTACATGCACATATACACATGTGTACATGCACATATACACATATGTGT 91397

Qy        649 ACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATAT 708
              ||||||| ||| ||||| ||   | || ||| |||||||||  |||| |  |||||| ||
Db      91398 ACATACACATATACACATATGTGTGTATATACATATACACACATATGTGTGTATATACAT 91457

Qy        709 GTATATATATATGCTCCAGAGTTCATAAG 737
              ||||  ||||||  |       | |||| |
Db      91458 ATATACATATATACATATATATACATATG 91486
```

RESULT 59
ADA02726
ID ADA02726 standard; DNA; 96595 BP.
XX
AC ADA02726;
XX
DT 06-NOV-2003 (first entry)
XX
DE Human SYK carcinoma associated gene, SEQ ID NO:1244.
XX
KW Human; carcinoma associated; oncogene; carcinoma; cancer; breast;
KW prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO2003057146-A2.
XX
PD 17-JUL-2003.
XX

```
PF   26-DEC-2002; 2002WO-US041414.
XX
PR   26-DEC-2001; 2001US-00035832.
XX
PA   (SAGR-) SAGRES DISCOVERY.
XX
PI   Morris DW;
XX
DR   WPI; 2003-587068/55.
XX
PT   New recombinant nucleic acid encoding carcinoma associated protein,
PT   useful for preparing compositions for treating carcinomas.
XX
PS   Claim 1; SEQ ID NO 1244; 245pp; English.
XX
CC   The invention relates to recombinant carcinoma associated (CA) nucleic
CC   acid sequences from mouse and human (ADA01482-ADA03094), and to
CC   recombinant carcinoma associated proteins (CAP) encoded by them. The
CC   invention also encompasses expression vectors and host cells comprising a
CC   CA nucleic acid, a polypeptide (especially an antibody) that specifically
CC   binds to the protein, and a biochip comprising CA nucleic acid or
CC   fragments thereof. The sequences of the invention were identified using
CC   oncogenic retroviruses, which insert into the genome of the host organism
CC   at random. Many of these do not carry transduced host oncogenes or
CC   pathogenic trans-acting viral genes, meaning that cancer incidence is a
CC   direct consequence of the effects of proviral integration into host
CC   protooncogenes. The CA nucleic acid sequences can be used to diagnose
CC   carcinoma (especially breast cancer, prostate cancer, lymphoma or
CC   leukaemia) or a propensity to carcinoma by determination of the sequence
CC   of a CA gene, or by determination of CA gene expression in particular
CC   tissues. CA nucleic acids, proteins and antibodies are also useful as
CC   therapeutic agents and in screening and evaluating drug candidates. The
CC   present sequence represents a specifically claimed human CA nucleic acid
CC   sequence of the invention. Note: The complete sequence data for this
CC   patent did not form part of the printed specification, but was obtained
CC   in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 96595 BP; 27524 A; 20559 C; 21158 G; 26915 T; 0 U; 439 Other;

  Query Match             5.1%;  Score 251;  DB 9;  Length 96595;
  Best Local Similarity   69.5%;  Pred. No. 4.4e-34;
  Matches  354;  Conservative    0;  Mismatches  152;  Indels    3;  Gaps    1;

Qy        229 TATATATATATATATATATATATATACACACACACACACATATATATATACACATATATATA 288
              || ||| | |||| | |||||||||| ||| | |||||||| | | |||| ||| ||| |
Db      90982 TACATACACATATGTGTATATATACATACATATACACATATGTGTGTACATATACATACA 91041

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              || | | |||  | | |||| ||| ||||||| | ||| ||| ||   ||  |  |||||
Db      91042 CATACACATATGTGTGTACATATACATACACACACATACACATATGTGTACATGTACATA 91101

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
               ||| ||| | |  |||    | || |||| | ||| ||| ||| ||||    | || ||
Db      91102 CATACACATACACATATGTGTACATGTACATACATACACATACACATATGTGTACATGTA 91161

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              || ||||| | | ||| ||||    | || |||| |||||||||||||| ||| || | ||
Db      91162 CATACATACACATACACATATGTGTACATGTACATATATACACATATACACATATGTGTA 91221

Qy        469 CATATATACACATATATATAATATACACACATATATATACACATATATACACACATATAT 528
              ||| || | | ||| |    |||||||| || | ||  |  | ||||| |||||||||
Db      91222 CATGTACATATATACAC---ATATACACATATGTGTACATGTACATATATACACATATAC 91278

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACATATAT 588
              ||| || | ||||  || ||||||||||||||| ||| || | ||||| || | | | | ||
Db      91279 ACATATGTGTACATGTACATATACACATATACACATATGTGTACATGTACATATACACAT 91338

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
              |||||  ||||    | ||  ||||||| |||  | || |   | ||| ||||||||  |
Db      91339 ATACACATATGTGTACATGCACATATACACATGTGTACATGCACATATACACATATGTGT 91398

Qy        649 ACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATAT 708
```

```
              ||||||| ||| ||||| ||   | || ||| |||||||| |||| |  ||||||| ||
Db      91399 ACATACACATATACACATATGTGTGTATATACATATACACACATATGTGTGTATATACAT 91458

Qy        709 GTATATATATATGCTCCAGAGTTCATAAG 737
              |||| |||||| |       | |||| |
Db      91459 ATATACATATATACATATATATACATATG 91487


RESULT 60
ADB72464
ID   ADB72464 standard; DNA; 96595 BP.
XX
AC   ADB72464;
XX
DT   04-DEC-2003  (first entry)
XX
DE   Human SYK gene.
XX
KW   human; ds; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;
KW   cancer; neoplasm; adenocarcinoma; sarcoma; gene.
XX
OS   Homo sapiens.
XX
PN   WO2003008583-A2.
XX
PD   30-JAN-2003.
XX
PF   26-DEC-2001; 2001WO-US051291.
XX
PR   02-MAR-2001; 2001US-00798586.
PR   23-OCT-2001; 2001US-00004113.
PR   08-NOV-2001; 2001US-00052482.
PR   30-NOV-2001; 2001US-00997722.
PR   20-DEC-2001; 2001US-00034650.
XX
PA   (SAGR-) SAGRES DISCOVERY.
XX
PI   Morris DW,  Engelhard EK;
XX
DR   WPI; 2003-239337/23.
XX
PT   New recombinant nucleic acid, useful for treating carcinomas, lymphomas,
PT   cancers, neoplasm, adenocarcinoma, or sarcomas.
XX
PS   Claim 1; SEQ ID NO 292; 2304pp; English.
XX
CC   The invention relates to a novel recombinant nucleic acid comprising a
CC   nucleotide sequence selected from any of the 660 sequences fully defined
CC   in the specification. A polynucleotide of the invention has cytostatic
CC   activity, and may have a use in gene therapy, or in a vaccine. The
CC   recombinant nucleic acids and polypeptides are useful for treating
CC   carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and
CC   sarcomas. The present sequence represents a human gene of the invention.
XX
SQ   Sequence 96595 BP; 27524 A; 20559 C; 21158 G; 26915 T; 0 U; 439 Other;

  Query Match             5.1%;  Score 251;  DB 10;  Length 96595;
  Best Local Similarity   69.5%;  Pred. No. 4.4e-34;
  Matches  354;  Conservative    0;  Mismatches  152;  Indels    3;  Gaps    1;

Qy        229 TATATATATATATATATATATATATACACACACACACACATATATATATACACATATATATA 288
              || ||| | |||| | |||||||||| ||| | |||||||| | | |||| ||| ||| |
Db      90982 TACATACACATATGTGTATATATACATACATATACACATATGTGTGTACATATACATACA 91041

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              || | | |||  | | |||| ||| ||||||| | ||| ||| ||   ||  |  |||||
Db      91042 CATACACATATGTGTGTACATATACATACACACACATACACATATGTGTACATGTACATA 91101

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
               ||| ||| | |  |||    | || |||| | ||| ||| ||| ||||    | || ||
Db      91102 CATACACATACACATATGTGTACATGTACATACATACACATACACATATGTGTACATGTA 91161

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
```

```
Db      91162 CATACATACACATACACATATGTGTACATGTACATATATACACATATACACATATGTGTA 91221

Qy        469 CATATATACACATATATATAATATACACACATATATATACACATATATACACACATATAT 528
              ||| || | | ||| |    |||||||||| || | || |    | ||||| |||||||||
Db      91222 CATGTACATATATACAC---ATATACACATATGTGTACATGTACATATATACACATATAC 91278

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACATATAT 588
              ||| || | ||||  || ||||||||||||||| ||| || | ||||| || | | | | ||
Db      91279 ACATATGTGTACATGTACATATACACATATACACATATGTGTACATGTACATATACACAT 91338

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
              ||||| ||||    | ||  ||||||| |||  | || |   | ||| |||||||| |
Db      91339 ATACACATATGTGTACATGCACATATACACATGTGTACATGCACATATACACATATGTGT 91398

Qy        649 ACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATAT 708
              ||||||| ||| ||||| ||   | || ||| |||||||||| |||| |  |||||| ||
Db      91399 ACATACACATATACACATATGTGTGTATATACATATACACACATATGTGTGTATATACAT 91458

Qy        709 GTATATATATATGCTCCAGAGTTCATAAG 737
              |||| ||||||| |       | |||| |
Db      91459 ATATACATATATACATATATATACATATG 91487


RESULT 61
ADQ97712/c
ID   ADQ97712 standard; DNA; 103661 BP.
XX
AC   ADQ97712;
XX
DT   07-OCT-2004  (first entry)
XX
DE   Mouse cancer associated sequence MD10-032, SEQ ID 689.
XX
KW   Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Mouse; ds.
XX
OS   Mus musculus.
XX
PN   WO2004060304-A2.
XX
PD   22-JUL-2004.
XX
PF   22-DEC-2003; 2003WO-US041389.
XX
PR   27-DEC-2002; 2002US-00330773.
XX
PA   (SAGR-) SAGRES DISCOVERY INC.
XX
PI   Morris DW,  Malandro MS;
XX
DR   WPI; 2004-543781/52.
XX
PT   New isolated cancer associated nucleic acids comprising at least 10
PT   contiguous nucleotides, useful for diagnosing, preventing and/or treating
PT   cancers such as leukemia and lymphoma.
XX
PS   Claim 1; SEQ ID NO 689; 199pp; English.
XX
CC   The present invention relates to cancer associated sequences (ADQ97025-
CC   ADQ98004). The sequences are useful for the diagnosis, prevention and/or
CC   treatment of cancer, such as leukemia and lymphoma. Note: The sequence
CC   data for this patent did not form part of the printed specification, but
CC   was obtained in electronic formate directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 103661 BP; 25663 A; 25358 C; 25585 G; 26717 T; 0 U; 338 Other;

  Query Match             5.1%;  Score 251;  DB 12;  Length 103661;
  Best Local Similarity   72.6%;  Pred. No. 4.5e-34;
  Matches  339;  Conservative    0;  Mismatches  125;  Indels    3;  Gaps    1;

Qy        230 ATATATATATATATATATATATACACACACACACATATATATATACACATATATATAC 289
              | | | | | | | | | | ||| | |||||||||| | | | | ||||| | | ||||

Db      73981 ACACACACACACACACACACACACATATACACACACACACACACACACACACACATAC 73922

Qy        290 ACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATAT 349
              ||| | ||| | | | ||||| | | ||||||| | | ||||| ||| |   ||||| ||
Db      73921 ACACACATATACACACACACACACACACACATACACACACACACACACACACACACAT 73862

Qy        350 ATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATAC 409
              ||| ||||| |   | ||||| | | ||| ||| | ||||||| | | ||||| ||||||
Db      73861 ATACACACACACACACACACACACACACATATACACACACACACACACACACATATAC 73802

Qy        410 ACACATATATACACATATATACACATATATACACATATACACATATATACACATATATAC 469
              ||||| | | ||||| | | ||||| ||||||||| | ||||| | | ||||||| | ||
Db      73801 ACACACACACACACACACACACACACATATACACACACACACACACACACATACACAC 73742

Qy        470 ATATATACACATATATATA---ATATACACACATATATATACACATATATACACACATAT 526
              | | | ||||| | |||||   | | ||||||| | | | ||||||||| ||||||| |
Db      73741 ACACACACACACACATATACACACACACACACACACACACACACATATACACACACAC 73682

Qy        527 ATACACATATATACACATATATATACACATATATACACATATATACATATATACACATAT 586
              | ||||| ||| ||||| ||| | ||||| | | ||||||| | ||| | | ||||| |
Db      73681 ACACACACATACACACACATACACACACACACACACATACACACACACACACACAC 73622

Qy        587 ATATACATATATACACATATATACATATATACACATATATACATATATACACACATATAT 646
              ||| ||| | | ||||||| | ||| |||||||||| | | ||| | | ||||||| ||||
Db      73621 ATACACACACACACACATACACACACATATACACACACACACACACACACACACATAT 73562

Qy        647 ACACATACATATACACACACATAGATATACATATATATACACATATA 693
              ||||| ||| | ||| ||||| | ||||||| | | | ||||| | |
Db      73561 ACACACACACACACATACACACACATATACACACACACACACACA 73515


RESULT 62
ABD32998/c
ID   ABD32998 standard; DNA; 113079 BP.
XX
AC   ABD32998;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Human cancer-associated genomic DNA HD21-021.
XX
KW   Human; ds; cancer-associated protein; gene; cytostatic; cancer;
KW   leukaemia; lymphoma; CAP.
XX
OS   Homo sapiens.
XX
PN   WO2004074320-A2.
XX
PD   02-SEP-2004.
XX
PF   17-FEB-2004; 2004WO-US004730.
XX
PR   14-FEB-2003; 2003US-00367094.
PR   14-MAR-2003; 2003US-00388838.
PR   15-APR-2003; 2003US-00417375.
PR   13-JUN-2003; 2003US-00461862.
PR   15-SEP-2003; 2003US-00663431.
PR   15-DEC-2003; 2003US-00737318.
XX
PA   (SAGR-) SAGRES DISCOVERY INC.
XX
PI   Morris DW,  Morris DW,  Malandro MS;
XX
DR   WPI; 2004-652914/63.
XX
PT   New isolated cancer-associated polynucleotides and polypeptides useful
PT   for diagnosing, preventing or treating cancers, especially lymphoma and
PT   leukemia, or in screening for agents that modulate cancer.
XX
PS   claim 16; seqid 764; 310pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising at least 10
CC   contiguous nucleotides of any of the 233 polynucleotide sequences given
```

CC in the specification, or its complement. The nucleic acids encode cancer-
CC associated proteins. Also included are an expression vector comprising
CC the isolated nucleic acid cited above, a host cell comprising the above
CC recombinant nucleic acid or expression vector, a microarray for detecting
CC a cancer-associated (CA) nucleic acid comprising at least one probe
CC comprising at least 10 contiguous nucleotides of any of the above-
CC mentioned nucleotide sequences, an isolated polypeptide (encoded within
CC an open reading frame of a CA sequence selected from any of the 95
CC polynucleotide sequences as mentioned in the specification, or its
CC complement), an isolated antibody, (or its antigen binding fragment) that
CC binds to the above polypeptide, a hybridoma that produces the above
CC monoclonal antibody, a pharmaceutical composition comprising the above
CC antibody and a pharmaceutical excipient, a kit for detecting cancer
CC cells(comprising the antibody cited above, methods for diagnosing cancer
CC or for detecting the presence or absence of cancer cells in an
CC individual, a method for inhibiting growth of cancer cells in an
CC individual, a method for delivering a therapeutic agent to cancer cells
CC in an individual, an electronic library comprising the above
CC polynucleotide or polypeptide (or their fragments), methods of screening
CC for anticancer activity or for a bioactive agent capable of modulating
CC the activity of a CA protein (CAP), methods for detecting cancer
CC associated with expression of a polypeptide in a test cell sample, a
CC method for treating cancers and a method for inhibiting the expression of
CC CA gene in a cell. The composition and methods are useful for detecting,
CC diagnosing, preventing and treating cancers, especially lymphoma and
CC leukaemia. These may also be used in screening for agents that modulate
CC cancer. The present sequence is a human CAP genomic sequence. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 113079 BP; 27712 A; 27684 C; 28156 G; 29527 T; 0 U; 0 Other;

Query Match 5.0%; Score 250.6; DB 13; Length 113079;
Best Local Similarity 73.3%; Pred. No. 5.3e-34;
Matches 335; Conservative 0; Mismatches 119; Indels 3; Gaps 1;

Qy 240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
Db 31278 ATATATATATATACAGACATATATATATATAGACATATATATATATATAGACATATATAT 31219

Qy 300 ATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACAT 359
Db 31218 ATATAGACATATATATATAGAGACATATATATATATAGACATATATATAGAGAGACATAT 31159

Qy 360 ATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATAT 419
Db 31158 ATATATAGACATATATATAGAGAGACATATATATAGAGACATATATATAGAGACATATAT 31099

Qy 420 ACACATATATACACATATATACACATATACACATATATACACATATATACATATATACAC 479
Db 31098 ATAGAGACATATATATAGACATATATATAGACATATATATAGACATATATATATAGACAT 31039

Qy 480 ATATATATA---ATATACACACATATATATACACATATATACACACATATATACACATAT 536
Db 31038 ATATAGACATATATATAGACATATATATAGACATATATATGCAGACATATATATAGACAT 30979

Qy 537 ATACACATATATATACACATATATACACATATATACATATATACACATATATATACATAT 596
Db 30978 ATATATATAGACATATATATGCAGACATATATATAGACATATATAGACATATATATAGAC 30919

Qy 597 ATACACATATATACATATATACACATATATACATATATACACACATATATACACATACAT 656
Db 30918 ATATATATATAGACATATATATAGACATATATATATAAACATATATATATATATATATAT 30859

Qy 657 ATACACACACATAGATATACATATATATACACATATA 693
Db 30858 ATATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA 30822

RESULT 63
ABL33999/c
ID ABL33999 standard; DNA; 21537 BP.
XX
AC ABL33999;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 1972.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 1972; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 21537 BP; 6538 A; 333 C; 4426 G; 10240 T; 0 U; 0 Other;

Query Match 5.0%; Score 250; DB 6; Length 21537;
Best Local Similarity 68.1%; Pred. No. 6.1e-34;
Matches 391; Conservative 0; Mismatches 180; Indels 3; Gaps 3;

Qy 181 CCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATA 240
Db 6005 CCAAATTAAAATACAATAACGCGATCTCGATTCACTACACTATATATATATATATATATA 5946

Qy 241 TATATATATATACACACACACACATATATATATACACATATATATACACATATATACA 300
Db 5945 TATATATATATATACACATACAC-CATATATATATATATATATATACATACATATACA 5887

Qy 301 TATATACACATATATACACATATATATACACATACATATGTACACATATATATACACATA 360
Db 5886 AAACAATTACTTACTAAATACACAAATCAATAAAATAAAAAATTCATAAACAAACATATC 5827

Qy 361 TGTATACACATATATACACATATATACACACATATATACACATATATACACACATATATA 420
Db 5826 TCTCTTTACATATATATATCTATATCGATCACACTATATATATATATATATATATATA 5767

Qy 421 CACATATATACACATATATACACATATACACATATATACACATATATACATATATACACA 480
Db 5766 CACACATATATATATATATATATACACACATATATATATATATATATACACACATATA 5707

```
Qy        481 TATATATAATATACACACATATATATACACATATATACACACATATATACACATATATAC 540
              ||||||| | |||||||||||||||||| | ||||||| ||||| ||||| | |||||||
Db       5706 TATATAT-ACATACACACATATATATATATATATATACACACATATATATATATATAT 5648

Qy        541 ACATATATATACACATATATACACATATATACATATATACACATATATATACATATATAC 600
              | ||||||||| | ||||||| | ||||||| ||||||| | ||||| |  ||||||||
Db       5647 ATATATATATATATATATATATATATATATATATATATATATATATATACACGCATATATAT 5588

Qy        601 ACATATATACATATATACACATATATACATATATACACACATATATACACATACATATAC 660
              || |||||||||||||| | |   ||| ||||||| | ||||||||| | ||| |||||
Db       5587 ACGTATATACATATATATAAACTCATATATATATATATACATATATATATATATATAT 5528

Qy        661 ACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATAT 720
              | | | ||| |   ||| |||||| |  | | | ||       || | | ||||  |  ||
Db       5527 ATATATATAAACTCACA-ATATATCCTTAAACAAATTATATAATCTCTATATACCTCAAT 5469

Qy        721 GCTCCAGAGTTCATAAGAGGTAGCAGTTGATTAC 754
                ||   | |  | || |   |  | |  | |||
Db       5468 TTTCTCAAATATAAAACAAAAATAATTATAATAC 5435
```

RESULT 64
ADA02930
ID ADA02930 standard; DNA; 96596 BP.
XX
AC ADA02930;
XX
DT 06-NOV-2003 (first entry)
XX
DE Human PIK3R1 carcinoma associated gene, SEQ ID NO:1448.
XX
KW Human; carcinoma associated; oncogene; carcinoma; cancer; breast;
KW prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO2003057146-A2.
XX
PD 17-JUL-2003.
XX
PF 26-DEC-2002; 2002WO-US041414.
XX
PR 26-DEC-2001; 2001US-00035832.
XX
PA (SAGR-) SAGRES DISCOVERY.
XX
PI Morris DW;
XX
DR WPI; 2003-587068/55.
XX
PT New recombinant nucleic acid encoding carcinoma associated protein,
PT useful for preparing compositions for treating carcinomas.
XX
PS Claim 1; SEQ ID NO 1448; 245pp; English.
XX
CC The invention relates to recombinant carcinoma associated (CA) nucleic
CC acid sequences from mouse and human (ADA01482-ADA03094), and to
CC recombinant carcinoma associated proteins (CAP) encoded by them. The
CC invention also encompasses expression vectors and host cells comprising a
CC CA nucleic acid, a polypeptide (especially an antibody) that specifically
CC binds to the protein, and a biochip comprising CA nucleic acid or
CC fragments thereof. The sequences of the invention were identified using
CC oncogenic retroviruses, which insert into the genome of the host organism
CC at random. Many of these do not carry transduced host oncogenes or
CC pathogenic trans-acting viral genes, meaning that cancer incidence is a
CC direct consequence of the effects of proviral integration into host
CC protooncogenes. The CA nucleic acid sequences can be used to diagnose
CC carcinoma (especially breast cancer, prostate cancer, lymphoma or
CC leukaemia) or a propensity to carcinoma by determination of the sequence
CC of a CA gene, or by determination of CA gene expression in particular
CC tissues. CA nucleic acids, proteins and antibodies are also useful as
CC therapeutic agents and in screening and evaluating drug candidates. The
CC present sequence represents a specifically claimed human CA nucleic acid
CC sequence of the invention. Note: The complete sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 96596 BP; 27639 A; 18627 C; 19687 G; 30643 T; 0 U; 0 Other;

```
  Query Match             5.0%;  Score 249.6;  DB 9;  Length 96596;
  Best Local Similarity   70.5%;  Pred. No. 7.8e-34;
  Matches  347;  Conservative    0;  Mismatches  144;  Indels    1;  Gaps    1;

Qy        229 TATATATATATATATATATATACACACACACACATATATATATACACATATATATA 288
              | | || ||||||| || || || ||   | | |||  | || ||||| |||| || ||
Db      57110 TGTGTACATATATACATGTAGATGCATGTAGATACATGTGTACATATATACATGTAGATG 57169

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              ||  || |||||| ||||||  || |||||  || ||| |   |||||| |  | ||||
Db      57170 CATGTAGATACATGTATACATGTAGATACATGTAGATACATGTATACATGTAGATACATG 57229

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              || |||||  |||  ||| | | || || | | || || | |||| |||||| ||||| |
Db      57230 TAGATACATGTATACATATATACATGTAGATACATGTAGATACATGTATACATATATACA 57289

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              | ||| ||  | | ||||| |||| || | |||| || ||||||||||  || |||
Db      57290 TGTAGATACATGTATACATATATACATGTAGATACATGTATACATATATACATGTAGATA 57349

Qy        469 CATATATACACATATATATAATATACACACATATATATACACATATATACACACATATAT 528
              ||| |||||| ||||| ||  || | |||  || ||| |   ||| ||| | |||| ||
Db      57350 CATGTATACATATATACAT-GTAGATACATGTAGATACATGTATACATATATACATGTAG 57408

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACATATAT 588
              | |||| || | |||| |||| | | ||| ||  | ||| ||  ||||| | ||||| ||
Db      57409 ATACATGTAGATACATGTATACATATATACATGTAGATACATGTATATAGATACATACAT 57468

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
              | | ||||| | ||||| |||||||||  | |||| |||||||||  ||| ||||||
Db      57469 AGATATATAGATACATAGATACATATATTTATATATGTACATATATCTACATATATATCT 57528

Qy        649 ACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATAT 708
              | ||| | ||| |   ||||| ||||  |||||||||  |  || |||||  | |||||||
Db      57529 ATATATACATATATCTACATATATATCTATATATATATATCTACATATATATCTATATAT 57588

Qy        709 GTATATATATAT 720
              ||| | ||||||
Db      57589 GTAGAAATATAT 57600
```

RESULT 65
ADB72668
ID ADB72668 standard; DNA; 96596 BP.
XX
AC ADB72668;
XX
DT 04-DEC-2003 (first entry)
XX
DE Human PIK3R1 gene.
XX
KW human; ds; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;
KW cancer; neoplasm; adenocarcinoma; sarcoma; gene.
XX
OS Homo sapiens.
XX
PN WO2003008583-A2.
XX
PD 30-JAN-2003.
XX
PF 26-DEC-2001; 2001WO-US051291.
XX
PR 02-MAR-2001; 2001US-00798586.
PR 23-OCT-2001; 2001US-00004113.

```
PR   08-NOV-2001; 2001US-00052482.
PR   30-NOV-2001; 2001US-00997722.
PR   20-DEC-2001; 2001US-00034650.
XX
PA   (SAGR-) SAGRES DISCOVERY.
XX
PI   Morris DW,  Engelhard EK;
XX
DR   WPI; 2003-239337/23.
XX
PT   New recombinant nucleic acid, useful for treating carcinomas, lymphomas,
PT   cancers, neoplasm, adenocarcinoma, or sarcomas.
XX
PS   Claim 1; SEQ ID NO 496; 2304pp; English.
XX
CC   The invention relates to a novel recombinant nucleic acid comprising a
CC   nucleotide sequence selected from any of the 660 sequences fully defined
CC   in the specification. A polynucleotide of the invention has cytostatic
CC   activity, and may have a use in gene therapy, or in a vaccine. The
CC   recombinant nucleic acids and polypeptides are useful for treating
CC   carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and
CC   sarcomas. The present sequence represents a human gene of the invention.
XX
SQ   Sequence 96596 BP; 27639 A; 18627 C; 19687 G; 30643 T; 0 U; 0 Other;

  Query Match             5.0%;  Score 249.6;  DB 10;  Length 96596;
  Best Local Similarity   70.5%;  Pred. No. 7.8e-34;
  Matches  347;  Conservative    0;  Mismatches  144;  Indels    1;  Gaps    1;

Qy        229 TATATATATATATATATATATATACACACACACACACATATATATATACACATATATATA 288
              | | || ||||||| || || || ||   | | |||  | || ||||| |||| || ||
Db      57110 TGTGTACATATATACATGTAGATGCATGTAGATACATGTGTACATATATACATGTAGATG 57169

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              || .|| |||||| ||||||  || |||||  || ||| |   |||||| |  | ||||
Db      57170 CATGTAGATACATGTATACATGTAGATACATGTAGATACATGTATACATGTAGATACATG 57229

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              || |||||  |||  ||| | | || || | | || || | |||| |||||| ||||| |
Db      57230 TAGATACATGTATACATATATACATGTAGATACATGTAGATACATGTATACATATATACA 57289

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              | ||| ||  | | ||||| |||| || | |||| || |||||||||||  || |||
Db      57290 TGTAGATACATGTATACATATATACATGTAGATACATGTATACATATATACATGTAGATA 57349

Qy        469 CATATATACACATATATATAATATACACACATATATATACACATATATACACACATATAT 528
              ||| |||||| ||||| ||  || | |||  || |||  |   ||| ||| | |||| ||
Db      57350 CATGTATACATATATACAT-GTAGATACATGTAGATACATGTATACATATATACATGTAG 57408

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACACATATAT 588
              | |||| || | |||| |||| | | ||| ||  | ||| ||  ||||| | ||||| ||
Db      57409 ATACATGTAGATACATGTATACATATATACATGTAGATACATGTATATAGATACATACAT 57468

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
              | | ||||| | ||||| ||||||||||  | |||| |||||||||  ||| ||||||
Db      57469 AGATATATAGATACATAGATACATATATTTATATATGTACATATATCTACATATATATCT 57528

Qy        649 ACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATAT 708
              | ||| | ||| |   ||||| ||||  |||||||| |  || |||||  | |||||||
Db      57529 ATATATACATATATCTACATATATATCTATATATATATATCTACATATATATCTATATAT 57588

Qy        709 GTATATATATAT 720
              ||| | ||||||
Db      57589 GTAGAAATATAT 57600


RESULT 66
ADC85410
ID   ADC85410 standard; DNA; 96596 BP.
XX
AC   ADC85410;
XX
DT   01-JAN-2004  (first entry)

XX
DE   Human Pik3r1 genomic sequence.
XX
KW   Cytostatic; gene therapy; vaccine; cancer; carcinoma-associated gene; CA;
KW   secreted; transmembrane; intracellular; ds.
XX
OS   Homo sapiens.
XX
PN   WO2003045230-A2.
XX
PD   05-JUN-2003.
XX
PF   02-DEC-2002; 2002WO-US038582.
XX
PR   30-NOV-2001; 2001US-00997722.
XX
PA   (SAGR-) SAGRES DISCOVERY.
XX
PI   Morris DW,  Engelhard EK;
XX
DR   WPI; 2003-513603/48.
XX
PT   New recombinant nucleic acid comprising a nucleotide sequence of any of
PT   the carcinoma-associated (CA) genes, useful for screening for drug
PT   candidates for diagnosing or treating carcinomas.
XX
PS   Claim 1; SEQ ID NO 196; 983pp; English.
XX
CC   The invention relates to a recombinant nucleic acid comprising a
CC   nucleotide sequence selected from any of the fully defined carcinoma-
CC   associated (CA) genes from the 50 tables given in the specification. The
CC   CA proteins are secreted, transmembrane or intracellular proteins. The
CC   recombinant nucleic acids are useful for screening for drug candidates
CC   for diagnosing or treating carcinomas. Sequences given in ADC85215-
CC   ADC85514 represent CA genes of the invention.
XX
SQ   Sequence 96596 BP; 27639 A; 18627 C; 19687 G; 30643 T; 0 U; 0 Other;

  Query Match             5.0%;  Score 249.6;  DB 10;  Length 96596;
  Best Local Similarity   70.5%;  Pred. No. 7.8e-34;
  Matches  347;  Conservative    0;  Mismatches  144;  Indels    1;  Gaps    1;

Qy        229 TATATATATATATATATATATATACACACACACACACATATATATATACACATATATATA 288
              | | || ||||||| || || || ||   | | |||  | || ||||| |||| || ||
Db      57110 TGTGTACATATATACATGTAGATGCATGTAGATACATGTGTACATATATACATGTAGATG 57169

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              || .|| |||||| ||||||  || |||||  || ||| |   |||||| |  | ||||
Db      57170 CATGTAGATACATGTATACATGTAGATACATGTAGATACATGTATACATGTAGATACATG 57229

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              || |||||  |||  ||| | | || || | | || || | |||| |||||| ||||| |
Db      57230 TAGATACATGTATACATATATACATGTAGATACATGTAGATACATGTATACATATATACA 57289

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              | ||| ||  | | ||||| |||| || | |||| || |||||||||||  || |||
Db      57290 TGTAGATACATGTATACATATATACATGTAGATACATGTATACATATATACATGTAGATA 57349

Qy        469 CATATATACACATATATATAATATACACACATATATATACACATATATACACACATATAT 528
              ||| |||||| ||||| ||  || | |||  || |||  |   ||| ||| | |||| ||
Db      57350 CATGTATACATATATACAT-GTAGATACATGTAGATACATGTATACATATATACATGTAG 57408

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACACATATAT 588
              | |||| || | |||| |||| | | ||| ||  | ||| ||  ||||| | ||||| ||
Db      57409 ATACATGTAGATACATGTATACATATATACATGTAGATACATGTATATAGATACATACAT 57468

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
              | | ||||| | ||||| ||||||||||  | |||| |||||||||  ||| ||||||
Db      57469 AGATATATAGATACATAGATACATATATTTATATATGTACATATATCTACATATATATCT 57528

Qy        649 ACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATAT 708
              | ||| | ||| |   ||||| ||||  |||||||| |  || |||||  | |||||||
Db      57529 ATATATACATATATCTACATATATATCTATATATATATATCTACATATATATCTATATAT 57588
```

```
Qy        709 GTATATATATAT 720
              ||| | ||||||
Db      57589 GTAGAAATATAT 57600


RESULT 67
ADM74525
ID   ADM74525 standard; DNA; 96596 BP.
XX
AC   ADM74525;
XX
DT   01-JUL-2004  (first entry)
XX
DE   Human carcinoma associated (CA) nucleic acid #97.
XX
KW   Human; carcinoma associated nucleic acid; CA nucleic acid; gene; ds;
KW   carcinoma associated protein; CAP; carcinoma; leukaemia; lymphoma;
KW   cytostatic.
XX
OS   Homo sapiens.
XX
PN   US2004072154-A1.
XX
PD   15-APR-2004.
XX
PF   30-NOV-2001; 2001US-00997722.
XX
PR   22-DEC-2000; 2000US-00747377.
PR   02-MAR-2001; 2001US-00798586.
XX
PA   (MORR/) MORRIS D W.
PA   (ENGE/) ENGELHARD E K.
XX
PI   Morris DW,  Engelhard EK;
XX
DR   WPI; 2004-328562/30.
XX
PT   New carcinoma associated gene or protein, useful for preparing a
PT   composition for diagnosing or treating carcinoma e.g., leukemia or
PT   lymphoma.
XX
PS   Claim 1; SEQ ID NO 196; 29pp; English.
XX
CC   The invention relates to new recombinant nucleic acids. The invention
CC   also relates to a host cell comprising a recombinant nucleic acid or
CC   expression vector, an expression vector comprising a recombinant nucleic
CC   acid, a recombinant protein, a method of screening for drug candidates, a
CC   method of screening for a bioactive agent capable of binding to a
CC   carcinoma associated protein (CAP) encoded by a nucleotide sequence, a
CC   method of screening for a bioactive agent capable of modulating the
CC   activity of a CAP, a method of evaluating the effect of a candidate
CC   carcinoma drug, a method of diagnosing carcinoma, a method for inhibiting
CC   the activity of a CAP, a method of treating carcinomas, a method of
CC   neutralising the effect of a CAP and a method of diagnosing carcinoma or
CC   propensity to carcinoma. A method of evaluating the effect of a candidate
CC   carcinoma drug comprises administering the drug to a patient, removing a
CC   cell sample from the patient and determining alterations in the
CC   expression or activation of a gene comprising the nucleotide sequence. A
CC   method of diagnosing carcinoma comprises determining the expression of
CC   one or more genes comprising the nucleic acid sequence in a first tissue
CC   type of a first individual and comparing the expression of the gene from
CC   a second normal tissue type from the first individual or a second
CC   unaffected individual, where a difference in the expression indicates
CC   that the first individual has carcinoma. A method of inhibiting the
CC   activity of a CAP comprises binding an inhibitor to the CAP. Treating
CC   carcinomas comprises administering to a patient an inhibitor of CAP.
CC   Neutralising the effect of a CAP comprises contacting an agent specific
CC   for the CAP. The polypeptide specifically binds to the protein encoded by
CC   the nucleic acid. It comprises an antibody that specifically binds to the
CC   protein encoded by the nucleic acid. The nucleic acids are useful for
CC   preparing a composition for diagnosing or treating carcinoma e.g.,
CC   leukaemia or lymphoma. This sequence represents a human carcinoma
CC   associated (CA) nucleic acid of the invention. Note: The sequence data
CC   for this patent did not form part of the printed specification but was
CC   obtained in electronic format directly from USPTO at
CC   seqdata.uspto.gov/sequence.html.
XX
SQ   Sequence 96596 BP; 27639 A; 18627 C; 19687 G; 30643 T; 0 U; 0 Other;

  Query Match             5.0%;  Score 249.6;  DB 12;  Length 96596;
  Best Local Similarity   70.5%;  Pred. No. 7.8e-34;
  Matches  347;  Conservative    0;  Mismatches  144;  Indels    1;  Gaps    1;

Qy        229 TATATATATATATATATATATACACACACACACACATATATATATACACATATATATA 288
              | | || ||||||| || || || ||   | | |||  | || ||||| |||| || ||
Db      57110 TGTGTACATATATACATGTAGATGCATGTAGATACATGTGTACATATATACATGTAGATG 57169

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              ||  || |||||| ||||||  || |||||  || ||| |   |||||| |  | ||||
Db      57170 CATGTAGATACATGTATACATGTAGATACATGTAGATACATGTATACATGTAGATACATG 57229

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              || |||||  |||  ||| | | || || | | || || | |||| |||||| ||||| |
Db      57230 TAGATACATGTATACATATATACATGTAGATACATGTAGATACATGTATACATATATACA 57289

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              | ||| ||  | | ||||| |||| || | |||| || ||||||||||  || |||
Db      57290 TGTAGATACATGTATACATATATACATGTAGATACATGTATACATATATACATGTAGATA 57349

Qy        469 CATATATACACATATATATAATATACACACATATATATACACATATATACACACATATAT 528
              ||| |||||| ||||| ||  || | |||  || ||| |   ||| ||| | |||| ||
Db      57350 CATGTATACATATATACAT-GTAGATACATGTAGATACATGTATACATATATACATGTAG 57408

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACATATAT 588
              | |||| || | |||| |||| | | ||| ||  | ||| ||  ||||| | ||||| ||
Db      57409 ATACATGTAGATACATGTATACATATATACATGTAGATACATGTATATAGATACATACAT 57468

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
              | | ||||| | ||||| ||||||||||  | |||| |||||||||  ||| ||||||
Db      57469 AGATATATAGATACATAGATACATATATTTATATATGTACATATATCTACATATATATCT 57528

Qy        649 ACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATAT 708
              | ||| | ||| |   ||||| ||||  ||||||||| |  || |||||  | |||||||
Db      57529 ATATATACATATATCTACATATATATCTATATATATATATCTACATATATATCTATATAT 57588

Qy        709 GTATATATATAT 720
              ||| | ||||||
Db      57589 GTAGAAATATAT 57600


RESULT 68
ADC87556/c
ID   ADC87556 standard; DNA; 1342 BP.
XX
AC   ADC87556;
XX
DT   01-JAN-2004  (first entry)
XX
DE   Human GPCR gene SEQ ID NO:2009.
XX
KW   ds; gene; human; GPCR;
KW   guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1270724-A2.
XX
PD   02-JAN-2003.
XX
PF   18-JUN-2002; 2002EP-00013517.
XX
PR   18-JUN-2001; 2001JP-00246789.
XX
PA   (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA   (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
```

XX
PI Suwa M, Asai K, Akiyama Y, Aburatani H;
XX
DR WPI; 2003-315783/31.
DR P-PSDB; ADC87557.
XX
PT New polynucleotide, useful for preparing a composition for treating a
PT patient in need of increased or suppressed activity or expression of the
PT guanosine triphosphate-binding protein coupled receptor.
XX
PS Claim 1; SEQ ID NO 2009; 28pp; English.
XX
CC The invention relates to a novel polynucleotide encoding a guanosine
CC triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC the invention may have a use in gene therapy. The polynucleotide and
CC polypeptide are useful for preparing a composition for treating a patient
CC in need of increased or suppressed activity or expression of the
CC guanosine triphosphate-binding protein coupled receptor. The
CC polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC invention.
XX
SQ Sequence 1342 BP; 397 A; 191 C; 193 G; 436 T; 0 U; 125 Other;

Query Match 5.0%; Score 248.4; DB 10; Length 1342;
Best Local Similarity 69.9%; Pred. No. 9.9e-34;
Matches 348; Conservative 0; Mismatches 143; Indels 7; Gaps 2;

Qy 229 TATATATATATATATATATATATATACACACACACACACATATATATATACACATATATATA 288
Db 891 TACACACACATATGTGTATATATACATATACACACCTACACACATATATATGTGTATATA 832

Qy 289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
Db 831 TACATATACACACACATATATGTGTATATATACATATACATACATATGTGTATACACATG 772

Qy 349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
Db 771 TATATGTGCACATATGTATACGTATATGCACATATACATACATATATGTATACATATATA 712

Qy 409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
Db 711 CACATATATGTGCATGTATGTATACGTATGTGCACATATGTACATGTGTATGCATAAATA 652

Qy 469 CATATATACACATATATATA-ATATACACACATATATATACA------CATATATACACA 521
Db 651 CATATATGTACATGTGTATACATAAATACACATATATGTACATGCGTGTACATAAATACA 592

Qy 522 CATATATACACATATATACACATATATATACACATATATACACATATATACATATATACA 581
Db 591 CATATATGTACATGCGTATACATAAATACACATATATGTACATGTGTATACATATATACA 532

Qy 582 CATATATATACATATATACACATATATACATATATACACATATATACATATATACACACA 641
Db 531 CATATATGTACATGCGTATACATATATACACATATGTACATGTGTGTACATAAATACACA 472

Qy 642 TATATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTA 701
Db 471 TATGTACATATGTGTACATAAATACACATATGTACATNNNNNNNNNNNNNNNNNNNNNG 412

Qy 702 TATATATGTATATATATA 719
Db 411 TACATGTATATATATATA 394

RESULT 69
ADC87320/c
ID ADC87320 standard; DNA; 1363 BP.
XX
AC ADC87320;
XX
DT 01-JAN-2004 (first entry)
XX
DE Human GPCR gene SEQ ID NO:1773.
XX
KW ds; gene; human; GPCR;
KW guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1270724-A2.
XX
PD 02-JAN-2003.
XX
PF 18-JUN-2002; 2002EP-00013517.
XX
PR 18-JUN-2001; 2001JP-00246789.
XX
PA (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI Suwa M, Asai K, Akiyama Y, Aburatani H;
XX
DR WPI; 2003-315783/31.
DR P-PSDB; ADC87321.
XX
PT New polynucleotide, useful for preparing a composition for treating a
PT patient in need of increased or suppressed activity or expression of the
PT guanosine triphosphate-binding protein coupled receptor.
XX
PS Claim 1; SEQ ID NO 1773; 28pp; English.
XX
CC The invention relates to a novel polynucleotide encoding a guanosine
CC triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC the invention may have a use in gene therapy. The polynucleotide and
CC polypeptide are useful for preparing a composition for treating a patient
CC in need of increased or suppressed activity or expression of the
CC guanosine triphosphate-binding protein coupled receptor. The
CC polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC invention.
XX
SQ Sequence 1363 BP; 424 A; 175 C; 269 G; 495 T; 0 U; 0 Other;

Query Match 5.0%; Score 248.2; DB 10; Length 1363;
Best Local Similarity 71.7%; Pred. No. 1.1e-33;
Matches 367; Conservative 0; Mismatches 139; Indels 6; Gaps 3;

Qy 227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
Db 1136 TGTACACATATATACATATATGTACATATATACGCACATACGCATATATATGTACATATA 1077

Qy 287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
Db 1076 TACACACATACGCATATATGTACATATATACACATATGCATATATGTACATATATACACA 1017

Qy 347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
Db 1016 TATGCATATATGTACATATATACACATATGCATATATGTACATATATACACACGCATATA 957

Qy 407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
Db 956 TGTACATATATACACACATAT----ACGTATATGTGTGTATACACACATACATATGTGTA 901

Qy 467 TACATATATACACATATATATAATATACACACATATATATACACATATATACACACATAT 526
Db 900 TACACACATACGTATATACGT-ATGTACACATATACATATATACGTATATATACACATAT 842

Qy 527 ATACACATATATACACATATATATACACATATATACACATATATACATATATACACATAT 586
Db 841 ACACCTATATACACACATATACATATATGTACATATACGCACATATACATATATGTGTAC 782

Qy 587 ATATACATATATACACATATATACATATATACACATATATACA-TATATACACACATATA 645
Db 781 ATATATACACATACATATATGCACATATATACACACATGTACACTATATACACACATGTA 722

Qy 646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705

```
Db        721 CACATATGCACATATATACACACATGCACACACGTACACACACGCACACATATGTACACA 662

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAG 737
              ||||| | || | | | |    |  |||| |
Db        661 CATGTACACATGTGTACACATATGCCCATATG 630


RESULT 70
ABT10748
ID   ABT10748 standard; cDNA; 121600 BP.
XX
AC   ABT10748;
XX
DT   04-DEC-2002  (first entry)
XX
DE   Human breast cancer associated coding sequence SEQ ID NO: 882.
XX
KW   Human; breast specific gene; breast cancer; differential expression;
KW   cytostatic; gene therapy; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200259271-A2.
XX
PD   01-AUG-2002.
XX
PF   25-JAN-2002; 2002WO-US002176.
XX
PR   25-JAN-2001; 2001US-0263757P.
PR   25-APR-2001; 2001US-0286090P.
PR   23-MAY-2001; 2001US-0292517P.
XX
PA   (GENE-) GENE LOGIC INC.
XX
PI   Orr MS,  Nation M,  Diggans JC,  Zeng W;
XX
DR   WPI; 2002-674803/72.
XX
PT   Diagnosing breast cancer in a patient comprises detecting the level of
PT   gene expression in cell or tissue samples, where a differential gene
PT   expression is indicative of breast cancer.
XX
PS   Claim 1; SEQ ID NO 882; 260pp + Sequence Listing; English.
XX
CC   The present invention relates to methods of diagnosing breast cancer in a
CC   patient, which comprise detecting the level of expression in a tissue
CC   sample of two or more genes selected from those shown in ABT09867-
CC   ABT11112, where a differential expression of the genes indicates breast
CC   cancer. The methods are useful in diagnosing, treating, detecting the
CC   progression, and in monitoring treatment of breast cancer in patients.
CC   The methods are also useful as a screening tool for agents that modulate
CC   the onset or progression of breast cancer. The breast cancer genes may be
CC   used as diagnostic markers for the prediction or identification of the
CC   malignant state of breast tissue, for confirming the type and progression
CC   of cancer, and for drug screening and assays. The present sequence is a
CC   coding sequence of the invention. Note: The sequence data for this patent
CC   did not form part of the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub.published_pct_sequences
XX
SQ   Sequence 121600 BP; 38372 A; 24652 C; 24048 G; 34528 T; 0 U; 0 Other;

  Query Match             5.0%;  Score 247.4;  DB 6;  Length 121600;
  Best Local Similarity   68.6%;  Pred. No. 1.9e-33;
  Matches  341;  Conservative    0;  Mismatches  156;  Indels    0;  Gaps    0;

Qy        224 CATTTTATATATATATATATATATATATACACACACACACACATATATATATACACATAT 283
              ||   |||||||||| ||||||| |||||| |    | | |    |||||| |||| | || |
Db      43894 CAGCCTATATATATCTATATATCTATATATATCTATATATCTATATATCTATATATATCT 43953

Qy        284 ATATACACATATATACATATATACACATATATACACATATATATACACATACATATGTAC 343
              |||||   |||||||  ||||||   ||||||  | || | |||| |   || |||| ||
Db      43954 ATATATCTATATATATCTATATATCTATATATCTATATCTCTATATATCTATATATCTAT 44013

Qy        344 ACATATATATACACATATGTATACACATATATACACATATATACACACATATATACACAT 403
              |  | ||||||   || | ||||    |||||    ||||||  | | || |||| | ||
Db      44014 ATCTCTATATATCTATCTATATATCTCTATATCTCTATATATCTATATATCTATATATAT 44073

Qy        404 ATATACACACATATATACACATATATACACATATATACACATATACACATATATACACAT 463
              ||||  |   ||||||  | |||| || | |||| || | || || | || |||| |  |
Db      44074 CTATATATCTATATATCTATATATCTATATATATCTATATATCTATATATCTATATATCT 44133

Qy        464 ATATACATATATACACATATATATAATATACACACATATATATACACATATATACACACA 523
              |||||  ||||||  |  ||||||| || || | |  |||||||  | |||| || | |
Db      44134 ATATATCTATATATATCTATATATCATCTATATATCTATATATCTATATATCTATATATC 44193

Qy        524 TATATACACATATATACACATATATATACACATATATACACATATATACATATATACACA 583
              ||||||   ||||||  | ||||||||  | ||||||  | |||| || ||||||  | |
Db      44194 TATATATCTATATATCTATATATATATCTATATATATCTATATATCTATATATATCTATA 44253

Qy        584 TATATATACATATATACACATATATACATATATACACATATATACATATATACACACATA 643
              ||| |||| || |||| | || |||| || |||| | || |||||||| |||| |   |||
Db      44254 TATCTATATATCTATATATATCTATATATCTATATATATCTATACATCTATATATCTATA 44313

Qy        644 TATACACATACATATATACACACACATAGATATACATATATATACACATATATATACGTATA 703
              ||||   ||| || || | |   ||| ||||  ||||||  || | |||| ||||  | ||
Db      44314 TATATCTATATATCTATATATCTATATATATCTATATATCTATATATATCTATATATCTA 44373

Qy        704 TATATGTATATATATAT 720
              ||||| | ||||||| |
Db      44374 TATATATCTATATATCT 44390


RESULT 71
ADQ18307
ID   ADQ18307 standard; DNA; 121600 BP.
XX
AC   ADQ18307;
XX
DT   26-AUG-2004  (first entry)
XX
DE   Human soft tissue sarcoma-upregulated DNA - SEQ ID 1125.
XX
KW   soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO2004048938-A2.
XX
PD   10-JUN-2004.
XX
PF   26-NOV-2003; 2003WO-US038193.
XX
PR   26-NOV-2002; 2002US-0429739P.
XX
PA   (PROT-) PROTEIN DESIGN LABS INC.
XX
PI   Aziz N,  Ginsburg WM,  Zlotnik A;
XX
DR   WPI; 2004-441208/41.
XX
PT   Early detection of soft tissue sarcoma comprises determining expression
PT   of a gene in a first soft tissue sample and a normal soft tissue sample
PT   and comparing the gene expression, also useful in treating soft tissue
PT   sarcoma.
XX
PS   Example 2; SEQ ID NO 1125; 210pp; English.
XX
CC   The invention relates to a novel method for detecting soft tissue sarcoma
CC   which comprises obtaining a first soft tissue sample from an individual
CC   and a normal soft tissue sample from the same or different individual,
CC   determining the expression of a gene in both samples and comparing the
CC   expression of the gene in both soft tissue samples, where a higher level
CC   of protein expression in the first soft tissue sample indicates the
CC   presence of soft tissue sarcoma. The method of the invention has
```

```
CC   cytostatic applications and may be useful for detecting soft tissue
CC   sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC   acid sequences may be useful in diagnostic and screening applications.
CC   The current sequence is that of a human soft tissue sarcoma-upregulated
CC   DNA of the invention. The current sequence is not shown within the
CC   specification per se but was submitted in CD format by the inventor.
XX
SQ   Sequence 121600 BP; 38372 A; 24652 C; 24048 G; 34528 T; 0 U; 0 Other;

  Query Match             5.0%;  Score 247.4;  DB 12;  Length 121600;
  Best Local Similarity   68.6%;  Pred. No. 1.9e-33;
  Matches  341;  Conservative    0;  Mismatches  156;  Indels    0;  Gaps    0;

Qy        224 CATTTTATATATATATATATATATATATACACACACACACACATATATATATACACATAT 283
              ||   ||||||||| ||||||| ||||||  |   | | |   |||||| |||| | || |
Db      43894 CAGCCTATATATATCTATATATCTATATATATCTATATATCTATATATCTATATATATCT 43953

Qy        284 ATATACACATATATACATATATACACATATATACACATATATATACACATACATATGTAC 343
              |||||   |||||||  ||||||   ||||||  | || | |||| |  || |||| ||
Db      43954 ATATATCTATATATATCTATATATCTATATATCTATATCTCTATATATCTATATATCTAT 44013

Qy        344 ACATATATATACACATATGTATACACATATATACACATATATACACACATATATACACAT 403
              |  | ||||||   || | ||||    |||||    ||||||  | | || |||| | ||
Db      44014 ATCTCTATATATCTATCTATATATCTCTATATCTCTATATATCTATATATCTATATATAT 44073

Qy        404 ATATACACACATATATACACATATATACACATATATACACATATACACATATATACACAT 463
              |||| |   ||||||  | |||| || | |||| || | || || | || |||| |  |
Db      44074 CTATATATCTATATATCTATATATCTATATATATCTATATATCTATATATCTATATATCT 44133

Qy        464 ATATACATATATACACATATATATAATATACACACATATATATACACATATATACACACA 523
              |||||  |||||| |  ||||||| || || | |  |||||||  | |||| || | |
Db      44134 ATATATCTATATATATCTATATATCATCTATATATCTATATATCTATATATCTATATATC 44193

Qy        524 TATATACACATATATACACATATATATACACATATATACACATATATACATATATACACA 583
              ||||||   ||||||  | ||||||||  | ||||||  | |||| || ||||||  | |
Db      44194 TATATATCTATATATCTATATATATATCTATATATATCTATATATCTATATATATCTATA 44253

Qy        584 TATATATACATATATACACATATATACATATATACACATATATACATATATACACACATA 643
              ||| |||| || |||| | || |||| || |||| | || |||||||  |||| |   |||
Db      44254 TATCTATATATCTATATATATCTATATATCTATATATATCTATACATCTATATATCTATA 44313

Qy        644 TATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATA 703
              ||||   ||| || || | |   ||| ||||  |||||| || | |||| ||||  | ||
Db      44314 TATATCTATATATCTATATATCTATATATATCTATATATCTATATATATCTATATATCTA 44373

Qy        704 TATATGTATATATATAT 720
              ||||| | ||||||| |
Db      44374 TATATATCTATATATCT 44390


RESULT 72
ACC79619
ID   ACC79619 standard; DNA; 71843 BP.
XX
AC   ACC79619;
XX
DT   27-OCT-2003  (revised)
DT   06-AUG-2003  (first entry)
XX
DE   Zebrafish protein kinase C lambda related genomic DNA SEQ ID NO:3.
XX
KW   Zebrafish; Danio rerio; protein kinase C lambda; PKC-lambda; enzyme;
KW   cardiant; cytostatic; nephrotropic; heart disease; cancer; organ failure;
KW   epithelial-epithelial cell interaction; epithelial cell polarity;
KW   heart failure; kidney failure; gene; ds.
XX
OS   Danio rerio.
XX
PN   WO2003023048-A2.
XX
PD   20-MAR-2003.
XX
PF   06-SEP-2002; 2002WO-US028410.
XX
PR   06-SEP-2001; 2001US-0317653P.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Peterson R,  Fishman MC;
XX
DR   WPI; 2003-313257/30.
XX
PT   Diagnosing or prognosing a disease or condition related to Protein Kinase
PT   C lambda, by analyzing sample nucleic acid obtained from the test subject
PT   to determine mutation in gene encoding Protein kinase C lambda.
XX
PS   Disclosure; Page 63-83; 88pp; English.
XX
CC   The present invention describes a method (M1) for determining whether a
CC   test subject has, or is at risk of developing, a disease or condition
CC   related to protein kinase C lambda (PKC-lambda) (I). M1 comprises
CC   analysing a nucleic acid molecule of a sample from the subject to
CC   determine whether the subject has a mutation in a gene encoding (I),
CC   where the presence of a mutation indicates that the subject has, or is at
CC   risk of developing, a disease or condition related to (I). (I) has
CC   cardiant, cytostatic and nephrotropic activities. M1 is useful for
CC   determining whether a test subject has, or is at risk of developing, a
CC   disease or condition related to (I), e.g. heart disease, cancer, or
CC   condition associated with epithelial-epithelial cell interactions or
CC   epithelial cell polarity. (I) is useful for treating organ failure such
CC   as heart or kidney failure. (I) is also useful for raising antibodies
CC   specific for various regions of (I). M1 is suitable for detecting an
CC   increased likelihood of diseases or conditions associated with (I), so
CC   that appropriate intervention can be instituted before any symptoms
CC   occur. This is useful with patients in high risk groups for such diseases
CC   or conditions. Diagnostic method facilitates determination of the
CC   etiology of such an existing disease or condition in a patient, so that
CC   an appropriate approach to treatment can be selected. The present
CC   sequence represents zebrafish genomic DNA for PKC-lambda, which is used
CC   in the exemplification of the present invention. (Updated on 27-OCT-2003
CC   to standardise OS field)
XX
SQ   Sequence 71843 BP; 24017 A; 12903 C; 12592 G; 21955 T; 0 U; 376 Other;

  Query Match             5.0%;  Score 246.6;  DB 8;  Length 71843;
  Best Local Similarity   63.1%;  Pred. No. 2.5e-33;
  Matches  313;  Conservative    0;  Mismatches  181;  Indels    2;  Gaps    2;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | |||||||||||||||||||||||| | | | | | ||||||||||| | |||||||
Db      68244 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 68303

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | |||||||  |||||||||||  | | |||||  | |  | |||||  ||| |
Db      68304 TATATATATATATATATATACACACACACACACACACACACACACACACACACANNNNNNNN 68363

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406

Db      68364 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 68423

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
                                                  |||||  | |||||||  | |||||
Db      68424 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNATATATATATATATATATATA 68483

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || |||||||  | |||||||||  |||||  | |  |||||||||  | |||||||  | |  |||
Db      68484 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 68543

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACACATATATACACATA 585
              ||||  | |||||||  | ||||  |||| | |||||||  | |||||||  |||||||  | |||
Db      68544 TATATATATATATATATATATNTATATATATATATATATATATATATATATATATATATA 68603

Qy        586 TATATACATATATACACATATATACATATATA-CACATATATACATATATACACACATAT 644
              ||||||  |||||||  | |||||||  |||||||     ||||||  |||||||  | |  ||||
Db      68604 TATATATATATATATATATATATATATATATAGTGTATATATGTATATATATATATATAT 68663
```

```
Qy        645 ATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATAT 704
              ||| | ||| ||||| |   | ||| ||||| ||||||||| | |||||||||  |||||
Db      68664 ATATATATATATATATATNTATATATATATATATATATATATATATATATATATATAT 68723

Qy        705 ATATGTATATATATAT 720
              |||| ||||||||||||
Db      68724 ATATATATATATATAT 68739


RESULT 73
AAS46591/c
ID   AAS46591 standard; DNA; 6286 BP.
XX
AC   AAS46591;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Tumour suppressor gene derived chemically modified sequence #313.
XX
KW   Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer;
KW   tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW   cytosine methylation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200168912-A2.
XX
PD   20-SEP-2001.
XX
PF   15-MAR-2001; 2001WO-EP002955.
XX
PR   15-MAR-2000; 2000DE-01013847.
PR   06-APR-2000; 2000DE-01019058.
PR   07-APR-2000; 2000DE-01019173.
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2001-602752/68.
XX
PT   Fragments of chemically modified genes associated with tumor suppressor
PT   genes and oncogenes, useful in designing primers and probes for analyzing
PT   diseases associated with cytosine methylation state e.g. cancer.
XX
PS   Claim 1; SEQ ID NO 313; 27pp; English.
XX
CC   The invention relates to a nucleic acid comprising a sequence of 18
CC   bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC   bisulphite, of genes associated with tumour suppression and oncogenes
CC   having a sequence taken from 536 (actually 533 since numbers 408, 458 and
CC   500 are missing from the sequence listing) sequences (Ss) and sequences
CC   complementary to (Ss). The nucleic acid may be a peptide nucleic acid-
CC   oligomer (PNA) of at least 9 nucleotides and may form part of a set of
CC   probes for detecting the cytosine methylation state and/or single
CC   nucleotide polymorphisms and also to be used in an array for analysing
CC   diseases associated with CpG dinucleotides e.g. cancers and tumours. The
CC   probes can also be used in a method for ascertaining genetic and/or
CC   epigenetic parameters for the diagnosis and/or therapy of existing
CC   diseases or the predisposition to specific diseases, by analysing
CC   cytosine methylations. The parameters may be compared to another set of
CC   genetic and/or epigenetic parameters, the differences serving as basis
CC   for diagnosis and/or prognosis events which are disadvantageous to
CC   patients. The present sequence is one of the 533 genomic sequences
CC   derived from tumour suppressor genes and oncogenes. Note: The sequence
CC   data for this patent did not form part of the printed specification, but
CC   was obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 6286 BP; 1508 A; 124 C; 1538 G; 3116 T; 0 U; 0 Other;

  Query Match             4.9%;  Score 245;  DB 4;  Length 6286;
  Best Local Similarity   57.3%;  Pred. No. 4.2e-33;
  Matches  443;  Conservative    0;  Mismatches  330;  Indels    0;  Gaps    0;

Qy        174 CTCCAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATAT 233
              | ||  |||| |       |   | || |  | | | | |||  || |   |  | |||
Db       4589 CACCCACCCATCCCCATTCATAAACATTTACTATAAAATAAATCAATAAATAAATAATAA 4530

Qy        234 ATATATATATATATATATACACACACACACACATATATATATACACATATATATACACAT 293
              ||  ||| ||| ||| ||| | | | | | ||||| ||| ||| | | ||| ||| | |
Db       4529 ATTAATAAATAAATAAATAAATAAATAAATACATAAATAAATAAATAAATAAATAAATAA 4470

Qy        294 ATATACATATATACACATATATACACATATATATACACATACATATGTACACATATATAT 353
              ||| | | ||| | | ||| ||||| | ||| ||| | | | | ||  | | ||| |||
Db       4469 ATAAATAAATAAATAAATAAATACATAAATAAATAAATAAATAAATAAATAAATAAATAA 4410

Qy        354 ACACATATGTATACACATATATACACATATATACACACATATATACACATATATACACAC 413
              | | |||  || | ||||| ||| | | ||| | | | ||| | | ||| ||| | |
Db       4409 ATAAATAAATAAATACATAAATAAATAAATAAATAAATAAATAAATACATAAATAAATAA 4350

Qy        414 ATATATACACATATATACACATATATACACATATACACATATATACACATATATACATAT 473
              ||| ||||| | ||| | | ||| ||| | | ||| | | ||| | | ||| ||| |||
Db       4349 ATAAATACATAAATAAATAAATAAATAAATAAATAAATAAATACATAAATAAATAAATAA 4290

Qy        474 ATACACATATATATAATATACACACATATATATACACATATATACACACATATATACACA 533
              ||| | | ||| ||||    |   ||| ||| ||| | | ||| | | | | ||| | | |
Db       4289 ATAAATAAATAAATAAATAAACAACAAATAAATAAATAAATAAATAAATAAATAAATAA 4230

Qy        534 TATATACACATATATATACACATATATACACATATATACATATATACACATATATATACA 593
              || ||| | | ||| ||| | | ||| | | ||| ||| ||| ||| | | ||| ||| |
Db       4229 TAAATAAATACATAAATAAATAAATAAATAAATACATAAATAAATAAATAAATAAATAAA 4170

Qy        594 TATATACACATATATACATATATACACATATATACATATATACACACATATATACACATA 653
              || ||| | | ||| | | ||| | | ||| ||| ||| ||| | | ||| ||| | | |
Db       4169 TAAATAAATAAATACATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAA 4110

Qy        654 CATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATA 713
              | ||| | ||| | | | ||| ||| ||| | | ||| ||| |   ||| |||  || |
Db       4109 TAAATAAATACATAAATAAATAAATAAATAAATAAATAAATAAATACATAAATAAA 4050

Qy        714 TATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAA 773
              || |||        | |  ||||         |       |  | |    | | |   |||
Db       4049 TAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATACATAAATAAATAAA 3990

Qy        774 GAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGG 833
              | |     | |  | |  ||   | |  |  | | |          |||||
Db       3989 TAAATAAATAAATAAATAAATAAATAAATACATAAATAAATAAATAAATACATAAA 3930

Qy        834 GAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATAT 893
              ||||| ||||||   |||  |||   ||   |||   | | || ||   |  | || ||
Db       3929 TAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAACAACAAATAAAT 3870

Qy        894 ATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATG 946
              | ||| |||   |||  | |  | |  |||| |       ||  |  |||| |
Db       3869 AAATAAATAAATAAATAAATAAATAAATAAATACATAAATAAATAAACG 3817


RESULT 74
ADC87574/c
ID   ADC87574 standard; DNA; 1309 BP.
XX
AC   ADC87574;
XX
DT   01-JAN-2004  (first entry)
XX
DE   Human GPCR gene SEQ ID NO:2027.
XX
KW   ds; gene; human; GPCR;
KW   guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1270724-A2.
```

XX
PD 02-JAN-2003.
XX
PF 18-JUN-2002; 2002EP-00013517.
XX
PR 18-JUN-2001; 2001JP-00246789.
XX
PA (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI Suwa M, Asai K, Akiyama Y, Aburatani H;
XX
DR WPI; 2003-315783/31.
DR P-PSDB; ADC87575.
XX
PT New polynucleotide, useful for preparing a composition for treating a
PT patient in need of increased or suppressed activity or expression of the
PT guanosine triphosphate-binding protein coupled receptor.
XX
PS Claim 1; SEQ ID NO 2027; 28pp; English.
XX
CC The invention relates to a novel polynucleotide encoding a guanosine
CC triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC the invention may have a use in gene therapy. The polynucleotide and
CC polypeptide are useful for preparing a composition for treating a patient
CC in need of increased or suppressed activity or expression of the
CC guanosine triphosphate-binding protein coupled receptor. The
CC polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC invention.
XX
SQ Sequence 1309 BP; 463 A; 205 C; 200 G; 441 T; 0 U; 0 Other;

Query Match 4.9%; Score 243.2; DB 10; Length 1309;
Best Local Similarity 70.7%; Pred. No. 7.9e-33;
Matches 355; Conservative 0; Mismatches 138; Indels 9; Gaps 2;

```
Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | || |||| || || ||| ||||||| || |    || |||| ||||||| ||||  ||
Db        553 TATACATATGTACATCTATGTATATACGCATATGTGCATATATGTATATACGCATACGTA 494

Qy        287 TACACATATATACATATATACACATATATACACATATATATACAC------ATACATATG 340
              ||    |||||||  ||||| ||||||  | || |||| |||| ||       |||||||||
Db        493 TATGTGTATATACGCATATATACATATGTGCATATATGTATATACGCATATATACATATG 434

Qy        341 TACACATATATATACACATATGTATACACATATATACACATATATACACACATATATACA 400
              | || |||| |||| ||  || |||||| || |  | | || |||| || ||||||||||
Db        433 TGCATATATGTATATACGCATATATACATATGTGCATATATGTATATACGCATATATACA 374

Qy        401 CATATATACACACATATATACACATATATACACATATATACACATATACACATATATACA 460
               || |  | | |  |||||||  |||||||||| | || | || || | | |||| | |
Db        373 TATGTGCATATATGTATATACGCATATATACATACATGTGCATATGTGCATATATTCATA 314

Qy        461 CATATATACATATATACACATATATATAATATACACACATATATATACACATATATACAC 520
                || ||  || ||||   | ||||   || |||| ||| |||| |||| |||| ||||
Db        313 AGTACATCTATGTATATGTACATAT---ATGTACATACACATATGTACATATATGTACAT 257

Qy        521 ACATATATACACATATATACACATATATATACACATATATACACATATATACATATATAC 580
              ||| ||||  ||||||||  ||||| | ||| ||||||||  |||||||  | | |||||
Db        256 ACACATATGTACATATATGTACATACACATATACATATATGTACATATACATACATATAT 197

Qy        581 ACATATATATACATATATACACATATATACATATATACACATATATACATATATACACAC 640
              ||| ||| || ||||  ||| |||  || ||||| | | || |||| ||| | | | ||
Db        196 GCATGTATGTATATATGCACATATACCTATATATACATATATGTATATATACACATATAC 137

Qy        641 ATATATACACATACATATACACACACATAGATATACATATATATACACATATATATACGT 700
              ||||||  |||||  | || | ||||| | ||  | |||| ||||| |||||||  |  |
Db        136 ATATATGTACATATGTGTATATACACACATATGCATATATGTATACGCATATATGCATAT 77

Qy        701 ATATATATGTATATATATATGC 722
              || |||||  | | ||||||||
Db         76 ATGTATATACACACATATATGC 55
```

RESULT 75
ABX08336_08
Continuation (9 of 17) of ABX08336 from base 800001 (Human phosphodiesterase 4D (PDE4D) s
WP Sequence split into 17 fragments LOCUS ABX08336 Accession Abx08336

WP	Fragment Name	Begin	End
WP	ABX08336_00	1	110000
WP	ABX08336_01	100001	210000
WP	ABX08336_02	200001	310000
WP	ABX08336_03	300001	410000
WP	ABX08336_04	400001	510000
WP	ABX08336_05	500001	610000
WP	ABX08336_06	600001	710000
WP	ABX08336_07	700001	810000
WP	ABX08336_08	800001	910000
WP	ABX08336_09	900001	1010000
WP	ABX08336_10	1000001	1110000
WP	ABX08336_11	1100001	1210000
WP	ABX08336_12	1200001	1310000
WP	ABX08336_13	1300001	1410000
WP	ABX08336_14	1400001	1510000
WP	ABX08336_15	1500001	1610000
WP	ABX08336_16	1600001	1691080

Query Match 4.9%; Score 242; DB 6; Length 110000;
Best Local Similarity 70.0%; Pred. No. 1.6e-32;
Matches 368; Conservative 1; Mismatches 151; Indels 6; Gaps 3;

```
Qy        196 GAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACAC 255
              ||  ||| |   | | | |||| | |   | ||||||| ||||||||  |||||
Db      67577 GAGTTGTGAATATATATATAAAATATGTGTGTGTATATATACATATATATGCATATATGT 67636

Qy        256 ACACACACACATATATATATACACATATATATACACATATATACATATATACACATATAT 315
              ||  | | |||||||| |||| | |  |||||  || ||||||||||||||||||||||
Db      67637 ACCTATATACATATAT-TATATATACCTATATGAACCTATATACATATATACACATATAT 67695

Qy        316 ACACATATATATACACATACATATGTACACATATAT----ATACACATATGTATACACAT 371
              || |||||| ||| ||||| || :  || |||||    ||| ||| || | |  ||
Db      67696 GAACCTATATACACATATACACATAYGTACCTATATACACATATACACATATGTGTACCC 67755

Qy        372 ATATACACATATATACACACATATATACACATATATACACACATATATACACATATATAC 431
              |||||||||||||| ||| |  || || | | | ||| |||||||| |  ||  ||||||
Db      67756 ATATACACATATACACATATGTACCTATATACACATATACACATATGTGTACCCATATAC 67815

Qy        432 ACATATATACACATATACACATATATACACATATATACATATATACACATATATATAATA 491
              ||||||| ||| || |  ||  |||||||||||||| |||||| |  ||  |||||  | |
Db      67816 ACATATACACATATGTGTACCCATATACACATATACACATATGTGTACCCATATA-CACA 67874

Qy        492 TACACACATATATATACACATATATACACACATATATACACATATATACACATATATATA 551
              || |||||||| | ||| |||||| ||| | | | |||    ||   | | | | |||||
Db      67875 TATACACATATGTGTACCCATATACACATATACACATATGTGTACCCATATACACATATA 67934

Qy        552 CACATATATACACATATATACATATATACACATATATATACATATATACACATATATACA 611
              ||||||| |  || |||||||  |||||||||||| | |||  |||||||||||||  |||
Db      67935 CACATATGTGTACCCATATACACATATACACATATGTGTACCCATATACACATATACACA 67994

Qy        612 TATATACACATATATACATATATACACACATATATACACATACATATACACACACATAGA 671
              | | ||| |||||| |||||||  |||     ||    | | | ||||||||||    ||
Db      67995 TGTGTACCCATATACACATATACACATGTGTACCCATATACACATATACACATGTGTACC 68054

Qy        672 TATACATATATATACACATATATATACGTATATATATGTATATATA 717
              ||| | | |||||  |||||  | |||  |||||     ||| | ||
Db      68055 CATATACACATATATGCATATGTGTACCCATATACGCATATGTGTA 68100
```

RESULT 76
ADJ25985_08
Continuation (9 of 17) of ADJ25985 from base 800001 (Human phosphodiesterase 4D (PDE4D)
WP Sequence split into 17 fragments LOCUS ADJ25985 Accession Adj25985

WP	Fragment Name	Begin	End
WP	ADJ25985_00	1	110000
WP	ADJ25985_01	100001	210000
WP	ADJ25985_02	200001	310000
WP	ADJ25985_03	300001	410000

```
WP      ADJ25985_04         400001    510000
WP      ADJ25985_05         500001    610000
WP      ADJ25985_06         600001    710000
WP      ADJ25985_07         700001    810000
WP      ADJ25985_08         800001    910000
WP      ADJ25985_09         900001   1010000
WP      ADJ25985_10        1000001   1110000
WP      ADJ25985_11        1100001   1210000
WP      ADJ25985_12        1200001   1310000
WP      ADJ25985_13        1300001   1410000
WP      ADJ25985_14        1400001   1510000
WP      ADJ25985_15        1500001   1610000
WP      ADJ25985_16        1600001   1691139

  Query Match             4.9%;  Score 242;  DB 12;  Length 110000;
  Best Local Similarity   70.0%;  Pred. No. 1.6e-32;
  Matches  368;  Conservative    1;  Mismatches  151;  Indels    6;  Gaps    3;

Qy       196 GAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACAC 255
             ||  ||| |    | | | |||| | |    | ||||||| ||||||||  |||||
Db     67636 GAGTTGTGAATATATATATAAAATATGTGTGTGTATATATACATATATATGCATATATGT 67695

Qy       256 ACACACACACATATATATATACACATATATATACACATATATACATATATACACATATAT 315
             ||  | | |||||||| |||| | |  |||||  || ||||||||||||||||||||||||
Db     67696 ACCTATATACATATAT-TATATATACCTATATGAACCTATATACATATATACACATATAT 67754

Qy       316 ACACATATATATACACATACATATGTACACATATAT----ATACACATATGTATACACAT 371
             || |||||| ||| ||||| || :  || |||||    ||| ||| || | |  ||
Db     67755 GAACCTATATACACATATACACATAYGTACCTATATACACATATACACATATGTGTACCC 67814

Qy       372 ATATACACATATATACACACATATATACACATATATACACACATATATACACATATATAC 431
             |||||||||||| ||| |  ||  || | | | ||| ||||||| |  ||  ||||||
Db     67815 ATATACACATATACACATATGTACCTATATACACATATACACATATGTGTACCCATATAC 67874

Qy       432 ACATATATACACATATACACATATATACACATATATACATATATACACATATATATAATA 491
             ||||||| ||| || |  ||  |||||||||||||| |||||| |  ||  |||||  | |
Db     67875 ACATATACACATATGTGTACCCATATACACATATACACATATGTGTACCCATATA-CACA 67933

Qy       492 TACACACATATATATACACATATATACACACATATATACACATATATACACATATATATA 551
             || |||||||| | ||| |||||| ||| | | | |||    ||   | | | | |||||
Db     67934 TATACACATATGTGTACCCATATACACATATACACATATGTGTACCCATATACACATATA 67993

Qy       552 CACATATATACACATATATACATATATACACATATATATACATATATACACATATATACA 611
             ||||||| |  ||  ||||||| |||||||||||| | |||  |||||||||||| |||
Db     67994 CACATATGTGTACCCATATACACATATACACATATGTGTACCCATATACACATATACACA 68053

Qy       612 TATATACACATATATACATATATACACACATATATACACATACATATACACACACATAGA 671
             | | ||| |||||| ||||||| |||    ||   | | | ||||||||||    ||
Db     68054 TGTGTACCCATATACACATATACACATGTGTACCCATATACACATATACACATGTGTACC 68113

Qy       672 TATACATATATATACACATATATATACGTATATATATGTATATATA 717
             ||| | | |||||  |||||| ||| |||||    ||| | ||
Db     68114 CATATACACATATATGCATATGTGTACCCATATACGCATATGTGTA 68159


RESULT 77
ADN97989_08
Continuation (9 of 17) of ADN97989 from base 800001 (Human phosphodiesterase 4D genomic
WP Sequence split into 17 fragments  LOCUS ADN97989 Accession Adn97989
WP      Fragment Name       Begin     End
WP      ADN97989_00             1    110000
WP      ADN97989_01        100001    210000
WP      ADN97989_02        200001    310000
WP      ADN97989_03        300001    410000
WP      ADN97989_04        400001    510000
WP      ADN97989_05        500001    610000
WP      ADN97989_06        600001    710000
WP      ADN97989_07        700001    810000
WP      ADN97989_08        800001    910000
WP      ADN97989_09        900001   1010000
WP      ADN97989_10       1000001   1110000
WP      ADN97989_11       1100001   1210000
WP      ADN97989_12       1200001   1310000
WP      ADN97989_13       1300001   1410000
WP      ADN97989_14       1400001   1510000
WP      ADN97989_15       1500001   1610000
WP      ADN97989_16       1600001   1691138

  Query Match             4.9%;  Score 242;  DB 12;  Length 110000;
  Best Local Similarity   70.0%;  Pred. No. 1.6e-32;
  Matches  368;  Conservative    1;  Mismatches  151;  Indels    6;  Gaps    3;

Qy       196 GAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACAC 255
             ||  ||| |    | | | |||| | |    | ||||||| ||||||||  |||||
Db     67635 GAGTTGTGAATATATATATAAAATATGTGTGTGTATATATACATATATATGCATATATGT 67694

Qy       256 ACACACACACATATATATATACACATATATATACACATATATACATATATACACATATAT 315
             ||  | | |||||||| |||| | |  |||||  || ||||||||||||||||||||||||
Db     67695 ACCTATATACATATAT-TATATATACCTATATGAACCTATATACATATATACACATATAT 67753

Qy       316 ACACATATATATACACATACATATGTACACATATAT----ATACACATATGTATACACAT 371
             || |||||| ||| ||||| || :  || |||||    ||| ||| || | |  ||
Db     67754 GAACCTATATACACATATACACATAYGTACCTATATACACATATACACATATGTGTACCC 67813

Qy       372 ATATACACATATATACACACATATATACACATATATACACACATATATACACATATATAC 431
             |||||||||||| ||| |  ||  || | | | ||| ||||||| |  ||  ||||||
Db     67814 ATATACACATATACACATATGTACCTATATACACATATACACATATGTGTACCCATATAC 67873

Qy       432 ACATATATACACATATACACATATATACACATATATACATATATACACATATATATAATA 491
             ||||||| ||| || |  ||  |||||||||||||| |||||| |  ||  |||||  | |
Db     67874 ACATATACACATATGTGTACCCATATACACATATACACATATGTGTACCCATATA-CACA 67932

Qy       492 TACACACATATATATACACATATATACACACATATATACACATATATACACATATATATA 551
             || |||||||| | ||| |||||| ||| | | | |||    ||   | | | | |||||
Db     67933 TATACACATATGTGTACCCATATACACATATACACATATGTGTACCCATATACACATATA 67992

Qy       552 CACATATATACACATATATACATATATACACATATATATACATATATACACATATATACA 611
             ||||||| |  ||  ||||||| |||||||||||| | |||  |||||||||||| |||
Db     67993 CACATATGTGTACCCATATACACATATACACATATGTGTACCCATATACACATATACACA 68052

Qy       612 TATATACACATATATACATATATACACACATATATACACATACATATACACACACATAGA 671
             | | ||| |||||| ||||||| |||    ||   | | | ||||||||||    ||
Db     68053 TGTGTACCCATATACACATATACACATGTGTACCCATATACACATATACACATGTGTACC 68112

Qy       672 TATACATATATATACACATATATATACGTATATATATGTATATATA 717
             ||| | | |||||  |||||| ||| |||||    ||| | ||
Db     68113 CATATACACATATATGCATATGTGTACCCATATACGCATATGTGTA 68158


RESULT 78
ADO50281_08
Continuation (9 of 17) of ADO50281 from base 800001 (Human phosphodiesterase 4D (PDE4D) g
WP Sequence split into 17 fragments  LOCUS ADO50281 Accession Ado50281
WP      Fragment Name       Begin     End
WP      ADO50281_00             1    110000
WP      ADO50281_01        100001    210000
WP      ADO50281_02        200001    310000
WP      ADO50281_03        300001    410000
WP      ADO50281_04        400001    510000
WP      ADO50281_05        500001    610000
WP      ADO50281_06        600001    710000
WP      ADO50281_07        700001    810000
WP      ADO50281_08        800001    910000
WP      ADO50281_09        900001   1010000
WP      ADO50281_10       1000001   1110000
WP      ADO50281_11       1100001   1210000
WP      ADO50281_12       1200001   1310000
WP      ADO50281_13       1300001   1410000
WP      ADO50281_14       1400001   1510000
WP      ADO50281_15       1500001   1610000
WP      ADO50281_16       1600001   1691134

  Query Match             4.9%;  Score 242;  DB 12;  Length 110000;
  Best Local Similarity   70.0%;  Pred. No. 1.6e-32;
  Matches  368;  Conservative    1;  Mismatches  151;  Indels    6;  Gaps    3;
```

```
Qy        196 GAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACAC 255
              ||  ||| |    | | | |||| | |    | |||||||| ||||||||  |||||
Db      67636 GAGTTGTGAATATATATATAAAATATGTGTGTGTATATATACATATATATGCATATATGT 67695

Qy        256 ACACACACACATATATATATACACATATATATACACATATATACATATATACACATATAT 315
              ||  | | |||||||| |||| | |  |||||  || ||||||||||||||||||||||
Db      67696 ACCTATATACATATAT-TATATATACCTATATGAACCTATATACATATATACACATATAT 67754

Qy        316 ACACATATATATACACATACATATGTACACATATAT----ATACACATATGTATACACAT 371
              || |||||| ||| ||||| || :  || |||||    ||| ||| || | |  ||
Db      67755 GAACCTATATACACATATACACATAYGTACCTATATACACATATACACATATGTGTACCC 67814

Qy        372 ATATACACATATATACACACATATATACACATATATACACACATATATACACATATATAC 431
              |||||||||||||| ||| |  ||  || | | ||| |||||||| |  ||  ||||||
Db      67815 ATATACACATATACACATATGTACCTATATACACATATACACATATGTGTACCCATATAC 67874

Qy        432 ACATATATACACATATACACATATATACACATATATACATATATACACATATATATAATA 491
              ||||||| ||| || |  ||  |||||||||||||| |||||| |  ||  |||||  | |
Db      67875 ACATATACACATATGTGTACCCATATACACATATACACATATGTGTACCCATATA-CACA 67933

Qy        492 TACACACATATATATACACATATATACACACATATATACACATATATACACATATATATA 551
              || |||||||| | ||| |||||| ||| | | | |||     ||    | | | |||||
Db      67934 TATACACATATGTGTACCCATATACACATATACACATATGTGTACCCATATACACATATA 67993

Qy        552 CACATATATACACATATATACATATATACACATATATATACATATATACACATATATACA 611
              |||||||  |   ||  ||||||| |||||||||||| | |||  |||||||||||| |||
Db      67994 CACATATGTGTACCCATATACACATATACACATATGTGTACCCATATACACATATACACA 68053

Qy        612 TATATACACATATATACATATATACACACATATATACACATACATATACACACACATAGA 671
              | | |||  |||||| |||||||  |||     ||    | | | ||||||||||     ||
Db      68054 TGTGTACCCATATACACATATACACATGTGTACCCATATACACATATACACATGTGTACC 68113

Qy        672 TATACATATATATACACATATATATACGTATATATATGTATATATA 717
              |||  | |  |||||  |||||  | |||   |||||     |||  | ||
Db      68114 CATATACACATATATGCATATGTGTACCCATATACGCATATGTGTA 68159
```

RESULT 79
ADQ97965
ID ADQ97965 standard; DNA; 152096 BP.
XX
AC ADQ97965;
XX
DT 07-OCT-2004 (first entry)
XX
DE Mouse cancer associated sequence MD11-032, SEQ ID 942.
XX
KW Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Mouse; ds.
XX
OS Mus musculus.
XX
PN WO2004060304-A2.
XX
PD 22-JUL-2004.
XX
PF 22-DEC-2003; 2003WO-US041389.
XX
PR 27-DEC-2002; 2002US-00330773.
XX
PA (SAGR-) SAGRES DISCOVERY INC.
XX
PI Morris DW, Malandro MS;
XX
DR WPI; 2004-543781/52.
XX
PT New isolated cancer associated nucleic acids comprising at least 10
PT contiguous nucleotides, useful for diagnosing, preventing and/or treating
PT cancers such as leukemia and lymphoma.
XX
PS Claim 1; SEQ ID NO 942; 199pp; English.
XX
CC The present invention relates to cancer associated sequences (ADQ97025-
CC ADQ98004). The sequences are useful for the diagnosis, prevention and/or
CC treatment of cancer, such as leukemia and lymphoma. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic formate directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 152096 BP; 42690 A; 31533 C; 31445 G; 45238 T; 0 U; 1190 Other;

  Query Match   4.9%; Score 241.4; DB 12; Length 152096;
  Best Local Similarity   68.5%; Pred. No. 2.1e-32;
  Matches 346; Conservative   0; Mismatches 158; Indels   1; Gaps   1;

```
Qy        230 ATATATATATATATATATATATACACACACACACATATATATATACACATATATATAC 289
              | | | | | | | | | | | | ||| |||||||||| | | | ||| | ||| | | | ||
Db      78632 ACACACACACACACACACACACACATACACACACATACACACATACATACACACACAC 78691

Qy        290 ACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATAT 349
              ||| | ||||||| | ||||| | | ||||||| | | |||    ||| |    ||| | |
Db      78692 ACACACATACATACACACACACACACACACATACACACACATGCACACACACACATACAC 78751

Qy        350 ATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATAC 409
              | ||||||| |    |||||||| | ||||||| |||||||||||| | | ||| | | | ||
Db      78752 ACATACACACACACATACACACACATACACACATATACACACACACACACATACACACAC 78811

Qy        410 ACACATATATACACATATATACACATATATACACATATACACATATATACACATATATAC 469
              | ||| | ||||||| | | ||| | | | ||||| ||||||| ||| ||||| ||| ||
Db      78812 ATACACACATACACACACACACATACACACACACACATACACACATACACACACATACAC 78871

Qy        470 ATATATACACATATATATAATATACACACATATATATACACATATATACACACATATATA 529
              | | | ||||| | |     | | ||||||| | ||| ||||| | | ||||||| | | |
Db      78872 ACACACACACACACACCACACACACACACACACATACACACACACACACACACACACACA 78931

Qy        530 CACATATATACACATATATATACACATATATACACATATATACATATATACACATATATA 589
              |||||| | ||||| ||| | ||||| | | ||| | | | ||| | | | ||| | | |
Db      78932 CACATACACACACACATACACACACACACACACATACACACACACACACATACACACACA 78991

Qy        590 TACATATATACACATATATACATATATACACATATATACATATATACACACATATATACA 649
              ||||| | ||||| | | ||| ||| ||||| | | ||| | |||||||||| | | |||
Db      78992 CACATACACACACACACACACACATACACACACACACACACACATACACACACACACACA 79051

Qy        650 CATACATATACACACACATAGATATACA-TATATATACACATATATATACGTATATATAT 708
              || ||||| ||||||||| | | | |||  | | | ||||| | | | ||  | | || |
Db      79052 CACACATACACACACACACACACACACACCACACACACACACACACACACACACACATGT 79111

Qy        709 GTATATATATATGCTCCAGAGTTCA 733
              | | | |||  |||  |  | ||
Db      79112 GCGTGTGCATACACTCGTGCATGCA 79136
```

RESULT 80
ADC87404/c
ID ADC87404 standard; DNA; 1441 BP.
XX
AC ADC87404;
XX
DT 01-JAN-2004 (first entry)
XX
DE Human GPCR gene SEQ ID NO:1857.
XX
KW ds; gene; human; GPCR;
KW guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1270724-A2.
XX
PD 02-JAN-2003.
XX
PF 18-JUN-2002; 2002EP-00013517.
XX
PR 18-JUN-2001; 2001JP-00246789.
XX
PA (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.

XX
PI Suwa M, Asai K, Akiyama Y, Aburatani H;
XX
DR WPI; 2003-315783/31.
DR P-PSDB; ADC87405.
XX
PT New polynucleotide, useful for preparing a composition for treating a
PT patient in need of increased or suppressed activity or expression of the
PT guanosine triphosphate-binding protein coupled receptor.
XX
PS Claim 1; SEQ ID NO 1857; 28pp; English.
XX
CC The invention relates to a novel polynucleotide encoding a guanosine
CC triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC the invention may have a use in gene therapy. The polynucleotide and
CC polypeptide are useful for preparing a composition for treating a patient
CC in need of increased or suppressed activity or expression of the
CC guanosine triphosphate-binding protein coupled receptor. The
CC polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC invention.
XX
SQ Sequence 1441 BP; 518 A; 214 C; 219 G; 490 T; 0 U; 0 Other;

Query Match 4.9%; Score 241.2; DB 10; Length 1441;
Best Local Similarity 69.5%; Pred. No. 1.8e-32;
Matches 357; Conservative 0; Mismatches 153; Indels 4; Gaps 2;

```
Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | || |||||||||| ||| | |||| |  | | | |||||||||| |||||||||| |||
Db        702 TATACATATATATAGATACACATATGCGTATATATACACATATATGTATACACATACATA 643

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              | |  ||||| ||||||||  | | | ||||| || | |||||| |||  |||  |||
Db        642 TGCGTATATACACATATATGTATACACATACATATGTGTATACATATATGTATACACATG 583

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
               |||| |  | |||| |||  |||||||||   ||||| ||| |    |||| |||||||
Db        582 CATATGTGTATATATTTATGAACATATATATGTATATACACATATGCGTATATACATATA 523

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              |  | ||| ||||  | |||| ||   | | ||| |  ||||||||| |||   ||||
Db        522 TGTATACACATATGTATATATTTATGTACACATATATGTATACACATACATATGTGTATA 463

Qy        467 TACATATATACACATATATATAATATACACACATATATATACACATATATACACACATAT 526
              ||  ||| ||||||||||    |||||||| | |||| |  | |||| |||||| || |
Db        462 TATTTATGTACACATATAT---GTATACACATACATATGTGTATATATTTACACATATGT 406

Qy        527 ATACACATATATACACATATATATACACATATATACACATATATACATATATACACATAT 586
              ||||||| ||| |  | ||||||   | | | ||| || |||||| ||| | |  | |
Db        405 GTACACATACATATATGTGTATATATTTACACACACATATGTATACACATACATATGTGT 346

Qy        587 ATATACATATATACACATATATACATATATACACATATATACATA-TATACACACATATA 645
              |||||  |||  |||||||| || | | ||||  || | || ||| |  ||||| || ||
Db        345 ATATATTTATGCACACATATGTATACACATACGTATGTGTATATATTGCACACATATGTA 286

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              |||||||||||||    | | ||  ||  ||| |||| || ||| ||| ||| || ||||
Db        285 TACACATACATATGTGTATATATTTATGCACACATATGTATACACATACATATGTGTATA 226

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAG 739
              | | |||||| | |||        | |||| | |
Db        225 TGTTTATATACACATGTATACACATGCATATGTG 192


RESULT 81
ADC87576
ID ADC87576 standard; DNA; 1276 BP.
XX
AC ADC87576;
XX
DT 01-JAN-2004 (first entry)
XX
DE Human GPCR gene SEQ ID NO:2029.
XX
KW ds; gene; human; GPCR;
KW guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1270724-A2.
XX
PD 02-JAN-2003.
XX
PF 18-JUN-2002; 2002EP-00013517.
XX
PR 18-JUN-2001; 2001JP-00246789.
XX
PA (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI Suwa M, Asai K, Akiyama Y, Aburatani H;
XX
DR WPI; 2003-315783/31.
DR P-PSDB; ADC87577.
XX
PT New polynucleotide, useful for preparing a composition for treating a
PT patient in need of increased or suppressed activity or expression of the
PT guanosine triphosphate-binding protein coupled receptor.
XX
PS Claim 1; SEQ ID NO 2029; 28pp; English.
XX
CC The invention relates to a novel polynucleotide encoding a guanosine
CC triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC the invention may have a use in gene therapy. The polynucleotide and
CC polypeptide are useful for preparing a composition for treating a patient
CC in need of increased or suppressed activity or expression of the
CC guanosine triphosphate-binding protein coupled receptor. The
CC polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC invention.
XX
SQ Sequence 1276 BP; 416 A; 198 C; 221 G; 441 T; 0 U; 0 Other;

Query Match 4.8%; Score 240.8; DB 10; Length 1276;
Best Local Similarity 69.9%; Pred. No. 2e-32;
Matches 341; Conservative 0; Mismatches 142; Indels 5; Gaps 1;

```
Qy   230 ATATATATATATATATATATATACACACACACACATATATATATACACATATATATAC 289
              | ||||| |||| || |||||||| | ||| | ||| |||||||| || || ||
Db   793 ACATATACATATGTACATATATACATATGTACATATATACGTATATACATATGTACATCT 852

Qy   290 ACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATAT 349
              | ||||||| |||| ||||||| | ||| |||| ||| | | || || ||
Db   853 ATGTATATACGCATATGTGCATATATGTATATACGCATACGTATATGTGTATATACGCAT 912

Qy   350 ATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATAC 409
              ||||||| || || ||| | ||||||| ||||||||| | | ||| | |||||||
Db   913 ATATACATATGTGCATATATGTATATACGCATATATACATATGTGCATATATGTATATAC 972

Qy   410 ACACATATATACACATATATACACATATATACACATATACACATATATACACATATATAC 469
              || ||||| | | | ||| | ||||||| |||||| |||||| | || |||| ||
Db   973 GCATATATACATATGTGCATATATGTATATACGCATATATACATATGTGCATATATGTAT 1032

Qy   470 ATATATACACATATATATAATATACACACATATATATACACATATATACACACATATATA 529
              ||| | | ||| ||| | ||||| | | | || ||| |||| |||
Db   1033 ATACGCATATATA-----CATACATGTGCATATGTGCATATATTCATAAGTACATCTATG 1087

Qy   530 CACATATATACACATATATATACACATATATACACATATATACATATATACACATATATA 589
              | || || | | || || |||||||||| |||| |||| |||||| | | | || |||
Db   1088 TATATGTACATATATGTACATACACATATGTACATATATGTACATACACATATGTACATA 1147

Qy   590 TACATATATACACATATATACATATATACACATATATACATATATACACACATATATACA 649
              || || |||||||||||| | |||| |||| ||| |||||||||| || || |||| |
Db   1148 TATGTACATACACATATACATATATGTACATATACATACATATATGCATGTATGTATATA 1207

Qy   650 CATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATATG 709
                 |||||||| | | ||| ||||| |||||||||||||||| ||| | || |||||
```
```

```
Db       1208 TGCACATATACCTATATATACATATATGTATATATACACATATACATATATGTACATATG 1267

Qy        710 TATATATA 717
              ||||||
Db       1268 TGTATATA 1275


RESULT 82
ABL33869/c
ID   ABL33869 standard; DNA; 13573 BP.
XX
AC   ABL33869;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 1842.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP007537.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful for
PT   diagnosis and treatment of diseases associated with abnormal cytosine
PT   methylation.
XX
PS   Claim 1; SEQ ID NO 1842; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention
XX
SQ   Sequence 13573 BP; 4074 A; 204 C; 2749 G; 6546 T; 0 U; 0 Other;

  Query Match             4.8%;  Score 240.4;  DB 6;  Length 13573;
  Best Local Similarity   69.2%;  Pred. No. 2.7e-32;
  Matches  537;  Conservative    0;  Mismatches  206;  Indels   33;  Gaps   14;

Qy          4 TCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAA--TACAAAATTAG 61
              | || |||| ||| | ||| ||    ||| ||| ||||||| |||||     |||||||
Db       6406 TTAAAACCAACCTAAACAATATAACAAAATCCCGTCTCTACAAAAAAATACAAAAATTAA 6347

Qy         62 ACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAATC 121
               ||  | |  | | |||| ||| |||||||| ||||||   |  |  |   |  | ||||
Db       6346 CCAAACTTAATAACACACACCTATAATCCCAACTACTTAAAAAACTAAAATAAAAAAATC 6287

Qy        122 ACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAGCC 181
              |||| ||||           | |       | ||  ||| | ||| | ||||||| ||
Db       6286 ACTTAAACCCAAAAAATAAAAAC-TACAATAAACCATAATCGTACCACTACACTCCAACC 6228

Qy        182 CAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATAT---A 238
              | ||| || |   ||   | |    | |||||| | |       | |||||||||   |
Db       6227 TAAACAATAAAACAAAACCCTAACAAAAAAAAAAAATATATATATATATATATATAATAA 6168

Qy        239 TATATATATATATACACACACACACATATATATAT-ACACATATATATACACATATAT 297
              || |||||||||||  | | |||  |||||||||| | ||||||||||| || |||||
Db       6167 TACATATATATATACGTATATACATTTATATATATATAATACATATATATATACGTATAT 6108

Qy        298 ACATATATACACATAT-ATACACATATATATACACATACATATGTACACATATATATACA 356
              |||| |||| | |||| ||||| |||||||||| ||| | || || | ||||||| | |
Db       6107 ACATTTATATATATATAATACATATATATATACGTATATACATTTATATATATATA-ATA 6049

Qy        357 CATATGTATACACATATATACACATATAT-------ACACACATATATACACATATATAC 409
              ||||| |||| || |||||||  |||||       | ||| ||||||| || |||||||
Db       6048 CATATATATATACGTATATACATTTATATATATATAATACATATATATATACGTATATAC 5989

Qy        410 ACACATAT-------ATACACATATATACACATATATACACATATACACATATATACACA 462
              |   ||||       ||||| ||||||| || ||||||||  |||| | ||||| | |||
Db       5988 ATTTATATATATATAATACATATATATATACGTATATACATTTATATATATATA-ATACA 5930

Qy        463 TATATACATATATACACATATATATAATATACACACATATATATACACATATATACACAC 522
              |||||| |||||||||  ||||||||     | ||| ||||||||||| ||||| |    |
Db       5929 TATATATATATATACATTTATATATATATAATACATATATATATACATATATACATTTAT 5870

Qy        523 ATATAT---ACACATATATACACATATATATACACATATATACACATATATACATATATA 579
              ||||||   | |||||||| | | ||||||||  |||||| | ||| ||||||||||
Db       5869 ATATATATAATACATATATATATACATATATACATTTATATATATATA-ATACATATATA 5811

Qy        580 CACATATATATACATATATACACAT---ATATACATATATACACATATATACATATATAC 636
              | | |||||||||| |||| | ||   ||| | ||||||| |||||||||||| ||||
Db       5810 TATACATATATACATTTATATATATATAATACATATATATATACATATATACATTTATAT 5751

Qy        637 ACACATATATACACATACATATACACACACATAGATATACATATAT-ATACACATATATA 695
              | | ||| |||| ||| |||||||| | | | |  |||| |||||| ||||| |||||||
Db       5750 ATATATA-ATACATATATATATACATATATACATTTATATATATATAATACATATATATA 5692

Qy        696 TACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGAT 751
              ||| |||||| || ||||||||||  | || |  |    | |  ||    || ||
Db       5691 TACATATATACATTTATATATATATAATACATATATATATACATATATACATTTAT 5636


RESULT 83
ABK83460/c
ID   ABK83460 standard; cDNA; 165199 BP.
XX
AC   ABK83460;
XX
DT   14-AUG-2002  (first entry)
XX
DE   Human cDNA differentially expressed in granulocytic cells #31.
XX
KW   Human; ss; granulocytic cell; DNA chip; bacterial infection;
KW   viral infection; parasitic infection; protozoal infection;
KW   fungal infection; sterile inflammatory disease; psoriasis;
KW   rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KW   cardiac reperfusion injury; renal reperfusion injury; ARDS;
KW   adult respiratory distress syndrome; inflammatory bowel disease;
KW   Crohn's disease; ulcerative colitis; periodontal disease;
KW   granulocyte activation; chronic inflammation; allergy.
XX
OS   Homo sapiens.
XX
PN   WO200228999-A2.
XX
PD   11-APR-2002.
XX
PF   03-OCT-2001; 2001WO-US030821.
XX
PR   03-OCT-2000; 2000US-0237189P.
XX
```

PA (GENE-) GENE LOGIC INC.
XX
PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
XX
DR WPI; 2002-435328/46.
XX
PT Detecting granulocyte activation by detecting differential expression of
PT genes associated with granulocyte activation, which serves as diagnostic
PT markers that is useful for monitoring disease states and drug toxicity.
XX
PS Claim 1; SEQ ID NO 31; 114pp; English.
XX
CC The invention relates to detecting (M1) granulocyte (GC) activation
CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
CC DNA chip analysis as given in the specification, and comparing the
CC expression level to an expression level in an unactivated GC, where
CC differential expression of Gs is indicative of GCA. Also included are
CC modulating (M2) GA by contacting GC with an agent that alters the
CC expression of at least one gene in Gs; (2) screening (M3) for an agent
CC capable of modulating GCA or an inflammation (especially chronic) in a
CC tissue, an allergic response in a subject, exposure of a subject to a
CC pathogen or sterile inflammatory disease using the gene expression
CC profile; (3) detecting (M4) an inflammation (especially chronic) in a
CC tissue, an allergic response in a subject, exposure of a subject to a
CC pathogen or sterile inflammatory disease, by detecting the level of
CC expression in a sample of the tissue of gene(s) from Gs, where the level
CC of expression of the gene is indicative of inflammation; (4) treating
CC (M5) an inflammation (especially chronic) or in a tissue, an allergic
CC response in a subject, exposure of a subject to a pathogen or sterile
CC inflammatory disease, by contacting a tissue having inflammation with an
CC agent that modulates the expression of gene(s) from Gs in the tissue. M1
CC is useful for detecting GCA; M2 is useful for modulating GA; M3 is useful
CC for screening an agent capable of modulating GCA preferably in an
CC inflammation in a tissue; M4 is useful for detecting an inflammation
CC (especially chronic) in a tissue, an allergic response in a subject,
CC exposure of a subject to a pathogen or sterile inflammatory disease (e.g.
CC psoriasis, rheumatoid arthritis, glomerulonephritis, asthma, thrombosis,
CC cardiac reperfusion injury, renal reperfusion injury, ARDS, adult
CC respiratory distress syndrome, inflammatory bowel disease, Crohn's
CC disease, ulcerative colitis, periodontal disease; also bacterial
CC infection, viral infection, parasitic infection, protozoal infection,
CC fungal infection and M5 is useful for treating one of the above
CC conditions. The present sequence represents a gene differentially
CC expressed in granulocytes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 165199 BP; 48510 A; 33223 C; 34406 G; 49060 T; 0 U; 0 Other;

Query Match 4.8%; Score 240.2; DB 6; Length 165199;
Best Local Similarity 66.3%; Pred. No. 3.4e-32;
Matches 377; Conservative 0; Mismatches 186; Indels 6; Gaps 2;

```
Qy      234 ATATATATATATATATATACACACACACACATATATATATACACATATATATACACAT 293
            ||  |||||||||||||| ||||||  | | | |||||||||||| | |||||||||  |||
Db   139452 ATGGATATATATATATACACACGGATATATATATATATATATAGATATATATATCCAT 139393

Qy      294 ATATACATATATACACATATATACACATATATATACACATACATATGTACACATATATAT 353
            ||||  |||||| |   ||||| | || ||||||   ||| |||| |  | ||||  ||
Db   139392 ATATGGATATATATATCGATATATATATCTATATATGGATATATATATCTATATATGGAT 139333

Qy      354 ACACATATGTATACACATATATACACATATATACACACATATATACACATATATACACAC 413
            | | ||||| ||| | |||| || | ||  ||| | |   |||| | ||  ||| | |
Db   139332 ATATATATGGATATATATATCTATATATGGATATATATGGATATATATATGGATATATAT 139273

Qy      414 ATATATACACATATATACACATATATACACATATACACATATATACACATATATACATAT 473
            ||||||  |||||| |   ||||| | || || | ||  ||| | ||||  | ||||
Db   139272 CTATATATGGATATATATATGGATATATATATCTATATATGGATATATATATGGATATAT 139213

Qy      474 ATACACATATATATA-ATATACACACATATATATACACATATATACACACATATATACAC 532
            |||  ||||||  | ||||| |  |||||||||  ||||||  | | ||||||  |
Db   139212 ATATCTATATATGGATATATATATGGATATATATATCTATATATGGATATATATATGGAT 139153

Qy      533 ATATATACACATATATATACACATATATACACATATATACATATATACACATATATATAC 592
            |||||||  ||||||  | | ||||||  | |||||||  ||||||  |||||||  | |
Db   139152 ATATATATCTATATATGGATATATATATGGATATATATAACTATATATGGATATATACAT 139093

Qy      593 ATATATACACATATATACATATATACACATATATACATATATACACACATATATACACAT 652
            ||||||  |||||||  ||  ||     ||||||  | |||||  | | |||||||   ||
Db   139092 CTATATATGGATATATATATTGAT-----TATATATAGATATATAGATATATATATGGAT 139038

Qy      653 ACATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTAT 712
            | |||||   | | ||  ||||||  ||  |||||  | ||  ||||||  | |||||||| ||
Db   139037 ATATATATCTATATATGGATATATATGGATATATATATCTATATATGGATATATATGGAT 138978

Qy      713 ATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAA 772
            |||||||||    | |  |    | |  |||   |  ||    | |    |||||   |
Db   138977 ATATATATGGATTATATATAGATATATATAGATATATATATGGATTATATATAGATATAT 138918

Qy      773 AGAGAGTTTGACAGCAGTGTATTGTGAGA 801
            | |||  |  | | |  | |||  |  | |
Db   138917 ATAGATATATATATCTATATATGGATATA 138889
```

RESULT 84
ABD32673
ID ABD32673 standard; DNA; 91359 BP.
XX
AC ABD32673;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human cancer-associated genomic DNA HD13-075.
XX
KW Human; ds; cancer-associated protein; gene; cytostatic; cancer;
KW leukaemia; lymphoma; CAP.
XX
OS Homo sapiens.
XX
PN WO2004074320-A2.
XX
PD 02-SEP-2004.
XX
PF 17-FEB-2004; 2004WO-US004730.
XX
PR 14-FEB-2003; 2003US-00367094.
PR 14-MAR-2003; 2003US-00388838.
PR 15-APR-2003; 2003US-00417375.
PR 13-JUN-2003; 2003US-00461862.
PR 15-SEP-2003; 2003US-00663431.
PR 15-DEC-2003; 2003US-00737318.
XX
PA (SAGR-) SAGRES DISCOVERY INC.
XX
PI Morris DW, Morris DW, Malandro MS;
XX
DR WPI; 2004-652914/63.
XX
PT New isolated cancer-associated polynucleotides and polypeptides useful
PT for diagnosing, preventing or treating cancers, especially lymphoma and
PT leukemia, or in screening for agents that modulate cancer.
XX
PS claim 16; seqid 221; 310pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising at least 10
CC contiguous nucleotides of any of the 233 polynucleotide sequences given
CC in the specification, or its complement. The nucleic acids encode cancer-
CC associated proteins. Also included are an expression vector comprising
CC the isolated nucleic acid cited above, a host cell comprising the above
CC recombinant nucleic acid or expression vector, a microarray for detecting
CC a cancer-associated (CA) nucleic acid comprising at least one probe
CC comprising at least 10 contiguous nucleotides of any of the above-
CC mentioned nucleotide sequences, an isolated polypeptide (encoded within
CC an open reading frame of a CA sequence selected from any of the 95
CC polynucleotide sequences as mentioned in the specification, or its

```
CC   complement), an isolated antibody, (or its antigen binding fragment) that
CC   binds to the above polypeptide, a hybridoma that produces the above
CC   monoclonal antibody, a pharmaceutical composition comprising the above
CC   antibody and a pharmaceutical excipient, a kit for detecting cancer
CC   cells(comprising the antibody cited above, methods for diagnosing cancer
CC   or for detecting the presence or absence of cancer cells in an
CC   individual, a method for inhibiting growth of cancer cells in an
CC   individual, a method for delivering a therapeutic agent to cancer cells
CC   in an individual, an electronic library comprising the above
CC   polynucleotide or polypeptide (or their fragments), methods of screening
CC   for anticancer activity or for a bioactive agent capable of modulating
CC   the activity of a CA protein (CAP), methods for detecting cancer
CC   associated with expression of a polypeptide in a test cell sample, a
CC   method for treating cancers and a method for inhibiting the expression of
CC   CA gene in a cell. The composition and methods are useful for detecting,
CC   diagnosing, preventing and treating cancers, especially lymphoma and
CC   leukaemia. These may also be used in screening for agents that modulate
CC   cancer. The present sequence is a human CAP genomic sequence. Note: The
CC   sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 91359 BP; 24503 A; 18193 C; 19029 G; 29634 T; 0 U; 0 Other;

  Query Match             4.8%;  Score 238.8;  DB 13;  Length 91359;
  Best Local Similarity   68.8%;  Pred. No. 5.7e-32;
  Matches  373;  Conservative    0;  Mismatches  162;  Indels    7;  Gaps    3;

Qy        204 ATCTCAGAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACAC 263
              | || || |  |   || ||  | |     || |||||| || |||| |   | | |.
Db      19526 AGCTGAGTAGCATTCAATGGTGTATATATAATCTATATACATCTATACATCTATATATCT 19585

Qy        264 ACATATATATATACACATATATATACACATATATACATATATACACATATATACACATAT 323
              | ||| |||||||||| || ||| || | || | || ||| ||| | | ||||  | ||||
Db      19586 ATATACATATATACATATGTATGTATATATTTCTATATACATATATACATATGTATATAT 19645

Qy        324 ATATACACATACATATGTACACATATATATACACATATGTATACACATATATACACATAT 383
              | ||  | |||||| | || |||||| |||||| | |||||||||  ||| || |||||
Db      19646 AGATTGATATACATCTATATACATATGTATACATAGATGTATACATCTATGTATACATAG 19705

Qy        384 ATACACACATATATACACATATATACACACATATATACACATATATACACATATATACAC 443
              ||  | |||||       || |||| ||| |||| || ||||||||||| || ||||||
Db      19706 ATGTATACATA-----CTATGTATATACATATATGTATACATATATACATATGTATACAT 19760

Qy        444 ATATACACATATATACACATATATACATATATACACATATATATA-ATATACACACATAT 502
              |||  | | ||||| |||||| || ||||||  | ||||| ||| || || | |  |||
Db      19761 CTATGTATACATATATACATATGTATATATATCTATATATACATATATGTATATATGTAT 19820

Qy        503 ATATACACATATATACACACATATATACACATATATACACATATATATACACATATATAC 562
              ||||  | |||  || | |  |||||||| ||||| | | || |||||||| || ||||
Db      19821 ATATCTATATACGTATATAGGTATATACATATATAAATATATCTATATACATATGTATAT 19880

Qy        563 ACATATATACATATATACACATATATATACATATATACACATATATACATATATACACAT 622
              |  |||||| ||||||  | |||||| |||||||||  | ||| ||| || |||| |  |
Db      19881 ATCTATATAAATATATGGATATATATCTACATATATCTATATACATATATCTATATATCT 19940

Qy        623 ATATACATATATACACACATATATACACAT-ACATATACACACACATAGATATACATATA 681
              |||||||||| |   | || |||  ||||  |||||  | | | ||| |||||||| ||
Db      19941 ATATACATATAGATGTATATGTATCTACATGGCATATGTATATAGATATATATACATCTA 20000

Qy        682 TATACACATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGT 741
              |||  | || ||||||| ||| |||||   | |||||||   ||| |  |   | || |
Db      20001 TATGTATATCTATATACATATCTATATAGCTGTATATATATGCCATATATATAGATAGAT 20060

Qy        742 AG 743
              ||
Db      20061 AG 20062


RESULT 85
ABL51954
ID   ABL51954 standard; DNA; 40023 BP.
XX
AC   ABL51954;
XX
DT   11-JUL-2002  (first entry)
XX
DE   Human solute carrier family 18 member 2 (SLC18A2) gene SEQ ID NO:1.
XX
KW   Human; solute carrier family 18 member 2; SLC18A2; vesicular monoamine;
KW   vesicular monoamine transporter; VMAT2; polymorphic site; SNP;
KW   single nucleotide polymorphism; antiinflammatory; neuroleptic;
KW   haplotyping; genotyping; respiratory inflammatory disease;
KW   neuropsychiatric disorder; monoaminergic brain system; gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    4024
FT                   /*tag=  a
FT                   /note= "PS1: polymorphic base C or T"
FT   misc_feature    6142
FT                   /*tag=  b
FT                   /note= "PS2: polymorphic base C or G"
FT   misc_feature    6211
FT                   /*tag=  c
FT                   /note= "PS3: polymorphic base G or A"
FT   misc_feature    6213
FT                   /*tag=  d
FT                   /note= "PS4: polymorphic base G or T"
FT   misc_feature    6397
FT                   /*tag=  e
FT                   /note= "PS5: polymorphic base C or T"
FT   misc_feature    6398
FT                   /*tag=  f
FT                   /note= "PS6: polymorphic base T or G"
FT   misc_feature    16326
FT                   /*tag=  g
FT                   /note= "PS7: polymorphic base T or C"
FT   misc_feature    16464
FT                   /*tag=  h
FT                   /note= "PS8: polymorphic base C or T"
FT   misc_feature    16855
FT                   /*tag=  i
FT                   /note= "PS9: polymorphic base T or C"
FT   misc_feature    16959
FT                   /*tag=  j
FT                   /note= "PS10: polymorphic base T or A"
FT   misc_feature    16997
FT                   /*tag=  k
FT                   /note= "PS11: polymorphic base T or C"
FT   misc_feature    17761
FT                   /*tag=  l
FT                   /note= "PS12: polymorphic base C or T"
FT   misc_feature    17763
FT                   /*tag=  m
FT                   /note= "PS13: polymorphic base T or C"
FT   misc_feature    17780
FT                   /*tag=  n
FT                   /note= "PS14: polymorphic base A or C"
FT   misc_feature    17897
FT                   /*tag=  o
FT                   /note= "PS15: polymorphic base G or C"
FT   misc_feature    20047
FT                   /*tag=  p
FT                   /note= "PS16: polymorphic base C or T"
FT   misc_feature    20070
FT                   /*tag=  q
FT                   /note= "PS17: polymorphic base T or A"
FT   misc_feature    29050
FT                   /*tag=  r
FT                   /note= "PS18: polymorphic base G or T"
FT   misc_feature    29230
FT                   /*tag=  s
FT                   /note= "PS19: polymorphic base C or T"
FT   misc_feature    29395
```

```
FT                   /*tag=  t
FT                   /note= "PS20: polymorphic base C or T"
FT   misc_feature    32343
FT                   /*tag=  u
FT                   /note= "PS21: polymorphic base C or T"
FT   misc_feature    32572
FT                   /*tag=  v
FT                   /note= "PS22: polymorphic base A or G"
FT   misc_feature    39464
FT                   /*tag=  w
FT                   /note= "PS23: polymorphic base T or C"
FT   misc_feature    39766
FT                   /*tag=  x
FT                   /note= "PS24: polymorphic base C or T"
FT   misc_feature    39848
FT                   /*tag=  y
FT                   /note= "PS25: polymorphic base A or C"
FT   misc_feature    39910
FT                   /*tag=  z
FT                   /note= "PS26: polymorphic base T or C"
FT   misc_feature    39984
FT                   /*tag=  aa
FT                   /note= "PS27: polymorphic base T or C"
XX
PN   WO200222652-A2.
XX
PD   21-MAR-2002.
XX
PF   17-SEP-2001; 2001WO-US042217.
XX
PR   15-SEP-2000; 2000US-0232895P.
XX
PA   (GENA-) GENAISSANCE PHARM INC.
XX
PI   Anastasio AE,  Han J,  Kliem SE,  Sausker EA;
XX
DR   WPI; 2002-393942/42.
XX
PT   Novel genetic variants of soluble carrier family 18 (vesicular
PT   monoamine), member 2 gene useful for screening drugs to treat diseases
PT   e.g. neuropsychiatric disorders involving monoaminergic brain systems.
XX
PS   Claim 15; Fig 1; 183pp; English.
XX
CC   The present invention describes an isolated polynucleotide (I) having a
CC   sequence (S1) comprising soluble carrier family 18 (vesicular monoamine),
CC   member 2 (SLC18A2) isogene selected from 49 isogenes with regions of a
CC   sequence (SS) of 40023 bp (see ABL51954), and defined by a corresponding
CC   set of polymorphisms whose locations and identities are given in the
CC   specification; or a sequence (S2) complementary to (S1). (I) has
CC   antiinflammatory and neuroleptic activities, and can be used in gene
CC   therapy. Methods from the present invention can be used for haplotyping
CC   and genotyping the SLC18A2 gene in an individual. SLC18A2 is also known
CC   as the vesicular monoamine transporter (VMAT2). (I) is useful in studying
CC   the expression and function of SLC18A2, and in expressing the SLC18A2
CC   protein for use in screening for candidate drugs to treat diseases
CC   related to SLC18A2 activity and in studying the effect of the variation
CC   on the biological activity of SLC18A2 as well as on the binding affinity
CC   of candidate drugs targeting SLC18A2 for the treatment of respiratory
CC   inflammatory diseases such as neuropsychiatric disorders involving
CC   monoaminergic brain systems. The present sequence represents the human
CC   SLC18A2 gene sequence, indicating the polymorphic sites in the sequence
CC   by the ambiguity code
XX
SQ   Sequence 40023 BP; 10020 A; 9128 C; 9557 G; 11291 T; 0 U; 27 Other;

  Query Match             4.8%;  Score 237.4;  DB 6;  Length 40023;
  Best Local Similarity   71.2%;  Pred. No. 9.6e-32;
  Matches  341;  Conservative    0;  Mismatches  136;  Indels    2;  Gaps    2;

Qy        230 ATATATATATATATATATATATACACACACACACACATATATATATACACATATATATAC 289
              | ||| | ||  ||| | | | |||||||| || ||| | ||| | ||||  | | |  |
Db      31298 ACATACACATTCATACACAAACACACACACTCATACACACATACACACACTGACACACGC 31357

Qy        290 ACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATAT 349
              | | | | ||| ||| ||||||| |  | |  | ||| ||||||||| ||  ||| ||||
Db      31358 ATACACACACACATACACACATACACCCCCCCACATAAACACATACACATACACAAATAT 31417

Qy        350 ATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATAC 409
              | ||||||| || |  | ||||| |||||| || ||||||| | | ||||| | |    ||
Db      31418 ACATACACACATATGCATACATACATACACGTAGATACACATACACATACAAACACCCAC 31477

Qy        410 ACACATATATACACATATATACACATATATACACATATACACATATATACACATATATAC 469
              |||||| |||  || | ||||||| | | | |||||||||||_||| ||||| |  |||
Db      31478 CCACATACATATGCACACATACACACACACATACATATACACACATACACACACCTAC 31537

Qy        470 ATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATATAT 528
              | | |||||||||| | | | | | | | |||  || ||||||| | ||| ||||| | |
Db      31538 ACACATACACATACACACACACACAAATACAAGTACATACACAAACATATACACAGACAC 31597

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACATATAT 588
              | ||| ||  | ||| || || |||||||||||||||||| ||||| | ||||| |||||
Db      31598 ATACACATGCAAACACATGTACACACATATATACACATACATACACAGATACATATATAC 31657

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
              ||| | ||| | ||| ||| | | | | |  |||| | ||||||| | | ||||| | |
Db      31658 ATATACATA-ATACACATAAATACACACATGCATACACACATATACATAGACATACACAG 31716

Qy        649 ACATACATATACACACACATAGATATACATATATATATACACATATATATACGTATATATA 707
              | | | ||||||| | ||||| ||| |||| ||| ||||||||| |||| | ||| |
Db      31717 ATACATATATACATATACATACATACTCATAAATACACACATATACATACAAACATACA 31775


RESULT 86
ADR07428
ID   ADR07428 standard; cDNA; 1821 BP.
XX
AC   ADR07428;
XX
DT   04-NOV-2004  (first entry)
XX
DE   Full length human cDNA useful for treating neurological disease Seq 934.
XX
KW   gene; ss; human; oligo-capping method; diagnostic marker; gene therapy;
KW   osteoporosis; neurological disease; Alzheimer's disease;
KW   Parkinson's disease; dementia; short memory; cancer;
KW   sense or motor function; emotional reaction; fear response; panic;
KW   osteopathic; neuroprotective; nootropic; antiparkinsonian; cytostatic;
KW   tranquiliser.
XX
OS   Homo sapiens.
XX
PN   EP1447413-A2.
XX
PD   18-AUG-2004.
XX
PF   12-FEB-2004; 2004EP-00003145.
XX
PR   14-FEB-2003; 2003JP-00102207.
PR   09-MAY-2003; 2003JP-00131452.
XX
PA   (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI   Isogai T,  Yamamoto J,  Nishikawa T,  Isono Y,  Sugiyama T,  Otsuki T;
PI   Wakamatsu A,  Ishii S,  Nagai K,  Irie R;
XX
DR   WPI; 2004-583265/57.
DR   P-PSDB; ADR09384.
XX
PT   New 1995 cDNA, useful for treating osteoporosis, neurological diseases,
PT   Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.
XX
PS   Claim 1; SEQ ID NO 934; 2686pp; English.
XX
CC   This invention relates to novel, isolated full length human cDNA
CC   molecules and the encoded proteins thereof. Specifically, it refers to
```

```
CC   cDNA clones obtained by an oligo-capping method, where none of these
CC   clones are identical to any known human mRNAs. The present invention
CC   describes an immunoassay to identify agonists and antagonists, as well as
CC   antibodies, antisense molecules and siRNAs that can all be used to bind
CC   to and modulate expression of the cDNA molecules. As such, these
CC   molecules are useful for diagnostic markers or therapeutic targets for
CC   the various diseases or morbid states. In particular, they are useful in
CC   gene therapy for treating osteoporosis, neurological disease, Alzheimer's
CC   disease, Parkinson's disease, dementia, short memory and various cancers,
CC   as well as for maintaining equilibrium of sense or motor function, and
CC   for treating emotional reaction, fear response and panic. Accordingly,
CC   they exhibit osteopathic, neuroprotective, nootropic, antiparkinsonian,
CC   cytostatic and tranquiliser activities. This polynucleotide is a full
CC   length human cDNA sequence of the invention. NOTE: This sequence is not
CC   given in the sequence listing of the specification but can be obtained on
CC   CD-ROM from the European Patent Office, Vienna Sub-office.
XX
SQ   Sequence 1821 BP; 752 A; 616 C; 149 G; 304 T; 0 U; 0 Other;

  Query Match             4.8%;  Score 236.6;  DB 13;  Length 1821;
  Best Local Similarity   69.0%;  Pred. No. 1.1e-31;
  Matches  339;  Conservative    0;  Mismatches  149;  Indels    3;  Gaps    1;

Qy        230 ATATATATATATATATATATACACACACACACATATATATATACACATATATATAC 289
              | | ||  | ||| |    | || |||||||| ||||||  | ||||||||||  | | ||
Db         44 ACACATGCACATACACTCACATGCACACACATACACATGCACATATACACATGCACACAC 103

Qy        290 ACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATAT 349
                |||  ||||||| | || || ||||||||||  | |||||||  ||||   ||||| ||
Db        104 GCATGCATACATACACACGCACATATACACATGCACATACACACTCATACACACACACAT 163

Qy        350 ATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATAC 409
              ||| |||||||   || || | | || |||  ||  ||||||| | ||||||||  | |
Db        164 ATACACACATACACATGCATACACATGCACTCATGCACACACACACATACACATGCACAT 223

Qy        410 ACACATATATACACATATATACACATATATACACATATACACATATATACACATATATAC 469
              ||||| | || || | |||| |||||||| |||   | | |||||  |  |||||  | ||
Db        224 ACACAAACATGCAGAGATATGCACATATACACATGCACATACATACACTCACATGCACAC 283

Qy        470 ATATATACACATATATAT---AATATACACACATATATATACACATATATACACACATAT 526
              ||| | ||||||| | ||    ||||||| |||| | ||||||| | || |||||| |
Db        284 ATAGACACACATACACATGCACATATACATACATGCACATACACAAACATGCACACAGAC 343

Qy        527 ATACACATATATACACATATATATACACATATATACACATATATACATATATACACATAT 586
              | ||| | | ||||||| ||  | | ||| | ||||||  |||| |||| ||| ||| ||
Db        344 ACACATACACATACACACATGCACATACACAGATACACTCATATGCATACATATATACAT 403

Qy        587 ATATACATATATACACATATATACATATATACACATATATACATATATACACACATATAT 646
              |   ||||||| |||   | ||||| ||| ||||||  || || ||||||||||||  |
Db        404 ACGCACATATACACATGAACATACACATACACACATGCATGCACATATACACACATGCAC 463

Qy        647 ACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATAT 706
              ||||| |   | | | ||||| |  ||||| | | || |||| | | | ||  | ||
Db        464 ACACACATGCACATATACACACATGCATACACACACATGCACACAGACACACACACATGC 523

Qy        707 ATGTATATATA 717
              |   | |||  |
Db        524 ACACACATACA 534


RESULT 87
ACC79619/c
ID   ACC79619 standard; DNA; 71843 BP.
XX
AC   ACC79619;
XX
DT   27-OCT-2003  (revised)
DT   06-AUG-2003  (first entry)
XX
DE   Zebrafish protein kinase C lambda related genomic DNA SEQ ID NO:3.
XX
KW   Zebrafish; Danio rerio; protein kinase C lambda; PKC-lambda; enzyme;
KW   cardiant; cytostatic; nephrotropic; heart disease; cancer; organ failure;
KW   epithelial-epithelial cell interaction; epithelial cell polarity;
KW   heart failure; kidney failure; gene; ds.
XX
OS   Danio rerio.
XX
PN   WO2003023048-A2.
XX
PD   20-MAR-2003.
XX
PF   06-SEP-2002; 2002WO-US028410.
XX
PR   06-SEP-2001; 2001US-0317653P.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Peterson R,  Fishman MC;
XX
DR   WPI; 2003-313257/30.
XX
PT   Diagnosing or prognosing a disease or condition related to Protein Kinase
PT   C lambda, by analyzing sample nucleic acid obtained from the test subject
PT   to determine mutation in gene encoding Protein kinase C lambda.
XX
PS   Disclosure; Page 63-83; 88pp; English.
XX
CC   The present invention describes a method (M1) for determining whether a
CC   test subject has, or is at risk of developing, a disease or condition
CC   related to protein kinase C lambda (PKC-lambda) (I). M1 comprises
CC   analysing a nucleic acid molecule of a sample from the subject to
CC   determine whether the subject has a mutation in a gene encoding (I),
CC   where the presence of a mutation indicates that the subject has, or is at
CC   risk of developing, a disease or condition related to (I). (I) has
CC   cardiant, cytostatic and nephrotropic activities. M1 is useful for
CC   determining whether a test subject has, or is at risk of developing, a
CC   disease or condition related to (I), e.g. heart disease, cancer, or
CC   condition associated with epithelial-epithelial cell interactions or
CC   epithelial cell polarity. (I) is useful for treating organ failure such
CC   as heart or kidney failure. (I) is also useful for raising antibodies
CC   specific for various regions of (I). M1 is suitable for detecting an
CC   increased likelihood of diseases or conditions associated with (I), so
CC   that appropriate intervention can be instituted before any symptoms
CC   occur. This is useful with patients in high risk groups for such diseases
CC   or conditions. Diagnostic method facilitates determination of the
CC   etiology of such an existing disease or condition in a patient, so that
CC   an appropriate approach to treatment can be selected. The present
CC   sequence represents zebrafish genomic DNA for PKC-lambda, which is used
CC   in the exemplification of the present invention. (Updated on 27-OCT-2003
CC   to standardise OS field)
XX
SQ   Sequence 71843 BP; 24017 A; 12903 C; 12592 G; 21955 T; 0 U; 376 Other;

  Query Match             4.7%;  Score 236;  DB 8;  Length 71843;
  Best Local Similarity   60.9%;  Pred. No. 1.7e-31;
  Matches  314;  Conservative    0;  Mismatches  198;  Indels    4;  Gaps    2;

Qy        209 AGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACATA 268
              | ||||||||||    |  | |||||||||||||||||||||||| | | | | | |||
Db      68772 AAAAAAAAAAAATNTCTATATATATATATATATATATATATATATATATATATATATA 68713

Qy        269 TATATATACACATATATATACACATATATACATATATACACATATATACACATATATATA 328
              |||||||| | |||||||| | ||||||| ||||||| | ||||||| | |||||||||
Db      68712 TATATATATATATATATATATATATATANATATATATATATATATATATATATATATA 68653

Qy        329 CACATACATATGTACAC---ATATATATACACATATGTATACACATATATACACATATAT 385
              | |||||||| |||||   |||||||| | |||| |||| | ||||||| | ||||||
Db      68652 TATATACATATATACACTATATATATATATATATATATATATATATATATATATATAT 68593

Qy        386 ACACACATATATACACATATATACACACATATATACACATATATACACATATATACACAT 445
              | | | ||||||| | ||||||| | | ||||||| | ||||||| | ||||||| | ||
Db      68592 ATATATATATATATATATATATATATANATATATATATATATATATATATATATATAT 68533

Qy        446 ATACACATATATACACATATATACATATATACACATATATATA-ATATACACACATATAT 504
              ||| | ||||||| | ||||||| ||||||| | ||||||||| ||||| | | ||||||
```

```
Db      68532 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 68473

Qy        505 ATACACATATATACACACATATATACACATATATACACATATATATACACATATATACAC 564
              ||| | ||||||
Db      68472 ATATATATATATNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 68413

Qy        565 ATATATACATATATACACATATATATACATATATACACATATATACATATATACACATAT 624
                                                                        | |
Db      68412 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTGT 68353

Qy        625 ATACATATATACACACATATATACACATACATATACACACACATAGATATACATATATAT 684
                |   | | |     | | |     |  ||||| | | | ||| ||||| ||||||||
Db      68352 GTGTGTGTGTGTGTGTGTGTGTGTGTATATATATATATATATATATATATATATAT 68293

Qy        685 ACACATATATATACGTATATATATGTATATATATAT 720
              | | |||||||||  |||||||||| ||||||||||
Db      68292 ATATATATATATATATATATATATATATATATATAT 68257


RESULT 88
ADC87374/c
ID   ADC87374 standard; DNA; 1192 BP.
XX
AC   ADC87374;
XX
DT   01-JAN-2004  (first entry)
XX
DE   Human GPCR gene SEQ ID NO:1827.
XX
KW   ds; gene; human; GPCR;
KW   guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1270724-A2.
XX
PD   02-JAN-2003.
XX
PF   18-JUN-2002; 2002EP-00013517.
XX
PR   18-JUN-2001; 2001JP-00246789.
XX
PA   (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA   (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI   Suwa M,  Asai K,  Akiyama Y,  Aburatani H;
XX
DR   WPI; 2003-315783/31.
DR   P-PSDB; ADC87375.
XX
PT   New polynucleotide, useful for preparing a composition for treating a
PT   patient in need of increased or suppressed activity or expression of the
PT   guanosine triphosphate-binding protein coupled receptor.
XX
PS   Claim 1; SEQ ID NO 1827; 28pp; English.
XX
CC   The invention relates to a novel polynucleotide encoding a guanosine
CC   triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC   the invention may have a use in gene therapy. The polynucleotide and
CC   polypeptide are useful for preparing a composition for treating a patient
CC   in need of increased or suppressed activity or expression of the
CC   guanosine triphosphate-binding protein coupled receptor. The
CC   polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC   invention.
XX
SQ   Sequence 1192 BP; 384 A; 185 C; 211 G; 412 T; 0 U; 0 Other;

  Query Match             4.7%;  Score 235.8;  DB 10;  Length 1192;
  Best Local Similarity   69.2%;  Pred. No. 1.5e-31;
  Matches  355;  Conservative    0;  Mismatches  147;  Indels   11;  Gaps    2;

Qy        214 AAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACATATATAT 273
              |||||| |     | |||| |||||||| | ||||| || | ||||||| |||| |||||

Db       1000 AAAAAATATTATATATATACATATATATGTTTATATGTATATACACACATACATGTATAT 941

Qy        274 ATACACATATATATACACATATATACATATATACACATATATACACATATATATACACAT 333
              ||  | ||||| || ||| |||||| | | ||| ||| || ||      ||||||||||
Db        940 ATGTATATATACATGCACGTATATATACACATATACACATGTATGTGCGTATATACACAT 881

Qy        334 ACATATGTACACATATATATACACATATGTATACACATATATACACATATATACACACAT 393
              |||| | || | | | |||||||||| | | | |  ||||||||    ||||| ||||||
Db        880 ACATGTATATATACACATATACACATGTATGTGCGTATATATAC----ATATATACACAT 825

Qy        394 ATATACACATATATACACACATATATACACATATATACACATATATACACATATACACAT 453
              |||   | | |||| ||| ||||| |||  | | | |  ||||||| ||||||||||
Db        824 GTATGTGCGTGTATATACATATATACACATGTGTGTGCGTGTATATACATATATACACAT 765

Qy        454 ATATACACATATATACATATATACACATATATAT-------AATATACACACATATATAT 506
              | |   | | |||| | ||||| ||| | ||||        | || | |||||||||
Db        764 GTGTGTGCGTGTATATACATATATACACAGATATGTGTGTGCGTGTATATACATATATAC 705

Qy        507 ACACATATATACACACATATATACACATATATACACATATATATACACATATATACACAT 566
              ||| |||| |       |||||||| ||| | ||| |||| | |    |||| |||| ||
Db        704 ACAGATATGTGTGTGTATATATACATATACACACAGATATGTGTGTGTATATGTACATAT 645

Qy        567 ATATACATATATACACATATATATACATATATACACATATATACATATATACACATATAT 626
              | | |||| | |  | ||||| ||| ||| ||| ||| | ||   | |||| | | ||||
Db        644 ACACACATGTGTGTATATATACATATATACATATACACACATGTGTGTATATATACATAT 585

Qy        627 ACATATATACACACATATATACACATACATATACACACACATAGATATACATATATATAC 686
              | | |||||||||| | | || | ||||||||| ||| | | | |  |  |||||||||
Db        584 ATACATATACACACGTGTGTATATATACATATATACATATACACACGTGTGTATATATAC 525

Qy        687 ACATATATATACGTATATATATGTATATATATA 719
              | ||||| | || || || || || ||||||||
Db        524 ATATATACACACATACATGTACGTGTATATATA 492


RESULT 89
ABL33998/c
ID   ABL33998 standard; DNA; 21537 BP.
XX
AC   ABL33998;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 1971.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP007537.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful for
PT   diagnosis and treatment of diseases associated with abnormal cytosine
```

```
PT   methylation.
XX
PS   Claim 1; SEQ ID NO 1971; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention
XX
SQ   Sequence 21537 BP; 5893 A; 333 C; 4680 G; 10631 T; 0 U; 0 Other;

  Query Match             4.7%;  Score 235.2;  DB 6;  Length 21537;
  Best Local Similarity   69.6%;  Pred. No. 2.2e-31;
  Matches  348;  Conservative    0;  Mismatches  148;  Indels    4;  Gaps    2;

Qy        197 AAATGTCATCTCAGAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACA 256
              ||||   | | || |||||  | |   | | |   | |      |||| ||  ||| | |
Db      16071 AAATTAAAACACACAAAAACTATATAACTTATCCAAAAACACACTATAAATCCATATATA 16012

Qy        257 CACACACACATATATATATACACATATATATACACATATATACATATATACACATATATA 316
              | | | | |||||||||||| | ||||| | | | ||| || |||||||| | |||||
Db      16011 TATATATATATATATATATATATATATACACATATATAAATCCATATATATATATATACG 15952

Qy        317 CACATATATATACACATACATATGTACACATATATATACACATATGTATACACATATATA 376
              || |||||||   | ||| ||||  ||||| |||||||||||||| || ||||||||||
Db      15951 CATATATATACGTATATATATATACACACACATATATACACATATATACACACATATATA 15892

Qy        377 CACATATATACACACATATATACACATATATACACACATATATACACATATATACACATA 436
              |||||||| ||| | ||||||| | |||||||||||||| | ||| | ||||||||||| |
Db      15891 CACATATACACATATATATATATATATATACACACACACATATATATATACACACA 15832

Qy        437 TATACACATATACACATATATACACATATATACATATATACACATATATATA---ATATA 493
              | | | ||||| | | | | |||||||||| |||||||||| | | |||   |||||
Db      15831 CACATATATATATATACACACACACATATATATATATACACACACACATATATATATA 15772

Qy        494 CACACATATATATACACATATATACACACATATATACACATATATACACATATATATACA 553
              | | ||| | | |||||||||| |     ||  | | | ||||| | ||||| | | |
Db      15771 TATATATACACACACACATATATATAATATAATCGATATAAATATATATATATAAAAAAA 15712

Qy        554 CATATATACACATATATACATATATACACATATATATACATATATACACATATATACATA 613
              |||||   |  |||| ||         |||  || || | || ||   |   ||   |
Db      15711 AATATACCTATTTATAAACTCTCTATTCCATTAATTTATACATCTAATAAACAATTATTT 15652

Qy        614 TATACACATATATACATATATACACACATATATACACATACATATACACACACATAGATA 673
              |||| | ||||||| |||||||||||||||||| | ||| ||||| | | ||||| |||
Db      15651 TATATATATATATATATATATATACACACATATATATA-ATATATATATATATACATACATA 15593

Qy        674 TACATATATATACACATATA 693
              | | ||| ||| |||||||
Db      15592 CATACATACATATACATATA 15573


RESULT 90
AAS46748
ID   AAS46748 standard; DNA; 26997 BP.
XX
AC   AAS46748;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Tumour suppressor gene derived chemically modified sequence #472.
XX
KW   Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer;
KW   tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW   cytosine methylation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200168912-A2.
XX
PD   20-SEP-2001.
XX
PF   15-MAR-2001; 2001WO-EP002955.
XX
PR   15-MAR-2000; 2000DE-01013847.
PR   06-APR-2000; 2000DE-01019058.
PR   07-APR-2000; 2000DE-01019173.
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2001-602752/68.
XX
PT   Fragments of chemically modified genes associated with tumor suppressor
PT   genes and oncogenes, useful in designing primers and probes for analyzing
PT   diseases associated with cytosine methylation state e.g. cancer.
XX
PS   Claim 1; SEQ ID NO 472; 27pp; English.
XX
CC   The invention relates to a nucleic acid comprising a sequence of 18
CC   bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC   bisulphite, of genes associated with tumour suppression and oncogenes
CC   having a sequence taken from 536 (actually 533 since numbers 408, 458 and
CC   500 are missing from the sequence listing) sequences (Ss) and sequences
CC   complementary to (Ss). The nucleic acid may be a peptide nucleic acid-
CC   oligomer (PNA) of at least 9 nucleotides and may form part of a set of
CC   probes for detecting the cytosine methylation state and/or single
CC   nucleotide polymorphisms and also to be used in an array for analysing
CC   diseases associated with CpG dinucleotides e.g. cancers and tumours. The
CC   probes can also be used in a method for ascertaining genetic and/or
CC   epigenetic parameters for the diagnosis and/or therapy of existing
CC   diseases or the predisposition to specific diseases, by analysing
CC   cytosine methylations. The parameters may be compared to another set of
CC   genetic and/or epigenetic parameters, the differences serving as basis
CC   for diagnosis and/or prognosis events which are disadvantageous to
CC   patients. The present sequence is one of the 533 genomic sequences
CC   derived from tumour suppressor genes and oncogenes. Sequences with even
CC   numbered Seq ID numbers are the complementary sequence of the
CC   corresponding odd numbered sequence (e.g. ID 2 and ID1, ID 536 and ID
CC   535, except for those whose partner sequence is missing). Note: The
CC   sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 26997 BP; 7650 A; 354 C; 5924 G; 13069 T; 0 U; 0 Other;

  Query Match             4.7%;  Score 234.2;  DB 4;  Length 26997;
  Best Local Similarity   69.7%;  Pred. No. 3.4e-31;
  Matches  345;  Conservative    0;  Mismatches  148;  Indels    2;  Gaps    2;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | |||||||| |||||||||||||| |   | | | | |||||| |||| |  ||||||
Db       5699 TATATATATACGTATATATATATATATATTTATATATATATATATGTATATATGTATATA 5758

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||  |||| || | || |||| | |||||||||| | |||  ||| || | |
Db       5759 TATATATATATGTATATTTATATATGTATATATATATATATATA-ATATGTATATATATA 5817

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              ||||||||   |||| |||| | || |||| || |||||| | | || | |  | ||||
Db       5818 TATATATATGTATATATATATATATGTATATACGTATATATAAATATGTGTGTATATATG 5877

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | || | |||| |  |||||| | ||||||    ||||| |  ||| || |  ||||
Db       5878 TATATACGTGTATATATGTATATATATATATATGTGTATATATATGTATGTATATGTATA 5937

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || |||||||    ||||||||  |||  | |  |||||||| | |||| |  | |  ||
Db       5938 TATATATATATGTGTATATATATGTATGTATATGTATATATATATATATGTGTATATGTA 5997
```

```
Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||  | ||||||  | |||||||||| |  | ||||   |||| || ||||||  | |||
Db       5998 TATGTATATATATGTATATATATATATATGTGTATATGTATATGTATATATATGTATATA 6057

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              ||||||  || |  ||  | ||||||  |||||||   |||||||  |||||||      |||||
Db       6058 TATATATATGTGTATATATATATGTATATATATGTATATATATATATATATGTGTATATA 6117

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              |  | |||  | |  | | |  || || |   | ||||||   ||||||||  || | ||
Db       6118 TGTATATATGTGTGTATATATGTATATGTGTGTGTATATATGTTTATATATATGTGTGTA 6177

Qy        706 TATGTATATATATAT 720
              |||  ||||||||||
Db       6178 TATATATATATATAT 6192


RESULT 91
AAV83939
ID   AAV83939 standard; DNA; 80595 BP.
XX
AC   AAV83939;
XX
DT   03-MAR-1999  (first entry)
XX
DE   HC-contig derived from normal human chromosome 10q25.2 region.
XX
KW   Yeast artificial chromosome; YAC; probe; eukaryotic chromosome;
KW   neocentromere; replication; extra-chromosomal element; segregation;
KW   cell division; artificial chromosome; gene therapy; mardel(10);
KW   human artificial chromosome; transgenic; chromosome 10; 10q25.2; ss.
XX
OS   Homo sapiens.
XX
PN   WO9851790-A1.
XX
PD   19-NOV-1998.
XX
PF   13-MAY-1998;   98WO-AU000352.
XX
PR   13-MAY-1997;   97AU-00006784.
PR   26-AUG-1997;   97AU-00008791.
XX
PA   (AMRA-) AMRAD OPERATIONS PTY LTD.
XX
PI   Choo K,  Du Sart D,  Cancilla MR;
XX
DR   WPI; 1999-009773/01.
XX
PT   New isolated nucleic acid comprising neocentromere sequences from
PT   eukaryotic chromosome - used to produce replicable, segregating
PT   artificial chromosomes that can carry large amounts of DNA for gene
PT   therapy.
XX
PS   Claim 8; Fig 6; 540pp; English.
XX
CC   The present sequence represents the HC-contig derived from normal human
CC   chromosome 10, 10q25.2 region. This region can be naturally mutated to
CC   produce an unusual chromosomal marker referred to as mardel(10). The
CC   mardel(10) marker is mitotically stable and contains a functional
CC   neocentromere at a location regarded as non-centromeric. This
CC   neocentromere maps to q25.2 on chromosome 10. The specification describes
CC   nucleic acid sequences derived from a eukaryotic chromosome, including a
CC   neocentromere or its functional derivative or hybrid, that are able, in a
CC   compatible cell, of replicating, acting as extra-chromosomal element and
CC   segregating during cell division. The sequences can be used to construct
CC   artificial chromosomes for use in gene therapy comprising a replicable,
CC   segregating nucleic acid that confers a specific phenotype on cells.
CC   Human artificial chromosomes can propagate in human cells and carry large
CC   amounts of DNA (e.g. therapeutic genes), and, being extra-chromosomal,
CC   they are not mutagenic. The artificial chromosomes are also useful for
CC   generation of transgenic plants and animals, in production of proteins
CC   and to make diagnostic reagents, e.g. for expression of cytokines,
CC   receptors and growth factors, or to increase the copy number of a gene in
CC   a cell. The constructs may also be used for functional and structural
CC   analysis of chromosomes
XX
SQ   Sequence 80595 BP; 23183 A; 16613 C; 16824 G; 23975 T; 0 U; 0 Other;

  Query Match             4.7%;  Score 234.2;  DB 2;  Length 80595;
  Best Local Similarity   66.3%;  Pred. No. 3.6e-31;
  Matches  352;  Conservative    0;  Mismatches  178;  Indels    1;  Gaps    1;

Qy        189 CAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATA 248
              || |  |||| || | | |    |||||     ||   | ||| ||||||||| |||
Db      15479 CATGTTTGAAATATCTTATAATTTTTAAAAATTAAAATGGGAGATACATATATATGTATT 15538

Qy        249 TATACACACACACACATATATATATACACATATATATACACATATATACATATATACA 308
              |||  |   | | |   || || |||||||||| |||||||||| | ||||| | ||| | |
Db      15539 TATGTATGTATATATGTATGTACATATACACACATATATACATAAATATATACATAAATA 15598

Qy        309 CATATATACACATATATATACACATACATATGTACACATATATATACACATATGTATACA 368
              ||||||    |||||| ||| | | || | || |  | |||||| |||||  |||  |
Db      15599 TGTATATATGTGTATATAGACATAAATATGTATATATGTGTATATATACATAAATATGTA 15658

Qy        369 CATATATACACATATATACACACATATATACACATATATACACACATATATACACATATA 428
              ||||  |  | ||| | | | | || |||| |  | |||| | ||||| |||   |||||
Db      15659 TATATGTGTATATAGACATAAATATGTATATATGTGTATATAGACATAAATATGTATATA 15718

Qy        429 TACACATATATACACATATACACATATATACACATATATACATATATACACATATATATA 488
              |    ||||| ||| | |||   ||||||    ||||| ||||| |||   ||||||  |
Db      15719 TGTGTATATAGACATAAATATGTATATATGTGTATATAGACATAAATATGTATATATGT- 15777

Qy        489 ATATACACACATATATATACACATATATACACACATATATACACATATATACACATATAT 548
              ||||  | |||||  ||||   | ||||  |   | | |  | ||| | ||  |||| |   | |||
Db      15778 GTATATAGACATAAATATGTATATATGTGTATATAGACATAAATATGTATATATGTGTAT 15837

Qy        549 ATACACATATATACACATATATACATATATACACATATATATACATATATACACATATAT 608
              |||  |||||  |||    ||||||   ||||||  |||||  ||||   ||||||    |||||
Db      15838 ATAGACATAAATATGTATATATGTGTATATAGACATAAATATGTATATATGTGTATATAG 15897

Qy        609 ACATATATACACATATATACATATATACACACATATATACACATACATATACACACACAT 668
              |||||  |||   ||||||   ||||||  |||  |  ||||   |  |||   |  ||  |  ||||
Db      15898 ACATAAATATGTATATATGTGTATATAGACATAAATATGTATATATGTGTATATAGACAT 15957

Qy        669 AGATATACATATATATACACATATATATACGTATATATATGTATATATATA 719
              |  ||||  ||||||  |   |  |||  |  |||   |||  |||||  |  |  ||||||
Db      15958 AAATATGTATATATGTGTATATAGACATAAATATGTATATATGTGTATATA 16008


RESULT 92
ABQ88198
ID   ABQ88198 standard; cDNA; 154902 BP.
XX
AC   ABQ88198;
XX
DT   18-SEP-2002  (first entry)
XX
DE   Human osteoblast differentiation related cDNA SEQ ID NO 105.
XX
KW   Human; osteoblast; stem cell differentiation; bone tissue deposition;
KW   osteoporosis; osteopathic; ss.
XX
OS   Homo sapiens.
XX
PN   WO200250301-A2.
XX
PD   27-JUN-2002.
XX
PF   18-DEC-2001; 2001WO-US048276.
XX
PR   18-DEC-2000; 2000US-0255882P.
PR   24-APR-2001; 2001US-0285691P.
XX
PA   (GENE-) GENE LOGIC INC.
```

PA (PROC) PROCTER & GAMBLE CO.
XX
PI Ji D, Axelrod DW, Cook JS, Jaiswal N, Einstein R, Houghton A;
PI Mertz L;
XX
DR WPI; 2002-557663/59.
XX
PT Use of genes and their expression profiles associated with osteoblast
PT differentiation for screening modulators bone formation, for diagnosing
PT or treating e.g. osteoporosis, or as markers for the differentiation
PT process.
XX
PS Claim 1; SEQ ID NO 105; 78pp + Sequence Listing; English.
XX
CC The invention relates to genes and their expression profiles are used
CC for: (a) screening modulators of precursor stem cell differentiation into
CC osteoblasts, or bone tissue deposition; (b) diagnosing abnormal
CC deposition of bone tissue, abnormal rate of osteoblast formation or
CC osteoporosis; or (c) treating or monitoring treatment of the conditions
CC cited in (b), or monitoring the progression of bone tissue deposition.
CC Specific conditions include postmenopausal osteoporosis, glucocorticoid
CC osteoporosis or male osteoporosis, osteopenia, osteodystrophy, drug-
CC induced abnormalities in bone formation or bone loss, conditions that
CC involve altered bone metabolism (e.g. idiopathic juvenile osteoporosis),
CC skeletal disease linked to breast cancer, mastocytosis, Fanconi syndrome
CC or fibrous dysplasia. The present sequence is that of an osteoblast
CC differentiation associated cDNA marker of the invention. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 154902 BP; 43917 A; 31458 C; 32848 G; 46679 T; 0 U; 0 Other;

Query Match 4.7%; Score 234.2; DB 6; Length 154902;
Best Local Similarity 75.4%; Pred. No. 3.7e-31;
Matches 304; Conservative 0; Mismatches 98; Indels 1; Gaps 1;

Qy 286 ATACACATATATACACATATATACACATATATACACATATATATACACATACATATGTACAC 345
Db 31565 ATACATATATATACACATATATATACATATATATACATATATACATATATACATGTATAC 31624

Qy 346 ATATATATACACATATGTATACACATATATACACATATATACACACATATATACACATAT 405
Db 31625 ACATACATATATACATATATACACACATTTATGCATATATACATAGACACACATTTATGC 31684

Qy 406 ATACACACATATATACACATATATACACATATATACACATATACACATATATACACATAT 465
Db 31685 ATACATATATACACACATGCATATATACCTATATACACATATATGCATATATACCTATAT 31744

Qy 466 ATACATATATACACATATATATAATATACACACATATATATACACATATATACACACATA 525
Db 31745 ACACATATATGCATATATACAT-ATATACACACATGCATATATACATATACACACATATA 31803

Qy 526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACACATA 585
Db 31804 TGCATATATACATATATACATATATACACACATATGCATATATACATATATACACATACA 31863

Qy 586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
Db 31864 TATATACACATATACACACATATATACATATATACACACATATATATACACATATATGTA 31923

Qy 646 TACACATACATATACACACACATAGATATACATATATATACAC 688
Db 31924 TACACACACATATATATATATATATTTTAGCTGGAGTCTCAC 31966

RESULT 93
ADC87318/c
ID ADC87318 standard; DNA; 1210 BP.
XX
AC ADC87318;
XX
DT 01-JAN-2004 (first entry)
XX
DE Human GPCR gene SEQ ID NO:1771.
XX
KW ds; gene; human; GPCR;
KW guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1270724-A2.
XX
PD 02-JAN-2003.
XX
PF 18-JUN-2002; 2002EP-00013517.
XX
PR 18-JUN-2001; 2001JP-00246789.
XX
PA (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI Suwa M, Asai K, Akiyama Y, Aburatani H;
XX
DR WPI; 2003-315783/31.
DR P-PSDB; ADC87319.
XX
PT New polynucleotide, useful for preparing a composition for treating a
PT patient in need of increased or suppressed activity or expression of the
PT guanosine triphosphate-binding protein coupled receptor.
XX
PS Claim 1; SEQ ID NO 1771; 28pp; English.
XX
CC The invention relates to a novel polynucleotide encoding a guanosine
CC triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC the invention may have a use in gene therapy. The polynucleotide and
CC polypeptide are useful for preparing a composition for treating a patient
CC in need of increased or suppressed activity or expression of the
CC guanosine triphosphate-binding protein coupled receptor. The
CC polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC invention.
XX
SQ Sequence 1210 BP; 391 A; 154 C; 200 G; 465 T; 0 U; 0 Other;

Query Match 4.7%; Score 233.6; DB 10; Length 1210;
Best Local Similarity 68.5%; Pred. No. 3.6e-31;
Matches 337; Conservative 0; Mismatches 154; Indels 1; Gaps 1;

Qy 230 ATATATATATATATATATATATACACACACACACATATATATATACACATATATATAC 289
Db 885 ATACATGTATACGAATATATATACACATATATGTATACGAATATATATACATGTATGTAT 826

Qy 290 ACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATAT 349
Db 825 ACGAATATATATACACATATATGTATACGAATATATATACACATGTATGTATACGAATAT 766

Qy 350 ATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATAC 409
Db 765 ATATACACATGTATGTATACGAATATATATACACATGTATATGTATACGAATATATATAC 706

Qy 410 ACACATATATACACATATATACACATATATACACATATACACATATATACACATATATAC 469
Db 705 ACATGTATATGTATACAAATATATATACACACACATGTATATGAATACACACACATGTAT 646

Qy 470 ATATATACACATATATAT-AATATACACACATATATATACACATATATACACACATATAT 528
Db 645 ATGAATACACACATATATGAATACACACACATATGTATATGAATACACACACACATATGT 586

Qy 529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACATATAT 588
Db 585 ATATGAATACACACACACGTATGTATATGAATACACACACACGTATGTATATGAATACAC 526

Qy 589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
Db 525 ACACACGTATGTATATGAATACACACACACGTATGTATATGAATACACACACACGTATGT 466

```
Qy        649 ACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATAT 708
              | ||  ||| | |||||| ||  ||||   ||| |||||||  ||| ||  |  ||| ||
Db        465 ATATGAATACATACACACGTATGTATATGAATACATACACACGTATGTATATGAATACAT 406

Qy        709 GTATATATATAT 720
                | | ||||||
Db        405 ACACACATATAT 394


RESULT 94
AAF84800/c
ID   AAF84800 standard; DNA; 110000 BP.
XX
AC   AAF84800;
XX
DT   09-JUL-2001  (first entry)
XX
DE   Nucleotide sequence of the human SPG4 gene.
XX
KW   Human; SPG4 gene; spastin; PSF-AD; gene therapy;
KW   autosomal dominant familial spastic paraplegia; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             9932. .102009
FT                   /*tag= a
FT                   /note= "contains introns"
FT   exon            9932. .10471
FT                   /*tag= b
FT                   /number= 1
FT   intron          10472. .33718
FT                   /*tag= c
FT                   /number= 1
FT   exon            33719. .33805
FT                   /*tag= d
FT                   /number= 2
FT   intron          33806. .35748
FT                   /*tag= e
FT                   /number= 2
FT   exon            35749. .35832
FT                   /*tag= f
FT                   /number= 3
FT   intron          35833. .45022
FT                   /*tag= g
FT                   /number= 3
FT   exon            45023. .45118
FT                   /*tag= h
FT                   /number= 4
FT   intron          45119. .60863
FT                   /*tag= i
FT                   /number= 4
FT   exon            60864. .61051
FT                   /*tag= j
FT                   /number= 5
FT   intron          61052. .61927
FT                   /*tag= k
FT                   /number= 5
FT   exon            61928. .62061
FT                   /*tag= l
FT                   /number= 6
FT   intron          62062. .62344
FT                   /*tag= m
FT                   /number= 6
FT   exon            62345. .62438
FT                   /*tag= n
FT                   /number= 7
FT   intron          62439. .73173
FT                   /*tag= o
FT                   /number= 7
FT   exon            73174. .73248
FT                   /*tag= p
FT                   /number= 8
FT   intron          73249. .74633
FT                   /*tag= q
FT                   /number= 8
FT   exon            74634. .74705
FT                   /*tag= r
FT                   /number= 9
FT   intron          74706. .82788
FT                   /*tag= s
FT                   /number= 9
FT   exon            82789. .82864
FT                   /*tag= t
FT                   /number= 10
FT   intron          82865. .83102
FT                   /*tag= u
FT                   /number= 10
FT   exon            83103. .83194
FT                   /*tag= v
FT                   /number= 11
FT   intron          83195. .83334
FT                   /*tag= w
FT                   /number= 11
FT   exon            83335. .83414
FT                   /*tag= x
FT                   /number= 12
FT   intron          83415. .88129
FT                   /*tag= y
FT                   /number= 12
FT   exon            88130. .88172
FT                   /*tag= z
FT                   /number= 13
FT   intron          88173. .89561
FT                   /*tag= aa
FT                   /number= 13
FT   exon            89562. .89641
FT                   /*tag= ab
FT                   /number= 14
FT   intron          89642. .91162
FT                   /*tag= ac
FT                   /number= 14
FT   exon            91163. .91233
FT                   /*tag= ad
FT                   /number= 15
FT   intron          91234. .93443
FT                   /*tag= ae
FT                   /number= 15
FT   exon            93444. .93484
FT                   /*tag= af
FT                   /number= 16
FT   intron          93485. .100599
FT                   /*tag= ag
FT                   /number= 16
FT   exon            100600. .102009
FT                   /*tag= ah
FT                   /number= 17
XX
PN   FR2798138-A1.
XX
PD   09-MAR-2001.
XX
PF   03-SEP-1999;   99FR-00011097.
XX
PR   03-SEP-1999;   99FR-00011097.
XX
PA   (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI   Weissenbach J,  Hazan J;
XX
DR   WPI; 2001-283966/30.
XX
PT   New human nucleic acid from the SPG4 gene, useful e.g. for diagnosis of
PT   autosomal dominant familial spastic paraplegia and in drug screening.
XX
PS   Claim 2; Page 45-106; 145pp; French.
```

```
XX
CC   The present sequence represents a human SPG4 gene. The SPG4 gene encodes
CC   a spastin polypeptide. Mutations in the SPG4 gene are responsible for
CC   autosomal dominant familial spastic paraplegia. SPG4 polynucleotides, and
CC   their fragments, are used to screen DNA banks for sequences that encode
CC   spastin (particularly sequences in other mammals, specifically mice); to
CC   identify SPG4 mutations, or other genetic anomalies, particularly for
CC   diagnosis of autosomal dominant familial spastic paraplegia (PSF-AD); to
CC   identify promoters and other regulatory elements of the SPG4 gene; for
CC   detection and amplification; for recombinant production of spastin; and
CC   for diagnostic genotyping of PSF-AD
XX
SQ   Sequence 110000 BP; 30622 A; 21640 C; 22817 G; 34921 T; 0 U; 0 Other;

  Query Match             4.7%;  Score 232.8;  DB 5;  Length 110000;
  Best Local Similarity   67.2%;  Pred. No. 6.3e-31;
  Matches  449;  Conservative    0;  Mismatches  207;  Indels   12;  Gaps    8;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC--AAAAT 58
              ||||  ||| |||||||  ||||||||||||||||||| |||||||| ||||  |||||
Db      56570 AGTTTGAGAACAGCCTGGCCAACATGGTGAAACCCCATTTCTACTAAAGATACAAAAAAT 56511

Qy         59 TAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGA 118
              ||| | ||| |||||  ||  |||||||||||||| ||||  |||| | || ||||||||
Db      56510 TAGCCTGGCGTGGTGCCACGTGCCTGTAATCCCAGGTACTCGGGAGACTGAAGCAGGAGA 56451

Qy        119 ATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCC 177
              ||  ||||||||||  ||  |  ||||||||||  || ||||||    ||| |||| |||
Db      56450 ATTGCTTGAACCTG--GGAAGCAGAGGTTGCGGTAAGCCAAGATAGCACCACTGCATTCC 56393

Qy        178 AGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATAT 237
              ||||  ||  || || || | |   ||||||  |||||||| ||    |  | | | |
Db      56392 AGCCTGGGGGAC-AGGGCAAGACTCCATCTCCAAAAAAAAACAAAAAAAACAAAAAAAAA 56334

Qy        238 ATATATATATATATACACACACACACATATATATATACACATATATATACACATATAT 297
              | | | | | |  || | |   | ||||| ||||| | | |||||||||     ||| ||
Db      56333 AAACAAAAAAAACTATATATGTATACACACATATACAAATACATATATAACTGTATACAT 56274

Qy        298 ACATATATACACATATATACACATATATATACACATACATATGTACACATATAT-ATACA 356
              | |  | || | ||| ||  |||  |||||| |||||  ||| |  | |||  | |||||
Db      56273 ATA-GTTTATATATACATTTACAGCTATATATACATAGTTATTTTGATATAACTAATACA 56215

Qy        357 CATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATA 416
               ||||   || | || |||| |||||| | | |||||| ||   ||||||  | |   |
Db      56214 TATATAACTATATATGTATATACATATACATATACATATGTAACTATATATGTATATGCA 56155

Qy        417 TATACACATATATACACATATATACACATATACACATATATACACATATATACATATATA 476
              |||||| || ||   | |||| |  |   ||| |||||| ||   ||||||  ||||  |
Db      56154 TATACATATGTAACTATATATGTGTATGCATATACATATGTAACTATATATGTATATGCA 56095

Qy        477 CACATATATATAA-TATACACACATATATATACACATATATACACACATATATACACATA 535
              | | |||| ||| |||| |   ||||  ||| |||||| ||   | |||| ||    |
Db      56094 TATAGATATGTAACTATATATGTATATGCATATACATATGTAACTATATATGTATGTGTG 56035

Qy        536 TATACACATATATATACACATATATACACATATATACATATATACACATATATATACATA 595
              |||||| || ||  || | || |||   |||||| | |||||   | ||| | |||  |
Db      56034 TATACATATGTAACTATATATGTATGTGCATATACATATATAACTATATACAGATAGCT- 55976

Qy        596 TATACACATATATACATATATACACATATATACATATATACACACATATATACACATACA 655
              || |||||||||  ||||||  |||||||   || | ||   | ||  | ||  |   |
Db      55975 --TATACATATATAGTTATATATACATATAGTTATTTTTAGTTATATCAAAACTAAACAA 55918

Qy        656 TATACACA 663
              ||||  | |
Db      55917 TATATATA 55910


RESULT 95
AAV83940
ID   AAV83940 standard; DNA; 80240 BP.
XX
AC   AAV83940;
XX
DT   03-MAR-1999  (first entry)
XX
DE   NC-contig derived from mardel(10) on chromosome 10q25.2.
XX
KW   Yeast artificial chromosome; YAC; probe; eukaryotic chromosome;
KW   neocentromere; replication; extra-chromosomal element; segregation;
KW   cell division; artificial chromosome; gene therapy; mardel(10);
KW   human artificial chromosome; transgenic; chromosome 10; 10q25.2; ss.
XX
OS   Homo sapiens.
XX
PN   WO9851790-A1.
XX
PD   19-NOV-1998.
XX
PF   13-MAY-1998;   98WO-AU000352.
XX
PR   13-MAY-1997;   97AU-00006784.
PR   26-AUG-1997;   97AU-00008791.
XX
PA   (AMRA-) AMRAD OPERATIONS PTY LTD.
XX
PI   Choo K,  Du Sart D,  Cancilla MR;
XX
DR   WPI; 1999-009773/01.
XX
PT   New isolated nucleic acid comprising neocentromere sequences from
PT   eukaryotic chromosome - used to produce replicable, segregating
PT   artificial chromosomes that can carry large amounts of DNA for gene
PT   therapy.
XX
PS   Claim 9; Fig 16A; 540pp; English.
XX
CC   The present sequence represents the NC-contig derived from a mutated
CC   human chromosome 10, 10q25.2 region. The sequence contains an unusual
CC   chromosomal marker referred to as mardel(10). The mardel(10) marker is
CC   mitotically stable and contains a functional neocentromere at a location
CC   regarded as non-centromeric. This neocentromere maps to q25.2 on
CC   chromosome 10. The specification describes nucleic acid sequences derived
CC   from a eukaryotic chromosome, including a neocentromere or its functional
CC   derivative or hybrid, that are able, in a compatible cell, of
CC   replicating, acting as extra-chromosomal element and segregating during
CC   cell division. The sequences can be used to construct artificial
CC   chromosomes for use in gene therapy comprising a replicable, segregating
CC   nucleic acid that confers a specific phenotype on cells. Human artificial
CC   chromosomes can propagate in human cells and carry large amounts of DNA
CC   (e.g. therapeutic genes), and, being extra-chromosomal, they are not
CC   mutagenic. The artificial chromosomes are also useful for generation of
CC   transgenic plants and animals, in production of proteins and to make
CC   diagnostic reagents, e.g. for expression of cytokines, receptors and
CC   growth factors, or to increase the copy number of a gene in a cell. The
CC   constructs may also be used for functional and structural analysis of
CC   chromosomes
XX
SQ   Sequence 80240 BP; 23102 A; 16537 C; 16747 G; 23846 T; 0 U; 8 Other;

  Query Match             4.7%;  Score 232.2;  DB 2;  Length 80240;
  Best Local Similarity   64.9%;  Pred. No. 7.9e-31;
  Matches  359;  Conservative    0;  Mismatches  193;  Indels    1;  Gaps    1;

Qy        217 AAAAAGGCATTTTATATATATATATATATATATACACACACACACATATATATATA 276
              |||||  |  |  | ||| |||||||| ||| ||  |  | | |  ||| || |||
Db      15240 AAAAATTAAAATGGGAGATACATATATATGTATTTATGTATGTATATATGTATGTACATA 15299

Qy        277 CACATATATATACACATATATACATATATACACATATATACACATATATATACACATACA 336
              ||| | ||||| ||||| ||| ||| ||| | || |||| |  | |||||| ||||| |
Db      15300 TACACACATATATACATAAATATATACATAAATATGTATATATGTGTATATAGACATAAA 15359

Qy        337 TATGTACACATATATATACACATATGTATACACATATATACACATATATACACACATATA 396
              |||||| | || | |||| | | ||  |||   ||||||    ||||| ||| | ||||
Db      15360 TATGTATATATGTGTATATATACATAAATATGTATATATGTGTATATAGACATAAATATG 15419

Qy        397 TACACATATATACACACATATATACACATATATACACATATATACACATATACACATATA 456
```

```
Db      15420 TATATATGTGTATATAGACATAAATATGTATATATGTGTATATAGACATAAATATGTATA 15479

Qy        457 TACACATATATACATATATACACATATATATAATATACACACATATATATACACATATAT 516

Db      15480 TGTGTATATAGACATAAATATGTATATATGT-GTATATAGACATAAATATGTATATATGT 15538

Qy        517 ACACACATATATACACATATATACACATATATATACACATATATACACATATATACATAT 576

Db      15539 GTATATAGACATAAATATGTATATATGTGTATATAGACATAAATATGTATATATGTGTAT 15598

Qy        577 ATACACATATATATACATATATACACATATATACATATATACACATATATACATATATAC 636

Db      15599 ATAGACATAAATATGTATATATGTGTATATAGACATAAATATGTATATATGTGTATATAG 15658

Qy        637 ACACATATATACACATACATATACACACACATAGATATACATATATATACACATATATAT 696

Db      15659 ACATAAATATGTATATATGTGTATATAGACATAAATATGTATATATGTGTATATAGACAT 15718

Qy        697 ACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCA 756

Db      15719 AAATATGTATATGTGTGTATATAGACATAAATATGTATATATGTGTGTATATAGACATAA 15778

Qy        757 CTGGGGATAGAGG 769

Db      15779 ATATGTATATATG 15791
```

RESULT 96
ACN44026/c
ID ACN44026 standard; DNA; 122923 BP.
XX
AC ACN44026;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human genomic sequence hCG24510.
XX
KW Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.
XX
OS Homo sapiens.
XX
PN WO2003073826-A2.
XX
PD 12-SEP-2003.
XX
PF 28-FEB-2003; 2003WO-US006235.
XX
PR 01-MAR-2002; 2002US-00087192.
XX
PA (SAGR-) SAGRES DISCOVERY.
XX
PI Morris DW;
XX
DR WPI; 2003-328604/31.
XX
PT Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
PT comprises a nucleotide sequence.
XX
PS Claim 1; SEQ ID NO 268; 0pp; English.
XX
CC The present invention relates to novel DNA and protein sequences which
CC are associated with carcinomas. The sequences are useful for: (i) for
CC screening drug candidates; (ii) for screening of bioactive agent capable
CC of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
CC a bioactive agent capable of modulating the activity of CAP; (iv) for
CC evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
CC carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
CC carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
CC (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
CC determining Carcinoma Associated (CA) gene copy number. In addition, the
CC CA genes are useful as DNA vaccines and the CAP are useful as markers of
CC carcinoma including lymphoma. The present sequence is one such CA coding
CC sequence. Note: This patent is an equivalent to basic patent
CC US2002182586A1, for which no sequence data was published
XX
SQ Sequence 122923 BP; 28628 A; 24391 C; 25218 G; 32559 T; 0 U; 12127 Other;

```
  Query Match             4.7%;  Score 231.8;  DB 11;  Length 122923;
  Best Local Similarity   62.3%;  Pred. No. 9.5e-31;
  Matches  447;  Conservative    0;  Mismatches  257;  Indels   13;  Gaps    6;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAATTA 60

Db      43687 AGTTTGAGACCAGCCTGGGCAACATGGCAAAACCCTATCTCTACAAAAAATACAAAAATG 43628

Qy         61 GA-CAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119

Db      43627 CACCAGGTGTAGTGGCATGCACCTGTAGTCTCAACTACTTGGGGGGCCAAGGCAGAAGGA 43568

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCCA 178

Db      43567 TCGCTTGAACC--CGGGAGGTCAAGGTTGCAGTGAGCCAAGACCACACTACTGCACTCCA 43510

Qy        179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATA 238

Db      43509 GTCTAGGTGATAA-AGTGAGACCCTGTTTCAGAAAAAAAAA-------TTTATATATATA 43458

Qy        239 TATATATATATATACACACACACACACATATATATATACACATATATATACACATATATA 298

Db      43457 TATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTATGTGTGTGTATGTGTATATA 43398

Qy        299 CATATATACACATATATACACATATATATACACATACATATGTACACA-TATATATACAC 357

Db      43397 TGTGTGTGTCTATGTGTGTACATATATATAAAATTGTACATATAAAAATTATATAAGTAT 43338

Qy        358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417

Db      43337 ATATATGCACGTGTATATACGTATGTATGTGTGTGTATGTATGTGTGTGTATATATGTAT 43278

Qy        418 ATACACATATATACACATATATACACATATACACATATATACACATATATACATATATAC 477

Db      43277 GTGTATGTATATATGTATGTGTGTATATATATGTATGCATGTGTGTGTATGTATGTATGN 43218

Qy        478 ACATATATATAATATACACACATATATATACACATATATACACACATATATACACATATA 537

Db      43217 NNNNNNNNNNNNNNNNNNNACACATATATATACACACATATATATACACACATATATATA 43158

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATACATATA 597

Db      43157 CACATATATATATACACATATATATACACATATATATACACACATATATATACACATATA 43098

Qy        598 TACACATATATACATATATACACATATATACATATATACACACATATATACACATACATA 657

Db      43097 TATACACACATACACACACACACACACACACACACACACACACACACACACATATATATATA 43038

Qy        658 TACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATAT 714

Db      43037 TATATGCCAAGAGTAAGTATTAGGGGAATAAAGCTTAAAATAAAAAAATAAGAAAAT 42981
```

RESULT 97
ADL13505/c
ID ADL13505 standard; DNA; 76146 BP.
XX
AC ADL13505;
XX
DT 06-MAY-2004 (first entry)
XX
DE Osteoarthritis-associated polymorphic nucleotide #37.
XX
KW ds; gene; osteopathic; antiinflammatory; antiarthritic; gene therapy;
KW joint space narrowing; osteophyte development; joint pain;
KW osteoarthritis; SNP; single nucleotide polymorphism.
XX
OS Homo sapiens.

XX
PN WO2003054166-A2.
XX
PD 03-JUL-2003.
XX
PF 19-DEC-2002; 2002WO-US041225.
XX
PR 20-DEC-2001; 2001US-0342603P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Jones KA, Schafer A;
XX
DR WPI; 2003-559141/52.
XX
PT Determining susceptibility of an individual to joint space narrowing,
PT osteophyte development and/or joint pain comprises identifying whether
PT the individual has at least one polymorphism in a polynucleotide encoding
PT a protein.
XX
PS Disclosure; SEQ ID NO 37; 297pp; English.
XX
CC The invention relates to a method of determining susceptibility of an
CC individual to joint space narrowing and/or osteophyte development and/or
CC joint pain comprising identifying whether the individual has at least one
CC polymorphism in a polynucleotide encoding at least one of the protein
CC listed in the specification. The methods, composition and agent are
CC useful for modulating the susceptibility of an individual to joint space
CC narrowing and/or osteophyte development and/or joint pain that is
CC associated with a disease, preferably osteoarthritis. The cell line and
CC the non-human animal are useful for screening for an agent for diagnosing
CC an individual having susceptibility to joint space narrowing and/or
CC osteophyte development and/or joint pain. This sequence corresponds to
CC the polynucleotide encoding a protein listed in the specification. (Note:
CC The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences).
XX
SQ Sequence 76146 BP; 14629 A; 22827 C; 22905 G; 15785 T; 0 U; 0 Other;

Query Match 4.6%; Score 231; DB 10; Length 76146;
Best Local Similarity 69.7%; Pred. No. 1.3e-30;
Matches 373; Conservative 0; Mismatches 150; Indels 12; Gaps 4;

```
Qy      209 AGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACATA 268
            | |||| | | | |       ||  | ||  ||  | ||| || ||||||  ||||||||
Db    14496 ACAAAACACATATGTACACACATCCACATGCATCCACATACATGCACACATCCACACATA 14437

Qy      269 TATATATACACATATATATACACATATATACATATAT------ACACATATATACACATA 322
            ||  |  |||||  |||| ||||||  || ||| ||||      |||||  || ||||  |
Db    14436 TACGTGCACACACCTATACACACATGTACACACATATACATGCACACACCTAGACACTCA 14377

Qy      323 TATATACACATACATATGTACACATATATATACACATATGTATACACATATATACACATA 382
             | || || |  ||| | |||||||  | ||||||||||| |||||||||| |||||||
Db    14376 CACATGCATACCCATGTATACACATTCACATACACATATG-ATACACATATACACACATA 14318

Qy      383 TATACACACATATATACACATATATACACACATATATACACATATATACACATATATACA 442
            |||||||| ||  | |||||   || ||||||| |||||||||| || ||||| || |
Db    14317 TATACACAGATGCACACACACTCATGCACACATCTATACACATACATGCACATGTACATC 14258

Qy      443 CATATACACATATATACACATATATACATATATACACATATATATAATATACACACATAT 502
            || || | || ||  |  ||| |||||| | || |||| ||  |  | ||||||||||  |
Db    14257 CACATGCGCACATGCATGCATGTATACACACATGCACACATGCA-CACATACACACATGT 14199

Qy      503 ATATACACATATATACACACATATATACACATATATACACATATATATACACATATATAC 562
            | | ||||  | | |||| || | |||||||| ||||||||  || | |||||    ||
Db    14198 ACACACACGCACACACACGCACACATACACATTTATACACACGTACACACACAC----AC 14143

Qy      563 ACATATATACATATATACACATATATATACATATATACACATATATACATATATACACAT 622
            ||||  ||||||  | ||| | ||| |||||||  | |||||||| || | |||||||
Db    14142 ACATGCATACATGCACACATACATACATACATACCTGCACATATACACCTGTATACACGC 14083

Qy      623 ATATACATATATACACACATATATACACATACATATACACACACATAGATATACATATAT 682
```

```
            ||| |||||| |||||||||| || |||| |  ||||||||||   | | | ||||  |
Db    14082 ATACACATATGTACACACATACATGCACACATCTATACACACATGCATACACACATTCAC 14023

Qy      683 ATACACATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAG 737
            ||| |||||| ||||| |||| | ||  ||| | ||||   |    |  |||||
Db    14022 ATATACATATGTATACATATACACATACATACACATATATCCACACATAGATAAG 13968
```

RESULT 98
ADJ95414
ID ADJ95414 standard; DNA; 9800 BP.
XX
AC ADJ95414;
XX
DT 06-MAY-2004 (first entry)
XX
DE Human nootropic protein coding sequence, SEQ ID 3.
XX
KW Nootropic; Gene therapy; human; mental illness; gene; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 2. .250
FT /*tag= a
FT /partial
FT /note= "No start codon given"
FT /product= "Human nootropic protein #3"
XX
PN JP2003310264-A.
XX
PD 05-NOV-2003.
XX
PF 19-APR-2002; 2002JP-00117143.
XX
PR 19-APR-2002; 2002JP-00117143.
XX
PA (KAZU-) ZH KAZUSA DNA KENKYUSHO.
XX
DR WPI; 2004-027143/03.
DR P-PSDB; ADJ95425.
XX
PT Novel DNA encoding recombinant protein from cDNA library derived from
PT human adult whole brain and human fetus whole brain, useful for diagnosis
PT of disease such as mental illness.
XX
PS Claim 1; SEQ ID NO 3; 91pp; Japanese.
XX
CC The present invention relates to human proteins (ADJ95423-ADJ95433) and
CC their coding sequences (ADJ95412-ADJ95422). The sequences are useful for
CC diagnosis and treatment of disease such as mental illness.
XX
SQ Sequence 9800 BP; 2648 A; 2439 C; 2333 G; 2380 T; 0 U; 0 Other;

Query Match 4.6%; Score 229.4; DB 12; Length 9800;
Best Local Similarity 68.5%; Pred. No. 2.2e-30;
Matches 379; Conservative 0; Mismatches 161; Indels 13; Gaps 4;

```
Qy      185 GCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATA 244
            ||  | ||  ||| || |   || |     |||||       || | ||| ||| | | |||
Db     5611 GCCTCCAGTTCGAGATATACCCTAAATTTAAAAATTGTATATATAATACATACACACATA 5670

Qy      245 TATATATACA-----CACACACACACATATATATATACACATATATATACACATATATAC 299
             |||||||  |     ||||||  ||||| | ||| | | | | | | || || ||||| ||
Db     5671 CATATATAGACACACCACACATACACACACATACACATATACACACATGCATATATACAC 5730

Qy      300 ATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACAT 359
             |||||  ||||||  |||||||  | ||||||||||  ||||   | | ||| | | ||
Db     5731 ATATACACACATGCATACACACACATATACACATGCATACACATATATACACACACCACA 5790

Qy      360 ATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATAT 419
                    ||||||| ||||||| ||| | ||| |||| ||   | |||||  ||||||
Db     5791 CATACACAACATATACACACATACATATATACACATATGCATGCACATACATGCATATAC 5850
```

```
Qy        420 ACACATATATACACATATATACACATATACACATATATACACATATATACATATATACAC 479
              |||||||||| |||| ||| |||||| || ||| | ||||||||||||||   ||||||
Db       5851 ACACATATATGCACACATACACACATGTATACACACATACACATATATACTCATATACCC 5910

Qy        480 ATATATATA-ATATACACACATATATATACACATATATACACACATATATACACATATAT 538
              | | ||| |||| |||||| ||||| | ||||||   |||||| | ||| | | | |
Db       5911 CCACACATATATATCCACACACATATAAATACATATCCCCACACAGACATATATACACAC 5970

Qy        539 AC---ACATATATATACACATATATACACATATATACATATATACACATATATATACATA 595
              ||   ||| | | |||||    | ||| | ||||| |||| ||||| || ||| | ||| |
Db       5971 ACTATACACACACATACATGCAGATATATATATACACATGTATACTCACATAAACACACA 6030

Qy        596 TATACACATATATACATATATACACATATATACATATATACACACATATATACACATACA 655
              |  |  ||||||| ||   ||| | ||| | |  |||||||||  |||| |||| ||||
Db       6031 TGCATGCATATATGCAGGCATATATATACACATGTATATACACAAGTATACACACGTACA 6090

Qy        656 TATACACA----CACATAGATATACATATATATACACATATATATACGTATATATATGTA 711
              ||||||||     | ||  |||||||  |  |||||||||| |||||  ||||||    |
Db       6091 TATACACATATTTAAATGTATATACACAGGTATACACATATGTATACACATATATGCACA 6150

Qy        712 TATATATATGCTC 724
              | || ||||| |
Db       6151 CACATGTATGCAC 6163
```

```
RESULT 99
ABD32843/c
ID   ABD32843 standard; DNA; 326002 BP.
XX
AC   ABD32843;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Human cancer-associated genomic DNA HD17-014.
XX
KW   Human; ds; cancer-associated protein; gene; cytostatic; cancer;
KW   leukaemia; lymphoma; CAP.
XX
OS   Homo sapiens.
XX
PN   WO2004074320-A2.
XX
PD   02-SEP-2004.
XX
PF   17-FEB-2004; 2004WO-US004730.
XX
PR   14-FEB-2003; 2003US-00367094.
PR   14-MAR-2003; 2003US-00388838.
PR   15-APR-2003; 2003US-00417375.
PR   13-JUN-2003; 2003US-00461862.
PR   15-SEP-2003; 2003US-00663431.
PR   15-DEC-2003; 2003US-00737318.
XX
PA   (SAGR-) SAGRES DISCOVERY INC.
XX
PI   Morris DW,  Morris DW,  Malandro MS;
XX
DR   WPI; 2004-652914/63.
XX
PT   New isolated cancer-associated polynucleotides and polypeptides useful
PT   for diagnosing, preventing or treating cancers, especially lymphoma and
PT   leukemia, or in screening for agents that modulate cancer.
XX
PS   claim 16; seqid 506; 310pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising at least 10
CC   contiguous nucleotides of any of the 233 polynucleotide sequences given
CC   in the specification, or its complement. The nucleic acids encode cancer-
CC   associated proteins. Also included are an expression vector comprising
CC   the isolated nucleic acid cited above, a host cell comprising the above
CC   recombinant nucleic acid or expression vector, a microarray for detecting
CC   a cancer-associated (CA) nucleic acid comprising at least one probe
CC   comprising at least 10 contiguous nucleotides of any of the above-
CC   mentioned nucleotide sequences, an isolated polypeptide (encoded within
CC   an open reading frame of a CA sequence selected from any of the 95
CC   polynucleotide sequences as mentioned in the specification, or its
CC   complement), an isolated antibody, (or its antigen binding fragment) that
CC   binds to the above polypeptide, a hybridoma that produces the above
CC   monoclonal antibody, a pharmaceutical composition comprising the above
CC   antibody and a pharmaceutical excipient, a kit for detecting cancer
CC   cells(comprising the antibody cited above, methods for diagnosing cancer
CC   or for detecting the presence or absence of cancer cells in an
CC   individual, a method for inhibiting growth of cancer cells in an
CC   individual, a method for delivering a therapeutic agent to cancer cells
CC   in an individual, an electronic library comprising the above
CC   polynucleotide or polypeptide (or their fragments), methods of screening
CC   for anticancer activity or for a bioactive agent capable of modulating
CC   the activity of a CA protein (CAP), methods for detecting cancer
CC   associated with expression of a polypeptide in a test cell sample, a
CC   method for treating cancers and a method for inhibiting the expression of
CC   CA gene in a cell. The composition and methods are useful for detecting,
CC   diagnosing, preventing and treating cancers, especially lymphoma and
CC   leukaemia. These may also be used in screening for agents that modulate
CC   cancer. The present sequence is a human CAP genomic sequence. Note: The
CC   sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 326002 BP; 84319 A; 72970 C; 76829 G; 91884 T; 0 U; 0 Other;

  Query Match             4.6%;  Score 229;  DB 13;  Length 326002;
  Best Local Similarity   70.1%;  Pred. No. 3.1e-30;
  Matches  323;  Conservative    0;  Mismatches  135;  Indels    3;  Gaps    1;

Qy        235 TATATATATATATATATACACACACACACATATATATATACACATATATATACACATA 294
              ||| | |||| | | | ||||||||||||||| | | | | || | || | |||||||
Db      46727 TATGTGTATAGACACACACACACACACACACACACACACACGCGTACACATACACACA 46668

Qy        295 TATACATATATACACATATATACACATATATATACACATACATATGTACACATATATATA 354
              || ||  | | | ||| | ||||||| | |||  || | ||| | | | ||| ||| |
Db      46667 TACACGCACACATACACACATACACACACATACGCATACACAGACGCATACACATACACG 46608

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              |||| ||  | ||||| | |||||||   | |||||||| || |||| ||| | |||||
Db      46607 CACACATACACACACACATACACACGCACACACACATACATGCACACATACATACACA 46548

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              |||  || | |||  || ||| | ||||||| ||||  ||| |||||  ||| ||  | |
Db      46547 CATACGCATACATACGCATATACACACACATACACACGCATACACACACATACACGCACA 46488

Qy        475 TACACATATATATAATATACACACATATATATACACATATATACACACATATATACACAT 534
              | ||| | ||   ||| ||||| | ||| | |||||||||| ||  ||| | |||||||
Db      46487 CATACACACAT---ATACACACAGAGATACACACACATATACACGGACACACATACACAC 46431

Qy        535 ATATACACATATATATACACATATATACACATATATACATATATACACATATATATACAT 594
              |   | ||| ||  ||||||| ||| |  |||||| ||| ||| | ||||  ||| ||||
Db      46430 ACTCATACACATGCATACACACATACATGCATATACACACATACATACATGCATACACAT 46371

Qy        595 ATATACACATATATACATATATACACATATATACATATATACACACATATATACACATAC 654
              | ||||||| ||| |||||  |  ||| |   ||| ||| ||||||||| | || |||||
Db      46370 ACATACACACATACACATACCTGGACACACGCACACATACACACACATACACACGCATAC 46311

Qy        655 ATATACACACACATAGATATACATATATATACACATATATA 695
              | |||| |||||| |  ||||| | | | | || | | ||
Db      46310 ACATACGCACACACATGCATACACACACACAGACCTCTCTA 46270
```

```
RESULT 100
ADC87570
ID   ADC87570 standard; DNA; 1672 BP.
XX
AC   ADC87570;
XX
DT   01-JAN-2004  (first entry)
XX
```

DE Human GPCR gene SEQ ID NO:2023.
XX
KW ds; gene; human; GPCR;
KW guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1270724-A2.
XX
PD 02-JAN-2003.
XX
PF 18-JUN-2002; 2002EP-00013517.
XX
PR 18-JUN-2001; 2001JP-00246789.
XX
PA (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI Suwa M, Asai K, Akiyama Y, Aburatani H;
XX
DR WPI; 2003-315783/31.
DR P-PSDB; ADC87571.
XX
PT New polynucleotide, useful for preparing a composition for treating a
PT patient in need of increased or suppressed activity or expression of the
PT guanosine triphosphate-binding protein coupled receptor.
XX
PS Claim 1; SEQ ID NO 2023; 28pp; English.
XX
CC The invention relates to a novel polynucleotide encoding a guanosine
CC triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC the invention may have a use in gene therapy. The polynucleotide and
CC polypeptide are useful for preparing a composition for treating a patient
CC in need of increased or suppressed activity or expression of the
CC guanosine triphosphate-binding protein coupled receptor. The
CC polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC invention.
XX
SQ Sequence 1672 BP; 499 A; 171 C; 300 G; 602 T; 0 U; 100 Other;

Query Match 4.6%; Score 228.8; DB 10; Length 1672;
Best Local Similarity 65.6%; Pred. No. 2.5e-30;
Matches 349; Conservative 0; Mismatches 182; Indels 1; Gaps 1;

```
Qy        204 ATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACAC 263
              || |  | | | | | |       | | |||||| | |||| | | |||| ||||| | |
Db        832 ATGTGTGTATATACATATATATGTGTGTATATACACATATGTGTGTATATACACATATAT 891

Qy        264 ACATATATATATACACATATATATACACATATATACATATATACACATATATACACATAT 323
              |||||||| | || | | | ||||| |||||||| ||||||||||||||||||  | |
Db        892 ACATATATGTGTATATACACATATATACATATATGTGTATATACACATATATACATGTGT 951

Qy        324 ATATACACATACATATGTACACATATATATACACATATGTATACACATATATACACATAT 383
              | || | | |||||| |      | ||||||||||||| | |     ||||||||||||
Db        952 GTGTATATACACATATATGTGTGTATATACACATATATGTGTGTATATACACATAT 1011

Qy        384 ATACACACATATATACACATATATACACACATATATACACATATATACACATATATACAC 443
              ||       ||||||||||||||       ||||||||||||||     |||||||||
Db       1012 ATGTGTGTGTATATACACATATATGTGTGTGTATATACACATATATGTGTGTATATACAC 1071

Qy        444 ATATACACATATATACACATATATACATATATACACATATATATAATATACACACATATA 503
              |||||    | | || | | | ||| ||| ||| |  ||| | |  |||| | ||| |||
Db       1072 ATATATGTGTGTGTATATACACATATATACATATATCTATGTGT-GTATATATACACATA 1130

Qy        504 TATACACATATATACACACATATATACACATATATACACATATATATACACATATATACA 563
              ||||||  | | |       ||||||| ||||||||    ||||||| ||| | |||||
Db       1131 TATACATGTCTGTGTGTGTATATATATACATATATGTGTATATATACACATACATATATG 1190

Qy        564 CATATATACATATATACACATATATATACATATATACACATATATACATATATACACATA 623
                | | || |||| ||||||||||  || ||| ||     ||||||  | || | | |||
Db       1191 TGTGTGTATATATGTACACATATACGTATATACATGTGTGTATATATGTGTACATATATA 1250

Qy        624 TATACATATATACACACATATATACACATACATATACACACACATAGATATACATATATA 683
```

```
              | |  |||||||     |||||||     || ||||    | | |||  | |  ||||||
Db       1251 TGTGTATATATATGTGTATATATATGTGTATATATGTGTATATATATGTGTGTATATATG 1310

Qy        684 TACACATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATA 735
              |  | |||| | |  || |||||||| | | |||||||||     | ||  |||
Db       1311 TGTATATATGTGTGTGTGTATATATATGTGTATATATGTGTGTGTGTGTATA 1362
```

Search completed: April 21, 2005, 00:48:26
Job time : 1646.07 secs

This Page Blank (uspto)

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 20, 2005, 22:52:20 ; Search time 9251.96 Seconds
(without alignments)
19011.657 Million cell updates/sec

Title: US-09-939-209A-3_COPY_1_4621
Perfect score: 4621
Sequence: 1 agttcaagaccagcctgagc.....ttgctttttagtcctaaaaa 4621

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_htc:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	347.8	7.5	976	9	CL830166	CL830166 OR_CBa005
	2	345.4	7.5	863	9	CL840224	CL840224 OR_CBa006
	3	342	7.4	835	9	CL800294	CL800294 OR_CBa001
	4	341.4	7.4	828	8	CC107667	CC107667 NDL.37F6.
c	5	340.4	7.4	976	9	CL830166	CL830166 OR_CBa005
c	6	338.8	7.3	835	9	CL800294	CL800294 OR_CBa001
c	7	338.4	7.3	863	9	CL840224	CL840224 OR_CBa006
c	8	338.2	7.3	828	8	CC107667	CC107667 NDL.37F6.
c	9	337.4	7.3	620	7	CF626074	CF626074 zmrws05_0
c	10	337.4	7.3	632	7	CF634146	CF634146 zmrww00_0
c	11	337.4	7.3	883	9	CW016544	CW016544 ZMMBLc000
	12	337.2	7.3	883	9	CW016544	CW016544 ZMMBLc000
	13	336.6	7.3	725	8	AQ917102	AQ917102 T233131b
	14	336.4	7.3	827	8	CC067623	CC067623 CSU-K33r.
	15	336.2	7.3	590	9	CR294134	CR294134 Medicago
c	16	335.8	7.3	827	8	CC067623	CC067623 CSU-K33r.
	17	335.2	7.3	620	7	CF626074	CF626074 zmrws05_0
	18	335.2	7.3	632	7	CF634146	CF634146 zmrww00_0
	19	335.2	7.3	717	4	BI505681	BI505681 BB170032B
c	20	334.8	7.2	590	9	CR294134	CR294134 Medicago
c	21	334.2	7.2	717	4	BI505681	BI505681 BB170032B
c	22	333.8	7.2	725	8	AQ917102	AQ917102 T233131b
c	23	333	7.2	744	9	CL623140	CL623140 OR_BBa001
c	24	332.6	7.2	512	8	BZ796807	BZ796807 PUGDF25TD
c	25	331.6	7.2	707	9	CL755178	CL755178 OR_BBa012
	26	331.2	7.2	784	9	CL934691	CL934691 OA_ABa004
	27	331	7.2	512	8	BZ796807	BZ796807 PUGDF25TD
c	28	330.4	7.1	668	9	BX145650	BX145650 Danio rer
	29	330.4	7.1	707	9	CL755178	CL755178 OR_BBa012
c	30	330.2	7.1	496	6	CA916667	CA916667 PCSC21302
c	31	329.8	7.1	751	7	CF924719	CF924719 DD005F11.
	32	329	7.1	668	9	BX145650	BX145650 Danio rer
c	33	328.4	7.1	784	9	CL934691	CL934691 OA_ABa004
c	34	328.2	7.1	587	9	CL776144	CL776144 OR_BBa009
	35	327.2	7.1	751	7	CF924719	CF924719 DD005F11.
	36	326.2	7.1	496	6	CA916667	CA916667 PCSC21302
	37	324.2	7.0	744	9	CL623140	CL623140 OR_BBa001
	38	323.8	7.0	587	9	CL776144	CL776144 OR_BBa009
c	39	322.6	7.0	491	9	CL830207	CL830207 OR_CBa005
	40	318.2	6.9	845	8	CC099198	CC099198 CSU-K34.1
c	41	317	6.9	845	8	CC099198	CC099198 CSU-K34.1
	42	315	6.8	816	9	CL855754	CL855754 OR_CBa008
	43	314.6	6.8	570	4	BI506858	BI506858 BB170024B
c	44	313.2	6.8	570	4	BI506858	BI506858 BB170024B
c	45	311.4	6.7	816	9	CL855754	CL855754 OR_CBa008

ALIGNMENTS

RESULT 1
CL830166
LOCUS CL830166 976 bp DNA linear GSS 09-AUG-2004
DEFINITION OR_CBa0052B14.f OR_CBa Oryza rufipogon genomic clone OR_CBa0052B14 5', genomic survey sequence.
ACCESSION CL830166
VERSION CL830166.1 GI:51075776
KEYWORDS GSS.
SOURCE Oryza rufipogon
ORGANISM Oryza rufipogon
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 976)
AUTHORS Kim,H., Yu,Y., Wissotski,M., Yost,D., Stum,D., Rao,K., Luo,M., Jetty,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
TITLE OMAP project
JOURNAL Unpublished (2004)
COMMENT Contact: Rod A. Wing
Arizona Genomics Institute
University of Arizona
Forbes Building Room 303, Tucson, AZ 85721-0036, USA
Tel: 520 626 9595
Fax: 520 621 1259
Email: http://genome.arizona.edu
PCR PRimers
FORWARD: TAA TAC GAC TCA CTA TAG GG
BACKWARD: CAC TCA TTA GGC ACC CCA
Plate: 0052 row: B column: 14
Seq primer: TAA TAC GAC TCA CTA TAG GG
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .976
/organism="Oryza rufipogon"
/mol_type="genomic DNA"
/db_xref="taxon:4529"
/clone="OR_CBa0052B14"
/tissue_type="young leaves"
/dev_stage="2 week old seedlings"
/lab_host="DH10B T1 phage resistant"
/clone_lib="OR_CBa"
/note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII; drk treated 36 hrs before harvest"
ORIGIN

Query Match 7.5%; Score 347.8; DB 9; Length 976;

```
  Best Local Similarity   69.5%;  Pred. No. 5e-49;
  Matches  486;  Conservative    0;  Mismatches  212;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||||| | | | | | ||||||||||| | |||||||
Db        188 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 247

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db        248 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 307

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        308 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 367

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db        368 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 427

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db        428 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 487

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db        488 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 547

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db        548 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 607

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db        608 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 667

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| |||||||||||      | |||    | |   ||  |      |      | ||
Db        668 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 727

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
              |  | | | |  | | |    | |||    | |   | || |       |      | |
Db        728 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 787

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885
              ||       | ||| ||| || | ||   | |  ||| | |||  ||  ||   | ||
Db        788 TATATATATATATATATATATATATATATATATATATATATAATATATATATATATATAT 847

Qy        886 CACTATATATATATAGTGAAAGCAGTGTAAACCTAAA 924
                ||||||||||||||   | |   | || |  || |
Db        848 ATATATATATATATATATATATATATATATATATATA 886


RESULT 2
CL840224
LOCUS       CL840224                863 bp    DNA     linear   GSS 09-AUG-2004
DEFINITION  OR_CBa0067C10.f OR_CBa Oryza rufipogon genomic clone OR_CBa0067C10
            5', genomic survey sequence.
ACCESSION   CL840224
VERSION     CL840224.1  GI:51085834
KEYWORDS    GSS.
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 863)
  AUTHORS   Kim,H., Yu,Y., Wissotski,M., Yost,D., Stum,D., Rao,K., Luo,M.,
            Jetty,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and
            Wing,R.
  TITLE     OMAP project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0067  row: C  column: 10
            Seq primer: TAA TAC GAC TCA CTA TAG GG
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .863
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
                     /clone="OR_CBa0067C10"
                     /tissue_type="young leaves"
                     /dev_stage="2 week old seedlings"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OR_CBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII;
                     drk treated 36 hrs before harvest"
ORIGIN

  Query Match             7.5%;  Score 345.4;  DB 9;  Length 863;
  Best Local Similarity   70.3%;  Pred. No. 1.3e-48;
  Matches  477;  Conservative    0;  Mismatches  201;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||||| | | | | | ||||||||||| | |||||||
Db        152 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 211

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db        212 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 271

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        272 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 331

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db        332 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 391

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db        392 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 451

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db        452 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 511

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db        512 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 571

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db        572 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 631

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| |||||||||||      | |||    | |   ||  |      |      | ||
Db        632 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 691

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
              |  | | | |  | | |    | |||    | |   | || |       |      | |
Db        692 TATATATATATATATATATATATATATATATATATATATATATATATATTTATATA 751

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885
              ||       | ||| ||| || | ||   | |  ||| | |||  ||  ||   | ||
Db        752 TATATATATATATATATATATATATATATATATATATATATATATATATATATAATATAT 811
```

```
Qy        886 CACTATATATATATATAGT 904
              ||||||||||||||| |
Db        812 ATATATATATATATATAAT 830


RESULT 3
CL800294
LOCUS       CL800294                835 bp    DNA     linear   GSS 06-AUG-2004
DEFINITION  OR_CBa0011N09.r OR_CBa Oryza rufipogon genomic clone OR_CBa0011N09
            3', genomic survey sequence.
ACCESSION   CL800294
VERSION     CL800294.1  GI:51025055
KEYWORDS    GSS.
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 835)
  AUTHORS   Kim,H., Yu,Y., Wissotski,M., Yost,D., Stum,D., Rao,K., Luo,M.,
            Jetty,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and
            Wing,R.
  TITLE     OMAP project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0011  row: N  column: 09
            Seq primer: CAC TCA TTA GGC ACC CCA
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .835
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
                     /clone="OR_CBa0011N09"
                     /tissue_type="young leaves"
                     /dev_stage="2 week old seedlings"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OR_CBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII;
                     drk treated 36 hrs before harvest"
ORIGIN

  Query Match             7.4%;  Score 342;  DB 9;  Length 835;
  Best Local Similarity   67.4%;  Pred. No. 5e-48;
  Matches  511;  Conservative    0;  Mismatches  245;  Indels    2;  Gaps    2;

Qy        212 AAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACATATAT 271
              || |  |  |  || | |||||||||||||||||||||||| | | | | | ||||||
Db         67 AACAGTATTAAACAGTATATATATATATATATATATATATATATATATATATATATAT 126

Qy        272 ATATACACATATATATACACATATATACATATATACACATATATACACATATATATACAC 331
              ||||| | |||||||||| | ||||||| ||||||| | ||||||| | |||||||||| |
Db        127 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 186

Qy        332 ATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACACAC 391
              ||| |||| || | ||||||||| | |||| |||| | ||||||| | ||||||| | ||
Db        187 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 246

Qy        392 ATATATACACATATATACACACATATATACACATATATACACATATATACACATATACAC 451
              ||||||| | ||||||| | | ||||||| | ||||||| | ||||||| | ||||| ||
Db        247 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 306

Qy        452 ATATATACACATATATACATATATACACATATATATA-ATATACACACATATATATACAC 510
              ||||||| | ||||||| ||||||| | ||||||||| ||||| | | |||||||||| |
Db        307 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 366

Qy        511 ATATATACACACATATATACACATATATACACATATATATACACATATATACACATATAT 570
              ||||||| | | ||||||| | ||||||| | |||||||||| | ||||||| | ||||||
Db        367 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 426

Qy        571 ACATATATACACATATATATACATATATACACATATATACATATATACACATATATACAT 630
              | ||||||| | ||||||||| ||||||| | ||||||| ||||||| | ||||||| ||
Db        427 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 486

Qy        631 ATATACACACATATATACACATACATATACACACACATAGATATACATATATATACACAT 690
              ||||| | | ||||||| | ||| ||||| | | | ||| ||||| ||||||||| | ||
Db        487 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 546

Qy        691 ATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGA 750
              ||||||| |||||||| |||||||||||          | |||          | |
Db        547 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 606

Qy        751 TTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTT 810
              ||    |   ||| |   | | | |  | | |   | |||   | |  | || |
Db        607 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 666

Qy        811 CAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA-AG 869
                  |   | |||     | ||| ||| || | ||  | | | ||| | || |
Db        667 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 726

Qy        870 GTAGTTAAGAAGGTAACACTATATATATATATAGTGAAAGCAGTGTAAACCTAAAGGATG 929
              ||  ||   |  ||    |||||||||||||   | |   | || |  || |
Db        727 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 786

Qy        930 GGCCAAGGATTTAAATGTTATAGAAGAATGGCTAAGAT 967
                |   || || || | ||| |   ||   || ||
Db        787 ATATATATATATATATAATATATATATAAAAAATAT 824


RESULT 4
CC107667
LOCUS       CC107667                828 bp    DNA     linear   GSS 16-APR-2003
DEFINITION  NDL.37F6.T7 Notre Dame Liverpool Aedes aegypti genomic clone
            NDL.37F6, genomic survey sequence.
ACCESSION   CC107667
VERSION     CC107667.1  GI:29976722
KEYWORDS    GSS.
SOURCE      Aedes aegypti (yellow fever mosquito)
  ORGANISM  Aedes aegypti
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Aedes;
            Stegomyia.
REFERENCE   1  (bases 1 to 828)
  AUTHORS   Loftus,B., Shetty,J., Knudson,D. and Severson,D.
  TITLE     BAC end sequencing of Aedes aegypti
  JOURNAL   Unpublished (2003)
COMMENT     Other_GSSs: NDL.37F6.SP6
            Contact: Brendan Loftus
            Department of Eukaryotic Genomics
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-3543
            Fax: 301-838-0208
            Email: enta@tigr.org
            Library was provided by David Severson
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .828
                     /organism="Aedes aegypti"
                     /mol_type="genomic DNA"
                     /strain="Liverpool"
                     /db_xref="taxon:7159"
                     /clone="NDL.37F6"
                     /clone_lib="Notre Dame Liverpool"
                     /note="Vector: pECBAC1; Site_1: Hind III; The library was
                     prepared from whole body tissue of newly hatched L1 larvae
```

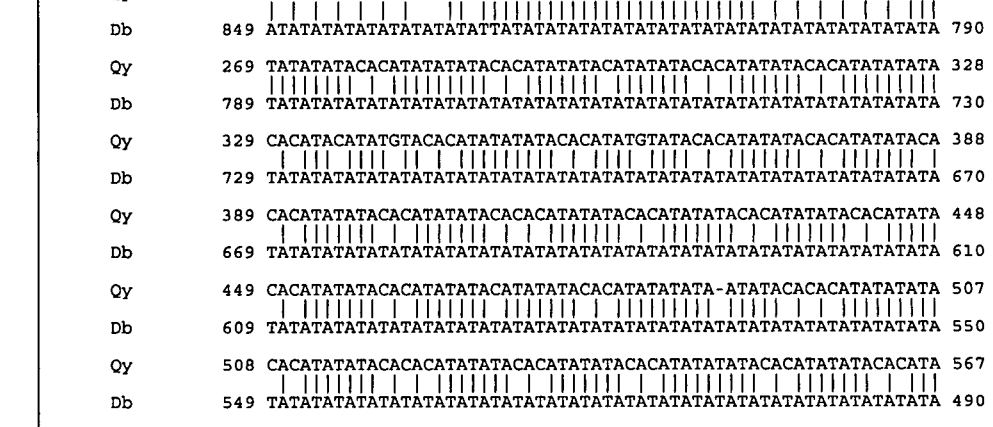
by David Severson at the University of Notre Dame and
Hongbin Zhang"
ORIGIN

Query Match 7.4%; Score 341.4; DB 8; Length 828;
Best Local Similarity 69.7%; Pred. No. 6.3e-48;
Matches 490; Conservative 0; Mismatches 211; Indels 2; Gaps 2;



RESULT 5
CL830166/c
LOCUS CL830166 976 bp DNA linear GSS 09-AUG-2004
DEFINITION OR_CBa0052B14.f OR_CBa Oryza rufipogon genomic clone OR_CBa0052B14 5', genomic survey sequence.
ACCESSION CL830166
VERSION CL830166.1 GI:51075776
KEYWORDS GSS.
SOURCE Oryza rufipogon
ORGANISM Oryza rufipogon
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 976)
AUTHORS Kim,H., Yu,Y., Wissotski,M., Yost,D., Stum,D., Rao,K., Luo,M., Jetty,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
TITLE OMAP project
JOURNAL Unpublished (2004)
COMMENT Contact: Rod A. Wing
Arizona Genomics Institute
University of Arizona
Forbes Building Room 303, Tucson, AZ 85721-0036, USA
Tel: 520 626 9595
Fax: 520 621 1259
Email: http://genome.arizona.edu
PCR PRimers
FORWARD: TAA TAC GAC TCA CTA TAG GG
BACKWARD: CAC TCA TTA GGC ACC CCA
Plate: 0052 row: B column: 14
Seq primer: TAA TAC GAC TCA CTA TAG GG
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .976
/organism="Oryza rufipogon"
/mol_type="genomic DNA"
/db_xref="taxon:4529"
/clone="OR_CBa0052B14"
/tissue_type="young leaves"
/dev_stage="2 week old seedlings"
/lab_host="DH10B T1 phage resistant"
/clone_lib="OR_CBa"
/note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII; drk treated 36 hrs before harvest"
ORIGIN

Query Match 7.4%; Score 340.4; DB 9; Length 976;
Best Local Similarity 68.9%; Pred. No. 9.1e-48;
Matches 495; Conservative 0; Mismatches 221; Indels 2; Gaps 2;



```
Qy        808 TTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA 867
              | |     |       | |||     | ||| ||| || | ||   | |  ||| | ||
Db        249 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 190

Qy        868 -AGGTAGTTAAGAAGGTAACACTATATATATATATAGTGAAAGCAGTGTAAACCTAAA 924
               |  ||  ||   |  ||    ||||||||||||||   | |    | || |  || |
Db        189 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 132


RESULT 6
CL800294/c
LOCUS       CL800294                835 bp    DNA     linear   GSS 06-AUG-2004
DEFINITION  OR_CBa0011N09.r OR_CBa Oryza rufipogon genomic clone OR_CBa0011N09
            3', genomic survey sequence.
ACCESSION   CL800294
VERSION     CL800294.1  GI:51025055
KEYWORDS    GSS.
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 835)
  AUTHORS   Kim,H., Yu,Y., Wissotski,M., Yost,D., Stum,D., Rao,K., Luo,M.,
            Jetty,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and
            Wing,R.
  TITLE     OMAP project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0011  row: N  column: 09
            Seq primer: CAC TCA TTA GGC ACC CCA
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .835
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
                     /clone="OR_CBa0011N09"
                     /tissue_type="young leaves"
                     /dev_stage="2 week old seedlings"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OR_CBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII;
                     drk treated 36 hrs before harvest"
ORIGIN

  Query Match             7.3%;  Score 338.8;  DB 9;  Length 835;
  Best Local Similarity   69.5%;  Pred. No. 1.7e-47;
  Matches  488;  Conservative    0;  Mismatches  212;  Indels    2;  Gaps    2;

Qy        225 ATTTTATATATATATATATATATATATACACACACACACACATATATATATACACATATA 284
              || |||||||||||||||||||||||| | | | | | | |||||||||||| | |||||
Db        809 ATATTATATATATATATATATATATATATATATATATATATATATATATATATATATATA 750

Qy        285 TATACACATATATACATATATACACATATATACACATATATATACACATACATATGTACA 344
              ||||  | |||||||  |||||||  | |||||||  | |||||||||  | |||  |||| || |
Db        749 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 690

Qy        345 CATATATATACACATATGTATACACATATATACACATATATACACACATATATACACATA 404
               |||||||||  | ||||  ||||  | |||||||  | |||||||  | |  |||||||  | |||
Db        689 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 630

Qy        405 TATACACACATATATACACATATATACACATATATACACATATACACATATATACACATA 464
```

```
              |||| | | |||||||  | |||||||  | |||||||  | |||||  | |||||||  | |||
Db        629 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 570

Qy        465 TATACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACA 523
              ||||  |||||||  | |||||||||  |||||  | | |||||||||  | |||||||  | | |
Db        569 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 510

Qy        524 TATATACACATATATACACATATATATACACATATATACACATATATACATATATACACA 583
              ||||||  | |||||||  | |||||||||  | |||||||  | |||||||  |||||||  | |
Db        509 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 450

Qy        584 TATATATACATATATACACATATATACATATATACACATATATACATATATACACACATA 643
              ||||||||  |||||||  | |||||||  |||||||  | |||||||  |||||||  | | |||
Db        449 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 390

Qy        644 TATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATA 703
              ||||  | |||  |||||  | |  | |||  |||||  |||||||||  | |||||||||  ||||
Db        389 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 330

Qy        704 TATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGA 763
              |||||  ||||||||||||          |  |||          |  |   ||    |   |
Db        329 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 270

Qy        764 TAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCA 823
              ||  |   | | |  |  | |  |  | |||    | |   | || |    |      |
Db        269 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 210

Qy        824 AATAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA-AGGTAGTTAAGAAGG 882
              |||      | ||| ||| || | ||   | |  ||| | || |  ||  ||   |
Db        209 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 150

Qy        883 TAACACTATATATATATAGTGAAAGCAGTGTAAACCTAAA 924
              ||    ||||||||||||||   | |    | || |  || |
Db        149 TATATATATATATATATATATATATATATATATATATATATA 108


RESULT 7
CL840224/c
LOCUS       CL840224                863 bp    DNA     linear   GSS 09-AUG-2004
DEFINITION  OR_CBa0067C10.f OR_CBa Oryza rufipogon genomic clone OR_CBa0067C10
            5', genomic survey sequence.
ACCESSION   CL840224
VERSION     CL840224.1  GI:51085834
KEYWORDS    GSS.
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 863)
  AUTHORS   Kim,H., Yu,Y., Wissotski,M., Yost,D., Stum,D., Rao,K., Luo,M.,
            Jetty,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and
            Wing,R.
  TITLE     OMAP project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0067  row: C  column: 10
            Seq primer: TAA TAC GAC TCA CTA TAG GG
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .863
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
```

```
                     /clone="OR_CBa0067C10"
                     /tissue_type="young leaves"
                     /dev_stage="2 week old seedlings"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OR_CBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII;
                     drk treated 36 hrs before harvest"
ORIGIN

  Query Match             7.3%;  Score 338.4;  DB 9;  Length 863;
  Best Local Similarity   69.1%;  Pred. No. 2e-47;
  Matches  492;  Conservative    0;  Mismatches  216;  Indels    4;  Gaps    2;

Qy        204 ATCTCAGAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACAC 263
              ||   | | | | | | |   || ||||||||||||||||||||||||| | | | | |
Db        830 ATTATATATATATATATATATATATTATATATATATATATATATATATATATATATAT 771

Qy        264 ACATATATATATACACATATATATACACATATATACATATATACACATATATACACATAT 323
              | |||||||||||| | ||||||||| | ||||||| ||||||| | ||||||| | ||||
Db        770 ATATATATATATATATATATATATATAAATATATATATATATATATATATATATATATAT 711

Qy        324 ATATACACATACATATGTACACATATATATACACATATGTATACACATATATACACATAT 383
              ||||| | ||| |||| || | ||||||||| | |||| |||| | ||||||| | ||||
Db        710 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 651

Qy        384 ATACACACATATATACACATATATACACACATATATACACATATATACACATATATACAC 443
              ||| | | ||||||| | ||||||| | | ||||||| | ||||||| | ||||||| |
Db        650 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 591

Qy        444 ATATACACATATATACACATATATACATATATACACATATATATA-ATATACACACATAT 502
              ||||| | ||||||| | ||||||| ||||||| | ||||||||| ||||| | | ||||
Db        590 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 531

Qy        503 ATATACACATATATACACACATATATACACATATATACACATATATATACACATATATAC 562
              ||||| | ||||||| | | ||||||| | ||||||| | ||||||||| | ||||||||
Db        530 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 471

Qy        563 ACATATATACATATATACACATATATATACATATATACACATATATACATATATACACAT 622
              | ||||||| ||||||| | ||||||||| ||||||| | ||||||| ||||||| | ||
Db        470 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 411

Qy        623 ATATACATATATACACACATATATACACATACATATACACACACATAGATATACATATAT 682
              ||||| ||||||| | | ||||||| | ||| ||||| | | | ||| ||||| ||||||
Db        410 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 351

Qy        683 ATACACATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTA 742
              ||| | ||||||||| |||||||||| ||||||||||          |   |||
Db        350 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 291

Qy        743 GCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAA 802
                | |   ||    |    ||| |   | | | |  | | |    | |||    | |
Db        290 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 231

Qy        803 GGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTAT 862
              | || |    |  |  | |||     | ||| ||| || | ||   | |  |||
Db        230 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 171

Qy        863 CTGTAAGGTAGTTAAGAAGGTAACACTATATATATATATAGTGAAAGCAGTG 914
              | ||   ||  ||   |  ||    |||||||||| | ||   | | |||
Db        170 ATATA---TATATATATATATATATATATATATGTTGTTTCACCTGTG 122


RESULT 8
CC107667/c
LOCUS       CC107667                828 bp    DNA     linear   GSS 16-APR-2003
DEFINITION  NDL.37F6.T7 Notre Dame Liverpool Aedes aegypti genomic clone
            NDL.37F6, genomic survey sequence.
ACCESSION   CC107667
VERSION     CC107667.1  GI:29976722
KEYWORDS    GSS.
SOURCE      Aedes aegypti (yellow fever mosquito)
  ORGANISM  Aedes aegypti
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Aedes;
            Stegomyia.
REFERENCE   1  (bases 1 to 828)
  AUTHORS   Loftus,B., Shetty,J., Knudson,D. and Severson,D.
  TITLE     BAC end sequencing of Aedes aegypti
  JOURNAL   Unpublished (2003)
COMMENT     Other_GSSs: NDL.37F6.SP6
            Contact: Brendan Loftus
            Department of Eukaryotic Genomics
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-3543
            Fax: 301-838-0208
            Email: enta@tigr.org
            Library was provided by David Severson
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .828
                     /organism="Aedes aegypti"
                     /mol_type="genomic DNA"
                     /strain="Liverpool"
                     /db_xref="taxon:7159"
                     /clone="NDL.37F6"
                     /clone_lib="Notre Dame Liverpool"
                     /note="Vector: pECBAC1; Site_1: Hind III; The library was
                     prepared from whole body tissue of newly hatched L1 larvae
                     by David Severson at the University of Notre Dame and
                     Hongbin Zhang"
ORIGIN

  Query Match             7.3%;  Score 338.2;  DB 8;  Length 828;
  Best Local Similarity   70.2%;  Pred. No. 2.2e-47;
  Matches  482;  Conservative    0;  Mismatches  203;  Indels    2;  Gaps    2;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | |||||||||||||||||||||||| | | | | | |||||||||| | |||||||
Db        817 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 758

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | |||||||||| | ||| |||| || | |
Db        757 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 698

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        697 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 638

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db        637 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 578

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | |||||||||| | ||||||| | | |||
Db        577 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 518

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db        517 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 458

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db        457 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 398

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| |||||||||| | |||||||||  ||||||
Db        397 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 338

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||          | |||          | |   ||    |    |||
Db        337 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 278

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
                |  | | | |  | | |   | |||   | |  | || |     |     | |
```

```
Db        277 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 218

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTA 884
              ||       | ||| ||| || | ||   | |  ||| | || |  ||  ||   |  ||
Db        217 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 158

Qy        885 ACACTATATATATATAGTGAAAGCA 911
                |||||||||||||||   || |||
Db        157 TATATATATATATATATATAATGCA 131
```

RESULT 9
CF626074/c
LOCUS CF626074 620 bp mRNA linear EST 02-OCT-2003
DEFINITION zmrws05_0B10-009-d11.s3 zmrws05 Zea mays cDNA 3', mRNA sequence.
ACCESSION CF626074
VERSION CF626074.1 GI:37378736
KEYWORDS EST.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 620)
AUTHORS Bohnert,H., Sharp,R.E., Springer,G.K., Poroyko,V., Fredricksen,M., Sharp,L.G., Spollen,W.G., Ries,J., Guillen,A., Khambati,A., Topinka,C., Davis,G.E., Schachtman,D., Wu,Y. and Nguyen,H.T.
TITLE NSF Grant DBI-0211842: Functional Genomics of Root Growth and Root Signaling Under Drought
JOURNAL Unpublished (2003)
COMMENT Contact: Hans Bohnert
University of Illinois, Urbana-Champaign
1201 West Gregory Drive, Urbana, IL 61801, USA
Tel: 217-265-5475
Fax: 217-333-5574
Email: bohnerth@life.uiuc.edu
POLYA=Yes.
FEATURES Location/Qualifiers
source 1. .620
/organism="Zea mays"
/mol_type="mRNA"
/db_xref="taxon:4577"
/clone_lib="zmrws05"
/note="Samples were collected in Robert E. Sharp's lab (University of Missouri-Columbia) to construct three normalized cDNA libraries. Dark-grown maize seedlings with primary roots 12-20 mm in length were transplanted to high (-0.03 MPa) or low water potential (-1.6 MPa) vermiculite, and harvested at 5 h and 48 h after transplanting. About 1,000 roots were used for each of the low water potential libraries (zmrws05 and zmrws48) while 500 roots were combined from each of the two time points at high water potential (zmrww00). Each root was divided into 4 segments (distances are from the junction of the root apex and root cap): segment 1, 0-3 mm plus the root cap; segment 2, 3-7 mm; segment 3, 7-12 mm; segment 4, 12-20 mm. (For detatils of conditions see (1) with nutrient modifications as in (2)). The three normalized cDNA libraries were constructed in the lab of Hans Bohnert (University of Illinois-UC). Total RNA was extracted by the 'hot Phenol' method (Plant Molecular Biology manual. D5: 1-13, 2nd ed., 1997). This method worked in eliminating carbohydrate material present in the root tips. The integrity of the RNA was verified by denaturing agarose gels and spectrophotometry (ratio A260/280). Poly(A)+mRNA was isolated twice from total RNA using the Oligotex Direct mRNA kit (Qiagen). Poly(A)+ mRNA was converted to double-stranded cDNA and tagged by using modified Oligo(dT) primers. One of 4 sequence tags corresponding to a different segment of the root was added to the 3'-end of the modified Oligo(dT) primers, including a Not1 site and used to reverse transcribe the segment-specific mRNAs into cDNAs. Each library contains all four tags. A suffix (s1, s2, s3, or s4) has been added to each sequence identifier to designate which region of the root (Root_segment_1, 2, 3, or 4) the sequence was found in based on the identification of the tag. A suffix of s0 indicates that the sequence tag, and hence the source segment, could not be identified. The double stranded cDNAs were size-selected (>450 bp). Size selected cDNAs were adaptored with EcoRI adaptors at both ends, and then digested with NotI. The cDNA was directionally cloned into EcoR1-NotI digested pBS II SK(+) phagemid vector (Stratagene) and electroporated into E.coli DH10B. The total number of white colony forming units (cfu) in the primary libraries before amplification was as follows: zmrws05: 3.37 x 106; zmrws48: 4.87 x 106; zmrww00: 3 x 106. The background of empty clones was less than 1%. Inserts ranged from ~0.5kb to >2.5 kb, as determined by PCR. Plasmid DNA from the primary libraries then was converted to single-stranded circles and used as a template for PCR amplification using the T7 and T3 priming sites that flank the cloned cDNA inserts. The purified PCR products, representing the entire cDNA population cloned in each library, were used as a driver for normalization. Hybridization between the single-stranded library and the PCR products was carried out for 44 hours at 30oC. Non-hybridized single-stranded DNA circles were separated from hybridized DNA rendered partially double-stranded and electroporated into DH10B. The total number of clones with insert was: zmrws05: 2.0x107; zmrws48: 4.2x107; zmrww00: 1.1x107. The background of empty clones was less than 2%. Insert size, determined by PCR of the entire library, ranged from 0.5kb to 2.5kb. (1) Sharp R E; Silk W K; Hsiao T C. Growth of the Maize Primary Root at Low Water Potentials I. Spatial Distribution of Expansive Growth. Plant Physiology (Rockville). 87(1). 1988. 50-57. (2) Spollen W G; LeNoble M E; Samuels T D; Bernstein N; Sharp R E. Abscisic acid accumulation maintains maize primary root elongation at low water potentials by restricting ethylene production. Plant Physiology (Rockville). 122(3). March, 2000. 967-976.
TAG_TISSUE=Root_segment_3
TAG_SEQ=TCGCA"
ORIGIN

Query Match 7.3%; Score 337.4; DB 7; Length 620;
Best Local Similarity 78.7%; Pred. No. 3.1e-47;
Matches 415; Conservative 0; Mismatches 111; Indels 1; Gaps 1;

```
Qy        227 TTTATATATATATATATATATATATACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||||| | | | | | ||||||||||| | |||||||
Db        566 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 507

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db        506 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 447

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        446 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 387

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db        386 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 327

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db        326 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 267

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db        266 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 207

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db        206 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 147
```

```
Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db        146 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 87

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATT 752
              ||| ||||||||||          ||  || ||    | ||| |||||
Db         86 TATATATATATATATATATATATGTAAATTAGCATCACCAGCTGATT 40


RESULT 10
CF634146/c
LOCUS       CF634146               632 bp    mRNA    linear   EST 02-OCT-2003
DEFINITION  zmrww00_0A10-008-g12.s3 zmrww00 Zea mays cDNA 3', mRNA sequence.
ACCESSION   CF634146
VERSION     CF634146.1  GI:37393761
KEYWORDS    EST.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 632)
  AUTHORS   Bohnert,H., Sharp,R.E., Springer,G.K., Poroyko,V., Fredricksen,M.,
            Sharp,L.G., Spollen,W.G., Ries,J., Guillen,A., Khambati,A.,
            Topinka,C., Davis,G.E., Schachtman,D., Wu,Y. and Nguyen,H.T.
  TITLE     NSF Grant DBI-0211842: Functional Genomics of Root Growth and Root
            Signaling Under Drought
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Hans Bohnert
            University of Illinois, Urbana-Champaign
            1201 West Gregory Drive, Urbana, IL 61801, USA
            Tel: 217-265-5475
            Fax: 217-333-5574
            Email: bohnerth@life.uiuc.edu
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .632
                     /organism="Zea mays"
                     /mol_type="mRNA"
                     /db_xref="taxon:4577"
                     /clone_lib="zmrww00"
                     /note="Samples were collected in Robert E. Sharp's lab
                     (University of Missouri-Columbia) to construct three
                     normalized cDNA libraries. Dark-grown maize seedlings with
                     primary roots 12-20 mm in length were transplanted to high
                     (-0.03 MPa) or low water potential (-1.6 MPa) vermiculite,
                     and harvested at 5 h and 48 h after transplanting. About
                     1,000 roots were used for each of the low water potential
                     libraries (zmrws05 and zmrws48) while 500 roots were
                     combined from each of the two time points at high water
                     potential (zmrww00). Each root was divided into 4 segments
                     (distances are from the junction of the root apex and root
                     cap): segment 1, 0-3 mm plus the root cap; segment 2, 3-7
                     mm; segment 3, 7-12 mm; segment 4, 12-20 mm. (For detatils
                     of conditions see (1) with nutrient modifications as in
                     (2)). The three normalized cDNA libraries were constructed
                     in the lab of Hans Bohnert (University of Illinois-UC).
                     Total RNA was extracted by the 'hot Phenol' method (Plant
                     Molecular Biology manual. D5: 1-13, 2nd ed., 1997). This
                     method worked in eliminating carbohydrate material present
                     in the root tips. The integrity of the RNA was verified by
                     denaturing agarose gels and spectrophotometry (ratio
                     A260/280). Poly(A)+mRNA was isolated twice from total RNA
                     using the Oligotex Direct mRNA kit (Qiagen). Poly(A)+ mRNA
                     was converted to double-stranded cDNA and tagged by using
                     modified Oligo(dT) primers. One of 4 sequence tags
                     corresponding to a different segment of the root was added
                     to the 3'-end of the modified Oligo(dT) primers, including
                     a Not1 site and used to reverse transcribe the
                     segment-specific mRNAs into cDNAs. Each library contains
                     all four tags. A suffix (s1, s2, s3, or s4) has been added
                     to each sequence identifier to designate which region of
                     the root (Root_segment_1, 2, 3, or 4) the sequence was
                     found in based on the identification of the tag. A suffix
                     of s0 indicates that the sequence tag, and hence the
                     source segment, could not be identified. The double
                     stranded cDNAs were size-selected (>450 bp). Size selected
                     cDNAs were adaptored with EcoRI adaptors at both ends, and
                     then digested with NotI. The cDNA was directionally cloned
                     into EcoR1-NotI digested pBS II SK(+) phagemid vector
                     (Stratagene) and electroporated into E.coli DH10B. The
                     total number of white colony forming units (cfu) in the
                     primary libraries before amplification was as follows:
                     zmrws05: 3.37 x 106; zmrws48: 4.87 x 106; zmrww00: 3 x
                     106. The background of empty clones was less than 1%.
                     Inserts ranged from ~0.5kb to >2.5 kb, as determined by
                     PCR. Plasmid DNA from the primary libraries then was
                     converted to single-stranded circles and used as a
                     template for PCR amplification using the T7 and T3 priming
                     sites that flank the cloned cDNA inserts. The purified PCR
                     products, representing the entire cDNA population cloned
                     in each library, were used as a driver for normalization.
                     Hybridization between the single-stranded library and the
                     PCR products was carried out for 44 hours at 30oC.
                     Non-hybridized single-stranded DNA circles were separated
                     from hybridized DNA rendered partially double-stranded and
                     electroporated into DH10B. The total number of clones with
                     insert was: zmrws05: 2.0x107; zmrws48: 4.2x107; zmrww00:
                     1.1x107. The background of empty clones was less than 2%.
                     Insert size, determined by PCR of the entire library,
                     ranged from 0.5kb to 2.5kb. (1) Sharp R E; Silk W K; Hsiao
                     T C. Growth of the Maize Primary Root at Low Water
                     Potentials I. Spatial Distribution of Expansive Growth.
                     Plant Physiology (Rockville). 87(1). 1988. 50-57. (2)
                     Spollen W G; LeNoble M E; Samuels T D; Bernstein N; Sharp
                     R E. Abscisic acid accumulation maintains maize primary
                     root elongation at low water potentials by restricting
                     ethylene production. Plant Physiology (Rockville). 122(3).
                     March, 2000. 967-976.
                     TAG_TISSUE=Root_segment_3
                     TAG_SEQ=TCGCA"
ORIGIN

  Query Match             7.3%;  Score 337.4;  DB 7;  Length 632;
  Best Local Similarity   78.7%;  Pred. No. 3.1e-47;
  Matches  415;  Conservative    0;  Mismatches  111;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACATATATATATACACATATATA 286
              | |||||||||||||||||||||||||| | | | | | |||||||||| | |||||||
Db        545 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 486

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db        485 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 426

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        425 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 366

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db        365 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 306

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db        305 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 246

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db        245 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 186

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              ||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db        185 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 126
```

```
Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db        125 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 66

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATT 752
              ||| ||||||||||||       ||  || ||    | ||| |||||
Db         65 TATATATATATATATATATATATATGTAAATTAGCATCACCAGCTGATT 19


RESULT 11
CW016544/c
LOCUS       CW016544                883 bp    DNA     linear   GSS 23-SEP-2004
DEFINITION  ZMMBLc0009C11.f ZMMBLc Zea mays genomic clone ZMMBLc0009C11 5',
            genomic survey sequence.
ACCESSION   CW016544
VERSION     CW016544.1  GI:52606273
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 883)
  AUTHORS   Wing,R., Luo,M., Soderlund,C. and Haller,K.
  TITLE     ZMMBL sequences
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            Plate: 0009  row: C  column: 11
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .883
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /cultivar="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBLc0009C11"
                     /tissue_type="immature ears"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="ZMMBLc"
                     /note="Vector: pAGIBAC1; Site_1: Hpa II; Site_2: Hpa II"
ORIGIN

  Query Match             7.3%;  Score 337.4;  DB 9;  Length 883;
  Best Local Similarity   81.4%;  Pred. No. 3e-47;
  Matches  403;  Conservative    0;  Mismatches   91;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | |||||||||||||||||||||||| | | ||| | | |||||||||||| |||||||||
Db        872 TATATATATATATATATATATATATATATACATATATATATATATATATACATATATA 813

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              |||| ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db        812 TACATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 753

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        752 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 693

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db        692 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 633

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db        632 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 573

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db        572 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 513

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db        512 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 453

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db        452 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 393

Qy        706 TATGTATATATATAT 720
              ||| ||||||||||||
Db        392 TATATATATATATAT 378


RESULT 12
CW016544
LOCUS       CW016544                883 bp    DNA     linear   GSS 23-SEP-2004
DEFINITION  ZMMBLc0009C11.f ZMMBLc Zea mays genomic clone ZMMBLc0009C11 5',
            genomic survey sequence.
ACCESSION   CW016544
VERSION     CW016544.1  GI:52606273
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 883)
  AUTHORS   Wing,R., Luo,M., Soderlund,C. and Haller,K.
  TITLE     ZMMBL sequences
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            Plate: 0009  row: C  column: 11
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .883
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /cultivar="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBLc0009C11"
                     /tissue_type="immature ears"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="ZMMBLc"
                     /note="Vector: pAGIBAC1; Site_1: Hpa II; Site_2: Hpa II"
ORIGIN

  Query Match             7.3%;  Score 337.2;  DB 9;  Length 883;
  Best Local Similarity   71.3%;  Pred. No. 3.2e-47;
  Matches  458;  Conservative    0;  Mismatches  183;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | |||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db        233 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 292

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db        293 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 352

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        353 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 412

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
```

```
Db        413 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 472

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db        473 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 532

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db        533 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 592

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db        593 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 652

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| |||||||| | |||||||| ||||||
Db        653 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 712

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||          | |||          | |   ||    |    |||
Db        713 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 772

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
              |   | | | |  | | |   | |||   | | | | || |  |  |      | |
Db        773 TATATATATATATATATATATATATATATATATATATGTATATATATGTATATATATATA 832

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA 867
              ||    | | ||| ||| || | ||   | |  ||| | ||
Db        833 TATATATGTATATATATATATATATATATATATATATATA 874


RESULT 13
AQ917102
LOCUS       AQ917102                725 bp    DNA     linear   GSS 07-DEC-1999
DEFINITION  T233131b Medicago truncatula BAC library Medicago truncatula
            genomic clone 24E12, genomic survey sequence.
ACCESSION   AQ917102
VERSION     AQ917102.1  GI:6536747
KEYWORDS    GSS.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 725)
  AUTHORS   Kim,D., Limpens,E., Peng,H., Ellis,L. and Cook,D.R.
  TITLE     BAC end sequencing of Medicago truncatula
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Cook DR
            The Crop Biotechnology Center
            Texas A&M University
            Department of Plant Pathology and Microbiology, Rm 120 L.F.
            Peterson Bldg, College Station, TX 77843-2132, USA
            Tel: 409 845 8743
            Fax: 409 862 4790
            Email: dcook@ppserver.tamu.edu
            Other name: DK277-L; date: 8/12/99; Submitted to the Database of
            Genome Survey Sequences (GSS) on 12/06/99; More information is
            available at 'http://chrysie.tamu.edu/medicago'.
            Seq primer: SQ-BAC-L (AACGCCAGGGTTTTCCCAGTCACGACG)
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .725
                     /organism="Medicago truncatula"
                     /mol_type="genomic DNA"
                     /cultivar="genotype A17"
                     /db_xref="taxon:3880"
                     /clone="24E12"
                     /clone_lib="Medicago truncatula BAC library"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII; Nam. Y-W, Penmetsa, R.V., Endre, G., Kim, D., and
                     Cook, D.R.  1999. Construction of a bacterial artificial
                     chromosome  library of Medicago truncatula and
                     identification of clones containing ethylene response
                     genes. Theor Appl Genet 98:  638-646. "
ORIGIN

  Query Match             7.3%;  Score 336.6;  DB 8;  Length 725;
  Best Local Similarity   77.7%;  Pred. No. 4.2e-47;
  Matches  419;  Conservative    0;  Mismatches  119;  Indels    1;  Gaps    1;

Qy        183 AGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATA 242
              ||  || ||   | |  |||  |   |   ||||   | | |||  |  ||||||||||
Db        158 AGCAAAGAAACCCAATTTGTGGTGGAAATGAAAAGGGATGCCATGATGGATATATATATA 217

Qy        243 TATATATATACACACACACACATATATATATACACATATATATACACATATATACATA 302
              |||||||||| | | | | | | |||||||||| | ||||||||| | ||||||| |||
Db        218 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 277

Qy        303 TATACACATATATACACATATATATACACATACATATGTACACATATATATACACATATG 362
              |||| | ||||||| | |||||||| | ||| |||| || | ||||||||| | ||||
Db        278 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 337

Qy        363 TATACACATATATACACATATATACACACATATATACACATATATACACACATATATACA 422
              |||| | ||||||| | ||||||| | | ||||||| | ||||||| | | ||||||| |
Db        338 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 397

Qy        423 CATATATACACATATATACACATATACACATATATACACATATATACATATATACACATA 482
              ||||||| | ||||||| | ||||| | ||||||| | ||||||| ||||||| | |||
Db        398 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 457

Qy        483 TATATA-ATATACACACATATATATACACATATATACACACATATATACACATATATACA 541
              |||||| ||||| | | |||||||||| | ||||||| | | ||||||| | ||||||| |
Db        458 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 517

Qy        542 CATATATATACACATATATACACATATATACATATATACACATATATATACATATATACA 601
              ||||||||| | ||||||| ||||||| | ||||||||| ||||||| |
Db        518 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 577

Qy        602 CATATATACATATATACACATATATACATATATACACACATATATACACATACATATACA 661
              ||||||| ||||||| | ||||||| ||||||| | | ||||||| | ||| ||||| |
Db        578 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 637

Qy        662 CACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATAT 720
              | | ||| ||||| ||||||||| | ||||||||  ||||||||| ||||||||||
Db        638 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 696


RESULT 14
CC067623
LOCUS       CC067623                827 bp    DNA     linear   GSS 16-APR-2003
DEFINITION  CSU-K33r.18M15.SP6 CSU-K33r Aedes aegypti genomic clone
            CSU-K33r.18M15, genomic survey sequence.
ACCESSION   CC067623
VERSION     CC067623.1  GI:29906129
KEYWORDS    GSS.
SOURCE      Aedes aegypti (yellow fever mosquito)
  ORGANISM  Aedes aegypti
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Aedes;
            Stegomyia.
REFERENCE   1  (bases 1 to 827)
  AUTHORS   Loftus,B., Shetty,J., Severson,D., Brown,S. and Knudson,D.
  TITLE     End sequencing of Aedes aegypti BACs
  JOURNAL   Unpublished (2003)
COMMENT     Other_GSSs: CSU-K33r.18M15.T7
            Contact: Brendan Loftus
            Department of Eukaryotic Genomics
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-3543
            Fax: 301-838-0208
            Email: enta@tigr.org
            Library was provided by Susan Brown and Dennis Knudson at Colorado
            State University.
```

```
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .827
                     /organism="Aedes aegypti"
                     /mol_type="genomic DNA"
                     /strain="Rexville"
                     /db_xref="taxon:7159"
                     /clone="CSU-K33r.18M15"
                     /clone_lib="CSU-K33r"
                     /note="Vector: pBeloBAC11; Site_1: HindIII"
ORIGIN

  Query Match             7.3%;  Score 336.4;  DB 8;  Length 827;
  Best Local Similarity   78.7%;  Pred. No. 4.4e-47;
  Matches  414;  Conservative    0;  Mismatches  111;  Indels    1;  Gaps    1;

Qy        196 GAAATGTCATCTCAGAAAAAAAAAAAGGCATTTTATATATATATATATATATATACAC 255
              ||| |  | ||  |  ||     |||| |     |||||||||||||||||||||||| |
Db        238 GAATTTCCTTCAAACGAATGCTAAAACGTGCGATATATATATATATATATATATATAT 297

Qy        256 ACACACACACATATATATATACACATATATATACACATATATACATATATACACATATAT 315
              | | | | | |||||||||||| | ||||||||| | ||||||| ||||||| | ||||||
Db        298 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 357

Qy        316 ACACATATATATACACATACATATGTACACATATATATACACATATGTATACACATATAT 375
              | | ||||||||| | ||| |||| || | ||||||||| | |||| |||| | ||||||
Db        358 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 417

Qy        376 ACACATATATACACACATATATACACATATATACACACATATATACACATATATACACAT 435
              | |.|||||||| | | ||||||| | ||||||| | | ||||||| | ||||||| | ||
Db        418 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 477

Qy        436 ATATACACATATACACATATATACACATATATACACATATATACACATATATATA-ATATAC 494
              ||||| | ||||| | ||||||| | ||||||| ||||||| | ||||||||| |||||
Db        478 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 537

Qy        495 ACACATATATATACACATATATACACACATATATACACATATATACACATATATATACAC 554
              | |.|||||||| | ||||||| | | ||||||| | ||||||| | ||||||||| |
Db        538 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 597

Qy        555 ATATATACACATATATACATATATACACATATATATACATATATACACATATATACATAT 614
              ||||||| | ||||||| ||||||| | ||||||||| ||||||| | |||||||| ||||
Db        598 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 657

Qy        615 ATACACATATATACATATATACACACATATATACACATACATATACACACACATAGATAT 674
              ||| | ||||||| ||||||| | | ||||||| | ||| ||||| | | | ||| ||||
Db        658 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 717

Qy        675 ACATATATATACACATATATATACGTATATATATGTATATATATAT 720
              | ||||||||| | |||||||||  ||||||||| ||||||||||
Db        718 ATATATATATATATATATATATATATATATATATATATATATAT 763


RESULT 15
CR294134
LOCUS       CR294134                590 bp    DNA     linear   GSS 28-FEB-2004
DEFINITION  Medicago truncatula BAC ends cultivar Jemalong A17 of Medicago
            truncatula, genomic survey sequence.
ACCESSION   CR294134
VERSION     CR294134.1  GI:44684558
KEYWORDS    GSS.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 590)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-FEB-2004) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
FEATURES             Location/Qualifiers
     source          1. .590
                     /organism="Medicago truncatula"
                     /mol_type="genomic DNA"
                     /cultivar="Jemalong A17"
                     /db_xref="taxon:3880"
                     /clone_lib="MTE1"
                     /note="Vector: pIndigoBAC ; Site_1: EcoRI ; Site_2: EcoRI
                     ; Debelle F. and Chalhoub B.-Genoscope sequence ID :
                     mte1-1J18FM1"
ORIGIN

  Query Match             7.3%;  Score 336.2;  DB 9;  Length 590;
  Best Local Similarity   79.7%;  Pred. No. 5e-47;
  Matches  409;  Conservative    0;  Mismatches  103;  Indels    1;  Gaps    1;

Qy        209 AGAAAAAAAAAAAGGCATTTTATATATATATATATATATATATACACACACACACATA 268
              || | | | | | | ||  |||||||||||||||||||||||| | | | | | |||
Db         15 AGGAGATATATATAGTATAGTATATATATATATATATATATATATATATATATATATA 74

Qy        269 TATATATACACATATATATACACATATATACATATATACACATATATACACATATATATA 328
              |||||||| | ||||||||| | ||||||| ||||||| | ||||||| | |||||||||
Db         75 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 134

Qy        329 CACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACA 388
              | ||| |||| || | ||||||||| | |||| |||| | ||||||| | ||||||| |
Db        135 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 194

Qy        389 CACATATATACACATATATACACACATATATACACATATATACACATATATACACATATA 448
              | ||||||| | ||||||| | | ||||||| | ||||||| | ||||||| | |||||
Db        195 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 254

Qy        449 CACATATATACACATATATACATATATACACATATATATA-ATATACACACATATATATA 507
              | ||||||| | ||||||| ||||||| | ||||||||| |||||.| | |||||||||
Db        255 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 314

Qy        508 CACATATATACACACATATATACACATATATACACATATATATACACATATATACACATA 567
              | ||||||| | | ||||||| | ||||||| | ||||||||| | ||||||| | |||
Db        315 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 374

Qy        568 TATACATATATACACATATATATACATATATACACATATATACATATATACACATATATA 627
              |||| ||||||| | ||||||||| ||||||| | ||||||| ||||||| | |||||||
Db        375 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 434

Qy        628 CATATATACACACATATATACACATACATATACACACACATAGATATACATATATATACA 687
              ||||||| | | ||||||| | ||| ||||| | | | ||| ||||| |||||||||  |
Db        435 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 494

Qy        688 CATATATATACGTATATATATGTATATATATAT 720
              |||||||||  ||||||||| ||||||||||
Db        495 TATATATATATATATATATATATATATATAT 527


RESULT 16
CC067623/c
LOCUS       CC067623                827 bp    DNA     linear   GSS 16-APR-2003
DEFINITION  CSU-K33r.18M15.SP6 CSU-K33r Aedes aegypti genomic clone
            CSU-K33r.18M15, genomic survey sequence.
ACCESSION   CC067623
VERSION     CC067623.1  GI:29906129
KEYWORDS    GSS.
SOURCE      Aedes aegypti (yellow fever mosquito)
  ORGANISM  Aedes aegypti
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Aedes;
            Stegomyia.
REFERENCE   1  (bases 1 to 827)
  AUTHORS   Loftus,B., Shetty,J., Severson,D., Brown,S. and Knudson,D.
  TITLE     End sequencing of Aedes aegypti BACs
  JOURNAL   Unpublished (2003)
COMMENT     Other_GSSs: CSU-K33r.18M15.T7
            Contact: Brendan Loftus
```

```
           Department of Eukaryotic Genomics
           TIGR
           9712 Medical Center Drive, Rockville, MD 20850, USA
           Tel: 301-838-3543
           Fax: 301-838-0208
           Email: enta@tigr.org
           Library was provided by Susan Brown and Dennis Knudson at Colorado
           State University.
           Seq primer: SP6
           Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .827
                     /organism="Aedes aegypti"
                     /mol_type="genomic DNA"
                     /strain="Rexville"
                     /db_xref="taxon:7159"
                     /clone="CSU-K33r.18M15"
                     /clone_lib="CSU-K33r"
                     /note="Vector: pBeloBAC11; Site_1: HindIII"
ORIGIN

  Query Match             7.3%;  Score 335.8;  DB 8;  Length 827;
  Best Local Similarity   76.2%;  Pred. No. 5.6e-47;
  Matches  426;  Conservative    0;  Mismatches  132;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db        766 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 707

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db        706 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 647

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | |||||||| | |||||||| | | ||||||| | |||||
Db        646 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 587

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db        586 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 527

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db        526 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 467

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db        466 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 407

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db        406 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 347

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db        346 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 287

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||     ||   || |  |   ||   |   | | |   | | |  |
Db        286 TATATATATATATATATCGCACGTTTTAGCATTCGTTTGAAGGAAATTCAGGAGAAGGGA 227

Qy        766 GAGGAAAAGAGAGTTTGAC 784
              ||| |    | | ||| ||
Db        226 GAGTATTGTATAATTTTAC 208


RESULT 17
CF626074
LOCUS       CF626074                620 bp    mRNA    linear   EST 02-OCT-2003
DEFINITION  zmrws05_0B10-009-d11.s3 zmrws05 Zea mays cDNA 3', mRNA sequence.
ACCESSION   CF626074
VERSION     CF626074.1  GI:37378736
KEYWORDS    EST.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 620)
  AUTHORS   Bohnert,H., Sharp,R.E., Springer,G.K., Poroyko,V., Fredricksen,M.,
            Sharp,L.G., Spollen,W.G., Ries,J., Guillen,A., Khambati,A.,
            Topinka,C., Davis,G.E., Schachtman,D., Wu,Y. and Nguyen,H.T.
  TITLE     NSF Grant DBI-0211842: Functional Genomics of Root Growth and Root
            Signaling Under Drought
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Hans Bohnert
            University of Illinois, Urbana-Champaign
            1201 West Gregory Drive, Urbana, IL 61801, USA
            Tel: 217-265-5475
            Fax: 217-333-5574
            Email: bohnerth@life.uiuc.edu
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .620
                     /organism="Zea mays"
                     /mol_type="mRNA"
                     /db_xref="taxon:4577"
                     /clone_lib="zmrws05"
                     /note="Samples were collected in Robert E. Sharp's lab
                     (University of Missouri-Columbia) to construct three
                     normalized cDNA libraries. Dark-grown maize seedlings with
                     primary roots 12-20 mm in length were transplanted to high
                     (-0.03 MPa) or low water potential (-1.6 MPa) vermiculite,
                     and harvested at 5 h and 48 h after transplanting. About
                     1,000 roots were used for each of the low water potential
                     libraries (zmrws05 and zmrws48) while 500 roots were
                     combined from each of the two time points at high water
                     potential (zmrww00). Each root was divided into 4 segments
                     (distances are from the junction of the root apex and root
                     cap): segment 1, 0-3 mm plus the root cap; segment 2, 3-7
                     mm; segment 3, 7-12 mm; segment 4, 12-20 mm. (For detatils
                     of conditions see (1) with nutrient modifications as in
                     (2)). The three normalized cDNA libraries were constructed
                     in the lab of Hans Bohnert (University of Illinois-UC).
                     Total RNA was extracted by the 'hot Phenol' method (Plant
                     Molecular Biology manual. D5: 1-13, 2nd ed., 1997). This
                     method worked in eliminating carbohydrate material present
                     in the root tips. The integrity of the RNA was verified by
                     denaturing agarose gels and spectrophotometry (ratio
                     A260/280). Poly(A)+mRNA was isolated twice from total RNA
                     using the Oligotex Direct mRNA kit (Qiagen). Poly(A)+ mRNA
                     was converted to double-stranded cDNA and tagged by using
                     modified Oligo(dT) primers. One of 4 sequence tags
                     corresponding to a different segment of the root was added
                     to the 3'-end of the modified Oligo(dT) primers, including
                     a Not1 site and used to reverse transcribe the
                     segment-specific mRNAs into cDNAs. Each library contains
                     all four tags. A suffix (s1, s2, s3, or s4) has been added
                     to each sequence identifier to designate which region of
                     the root (Root_segment_1, 2, 3, or 4) the sequence was
                     found in based on the identification of the tag. A suffix
                     of s0 indicates that the sequence tag, and hence the
                     source segment, could not be identified. The double
                     stranded cDNAs were size-selected (>450 bp). Size selected
                     cDNAs were adaptored with EcoRI adaptors at both ends, and
                     then digested with NotI. The cDNA was directionally cloned
                     into EcoR1-NotI digested pBS II SK(+) phagemid vector
                     (Stratagene) and electroporated into E.coli DH10B. The
                     total number of white colony forming units (cfu) in the
                     primary libraries before amplification was as follows:
                     zmrws05: 3.37 x 106; zmrws48: 4.87 x 106; zmrww00: 3 x
                     106. The background of empty clones was less than 1%.
                     Inserts ranged from ~0.5kb to >2.5 kb, as determined by
                     PCR. Plasmid DNA from the primary libraries then was
                     converted to single-stranded circles and used as a
                     template for PCR amplification using the T7 and T3 priming
```

sites that flank the cloned cDNA inserts. The purified PCR products, representing the entire cDNA population cloned in each library, were used as a driver for normalization. Hybridization between the single-stranded library and the PCR products was carried out for 44 hours at 30oC. Non-hybridized single-stranded DNA circles were separated from hybridized DNA rendered partially double-stranded and electroporated into DH10B. The total number of clones with insert was: zmrws05: 2.0x107; zmrws48: 4.2x107; zmrww00: 1.1x107. The background of empty clones was less than 2%. Insert size, determined by PCR of the entire library, ranged from 0.5kb to 2.5kb. (1) Sharp R E; Silk W K; Hsiao T C. Growth of the Maize Primary Root at Low Water Potentials I. Spatial Distribution of Expansive Growth. Plant Physiology (Rockville). 87(1). 1988. 50-57. (2) Spollen W G; LeNoble M E; Samuels T D; Bernstein N; Sharp R E. Abscisic acid accumulation maintains maize primary root elongation at low water potentials by restricting ethylene production. Plant Physiology (Rockville). 122(3). March, 2000. 967-976.
TAG_TISSUE=Root_segment_3
TAG_SEQ=TCGCA"

ORIGIN

Query Match 7.3%; Score 335.2; DB 7; Length 620;
Best Local Similarity 78.4%; Pred. No. 7.4e-47;
Matches 414; Conservative 0; Mismatches 113; Indels 1; Gaps 1;

```
Qy      194 GCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATATAC 253
            || || |  ||| |    | | | |      | ||||||||||||||||||||||||
Db       54 GCTAATTTACATATATATATATATATATATATATATATATATATATATATATATATATAT 113

Qy      254 ACACACACACACATATATATATACACATATATATACACATATATACATATATACACATAT 313
            | | | | | ||||||||||| | ||||||||| | ||||||| ||||||| | ||||
Db      114 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 173

Qy      314 ATACACATATATATACACATACATATGTACACATATATATACACATATGTATACACATAT 373
            ||| | |||||||| | ||| |||| || | ||||||||| | |||| |||| | ||||
Db      174 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 233

Qy      374 ATACACATATATACACACATATATACACATATATACACACATATATACACATATATACAC 433
            ||| | ||||||| | | ||||||| | ||||||| | | ||||||| | ||||||| |
Db      234 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 293

Qy      434 ATATATACACATATACACATATATACACATATATACATATATACACATATATATA-ATAT 492
            ||||||| | ||||| | ||||||| | ||||||| ||||||| | ||||||||| ||||
Db      294 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 353

Qy      493 ACACACATATATATACACATATATACACACATATATACACATATATACACATATATATAC 552
            | | | ||||||||| | ||||||| | | ||||||| | ||||||| | |||||||||
Db      354 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 413

Qy      553 ACATATATACACATATATACATATATACACATATATATACATATATACACATATATACAT 612
            | ||||||| | ||||||| ||||||| | ||||||||| ||||||| | ||||||| ||
Db      414 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 473

Qy      613 ATATACACATATATACATATATACACACATATATACACATACATATACACACACATAGAT 672
            ||||| | ||||||| ||||||| | | ||||||| | ||| ||||| | | | ||| ||
Db      474 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 533

Qy      673 ATACATATATATACACATATATATACGTATATATATGTATATATATAT 720
            ||| ||||||||| | |||||||||  ||||||||| ||||||||||
Db      534 ATATATATATATATATATATATATATATATATATATATATATATATAT 581
```

RESULT 18
CF634146
LOCUS CF634146 632 bp mRNA linear EST 02-OCT-2003
DEFINITION zmrww00_0A10-008-g12.s3 zmrww00 Zea mays cDNA 3', mRNA sequence.
ACCESSION CF634146
VERSION CF634146.1 GI:37393761
KEYWORDS EST.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 632)
AUTHORS Bohnert,H., Sharp,R.E., Springer,G.K., Poroyko,V., Fredricksen,M., Sharp,L.G., Spollen,W.G., Ries,J., Guillen,A., Khambati,A., Topinka,C., Davis,G.E., Schachtman,D., Wu,Y. and Nguyen,H.T.
TITLE NSF Grant DBI-0211842: Functional Genomics of Root Growth and Root Signaling Under Drought
JOURNAL Unpublished (2003)
COMMENT Contact: Hans Bohnert
University of Illinois, Urbana-Champaign
1201 West Gregory Drive, Urbana, IL 61801, USA
Tel: 217-265-5475
Fax: 217-333-5574
Email: bohnerth@life.uiuc.edu
POLYA=Yes.
FEATURES Location/Qualifiers
source 1. .632
/organism="Zea mays"
/mol_type="mRNA"
/db_xref="taxon:4577"
/clone_lib="zmrww00"
/note="Samples were collected in Robert E. Sharp's lab (University of Missouri-Columbia) to construct three normalized cDNA libraries. Dark-grown maize seedlings with primary roots 12-20 mm in length were transplanted to high (-0.03 MPa) or low water potential (-1.6 MPa) vermiculite, and harvested at 5 h and 48 h after transplanting. About 1,000 roots were used for each of the low water potential libraries (zmrws05 and zmrws48) while 500 roots were combined from each of the two time points at high water potential (zmrww00). Each root was divided into 4 segments (distances are from the junction of the root apex and root cap): segment 1, 0-3 mm plus the root cap; segment 2, 3-7 mm; segment 3, 7-12 mm; segment 4, 12-20 mm. (For detatils of conditions see (1) with nutrient modifications as in (2)). The three normalized cDNA libraries were constructed in the lab of Hans Bohnert (University of Illinois-UC). Total RNA was extracted by the 'hot Phenol' method (Plant Molecular Biology manual. D5: 1-13, 2nd ed., 1997). This method worked in eliminating carbohydrate material present in the root tips. The integrity of the RNA was verified by denaturing agarose gels and spectrophotometry (ratio A260/280). Poly(A)+mRNA was isolated twice from total RNA using the Oligotex Direct mRNA kit (Qiagen). Poly(A)+ mRNA was converted to double-stranded cDNA and tagged by using modified Oligo(dT) primers. One of 4 sequence tags corresponding to a different segment of the root was added to the 3'-end of the modified Oligo(dT) primers, including a NotI site and used to reverse transcribe the segment-specific mRNAs into cDNAs. Each library contains all four tags. A suffix (s1, s2, s3, or s4) has been added to each sequence identifier to designate which region of the root (Root_segment_1, 2, 3, or 4) the sequence was found in based on the identification of the tag. A suffix of s0 indicates that the sequence tag, and hence the source segment, could not be identified. The double stranded cDNAs were size-selected (>450 bp). Size selected cDNAs were adaptored with EcoRI adaptors at both ends, and then digested with NotI. The cDNA was directionally cloned into EcoR1-NotI digested pBS II SK(+) phagemid vector (Stratagene) and electroporated into E.coli DH10B. The total number of white colony forming units (cfu) in the primary libraries before amplification was as follows: zmrws05: 3.37 x 106; zmrws48: 4.87 x 106; zmrww00: 3 x 106. The background of empty clones was less than 1%. Inserts ranged from ~0.5kb to >2.5 kb, as determined by PCR. Plasmid DNA from the primary libraries then was converted to single-stranded circles and used as a template for PCR amplification using the T7 and T3 priming sites that flank the cloned cDNA inserts. The purified PCR

products, representing the entire cDNA population cloned in each library, were used as a driver for normalization. Hybridization between the single-stranded library and the PCR products was carried out for 44 hours at 30oC. Non-hybridized single-stranded DNA circles were separated from hybridized DNA rendered partially double-stranded and electroporated into DH10B. The total number of clones with insert was: zmrws05: 2.0x107; zmrws48: 4.2x107; zmrww00: 1.1x107. The background of empty clones was less than 2%. Insert size, determined by PCR of the entire library, ranged from 0.5kb to 2.5kb. (1) Sharp R E; Silk W K; Hsiao T C. Growth of the Maize Primary Root at Low Water Potentials I. Spatial Distribution of Expansive Growth. Plant Physiology (Rockville). 87(1). 1988. 50-57. (2) Spollen W G; LeNoble M E; Samuels T D; Bernstein N; Sharp R E. Abscisic acid accumulation maintains maize primary root elongation at low water potentials by restricting ethylene production. Plant Physiology (Rockville). 122(3). March, 2000. 967-976.
TAG_TISSUE=Root_segment_3
TAG_SEQ=TCGCA"

ORIGIN

Query Match 7.3%; Score 335.2; DB 7; Length 632;
Best Local Similarity 78.4%; Pred. No. 7.3e-47;
Matches 414; Conservative 0; Mismatches 113; Indels 1; Gaps 1;

```
Qy      194 GCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATATAC 253
            || || |  ||| |    | | | |      | ||||||||||||||||||||||||||
Db       33 GCTAATTTACATATATATATATATATATATATATATATATATATATATATATATATATAT 92

Qy      254 ACACACACACACATATATATATACACATATATATACACATATATACATATATACACATAT 313
            | | | | | | |||||||||||| | ||||||||| | ||||||| ||||||| | ||||
Db       93 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 152

Qy      314 ATACACATATATATACACATACATATGTACACATATATATACACATATGTATACACATAT 373
            ||| | ||||||||| | ||| |||| || | ||||||||| | |||| |||| | ||||
Db      153 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 212

Qy      374 ATACACATATATACACACATATATACACATATATACACACATATATACACATATATACAC 433
            ||| | ||||||| | | ||||||| | ||||||| | | ||||||| | ||||||| |
Db      213 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 272

Qy      434 ATATATACACATATACACATATATACACATATATACATATATACACATATATATA-ATAT 492
            ||||||| | ||||| | ||||||| | ||||||| ||||||| | ||||||||| ||||
Db      273 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 332

Qy      493 ACACACATATATATACACATATATACACACATATATACACATATATACACATATATATAC 552
            | | | ||||||||| | ||||||| | | ||||||| | ||||||| | |||||||||
Db      333 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 392

Qy      553 ACATATATACACATATATACATATATACACATATATATACATATATACACATATATACAT 612
            | ||||||| | ||||||| ||||||| | ||||||||| ||||||| | ||||||| ||
Db      393 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 452

Qy      613 ATATACACATATATACATATATACACACATATATACACATACATATACACACACATAGAT 672
            ||||| | ||||||| ||||||| | | ||||||| | ||| ||||| | | | ||| ||
Db      453 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 512

Qy      673 ATACATATATATACACATATATATACGTATATATATGTATATATATAT 720
            ||| ||||||||| | |||||||||  ||||||||| ||||||||||
Db      513 ATATATATATATATATATATATATATATATATATATATATATATATAT 560
```

RESULT 19
BI505681
LOCUS BI505681 717 bp mRNA linear EST 08-APR-2002
DEFINITION BB170032B10E06.5 Bee Brain Normalized/Subtracted Library, BB17 Apis mellifera cDNA clone BB170032B10E06 5', mRNA sequence.
ACCESSION BI505681
VERSION BI505681.1 GI:15356055
KEYWORDS EST.
SOURCE Apis mellifera (honey bee)
ORGANISM Apis mellifera
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Apoidea; Apidae; Apis.
REFERENCE 1 (bases 1 to 717)
AUTHORS Whitfield,C.W., Band,M.R., Bonaldo,M.F., Kumar,C.G., Liu,L., Pardinas,J., Robertson,H.M., Soares,B. and Robinson,G.E.
TITLE Annotated expressed sequence tags and cDNA microarrays for studies of brain and behavior in the honey bee
JOURNAL Genome Res. 12 (4), 555-566 (2002)
MEDLINE 21929762
PUBMED 11932240
COMMENT Contact: Gene E. Robinson
Department of Entomology
University of Illinois
505 S. Goodwin Ave., Urbana, IL 61801, USA
Tel: 217 265 0309
Fax: 217 244 3499
Email: generobi@life.uiuc.edu
This research was funded by the University of Illinois Critical Research Initiatives Fund and a Burroughs-Wellcome Trust Innovation Award in Functional Genomics to G.E. Robinson and an NSF Postdoctoral Fellowship in Bioinformatics to C.W. Whitfield.
REPEATS IN THE SEQUENCE
Low_complexity STRAND (+) ELEMENT AT_rich LOCATION [101,132];
Simple_repeat STRAND (+) ELEMENT (TA)n LOCATION [133,717].
PCR PRIMERS
FORWARD: TAATACGACTCACTATAGGG
BACKWARD: ATTAACCCTCACTAAAG
Plate: BB170032B10 row: E column: 06
Seq primer: AGCGGATAACAATTTCACACAGGA
High quality sequence stop: 717.
FEATURES Location/Qualifiers
source 1. .717
/organism="Apis mellifera"
/mol_type="mRNA"
/strain="mixed strains of European bees, predominantly A.m. ligustica"
/db_xref="taxon:7460"
/clone="BB170032B10E06"
/sex="female"
/tissue_type="brain"
/dev_stage="adult worker honey bee"
/lab_host="DH10B"
/clone_lib="Bee Brain Normalized/Subtracted Library, BB17"
/note="Organ: brain; Vector: pT7T3-Pac; Site_1: EcoR1; Site_2: Not1; This BB17 cDNA library was generated by subtraction of the BB16 library with 4000 previously sequenced clones. The BB16 library was contributed by the Soares laboratory and it was constructed and normalized as described by Bonaldo, M.F., Lennon, G. and Soares, M.B. (1996), Genome Research 6(9): 791-806. RNA was prepared from dissected brains of adult worker bees of various ages and various behavioral groups. "

ORIGIN

Query Match 7.3%; Score 335.2; DB 4; Length 717;
Best Local Similarity 77.2%; Pred. No. 7.2e-47;
Matches 420; Conservative 0; Mismatches 123; Indels 1; Gaps 1;

```
Qy      178 AGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATAT 237
            | | ||   || | ||  | || |       | | || | | |       | |||||||||
Db       96 AACTCAATTAAAATGATTTATATATATATATATATAATATATATATATATATATATAT 155

Qy      238 ATATATATATATATACACACACACACATATATATATACACATATATATACACATATAT 297
            |||||||||||||| | | | | | |||||||||||| | ||||||||| | ||||||
Db      156 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 215

Qy      298 ACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACAC 357
            | ||||||| | ||||||| | ||||||||| | ||| |||| || | ||||||||| |
Db      216 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 275

Qy      358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417
```

```
Db        276 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 335

Qy        418 ATACACATATATACACATATATACACATATACACATATATACACATATATACATATATAC 477

Db        336 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 395

Qy        478 ACATATATATA-ATATACACACATATATATACACATATATACACACATATATACACATAT 536

Db        396 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 455

Qy        537 ATACACATATATATACACATATATACACATATATACATATATACACATATATATACATAT 596

Db        456 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 515

Qy        597 ATACACATATATACATATATACACATATATACATATATACACACATATATACACATACAT 656

Db        516 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 575

Qy        657 ATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATAT 716

Db        576 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 635

Qy        717 ATAT 720

Db        636 ATAT 639
```

```
RESULT 20
CR294134/c
LOCUS       CR294134                590 bp    DNA     linear   GSS 28-FEB-2004
DEFINITION  Medicago truncatula BAC ends cultivar Jemalong A17 of Medicago
            truncatula, genomic survey sequence.
ACCESSION   CR294134
VERSION     CR294134.1  GI:44684558
KEYWORDS    GSS.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 590)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-FEB-2004) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
FEATURES             Location/Qualifiers
     source          1. .590
                     /organism="Medicago truncatula"
                     /mol_type="genomic DNA"
                     /cultivar="Jemalong A17"
                     /db_xref="taxon:3880"
                     /clone_lib="MTE1"
                     /note="Vector: pIndigoBAC ; Site_1: EcoRI ; Site_2: EcoRI
                     ; Debelle F. and Chalhoub B.~Genoscope sequence ID :
                     mte1-1J18FM1"
ORIGIN

  Query Match             7.2%;  Score 334.8;  DB 9;  Length 590;
  Best Local Similarity   79.8%;  Pred. No. 8.7e-47;
  Matches  407;  Conservative    0;  Mismatches  102;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACATATATATATACACATATATA 286

Db        532 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 473

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346

Db        472 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 413

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406

Db        412 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 353

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466

Db        352 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 293

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525

Db        292 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 233

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585

Db        232 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 173

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645

Db        172 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 113

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705

Db        112 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 53

Qy        706 TATGTATATATATATGCTCCAGAGTTCATA 735

Db         52 TATATATATATATATATACTATACTATATA 23
```

```
RESULT 21
BI505681/c
LOCUS       BI505681                717 bp    mRNA    linear   EST 08-APR-2002
DEFINITION  BB170032B10E06.5 Bee Brain Normalized/Subtracted Library, BB17 Apis
            mellifera cDNA clone BB170032B10E06 5', mRNA sequence.
ACCESSION   BI505681
VERSION     BI505681.1  GI:15356055
KEYWORDS    EST.
SOURCE      Apis mellifera (honey bee)
  ORGANISM  Apis mellifera
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Apoidea;
            Apidae; Apis.
REFERENCE   1  (bases 1 to 717)
  AUTHORS   Whitfield,C.W., Band,M.R., Bonaldo,M.F., Kumar,C.G., Liu,L.,
            Pardinas,J., Robertson,H.M., Soares,B. and Robinson,G.E.
  TITLE     Annotated expressed sequence tags and cDNA microarrays for studies
            of brain and behavior in the honey bee
  JOURNAL   Genome Res. 12 (4), 555-566 (2002)
  MEDLINE   21929762
   PUBMED   11932240
COMMENT     Contact: Gene E. Robinson
            Department of Entomology
            University of Illinois
            505 S. Goodwin Ave., Urbana, IL 61801, USA
            Tel: 217 265 0309
            Fax: 217 244 3499
            Email: generobi@life.uiuc.edu
            This research was funded by the University of Illinois Critical
            Research Initiatives Fund and a Burroughs-Wellcome Trust Innovation
            Award in Functional Genomics to G.E. Robinson and an NSF
            Postdoctoral Fellowship in Bioinformatics to C.W. Whitfield.
            REPEATS IN THE SEQUENCE
            Low_complexity STRAND (+) ELEMENT AT_rich LOCATION [101,132];
            Simple_repeat STRAND (+) ELEMENT (TA)n LOCATION [133,717].
            PCR PRImers
            FORWARD: TAATACGACTCACTATAGGG
            BACKWARD: ATTAACCCTCACTAAAG
            Plate: BB170032B10  row: E  column: 06
            Seq primer: AGCGGATAACAATTTCACACAGGA
            High quality sequence stop: 717.
FEATURES             Location/Qualifiers
     source          1. .717
                     /organism="Apis mellifera"
                     /mol_type="mRNA"
```

/strain="mixed strains of European bees, predominantly A.m. ligustica"
/db_xref="taxon:7460"
/clone="BB170032B10E06"
/sex="female"
/tissue_type="brain"
/dev_stage="adult worker honey bee"
/lab_host="DH10B"
/clone_lib="Bee Brain Normalized/Subtracted Library, BB17"
/note="Organ: brain; Vector: pT7T3-Pac; Site_1: EcoR1; Site_2: Not1; This BB17 cDNA library was generated by subtraction of the BB16 library with 4000 previously sequenced clones. The BB16 library was contributed by the Soares laboratory and it was constructed and normalized as described by Bonaldo, M.F., Lennon, G. and Soares, M.B. (1996), Genome Research 6(9): 791-806. RNA was prepared from dissected brains of adult worker bees of various ages and various behavioral groups. "

ORIGIN

Query Match 7.2%; Score 334.2; DB 4; Length 717;
Best Local Similarity 78.4%; Pred. No. 1.1e-46;
Matches 413; Conservative 0; Mismatches 113; Indels 1; Gaps 1;

```
Qy      227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
            | ||||||||||||||||||||||| | | | | | |||||||||| | |||||||
Db      628 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 569

Qy      287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
            || | |||||| |||||| | ||||||| | |||||||| | ||| |||| || | |
Db      568 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 509

Qy      347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
            |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db      508 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 449

Qy      407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
            || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db      448 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 389

Qy      467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
            || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db      388 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 329

Qy      526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
            |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db      328 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 269

Qy      586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
            |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db      268 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 209

Qy      646 TACACATACATATATACACACATAGATATACATATATATACACATATATATACGTATATA 705
            || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db      208 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 149

Qy      706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATT 752
            ||| |||||||||||  |  | |  |  |  | || || | |
Db      148 TATATATATATATATATTATATATATATATATATAAATCATTTTAAT 102
```

RESULT 22
AQ917102/c
LOCUS AQ917102 725 bp DNA linear GSS 07-DEC-1999
DEFINITION T233131b Medicago truncatula BAC library Medicago truncatula genomic clone 24E12, genomic survey sequence.
ACCESSION AQ917102
VERSION AQ917102.1 GI:6536747
KEYWORDS GSS.
SOURCE Medicago truncatula (barrel medic)
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.
REFERENCE 1 (bases 1 to 725)
AUTHORS Kim,D., Limpens,E., Peng,H., Ellis,L. and Cook,D.R.
TITLE BAC end sequencing of Medicago truncatula
JOURNAL Unpublished (1999)
COMMENT Contact: Cook DR
The Crop Biotechnology Center
Texas A&M University
Department of Plant Pathology and Microbiology, Rm 120 L.F. Peterson Bldg, College Station, TX 77843-2132, USA
Tel: 409 845 8743
Fax: 409 862 4790
Email: dcook@ppserver.tamu.edu
Other name: DK277-L; date: 8/12/99; Submitted to the Database of Genome Survey Sequences (GSS) on 12/06/99; More information is available at 'http://chrysie.tamu.edu/medicago'.
Seq primer: SQ-BAC-L (AACGCCAGGGTTTTCCCAGTCACGACG)
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .725
/organism="Medicago truncatula"
/mol_type="genomic DNA"
/cultivar="genotype A17"
/db_xref="taxon:3880"
/clone="24E12"
/clone_lib="Medicago truncatula BAC library"
/note="Vector: pBeloBAC11; Site_1: HindIII; Site_2: HindIII; Nam. Y-W, Penmetsa, R.V., Endre, G., Kim, D., and Cook, D.R. 1999. Construction of a bacterial artificial chromosome library of Medicago truncatula and identification of clones containing ethylene response genes. Theor Appl Genet 98: 638-646. "

ORIGIN

Query Match 7.2%; Score 333.8; DB 8; Length 725;
Best Local Similarity 80.4%; Pred. No. 1.2e-46;
Matches 403; Conservative 0; Mismatches 97; Indels 1; Gaps 1;

```
Qy      227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
            | ||||||||||||||||||||||| | | | | | |||||||||| | |||||||
Db      701 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 642

Qy      287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
            || | |||||| |||||| | ||||||| | |||||||| | ||| |||| || | |
Db      641 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 582

Qy      347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
            |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db      581 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 522

Qy      407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
            || | | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db      521 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 462

Qy      467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
            || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db      461 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 402

Qy      526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
            |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db      401 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 342

Qy      586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
            |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db      341 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 282

Qy      646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
            || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db      281 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 222

Qy      706 TATGTATATATATATGCTCCA 726
            ||| |||||||||||  |  ||
```

```
Db        221 TATATATATATATATCCATCA 201


RESULT 23
CL623140/c
LOCUS       CL623140                744 bp    DNA     linear   GSS 01-JUL-2004
DEFINITION  OR_BBa0018D11.r OR_BBa Oryza rufipogon genomic clone OR_BBa0018D11
            3', genomic survey sequence.
ACCESSION   CL623140
VERSION     CL623140.1  GI:49121028
KEYWORDS    GSS.
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 744)
  AUTHORS   Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R.,
            Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
  TITLE     OMAP Project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Insert Length: 161    Std Error: 0.00
            Plate: 0018  row: D  column: 11
            Seq primer: CAC TCA TTA GGC ACC CCA
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .744
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
                     /clone="OR_BBa0018D11"
                     /tissue_type="young leaves"
                     /lab_host="DH10B-T1 phage resistant"
                     /clone_lib="OR_BBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

  Query Match             7.2%;  Score 333;  DB 9;  Length 744;
  Best Local Similarity   79.3%;  Pred. No. 1.7e-46;
  Matches  407;  Conservative    0;  Mismatches  105;  Indels    1;  Gaps    1;

Qy        209 AGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACACATA 268
              | | | | | | |   || |||||||| ||||||||||||||||| | | | | | |||
Db        702 ATATATATATATATATATATTATATATCTATATATATATATATATATATATATATATA 643

Qy        269 TATATATACACATATATATACACATATATACATATATACACATATATACACATATATATA 328
              |||||||| | ||||||||| | ||||||| ||||||| | ||||||| | |||||||||
Db        642 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 583

Qy        329 CACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACA 388
              | ||| |||| || | ||||||||| | |||| |||| | ||||||| | ||||||| |
Db        582 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 523

Qy        389 CACATATATACACATATATACACACATATATACACATATATACACATATATACACATATA 448
              | ||||||| | ||||||| | | ||||||| | ||||||| ||||||| | |||||
Db        522 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 463

Qy        449 CACATATATACACATATATACATATATACACATATATATAATATACACACATATATATAC 508
              | |||||||| | ||||||| ||||||| | |||||||| ||||| | | |||||||||
Db        462 TATATATATATATATATATATATATATATATATATATAT-ATATATATATATATATATAT 404

Qy        509 ACATATATACACACATATATACACATATATACACATATATATACACATATATACACATAT 568
              | ||||||| | | ||||||| | ||||||| | ||||||||| | ||||||| | ||||
Db        403 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 344

Qy        569 ATACATATATACACATATATATACATATATACACATATATACATATATACACATATATAC 628
              ||| ||||||| | ||||||||| ||||||| | ||||||| ||||||| | |||||||
Db        343 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 284

Qy        629 ATATATACACACATATATACACATACATATACACACACATAGATATACATATATATACAC 688
              ||||||| | | ||||||| | ||| ||||| | | | ||| ||||| ||||||||| |
Db        283 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 224

Qy        689 ATATATATACGTATATATATGTATATATATATG 721
              |||||||||  |||||||||||||| ||| ||||
Db        223 ATATATATATATATATATATGTATGTATGTATG 191


RESULT 24
BZ796807/c
LOCUS       BZ796807                512 bp    DNA     linear   GSS 17-MAR-2003
DEFINITION  PUGDF25TD ZM_0.6_1.0_KB Zea mays genomic clone ZMMBTa341F02,
            genomic survey sequence.
ACCESSION   BZ796807
VERSION     BZ796807.1  GI:28998547
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 512)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
            Bennetzen,J.
  TITLE     Maize Genomics Consortium
  JOURNAL   Unpublished (2003)
COMMENT     Other_GSSs: PUGDF25TB
            Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .512
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBTa341F02"
                     /clone_lib="ZM_0.6_1.0_KB"
                     /note="Vector: pCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high
                     CoT selected genomic DNA library"
ORIGIN

  Query Match             7.2%;  Score 332.6;  DB 8;  Length 512;
  Best Local Similarity   80.8%;  Pred. No. 2.1e-46;
  Matches  400;  Conservative    0;  Mismatches   94;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATA 286
              | |||||||||||||||||||||||||| | | | | | |||||||||| | |||||||
Db        508 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 449

Qy        287 TACACATATATACACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| ||||||||| | ||| |||| || | |
Db        448 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 389

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        388 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 329

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
```

```
Db        328 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 269

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db        268 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 209

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db        208 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 149

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db        148 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 89

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | ||||||||  ||||||
Db         88 TATATATATATATATATATATATATATATATATATATATATATATATATATATAAATATATA 29

Qy        706 TATGTATATATATAT 720
              ||| ||||||||||||
Db         28 TATATATATATATAT 14


RESULT 25
CL755178/c
LOCUS       CL755178                 707 bp    DNA     linear   GSS 27-JUL-2004
DEFINITION  OR_BBa0122N12.r OR_BBa Oryza rufipogon genomic clone OR_BBa0122N12
            3', genomic survey sequence.
ACCESSION   CL755178
VERSION     CL755178.1  GI:50706035
KEYWORDS    GSS.
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 707)
  AUTHORS   Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R.,
            Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
  TITLE     OMAP Project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Insert Length: 161   Std Error: 0.00
            Plate: 0122  row: N  column: 12
            Seq primer: CAC TCA TTA GGC ACC CCA
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .707
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
                     /clone="OR_BBa0122N12"
                     /tissue_type="young leaves"
                     /lab_host="DH10B-T1 phage resistant"
                     /clone_lib="OR_BBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

  Query Match             7.2%;  Score 331.6;  DB 9;  Length 707;
  Best Local Similarity   79.4%;  Pred. No. 3e-46;
  Matches  405;  Conservative    0;  Mismatches  104;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||| | |||||||||||||| | | | | | |||||||||||| | |||||||

Db        563 TATATATATCTTTATATATATATATATATATATATATATATATATATATATATATATA 504

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||  |||| || | |
Db        503 TATATATATATATATATATATATATATATATATATATATATATATATATATCTATATTTATATA 444

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        443 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 384

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db        383 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 324

Qy        467 TACATATATACACATATATATAATATACACACATATATATACACATATATACACACATAT 526
              || ||||||| | |||||||| ||||| | | ||||||||| | ||||||| | | ||||
Db        323 TATATATATATATATATATAT-ATATATATATATATATATATATATATATATATATATAT 265

Qy        527 ATACACATATATACACATATATATACACATATATACACATATATACACATATATACACATAT 586
              ||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | ||||
Db        264 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 205

Qy        587 ATATACATATATACACATATATACATATATACACATATATACATATATACACACATATAT 646
              ||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | ||||||
Db        204 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 145

Qy        647 ACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATAT 706
              | | ||| ||||| | | | ||| ||||| |||||||||| | |||||||||  |||||||
Db        144 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 85

Qy        707 ATGTATATATATATGCTCCAGAGTTCATAA 736
              || |||||||||||| | | ||  | || |
Db         84 ATATATATATATATGATGCTGATATGATTA 55


RESULT 26
CL934691
LOCUS       CL934691                 784 bp    DNA     linear   GSS 14-SEP-2004
DEFINITION  OA_ABa0044J13.f OA_ABa Oryza australiensis genomic clone
            OA_ABa0044J13 5', genomic survey sequence.
ACCESSION   CL934691
VERSION     CL934691.1  GI:52063483
KEYWORDS    GSS.
SOURCE      Oryza australiensis
  ORGANISM  Oryza australiensis
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 784)
  AUTHORS   Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R.,
            Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
  TITLE     OMAP Project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0044  row: J  column: 13
            Seq primer: TAA TAC GAC TCA CTA TAG GG
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .784
                     /organism="Oryza australiensis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4532"
                     /clone="OA_ABa0044J13"
                     /tissue_type="young leaves"
```

```
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OA_ABa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

  Query Match             7.2%;  Score 331.2;  DB 9;  Length 784;
  Best Local Similarity   78.9%;  Pred. No. 3.4e-46;
  Matches  407;  Conservative    0;  Mismatches  108;  Indels    1;  Gaps    1;

Qy        205 TCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATATACACACACACACA 264
              ||||    | | | | |       | |||||||||||||||||||||||| | | | | | |
Db        201 TCTCTCTATATATATATATATATATATATATATATATATATATATATATATATATATATATA 260

Qy        265 CATATATATATACACATATATATACACATATATACATATATACACATATATACACATATA 324
               |||||||||||| | |||||||||| | ||||||| ||||||| | ||||||| | |||||
Db        261 TATATATATATATATATATATATATATATATATATATATATATATAGATATATATATATATA 320

Qy        325 TATACACATACATATGTACACATATATATACACATATGTATACACATATATACACATATA 384
              |||| | ||| |||| || | |||||||||| | |||| |||| | ||||||| | |||||
Db        321 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 380

Qy        385 TACACACATATATACACATATATACACACATATATACACATATATACACATATATACACA 444
              || | | ||||||| | ||||||| | | ||||||| | ||||||| | ||||||| | |
Db        381 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 440

Qy        445 TATACACATATATACACATATATACATATATACACATATATAATATACACACATATAT 504
              |||| | ||||||| | ||||||| ||||||| | |||||||| ||||| | | |||||||
Db        441 TATATATATATATATATATATATATATATATATATATATATAT-ATATATATATATATAT 499

Qy        505 ATACACATATATACACACATATATACACATATATACACATATATATACACATATATACAC 564
              ||| | ||||| | | | |||||||| | ||||||| | |||||||| | ||||||| | |
Db        500 ATAGATATATAGATATATATATATATATATATATATATATATATATATATATATATATAT 559

Qy        565 ATATATACATATATACACATATATATACATATATACACATATATACACATATATACACATAT 624
              ||||||| ||||||| | ||||||||| ||||||| | ||||||| ||||||| | ||||
Db        560 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 619

Qy        625 ATACATATATACACACATATATACACATACATATACACACACATAGATATACATATATAT 684
              ||| ||||||| | | ||||||| | ||| ||||| | | | ||||||||| ||||||||
Db        620 ATATATATATATATATATATATATATATATATATATATATATATAGATATATATATATAT 679

Qy        685 ACACATATATATACGTATATATATGTATATATATAT 720
              | | |||||||||  ||||||||  ||||||||||
Db        680 ATATATATATATATATATATATAAATATATATATAT 715


RESULT 27
BZ796807
LOCUS       BZ796807                512 bp    DNA     linear   GSS 17-MAR-2003
DEFINITION  PUGDF25TD ZM_0.6_1.0_KB Zea mays genomic clone ZMMBTa341F02,
            genomic survey sequence.
ACCESSION   BZ796807
VERSION     BZ796807.1  GI:28998547
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 512)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
            Bennetzen,J.
  TITLE     Maize Genomics Consortium
  JOURNAL   Unpublished (2003)
COMMENT     Other_GSSs: PUGDF25TB
            Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .512
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBTa341F02"
                     /clone_lib="ZM_0.6_1.0_KB"
                     /note="Vector: pCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high
                     CoT selected genomic DNA library"
ORIGIN

  Query Match             7.2%;  Score 331;  DB 8;  Length 512;
  Best Local Similarity   80.6%;  Pred. No. 3.9e-46;
  Matches  399;  Conservative    0;  Mismatches   95;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACATATATATATACACATATATA 286
              | |||||||||||||||||||||||||| | | | |    | |||||||||||| | |||||||
Db          1 TATATATATATATATATATATATATATATATATATTTATATATATATATATATATATA 60

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db         61 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 120

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        121 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 180

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db        181 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 240

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db        241 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 300

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db        301 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 360

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db        361 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 420

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db        421 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 480

Qy        706 TATGTATATATATAT 720
              ||| ||||||||||||
Db        481 TATATATATATATAT 495


RESULT 28
BX145650/c
LOCUS       BX145650                668 bp    DNA     linear   GSS 28-JAN-2003
DEFINITION  Danio rerio genomic clone DKEY-109N18, genomic survey sequence.
ACCESSION   BX145650
VERSION     BX145650.1  GI:27977003
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 668)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the T7 end of BAC 109N18. 109N18
```

```
          is part of the Daniokey BAC Library created by R. Plasterk and N.V.
          Keygene. Further details:
          http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .668
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-109N18"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             7.1%;  Score 330.4;  DB 9;  Length 668;
  Best Local Similarity   78.5%;  Pred. No. 4.8e-46;
  Matches  408;  Conservative    0;  Mismatches  111;  Indels    1;  Gaps    1;

Qy        229 TATATATATATATATATATATATACACACACACACATATATATATACACATATATATA 288
              ||||||||||||||||||||||| | | | | | |||||||||||| | |||||||||
Db        668 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 609

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              | ||||||| ||||||| | |||||||| | ||||||||| | ||| |||| || | |||
Db        608 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 549

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              |||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||||
Db        548 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 489

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||||
Db        488 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 429

Qy        469 CATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATATA 527
              |||||||| | |||||||||| ||||| | | |||||||||| | ||||||| | | |||||
Db        428 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 369

Qy        528 TACACATATATACACATATATATACACATATATACACATATATACACATATATACACATATA 587
              || | ||||||| | |||||||||| | ||||||| | ||||||| ||||||| | |||||
Db        368 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 309

Qy        588 TATACATATATACACATATATACATATATACACATATATACATATATACACACATATATA 647
              |||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||||
Db        308 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 249

Qy        648 CACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATA 707
              | ||| ||||| | | | ||| ||||| |||||||||| | |||||||||  ||||||||
Db        248 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 189

Qy        708 TGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGT 747
              | |||| | | |  ||  ||  || || |   | |||
Db        188 TATATAAACACTTCATCTCTAGACCAAAACACCCATGAGT 149


RESULT 29
CL755178
LOCUS       CL755178                707 bp    DNA     linear   GSS 27-JUL-2004
DEFINITION  OR_BBa0122N12.r OR_BBa Oryza rufipogon genomic clone OR_BBa0122N12
            3', genomic survey sequence.
ACCESSION   CL755178
VERSION     CL755178.1  GI:50706035
KEYWORDS    GSS.
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 707)
  AUTHORS   Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R.,
            Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
  TITLE     OMAP Project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Insert Length: 161   Std Error: 0.00
            Plate: 0122  row: N  column: 12
            Seq primer: CAC TCA TTA GGC ACC CCA
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .707
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
                     /clone="OR_BBa0122N12"
                     /tissue_type="young leaves"
                     /lab_host="DH10B-T1 phage resistant"
                     /clone_lib="OR_BBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

  Query Match             7.1%;  Score 330.4;  DB 9;  Length 707;
  Best Local Similarity   70.8%;  Pred. No. 4.7e-46;
  Matches  453;  Conservative    0;  Mismatches  186;  Indels    1;  Gaps    1;

Qy        228 TTATATATATATATATATATATATACACACACACACATATATATATACACATATATAT 287
              | ||||||||||||||||||||||| | | | | | |||||||||||| | ||||||||
Db         69 TCATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 128

Qy        288 ACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACAT 347
              | | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | ||
Db        129 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 188

Qy        348 ATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATAT 407
              ||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | ||||||
Db        189 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 248

Qy        408 ACACACATATATACACATATATACACATATATACACATATACACATATATACACATATAT 467
              | | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | ||||||
Db        249 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 308

Qy        468 ACATATATACACATATATATAATATACACACATATATATACACATATATACACACATATA 527
              | ||||||| | |||||||| ||||| | | ||||||||| | ||||||| | | |||||
Db        309 ATATATATATATATATATAT-ATATATATATATATATATATATATATATATATATATA 367

Qy        528 TACACATATATACACATATATATACACATATATACACATATATACATATATACACACATATA 587
              || | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||||
Db        368 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 427

Qy        588 TATACATATATACACATATATACATATATACACATATATACATATATACACACATATATA 647
              |||| ||||||| | ||||||| ||||| | | ||||||| ||||||| | | |||||||
Db        428 TATATATATATATATATATATATAAATATAGATATATATATATATATATATATATATA 487

Qy        648 CACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATA 707
              | ||| ||||| | | | ||| ||||| |||||||||| | ||||||||  ||||||||
Db        488 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 547

Qy        708 TGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGA 767
              | || | ||||||          |  |||          |  |   ||  | |    ||| |
Db        548 TATAAAGATATATATAGATATATATATATATATATATATATATATATAAATATATATATA 607

Qy        768 GGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATA 827
              | | | |  | | |  | | |||    | |   | || |     |      | |||
Db        608 TATATATATATATATATATATAAATATATATATATATATATATATATATATATATATA 667

Qy        828 GTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA 867
                |  ||||| | | || | ||   | |  ||| | ||
Db        668 TATAGACAAATAGAGATATATATATATATATATATATA 707
```

```
RESULT 30
CA916667/c
LOCUS       CA916667                496 bp    mRNA    linear   EST 27-DEC-2002
DEFINITION  PCSC21302 Scarlet Runner Bean Suspensor Region TriplEx2 Phaseolus
            coccineus cDNA 5' similar to No significant hits, mRNA sequence.
ACCESSION   CA916667
VERSION     CA916667.1  GI:27403659
KEYWORDS    EST.
SOURCE      Phaseolus coccineus
  ORGANISM  Phaseolus coccineus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Phaseolus.
REFERENCE   1  (bases 1 to 496)
  AUTHORS   Bui,A.Q., Le,B.H., Weterings,K., Bi,Y.-P., Choi,J.-S.,
            McElroy,K.E., Choi,P.S., Harada,J.J., Fischer,R.L. and
            Goldberg,R.B.
  TITLE     Gene Activity in Different Regions of a Post-Fertilization Plant
            Embryo by EST Analysis
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Goldberg, R.B.
            Department of Molecular, Cell, & Developmental Biology
            University of California, Los Angeles
            621 Charles E. Young Drive South, Los Angeles, CA 90095-1606, USA
            Tel: 310 825 3270
            Fax: 310 825 8201
            Email: bobg@ucla.edu
            Seq primer: 5' TriplEx
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .496
                     /organism="Phaseolus coccineus"
                     /mol_type="mRNA"
                     /cultivar="Hammond's Dwarf Scarlet"
                     /db_xref="taxon:3886"
                     /dev_stage="6-days post-pollination"
                     /clone_lib="Scarlet Runner Bean Suspensor Region TriplEx2"
                     /note="Organ: Suspensor Region of Globular-Stage Embryos;
                     Vector: TriplEx2; Site_1: SfiIA; Site_2: SfiIB; Suspensor
                     regions were micro-dissected from globular-stage embryos
                     six days after pollination from greenhouse-grown plants
                     [Weterings et al., Plant Cell 13, 2409-2425 (2001)].
                     Double-stranded cDNA was synthesized from suspensor mRNA
                     using the SMART cDNA Library Construction Kit according to
                     the manufacturer (Clontech). The suspensor cDNA fragments
                     were directionally ligated into the SfiI restriction site
                     of the lambda TriplEx2 vector (Clontech), and the
                     recombinant cDNAs were transformed into E. coli XL1-Blue
                     cells (Clontech). Suspensor cDNA plasmids used for
                     directional sequencing were obtained by in vivo excision
                     from the lambda TriplEx2 recombinants in E. coli BM25.8
                     cells (Clontech)."
ORIGIN

  Query Match             7.1%;  Score 330.2;  DB 6;  Length 496;
  Best Local Similarity   80.9%;  Pred. No. 5.4e-46;
  Matches  397;  Conservative    0;  Mismatches   93;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db        491 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 432

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | |||||||| | ||||||||| | ||| |||| || | |
Db        431 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 372

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        371 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 312

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db        311 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 252

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db        251 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 192

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db        191 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 132

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db        131 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 72

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||||||||||||||| |||||||||  |||| |
Db         71 TATATATATATATATATATATATATATATACATATATATACATATATATATATATACA 12

Qy        706 TATGTATATAT 716
              ||| |||||||
Db         11 TATCTATATAT 1


RESULT 31
CF924719/c
LOCUS       CF924719                751 bp    mRNA    linear   EST 10-NOV-2003
DEFINITION  DD005F11.ab1 Codonopsis lanceolata cDNA library Codonopsis
            lanceolata cDNA, mRNA sequence.
ACCESSION   CF924719
VERSION     CF924719.1  GI:38243245
KEYWORDS    EST.
SOURCE      Codonopsis lanceolata
  ORGANISM  Codonopsis lanceolata
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            asterids; campanulids; Asterales; Campanulaceae; Codonopsis.
REFERENCE   1  (bases 1 to 751)
  AUTHORS   Yang,D.C., In,J.G. and Lee,B.S.
  TITLE     Codonopsis lanceolata
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Deok Chun Yang
            BioPia
            1 Seochonri, Kihung, Yongin, Kyungki-do, 449-701, Korea
            Tel: 82 31 201 2688
            Fax: 82 31 205 6323
            Email: dcyang@ibiopia.com.
FEATURES             Location/Qualifiers
     source          1. .751
                     /organism="Codonopsis lanceolata"
                     /mol_type="mRNA"
                     /db_xref="taxon:103999"
                     /clone_lib="Codonopsis lanceolata cDNA library"
                     /note="Organ: root; Vector: pTriplEx2; Site_1: Sfi IA;
                     Site_2: Sfi IB"
ORIGIN

  Query Match             7.1%;  Score 329.8;  DB 7;  Length 751;
  Best Local Similarity   80.3%;  Pred. No. 6e-46;
  Matches  399;  Conservative    0;  Mismatches   97;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db        749 TATATATATATATATATATATATATAGATATATATATATATATATATATATATATATATA 690

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | |||||||| | ||||||||| | ||| |||| || | |
Db        689 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 630

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        629 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 570
```

```
Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db        569 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 510

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db        509 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 450

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db        449 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 390

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db        389 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 330

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | ||| |||||  ||||||
Db        329 TATATATATATATATATATATATATATATATATATATATATATATATAGATATATATATA 270

Qy        706 TATGTATATATATATGC 722
              ||  ||| |||||||||
Db        269 TAGATATGTATATATGC 253


RESULT 32
BX145650
LOCUS       BX145650                668 bp    DNA     linear   GSS 28-JAN-2003
DEFINITION  Danio rerio genomic clone DKEY-109N18, genomic survey sequence.
ACCESSION   BX145650
VERSION     BX145650.1  GI:27977003
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 668)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the T7 end of BAC 109N18. 109N18
            is part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .668
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-109N18"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             7.1%;  Score 329;  DB 9;  Length 668;
  Best Local Similarity   79.8%;  Pred. No. 8.3e-46;
  Matches  400;  Conservative    0;  Mismatches  100;  Indels    1;  Gaps    1;

Qy        217 AAAAAGGCATTTTATATATATATATATATATATACACACACACACACATATATATATA 276
              | | | | | | ||||||||||||||||||||||| | | | | | ||||||||||||
Db        169 AGAGATGAAGTGTTTATATATATATATATATATATATATATATATATATATATATATATA 228

Qy        277 CACATATATATACACATATATACATATATACACATATATACACATATATATACACATACA 336
              | ||||||||| | ||||||| ||||||| | ||||||| | ||||||||| | ||| |
Db        229 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 288

Qy        337 TATGTACACATATATATACACATATGTATACACATATATACACATATATACACACATATA 396
              ||| || | ||||||||| | |||| |||| | ||||||| | ||||||| | | |||||
Db        289 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 348
```

```
Qy        397 TACACATATATACACACATATATACACATATATACACATATATACACATATACACATATA 456
              || | ||||||| | | ||||||| | ||||||| | ||||||| | ||||| | |||||
Db        349 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 408

Qy        457 TACACATATATACATATATACACATATATATAATATACACACATATATATACACATATAT 516
              || | ||||||| ||||||| | |||||||| ||||| | | ||||||||| | |||||||
Db        409 TATATATATATATATATATATATATATATAT-ATATATATATATATATATATATATATAT 467

Qy        517 ACACACATATATACACATATATACACATATATATACACATATATACACATATATACATAT 576
              | | | ||||||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||
Db        468 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 527

Qy        577 ATACACATATATATACATATATACACATATATACATATATACACATATATACATATATAC 636
              ||| | ||||||||| ||||||| | ||||||| ||||||| | ||||||| |||||||
Db        528 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 587

Qy        637 ACACATATATACACATACATATACACACACATAGATATACATATATATACACATATATAT 696
              | | ||||||| | ||| ||||| | | | ||| ||||| |||||||||| | ||||||||
Db        588 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 647

Qy        697 ACGTATATATATGTATATATA 717
              |  |||||||| ||||||||
Db        648 ATATATATATATATATATATA 668


RESULT 33
CL934691/c
LOCUS       CL934691                784 bp    DNA     linear   GSS 14-SEP-2004
DEFINITION  OA_ABa0044J13.f OA_ABa Oryza australiensis genomic clone
            OA_ABa0044J13 5', genomic survey sequence.
ACCESSION   CL934691
VERSION     CL934691.1  GI:52063483
KEYWORDS    GSS.
SOURCE      Oryza australiensis
  ORGANISM  Oryza australiensis
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 784)
  AUTHORS   Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R.,
            Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
  TITLE     OMAP Project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0044  row: J  column: 13
            Seq primer: TAA TAC GAC TCA CTA TAG GG
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .784
                     /organism="Oryza australiensis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4532"
                     /clone="OA_ABa0044J13"
                     /tissue_type="young leaves"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OA_ABa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

  Query Match             7.1%;  Score 328.4;  DB 9;  Length 784;
  Best Local Similarity   79.0%;  Pred. No. 1e-45;
  Matches  403;  Conservative    0;  Mismatches  106;  Indels    1;  Gaps    1;

Qy        212 AAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACATATAT 271
```

```
Db        719 ATAAATATATATATATTTATATATATATATATATATATATATATATATATATATCTATAT 660

Qy        272 ATATACACATATATATACACATATATACATATATACACATATATACACATATATATACAC 331
Db        659 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 600

Qy        332 ATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACACAC 391
Db        599 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 540

Qy        392 ATATATACACATATATACACACATATATACACATATATACACATATATACACATATACAC 451
Db        539 ATATATATATATATATATATATATATATCTATATATCTATATATATATATATATATATAT 480

Qy        452 ATATATACACATATATACATATATACACATATATATA-ATATACACACATATATATACAC 510
Db        479 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 420

Qy        511 ATATATACACACATATATACACATATATACACATATATATACACATATATACACATATAT 570
Db        419 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 360

Qy        571 ACATATATACACATATATATACATATATACACATATATACATATATACACATATATACAT 630
Db        359 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATCTAT 300

Qy        631 ATATACACACATATATACACATACATATACACACACATAGATATACATATATATACACAT 690
Db        299 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 240

Qy        691 ATATATACGTATATATATGTATATATATAT 720
Db        239 ATATATATATATATATATATATATATAT 210


RESULT 34
CL776144/c
LOCUS       CL776144               587 bp    DNA     linear   GSS 30-JUL-2004
DEFINITION  OR_BBa0090B12.r OR_BBa Oryza rufipogon genomic clone OR_BBa0090B12
            3', genomic survey sequence.
ACCESSION   CL776144
VERSION     CL776144.1  GI:50850530
KEYWORDS    GSS.
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 587)
  AUTHORS   Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R.,
            Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
  TITLE     OMAP Project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Insert Length: 161   Std Error: 0.00
            Plate: 0090  row: B  column: 12
            Seq primer: CAC TCA TTA GGC ACC CCA
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .587
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
                     /clone="OR_BBa0090B12"
                     /tissue_type="young leaves"
                     /lab_host="DH10B-T1 phage resistant"
                     /clone_lib="OR_BBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

  Query Match             7.1%;  Score 328.2;  DB 9;  Length 587;
  Best Local Similarity   77.5%;  Pred. No. 1.1e-45;
  Matches  410;  Conservative    0;  Mismatches  118;  Indels    1;  Gaps    1;

Qy        209 AGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACATA 268
Db        537 ATATATATATATATATATATAATATATATATATATATATATATATATATATATATATA 478

Qy        269 TATATATACACATATATATACACATATATACATATATACACATATATACACATATATATA 328
Db        477 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 418

Qy        329 CACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACA 388
Db        417 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 358

Qy        389 CACATATATACACATATATACACACATATATACACATATATACACATATATACACATATA 448
Db        357 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 298

Qy        449 CACATATATACACATATATACATATATACACATATATATA-ATATACACACATATATATA 507
Db        297 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 238

Qy        508 CACATATATACACACATATATACACATATATACACATATATATACACATATATACACATA 567
Db        237 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 178

Qy        568 TATACATATATACACATATATATACATATATACACATATATACATATATACACATATATA 627
Db        177 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 118

Qy        628 CATATATACACACATATATACACATACATATACACACACATAGATATACATATATATACA 687
Db        117 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 58

Qy        688 CATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAA 736
Db         57 TATATATATATATATATATGCATATGCGCCTGTGTATAAGAAATCATAA 9


RESULT 35
CF924719
LOCUS       CF924719               751 bp    mRNA    linear   EST 10-NOV-2003
DEFINITION  DD005F11.ab1 Codonopsis lanceolata cDNA library Codonopsis
            lanceolata cDNA, mRNA sequence.
ACCESSION   CF924719
VERSION     CF924719.1  GI:38243245
KEYWORDS    EST.
SOURCE      Codonopsis lanceolata
  ORGANISM  Codonopsis lanceolata
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            asterids; campanulids; Asterales; Campanulaceae; Codonopsis.
REFERENCE   1  (bases 1 to 751)
  AUTHORS   Yang,D.C., In,J.G. and Lee,B.S.
  TITLE     Codonopsis lanceolata
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Deok Chun Yang
            BioPia
            1 Seochonri, Kihung, Yongin, Kyungki-do, 449-701, Korea
            Tel: 82 31 201 2688
            Fax: 82 31 205 6323
            Email: dcyang@ibiopia.com.
FEATURES             Location/Qualifiers
     source          1. .751
```

/organism="Codonopsis lanceolata"
/mol_type="mRNA"
/db_xref="taxon:103999"
/clone_lib="Codonopsis lanceolata cDNA library"
/note="Organ: root; Vector: pTriplEx2; Site_1: Sfi IA;
Site_2: Sfi IB"

ORIGIN

Query Match 7.1%; Score 327.2; DB 7; Length 751;
Best Local Similarity 80.0%; Pred. No. 1.6e-45;
Matches 397; Conservative 0; Mismatches 98; Indels 1; Gaps 1;

```
Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | |||| |||| |||||||||||||||   | | | | ||||||||||| | |||||||
Db        256 TATATACATATCTATATATATATATATCTATATATATATATATATATATATATATATATA 315

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db        316 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 375

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        376 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 435

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db        436 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 495

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db        496 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 555

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db        556 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 615

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db        616 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 675

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||| |||| ||||||
Db        676 TATATATATATATATATATATATATATATATATATATATATATATATATATCTATATATATA 735

Qy        706 TATGTATATATATATG 721
              ||| ||||||||||||
Db        736 TATATATATATATATG 751
```

RESULT 36
CA916667
LOCUS CA916667 496 bp mRNA linear EST 27-DEC-2002
DEFINITION PCSC21302 Scarlet Runner Bean Suspensor Region TriplEx2 Phaseolus coccineus cDNA 5' similar to No significant hits, mRNA sequence.
ACCESSION CA916667
VERSION CA916667.1 GI:27403659
KEYWORDS EST.
SOURCE Phaseolus coccineus
ORGANISM Phaseolus coccineus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Phaseolus.
REFERENCE 1 (bases 1 to 496)
AUTHORS Bui,A.Q., Le,B.H., Weterings,K., Bi,Y.-P., Choi,J.-S., McElroy,K.E., Choi,P.S., Harada,J.J., Fischer,R.L. and Goldberg,R.B.
TITLE Gene Activity in Different Regions of a Post-Fertilization Plant Embryo by EST Analysis
JOURNAL Unpublished (2002)
COMMENT Contact: Goldberg, R.B.
Department of Molecular, Cell, & Developmental Biology
University of California, Los Angeles
621 Charles E. Young Drive South, Los Angeles, CA 90095-1606, USA
Tel: 310 825 3270
Fax: 310 825 8201
Email: bobg@ucla.edu
Seq primer: 5' TriplEx
POLYA=No.

FEATURES Location/Qualifiers
source 1. .496
/organism="Phaseolus coccineus"
/mol_type="mRNA"
/cultivar="Hammond's Dwarf Scarlet"
/db_xref="taxon:3886"
/dev_stage="6-days post-pollination"
/clone_lib="Scarlet Runner Bean Suspensor Region TriplEx2"
/note="Organ: Suspensor Region of Globular-Stage Embryos; Vector: TriplEx2; Site_1: SfiIA; Site_2: SfiIB; Suspensor regions were micro-dissected from globular-stage embryos six days after pollination from greenhouse-grown plants [Weterings et al., Plant Cell 13, 2409-2425 (2001)]. Double-stranded cDNA was synthesized from suspensor mRNA using the SMART cDNA Library Construction Kit according to the manufacturer (Clontech). The suspensor cDNA fragments were directionally ligated into the SfiI restriction site of the lambda TriplEx2 vector (Clontech), and the recombinant cDNAs were transformed into E. coli XL1-Blue cells (Clontech). Suspensor cDNA plasmids used for directional sequencing were obtained by in vivo excision from the lambda TriplEx2 recombinants in E. coli BM25.8 cells (Clontech)."

ORIGIN

Query Match 7.1%; Score 326.2; DB 6; Length 496;
Best Local Similarity 80.0%; Pred. No. 2.6e-45;
Matches 396; Conservative 0; Mismatches 98; Indels 1; Gaps 1;

```
Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | |||| |||| |||||||||||||||   | | | | ||||||||||| | |||||||
Db          2 TATATAGATATGTATATATATATATATATGTATATATATGTATATATATATATATATATA 61

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db         62 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 121

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        122 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 181

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db        182 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 241

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db        242 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 301

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db        302 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 361

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db        362 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 421

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db        422 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 481

Qy        706 TATGTATATATATAT 720
              ||| ||||||||||||
Db        482 TATATATATATATAT 496
```

RESULT 37

```
CL623140
LOCUS       CL623140                744 bp    DNA     linear   GSS 01-JUL-2004
DEFINITION  OR_BBa0018D11.r OR_BBa Oryza rufipogon genomic clone OR_BBa0018D11
            3', genomic survey sequence.
ACCESSION   CL623140
VERSION     CL623140.1  GI:49121028
KEYWORDS    GSS.
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 744)
  AUTHORS   Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R.,
            Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
  TITLE     OMAP Project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Insert Length: 161   Std Error: 0.00
            Plate: 0018  row: D  column: 11
            Seq primer: CAC TCA TTA GGC ACC CCA
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .744
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
                     /clone="OR_BBa0018D11"
                     /tissue_type="young leaves"
                     /lab_host="DH10B-T1 phage resistant"
                     /clone_lib="OR_BBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

  Query Match             7.0%;  Score 324.2;  DB 9;  Length 744;
  Best Local Similarity   79.2%;  Pred. No. 5.3e-45;
  Matches  397;  Conservative    0;  Mismatches  103;  Indels    1;  Gaps    1;

Qy        217 AAAAAGGCATTTTATATATATATATATATATATACACACACACACATATATATATA 276
              |||||      |  ||| ||| ||||||||||||||||| | | | | | ||||||||||
Db        183 AAAAAATACATACATACATACATATATATATATATATATATATATATATATATATATA 242

Qy        277 CACATATATATACACATATATACATATATACACATATATACACATATATATACACATACA 336
              | ||||||||| | ||||||| ||||||| | ||||||| | ||||||||| | ||| |
Db        243 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 302

Qy        337 TATGTACACATATATATACACATATGTATACACATATATACACATATATACACACATATA 396
              ||| || | ||||||||| | |||| |||| | ||||||| | ||||||| | | |||||
Db        303 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 362

Qy        397 TACACATATATACACACATATATACACATATATACACATATATACACATATACACATATA 456
              || | ||||||| | | ||||||| | ||||||| | ||||||| | ||||| | |||||
Db        363 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 422

Qy        457 TACACATATATACATATATACACATATATATAATATACACACATATATATACACATATAT 516
              || | ||||||| ||||||| | |||||||| ||||| | | ||||||||| | ||||||
Db        423 TATATATATATATATATATATATATATATAT-ATATATATATATATATATATATATAT 481

Qy        517 ACACACATATATACACATATATACACATATATATACACATATATACACATATATACATAT 576
              | | | ||||||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||
Db        482 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 541

Qy        577 ATACACATATATATACATATATACACATATATACATATATACACATATATACATATATAC 636
              ||| | ||||||||| ||||||| | ||||||| ||||||| | ||||||| |||||||

Db        542 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 601

Qy        637 ACACATATATACACATACATATACACACACATAGATATACATATATATACACATATATAT 696
              | | ||||||| | ||| ||||| | | | ||| ||||| ||||||||| | ||||||||
Db        602 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 661

Qy        697 ACGTATATATATGTATATATA 717
              |  |||||||||  |||||||
Db        662 ATATATATATATAGATATATA 682


RESULT 38
CL776144
LOCUS       CL776144                587 bp    DNA     linear   GSS 30-JUL-2004
DEFINITION  OR_BBa0090B12.r OR_BBa Oryza rufipogon genomic clone OR_BBa0090B12
            3', genomic survey sequence.
ACCESSION   CL776144
VERSION     CL776144.1  GI:50850530
KEYWORDS    GSS.
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 587)
  AUTHORS   Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R.,
            Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
  TITLE     OMAP Project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Insert Length: 161   Std Error: 0.00
            Plate: 0090  row: B  column: 12
            Seq primer: CAC TCA TTA GGC ACC CCA
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .587
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
                     /clone="OR_BBa0090B12"
                     /tissue_type="young leaves"
                     /lab_host="DH10B-T1 phage resistant"
                     /clone_lib="OR_BBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

  Query Match             7.0%;  Score 323.8;  DB 9;  Length 587;
  Best Local Similarity   77.8%;  Pred. No. 6.4e-45;
  Matches  416;  Conservative    0;  Mismatches  117;  Indels    2;  Gaps    2;

Qy        188 ACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATAT 247
              |||   ||| |     ||| |     | | | | |       | ||||||||||||||||||
Db         23 ACACAGGCGCATATGCATATATATATATATATATATATATATATATATATATATATAT 82

Qy        248 ATATACACACACACACACATATATATATACACATATATATACACATATATACATATATAC 307
              ||||| | | | | | | |||||||||||| | ||||||||| | ||||||| |||||||
Db         83 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 142

Qy        308 ACATATATACACATATATATACACATACATATGTACACATATATATACACATATGTATAC 367
              | ||||||| | ||||||||| | ||| |||| || | ||||||||| | |||| ||||
Db        143 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 202

Qy        368 ACATATATACACATATATACACACATATATACACATATATACACACATATATACACATAT 427
              | ||||||| | ||||||| | | ||||||| | ||||||| | | ||||||| | ||||
```

```
Db        203 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 262

Qy        428 ATACACATATATACACATATACACATATATACACATATATACATATATACACATATATAT 487
              ||| | ||||||| | ||||| | ||||||| | ||||||| ||||||| | ||||||||
Db        263 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 322

Qy        488 A-ATATACACACATATATATACACATATATACACACATATATACACATATATACACACATAT 546
              | ||||| | | ||||||||| | ||||||| | | ||||||| | ||||||| | ||||
Db        323 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 382

Qy        547 ATATACACATATATACACATATATACATATATACACATATATATACATATATACACATAT 606
              ||||| | ||||||| | ||||||| ||||||| | ||||||||| ||||||| | ||||
Db        383 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 442

Qy        607 ATACATATATACACATATATACATATATACACACATATATACACATACATATACACACAC 666
              ||| ||||||| | ||||||| ||||||| | | ||||||| | ||| ||||| | | |
Db        443 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 502

Qy        667 ATAGATATACATA-TATATACACATATATATACGTATATATATGTATATATATAT 720
              ||| ||||| ||| |||||| | ||||||||| |||||||| ||||||||||
Db        503 ATATATATATATATTATATATATATATATATATATATATATATATATATATATATATAT 557


RESULT 39
CL830207/c
LOCUS       CL830207                491 bp    DNA     linear   GSS 09-AUG-2004
DEFINITION  OR_CBa0052C06.f OR_CBa Oryza rufipogon genomic clone OR_CBa0052C06
            5', genomic survey sequence.
ACCESSION   CL830207
VERSION     CL830207.1  GI:51075808
KEYWORDS    GSS.
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 491)
  AUTHORS   Kim,H., Yu,Y., Wissotski,M., Yost,D., Stum,D., Rao,K., Luo,M.,
            Jetty,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and
            Wing,R.
  TITLE     OMAP project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0052  row: C  column: 06
            Seq primer: TAA TAC GAC TCA CTA TAG GG
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .491
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
                     /clone="OR_CBa0052C06"
                     /tissue_type="young leaves"
                     /dev_stage="2 week old seedlings"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OR_CBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII;
                     drk treated 36 hrs before harvest"
ORIGIN

  Query Match             7.0%;  Score 322.6;  DB 9;  Length 491;
  Best Local Similarity   80.2%;  Pred. No. 1.1e-44;
  Matches  392;  Conservative    0;  Mismatches   94;  Indels    3;  Gaps    1;

Qy        232 ATATATATATATATATATATACACACACACACATATATATATACACATATATATACAC 291
              ||||| |||||||||||||||| | | | | | |||||||||||| |  ||| |||||||
Db        491 ATATACATATATATATATATATATATATATATATATATATATAGATCTATGTATACAC 432

Qy        292 ATATATACATATATACACATATATACACATATATATACACATACATATGTACACATATAT 351
              ||||||| ||||||| | ||| | | |||| |||  | ||| ||| ||| |||||||||
Db        431 ATATATATATATATATATATACAGATAGATATGTATGTATATATATAGGTATACATATAT 372

Qy        352 ATACACATATGTATACACATATATACACATATATACACACATATATACACATATATACAC 411
              ||| |||||| |||| | ||||||| | ||||||  | | ||||||| | ||||||| |
Db        371 ATATACATATATATATATATATATATATATATATCTATATATATATAGATATATATAT 312

Qy        412 ACATATATACACATATATACACATATATACACATATACACATATATACACATATATACAT 471
              | ||||| | | ||||||| | ||||||| | ||||| | ||||||| | ||||||| ||
Db        311 ATATATAGAGATATATATATATATATATATATATATATATATATATATATATATATATAT 252

Qy        472 ATATACACATATATATAATATACACACATATATATACACATATATACACACATATATACA 531
              ||||| | ||||||   ||||| | | ||||||||| | ||||||| | | ||||||| |
Db        251 ATATATATATATAT---ATATATATATATATATATATATATATATAGATATATATATATA 195

Qy        532 CATATATACACATATATATACACATATATACACATATATACATATATACACATATATATA 591
              ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||||||||
Db        194 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 135

Qy        592 CATATATACACATATATACATATATACACATATATACATATATACACACATATATACACA 651
              ||||||| | ||||||| ||||||| |||||||||||||||||||||||| ||||||| | |
Db        134 TATATATATATATATATATATATATATACATATATACATATATACACATATATATATA 75

Qy        652 TACATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTA 711
              || ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||||| ||
Db         74 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 15

Qy        712 TATATATAT 720
              |||||||||
Db         14 TATATATAT 6


RESULT 40
CC099198
LOCUS       CC099198                845 bp    DNA     linear   GSS 16-APR-2003
DEFINITION  CSU-K34.116B12.T7 CSU-K34 Aedes aegypti genomic clone
            CSU-K34.116B12, genomic survey sequence.
ACCESSION   CC099198
VERSION     CC099198.1  GI:29965572
KEYWORDS    GSS.
SOURCE      Aedes aegypti (yellow fever mosquito)
  ORGANISM  Aedes aegypti
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Aedes;
            Stegomyia.
REFERENCE   1  (bases 1 to 845)
  AUTHORS   Loftus,B., Shetty,J., Severson,D., Brown,S. and Knudson,D.
  TITLE     End sequencing of Aedes aegypti BACs
  JOURNAL   Unpublished (2003)
COMMENT     Other_GSSs: CSU-K34.116B12.SP6
            Contact: Brendan Loftus
            Department of Eukaryotic Genomics
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-3543
            Fax: 301-838-0208
            Email: enta@tigr.org
            Library was provided by Susan Brown and Dennis Knudson at Colorado
            State University.
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .845
                     /organism="Aedes aegypti"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7159"
                     /clone="CSU-K34.116B12"
                     /clone_lib="CSU-K34"
```

/note="Vector: pBACe3.6; Site_1: EcoRI; Source DNA: Aedes aegypti; strain unknown [derived from freshly hatched larvae at the Virus Research Centre, Poona, India. Reference: SINGH, K. R. P., 1967 Cell cultures derived from larvae of Aedes albopictus (Skuse) and Aedes aegypti (L.). Current Science 36: 506-508]; ATC-10 cell line ATCC CCL-125"

ORIGIN

Query Match 6.9%; Score 318.2; DB 8; Length 845;
Best Local Similarity 79.7%; Pred. No. 5.5e-44;
Matches 388; Conservative 0; Mismatches 98; Indels 1; Gaps 1;

```
Qy        211 AAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACATATA 270
              |||  || |||   | | |||||||||||||||||||||||| | | | | | | |||||
Db        360 AAACTAACAAATTAAATATATATATATATATATATATATATATATATATATATATATA 419

Qy        271 TATATACACATATATATACACATATATACATATATACACATATATACACATATATATACA 330
              |||||| | |||||||||| | ||||||| ||||||| | ||||||| | ||||||||| |
Db        420 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 479

Qy        331 CATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACACA 390
              ||| |||| || | ||||||||| | |||| |||| | ||||||| | ||||||| | |
Db        480 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 539

Qy        391 CATATATACACATATATACACACATATATACACATATATACACATATATACACATATACA 450
              ||||||| | ||||||| | | ||||||| | ||||||| | ||||||| | ||||| |
Db        540 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 599

Qy        451 CATATATACACATATATACATATATACACATATATATAATATACACACATATATATACAC 510
              ||||||| | ||||||| ||||||| | |||||||| ||||| | | |||||||| |
Db        600 TATATATATATATATATATATATATATATATATATAT-ATATATATATATATATATAT 658

Qy        511 ATATATACACACATATATACACATATATACACATATATATACACATATATACACATATAT 570
              ||||||| | | ||||||| | ||||||| | |||||||| | ||||||| | ||||||
Db        659 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 718

Qy        571 ACATATATACACATATATATACATATATACACATATATACATATATACACATATATACAT 630
              | ||||||| | |||||||| ||||||| | ||||||| ||||||| | ||||||| ||
Db        719 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 778

Qy        631 ATATACACACATATATACACATACATATACACACACATAGATATACATATATATACACAT 690
              ||||| | | ||||||| | ||| ||||| | | | ||| ||||| ||||||||| | ||
Db        779 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 838

Qy        691 ATATATA 697
              |||||||
Db        839 ATATATA 845
```

RESULT 41
CC099198/c
LOCUS CC099198 845 bp DNA linear GSS 16-APR-2003
DEFINITION CSU-K34.116B12.T7 CSU-K34 Aedes aegypti genomic clone CSU-K34.116B12, genomic survey sequence.
ACCESSION CC099198
VERSION CC099198.1 GI:29965572
KEYWORDS GSS.
SOURCE Aedes aegypti (yellow fever mosquito)
ORGANISM Aedes aegypti
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Aedes; Stegomyia.
REFERENCE 1 (bases 1 to 845)
AUTHORS Loftus,B., Shetty,J., Severson,D., Brown,S. and Knudson,D.
TITLE End sequencing of Aedes aegypti BACs
JOURNAL Unpublished (2003)
COMMENT Other_GSSs: CSU-K34.116B12.SP6
Contact: Brendan Loftus
Department of Eukaryotic Genomics
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-3543
Fax: 301-838-0208
Email: enta@tigr.org
Library was provided by Susan Brown and Dennis Knudson at Colorado State University.
Seq primer: T7
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .845
/organism="Aedes aegypti"
/mol_type="genomic DNA"
/db_xref="taxon:7159"
/clone="CSU-K34.116B12"
/clone_lib="CSU-K34"
/note="Vector: pBACe3.6; Site_1: EcoRI; Source DNA: Aedes aegypti; strain unknown [derived from freshly hatched larvae at the Virus Research Centre, Poona, India. Reference: SINGH, K. R. P., 1967 Cell cultures derived from larvae of Aedes albopictus (Skuse) and Aedes aegypti (L.). Current Science 36: 506-508]; ATC-10 cell line ATCC CCL-125"

ORIGIN

Query Match 6.9%; Score 317; DB 8; Length 845;
Best Local Similarity 78.0%; Pred. No. 8.8e-44;
Matches 394; Conservative 0; Mismatches 110; Indels 1; Gaps 1;

```
Qy        229 TATATATATATATATATATATACACACACACACATATATATATACACATATATATA 288
              |||||||||||||||||||||||| | | | | | ||||||||| | |||||||||
Db        845 TATATATATATATATATATATATATATATATATATATATATATATATATATATATA 786

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |||
Db        785 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 726

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              |||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||||
Db        725 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 666

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||||
Db        665 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 606

Qy        469 CATATATACACATATATATAATATACACACATATATATACACATATATACACACATATAT 528
              ||||||| | |||||||| ||||| | | ||||||||| | ||||||| | | ||||||
Db        605 TATATATATATATATATAT-ATATATATATATATATATATATATATATATATATATAT 547

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACATATAT 588
              | | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | ||||||
Db        546 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 487

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
              ||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||||
Db        486 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 427

Qy        649 ACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATAT 708
              | ||| ||||| | | | ||| ||||| ||||||||| | ||||||||  | ||  |
Db        426 ATATATATATATATATATATATATATATATATATATATATATATATATATTTAATTTG 367

Qy        709 GTATATATATATGCTCCAGAGTTCA 733
              ||  |  | ||  |  |  || ||
Db        366 TTAGTTTGAAATCATTGACTGTGCA 342
```

RESULT 42
CL855754
LOCUS CL855754 816 bp DNA linear GSS 16-AUG-2004
DEFINITION OR_CBa0085L12.r OR_CBa Oryza rufipogon genomic clone OR_CBa0085L12 3', genomic survey sequence.
ACCESSION CL855754
VERSION CL855754.1 GI:51264405
KEYWORDS GSS.
SOURCE Oryza rufipogon
ORGANISM Oryza rufipogon

```
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 816)
  AUTHORS   Kim,H., Yu,Y., Wissotski,M., Yost,D., Stum,D., Rao,K., Luo,M.,
            Jetty,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and
            Wing,R.
  TITLE     OMAP project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0085  row: L  column: 12
            Seq primer: CAC TCA TTA GGC ACC CCA
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .816
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
                     /clone="OR_CBa0085L12"
                     /tissue_type="young leaves"
                     /dev_stage="2 week old seedlings"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OR_CBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII;
                     drk treated 36 hrs before harvest"
ORIGIN

  Query Match             6.8%;  Score 315;  DB 9;  Length 816;
  Best Local Similarity   79.2%;  Pred. No. 1.9e-43;
  Matches  411;  Conservative    0;  Mismatches  105;  Indels    3;  Gaps    3;

Qy        204 ATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATATACACACACACAC 263
              || | |  | | || | |   || | |||||||||||||||||||||||| | | | | |
Db        250 ATATAATCATATAATATATATATATCATATATATATATATATATATATATATATATATAT 309

Qy        264 ACATATATATATACACATATATATACACATATATACATATATACACATATATACACATAT 323
              | |||||||||||| | ||||||||| | ||||||| ||||||| | ||||||| | ||||
Db        310 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 369

Qy        324 ATATACACATACATATGTACACATATATATACACATATGTATACACATATATACACATAT 383
              ||||| | ||| |||| || | ||||||||| | |||| |||| | ||||||| | ||||
Db        370 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 429

Qy        384 ATACACACATATATACACATATATACACA-CATATATACACATATATACACATATATACA 442
              ||| | | ||||||| | ||||||| | | |||||||| | ||||||| | ||||||| |
Db        430 ATATATATATATATATATATATATATATATCATATATATATATATATATATATATATATA 489

Qy        443 CATATACACATATATACACATATATACATATATACACATATATATA-ATATACACACATA 501
              |||||  | ||||||| | ||||||| ||||||| | ||||||||| ||||| | | |||
Db        490 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 549

Qy        502 TATATACACATATATACACACATATATACACATATATACACATATATATACACATATATA 561
              |||||| | ||||||| | | |||||||| | ||||||| | ||||||||| | |||||||
Db        550 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 609

Qy        562 CACATATATACATATATACACATATATATACATATATACACATATATACATATATACACA 621
              | ||||||| ||||||| | ||||||||| ||||||| | ||||||| ||||||| | |
Db        610 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 669

Qy        622 TATATACATATATACACACATATATACACATACATATACACACACATAGATATACATATA 681
              |||||| ||||||| | | ||||||| | ||| ||||| | | | ||| ||||| |||||
Db        670 TATATATATATATATA-ATATATATATATATATATATATATATATATATATATATATATA 728

Qy        682 TATACACATATATATACGTATATATATGTATATATATAT 720
```

```
              |||| | |||||||||  ||||||||| |||||||| ||
Db        729 TATATATATATATATATATATATATATATATATATATAAAT 767


RESULT 43
BI506858
LOCUS       BI506858                570 bp    mRNA    linear   EST 08-APR-2002
DEFINITION  BB170024B20C02.5 Bee Brain Normalized/Subtracted Library, BB17 Apis
            mellifera cDNA clone BB170024B20C02 5', mRNA sequence.
ACCESSION   BI506858
VERSION     BI506858.1  GI:15357232
KEYWORDS    EST.
SOURCE      Apis mellifera (honey bee)
  ORGANISM  Apis mellifera
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Apoidea;
            Apidae; Apis.
REFERENCE   1  (bases 1 to 570)
  AUTHORS   Whitfield,C.W., Band,M.R., Bonaldo,M.F., Kumar,C.G., Liu,L.,
            Pardinas,J., Robertson,H.M., Soares,B. and Robinson,G.E.
  TITLE     Annotated expressed sequence tags and cDNA microarrays for studies
            of brain and behavior in the honey bee
  JOURNAL   Genome Res. 12 (4), 555-566 (2002)
  MEDLINE   21929762
   PUBMED   11932240
COMMENT     Contact: Gene E. Robinson
            Department of Entomology
            University of Illinois
            505 S. Goodwin Ave., Urbana, IL 61801, USA
            Tel: 217 265 0309
            Fax: 217 244 3499
            Email: generobi@life.uiuc.edu
            This research was funded by the University of Illinois Critical
            Research Initiatives Fund and a Burroughs-Wellcome Trust Innovation
            Award in Functional Genomics to G.E. Robinson and an NSF
            Postdoctoral Fellowship in Bioinformatics to C.W. Whitfield.
            REPEAT IN THE SEQUENCE
            Simple_repeat STRAND (+) ELEMENT (TA)n LOCATION [90,570].
            PCR PRimers
            FORWARD: TAATACGACTCACTATAGGG
            BACKWARD: ATTAACCCTCACTAAAG
            Plate: BB170024B20  row: C  column: 02
            Seq primer: AGCGGATAACAATTTCACACAGGA
            High quality sequence stop: 570.
FEATURES             Location/Qualifiers
     source          1. .570
                     /organism="Apis mellifera"
                     /mol_type="mRNA"
                     /strain="mixed strains of European bees, predominantly
                     A.m. ligustica"
                     /db_xref="taxon:7460"
                     /clone="BB170024B20C02"
                     /sex="female"
                     /tissue_type="brain"
                     /dev_stage="adult worker honey bee"
                     /lab_host="DH10B"
                     /clone_lib="Bee Brain Normalized/Subtracted Library, BB17"
                     /note="Organ: brain; Vector: pT7T3-Pac; Site_1: EcoR1;
                     Site_2: Not1; This BB17 cDNA library was generated by
                     subtraction of the BB16  library with 4000 previously
                     sequenced clones. The BB16 library was  contributed by the
                     Soares laboratory and it was constructed and  normalized
                     as described by Bonaldo, M.F., Lennon, G. and  Soares,
                     M.B. (1996), Genome Research 6(9): 791-806. RNA was
                     prepared  from dissected brains of adult worker bees of
                     various ages and  various behavioral groups. "
ORIGIN

  Query Match             6.8%;  Score 314.6;  DB 4;  Length 570;
  Best Local Similarity   77.7%;  Pred. No. 2.4e-43;
  Matches  405;  Conservative    0;  Mismatches  114;  Indels    2;  Gaps    2;

Qy        201 GTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATATACACACACA 260
```

```
Db         48 GTCTCCACAGATAGAGAGGGGTGTAACTCAATTAAAATGATTTATATATATATATATATA 107

Qy        261 CACACATATATATATACACATATATATACACATATATACATATATACACATATATACACA 320
Db        108 -ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 166

Qy        321 TATATATACACATACATATGTACACATATATATACACATATGTATACACATATATACACA 380
Db        167 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 226

Qy        381 TATATACACACATATATACACATATATACACACATATATACACATATATACACATATATA 440
Db        227 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 286

Qy        441 CACATATACACATATATACACATATATACATATATACACATATATATA-ATATACACACA 499
Db        287 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 346

Qy        500 TATATATACACATATATACACACATATATACACATATATACACATATATATACACATATA 559
Db        347 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 406

Qy        560 TACACATATATACATATATACACATATATATACATATATACACATATATACATATATACA 619
Db        407 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 466

Qy        620 CATATATACATATATACACACATATATACACATACATATACACACACATAGATATACATA 679
Db        467 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 526

Qy        680 TATATACACATATATATACGTATATATATGTATATATATAT 720
Db        527 TATATATATATATATATATATATATATATATATATATAT 567
```

RESULT 44
BI506858/c

```
LOCUS       BI506858                570 bp    mRNA    linear   EST 08-APR-2002
DEFINITION  BB170024B20C02.5 Bee Brain Normalized/Subtracted Library, BB17 Apis
            mellifera cDNA clone BB170024B20C02 5', mRNA sequence.
ACCESSION   BI506858
VERSION     BI506858.1  GI:15357232
KEYWORDS    EST.
SOURCE      Apis mellifera (honey bee)
  ORGANISM  Apis mellifera
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Apoidea;
            Apidae; Apis.
REFERENCE   1  (bases 1 to 570)
  AUTHORS   Whitfield,C.W., Band,M.R., Bonaldo,M.F., Kumar,C.G., Liu,L.,
            Pardinas,J., Robertson,H.M., Soares,B. and Robinson,G.E.
  TITLE     Annotated expressed sequence tags and cDNA microarrays for studies
            of brain and behavior in the honey bee
  JOURNAL   Genome Res. 12 (4), 555-566 (2002)
  MEDLINE   21929762
   PUBMED   11932240
COMMENT     Contact: Gene E. Robinson
            Department of Entomology
            University of Illinois
            505 S. Goodwin Ave., Urbana, IL 61801, USA
            Tel: 217 265 0309
            Fax: 217 244 3499
            Email: generobi@life.uiuc.edu
            This research was funded by the University of Illinois Critical
            Research Initiatives Fund and a Burroughs-Wellcome Trust Innovation
            Award in Functional Genomics to G.E. Robinson and an NSF
            Postdoctoral Fellowship in Bioinformatics to C.W. Whitfield.
            REPEAT IN THE SEQUENCE
            Simple_repeat STRAND (+) ELEMENT (TA)n LOCATION [90,570].
            PCR PRimers
            FORWARD: TAATACGACTCACTATAGGG
            BACKWARD: ATTAACCCTCACTAAAG
            Plate: BB170024B20  row: C  column: 02
            Seq primer: AGCGGATAACAATTTCACACAGGA
            High quality sequence stop: 570.
FEATURES            Location/Qualifiers
     source         1. .570
                    /organism="Apis mellifera"
                    /mol_type="mRNA"
                    /strain="mixed strains of European bees, predominantly
                    A.m. ligustica"
                    /db_xref="taxon:7460"
                    /clone="BB170024B20C02"
                    /sex="female"
                    /tissue_type="brain"
                    /dev_stage="adult worker honey bee"
                    /lab_host="DH10B"
                    /clone_lib="Bee Brain Normalized/Subtracted Library, BB17"
                    /note="Organ: brain; Vector: pT7T3-Pac; Site_1: EcoR1;
                    Site_2: Not1; This BB17 cDNA library was generated by
                    subtraction of the BB16  library with 4000 previously
                    sequenced clones. The BB16 library was  contributed by the
                    Soares laboratory and it was constructed and  normalized
                    as described by Bonaldo, M.F., Lennon, G. and  Soares,
                    M.B. (1996), Genome Research 6(9): 791-806. RNA was
                    prepared  from dissected brains of adult worker bees of
                    various ages and  various behavioral groups. "
ORIGIN
```

```
  Query Match             6.8%;  Score 313.2;  DB 4;  Length 570;
  Best Local Similarity   80.5%;  Pred. No. 4.1e-43;
  Matches  391;  Conservative    0;  Mismatches   93;  Indels    2;  Gaps    2;

Qy        235 TATATATATATATATATACACACACACACATATATATATACACATATATATACACATA 294
Db        570 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 511

Qy        295 TATACATATATACACATATATACACATATATATACACATACATATGTACACATATATATA 354
Db        510 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 451

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
Db        450 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 391

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
Db        390 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 331

Qy        475 TACACATATATATAATATACACACATATATATACACATATATACACACATATATACACAT 534
Db        330 TATATATATATAT-ATATATATATATATATATATATATATATATATATATATATATAT 272

Qy        535 ATATACACATATATATACACATATATACACATATATACATATATACACATATATATACAT 594
Db        271 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 212

Qy        595 ATATACACATATATACATATATACACATATATACATATATACACACATATATACACATAC 654
Db        211 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 152

Qy        655 ATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATAT 714
Db        151 ATATATATATATATATATATATATATATATATATATATATA-TTATATATATATATAT 93

Qy        715 ATATAT 720
Db         92 ATAAAT 87
```

RESULT 45
CL855754/c

```
LOCUS       CL855754                816 bp    DNA     linear   GSS 16-AUG-2004
DEFINITION  OR_CBa0085L12.r OR_CBa Oryza rufipogon genomic clone OR_CBa0085L12
            3', genomic survey sequence.
```

```
ACCESSION   CL855754
VERSION     CL855754.1  GI:51264405
KEYWORDS    GSS.
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 816)
  AUTHORS   Kim,H., Yu,Y., Wissotski,M., Yost,D., Stum,D., Rao,K., Luo,M.,
            Jetty,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and
            Wing,R.
  TITLE     OMAP project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0085  row: L  column: 12
            Seq primer: CAC TCA TTA GGC ACC CCA
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .816
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
                     /clone="OR_CBa0085L12"
                     /tissue_type="young leaves"
                     /dev_stage="2 week old seedlings"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OR_CBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII;
                     drk treated 36 hrs before harvest"
ORIGIN

  Query Match             6.7%;  Score 311.4;  DB 9;  Length 816;
  Best Local Similarity   77.6%;  Pred. No. 7.9e-43;
  Matches  413;  Conservative    0;  Mismatches  116;  Indels    3;  Gaps    3;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||| |||||||||||||||||| | | | | | |||||||||| | |||||||
Db        770 TATATTTATATATATATATATATATATATATATATATATATATATATATATATATATATA 711

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | || |||| | ||||||||| | ||| |||| || | |
Db        710 TATATATATATATATATATATATAT-TATATATATATATATATATATATATATATATATA 652

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        651 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 592

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db        591 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 532

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db        531 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 472

Qy        526 TATACACATAT-ATACACATATATATACACATATATACACATATATACATATATACACAT 584
              |||| | |||| ||| | ||||||||| | ||||||| | ||||||| ||||||| | ||
Db        471 TATATATATATGATATATATATATATATATATATATATATATATATATATATATATATAT 412

Qy        585 ATATATACATATATACACATATATACATATATACACATATATACATATATACACACATAT 644
              ||||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | ||||
Db        411 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 352

Qy        645 ATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATAT 704
              ||| | ||| ||||| | | | ||| ||||| |||||||||| | ||||||||  |||||
Db        351 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 292

Qy        705 ATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCA 756
              |||| |||||||||||| |  | |  |  ||   | |   |  |   || ||
Db        291 ATATATATATATATATGATATATATATATTATATGATTATATATATATATCA 240


RESULT 46
CL858220/c
LOCUS       CL858220                546 bp    DNA     linear   GSS 16-AUG-2004
DEFINITION  OR_CBa0089A17.f OR_CBa Oryza rufipogon genomic clone OR_CBa0089A17
            5', genomic survey sequence.
ACCESSION   CL858220
VERSION     CL858220.1  GI:51269336
KEYWORDS    GSS.
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 546)
  AUTHORS   Kim,H., Yu,Y., Wissotski,M., Yost,D., Stum,D., Rao,K., Luo,M.,
            Jetty,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and
            Wing,R.
  TITLE     OMAP project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0089  row: A  column: 17
            Seq primer: TAA TAC GAC TCA CTA TAG GG
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .546
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
                     /clone="OR_CBa0089A17"
                     /tissue_type="young leaves"
                     /dev_stage="2 week old seedlings"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OR_CBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII;
                     drk treated 36 hrs before harvest"
ORIGIN

  Query Match             6.7%;  Score 310.6;  DB 9;  Length 546;
  Best Local Similarity   76.0%;  Pred. No. 1.1e-42;
  Matches  396;  Conservative    0;  Mismatches  124;  Indels    1;  Gaps    1;

Qy        242 ATATATATATACACACACACACATATATATATACACATATATATACACATATATACAT 301
              |||||||||| | | | |  | |||||           |||||| ||||||||| ||
Db        546 ATATATATATATATATATATCTATATATAATATATATATCATATATATACATATATAT 487

Qy        302 ATATACACATATATACACATATATATACACATACATATGTACACATATATATACACATAT 361
              ||||| | ||||||| | ||||||||| | |||  ||| || | |||||||||| | ||||
Db        486 ATATATATATATATATATATATATATATATATATCTATATATATATATATATATATAT 427

Qy        362 GTATACACATATATACACATATATACACACATATATACACATATATACACACATATATAC 421
              |||| | ||||||| | ||||||||| | ||||| | | ||||||  | | |||||||
Db        426 ATATATATATATATATATATATATACATATATATACATATATATATCTATATATATAT 367

Qy        422 ACATATATACACATATATACACATATACACATATATACACATATATACATATATACACAT 481
              | ||||||| | ||||||| | ||||| | ||||||| | ||||||| ||||||| | ||
```

```
Db        366 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 307

Qy        482 ATATATAATATACACACATATATATACACATATATACACACATATATACACATATATACA 541
              |||||| ||||| | | ||||||||| | ||||||| | | ||||||| | |||||||| |
Db        306 ATATAT-ATATATATATATATATATATATATATATATATATATATATATATATATATATA 248

Qy        542 CATATATATACACATATATACACATATATACATATATACACATATATATACATATATACA 601
              ||||||||| | ||||||| | ||||||| ||||||| | |||||||| ||||||| |
Db        247 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 188

Qy        602 CATATATACATATATACACATATATACATATATACACACATATATACACATACATATACA 661
              ||||||| ||||||| | ||||||| ||||||| | | ||||||| | ||| ||||| |
Db        187 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 128

Qy        662 CACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATATG 721
              | | ||| ||||| ||||||||| | |||||||||  |||||||| ||||||||||||
Db        127 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATG 68

Qy        722 CTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGG 762
                  |  ||  || |||  |  | |   |||||| ||||
Db         67 AGGTGAAACTCTAAAAAGGATGACGATCCATACCACAGGGG 27


RESULT 47
BX145980
LOCUS       BX145980                709 bp    DNA     linear   GSS 28-JAN-2003
DEFINITION  Danio rerio genomic clone DKEY-104E13, genomic survey sequence.
ACCESSION   BX145980
VERSION     BX145980.1  GI:27977335
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 709)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the T7 end of BAC 104E13. 104E13
            is part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .709
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-104E13"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             6.7%;  Score 308;  DB 9;  Length 709;
  Best Local Similarity   79.7%;  Pred. No. 3e-42;
  Matches  376;  Conservative    0;  Mismatches   95;  Indels    1;  Gaps    1;

Qy        250 ATACACACACACACACATATATATATACACATATATATACACATATATACATATATACAC 309
              | |  || |    |  |||||||||||| | |||||||| | |||||||| ||||||| |
Db        238 AAAGTCAAATTTATGCATATATATATATATATATATATATATATATATATATATATATAT 297

Qy        310 ATATATACACATATATATACACATACATATGTACACATATATATACACATATGTATACAC 369
              ||||||| | ||| | ||| | ||| |||| |||||||||||||| | |||| |||||||
Db        298 ATATATATATATACACATATATATATATATACACATATATATATATATATATACAC 357

Qy        370 ATATATACACATATATACACACATATATACACATATATACACACATATATACACATATAT 429
              | ||||| | ||||||| | | ||||||| | ||||||| | | ||||||| | ||||||
Db        358 ACATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 417

Qy        430 ACACATATATACACATATACACATATATACACATATATACATATATACACATATATATA- 488
              | | |||||| | | |||||| | ||||||| | ||| ||| ||| ||| | |||||||||
Db        418 ATATATATACATATATATATATATATATATATATATACATATATACATATATATATATAT 477

Qy        489 ATATACACACATATATATACACATATATACACACATATATACACATATATACACATATAT 548
              ||||||| |||||||||| | ||||||| | | |||||||| | ||||||| | ||||||
Db        478 ATATACATACATATATATATATATATATATATATATATATATATATATATATATATATAT 537

Qy        549 ATACACATATATACACATATATACATATATACACATATATATACATATATACACATATAT 608
              ||| | ||||||| | ||||||| ||||||| ||| ||||||| ||||||| ||||||||
Db        538 ATATATATATATATATATATATATATATATACACATATATATATATATACATATAT 597

Qy        609 ACATATATACACATATATACATATATACACACATATATACACATACATATACACACACAT 668
              | ||||||| | ||||||| ||||||| | | ||||||| | ||| ||||| | | |  |
Db        598 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATCT 657

Qy        669 AGATATACATATATATACACATATATATACGTATATATATGTATATATATAT 720
              | ||||| |||||||||| | |||||||||  |||||||| ||||||||||
Db        658 ATATATATATATATATATATATATATATATATATATATATATATATATATATAT 709


RESULT 48
CR177142
LOCUS       CR177142                648 bp    DNA     linear   GSS 06-JUL-2004
DEFINITION  Reverse strand read from insert in 5'HPRT insertion targeting and
            chromosome engineering clone MHPN105k12, genomic survey sequence.
ACCESSION   CR177142
VERSION     CR177142.1  GI:49955991
KEYWORDS    GSS; genome survey sequence; MICER.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 648)
  AUTHORS   Adams,D.J., Biggs,P.J., Cox,A.V., Davies,R.M., van der Weyden,L.,
            Jonkers,J., Smith,J., Plumb,R.W., Taylor,R.G., Nishijima,I., Yu,Y.,
            Rogers,J. and Bradley,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-FEB-2004) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. http://www.sanger.ac.uk/MICER
FEATURES             Location/Qualifiers
     source          1. .648
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="MHPN105k12"
                     /clone_lib="MHPN"
ORIGIN

  Query Match             6.6%;  Score 307.2;  DB 9;  Length 648;
  Best Local Similarity   75.4%;  Pred. No. 4.2e-42;
  Matches  395;  Conservative    0;  Mismatches  128;  Indels    1;  Gaps    1;

Qy        197 AAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACA 256
              |||   ||    | |||||  |||   |||| |||||||||||| | ||||||| | | |
Db        107 AAACTACAAGCAATAAAAATCAAATAACATTATATATATATATACACATATATACATATA 166

Qy        257 CACACACACATATATATATACACATATATATACACATATATACATATATACACATATATA 316
              ||| | | | ||||| | | | | | ||||| ||||||| ||||||| ||| ||||| |
Db        167 TACATATATACATATACACATATACACATATATACATATACACATATACACATATATACA 226

Qy        317 CACATATATATACACATACATATGTACACATATATATACACATATGTATACACATATATA 376
              | | | |||||||||||| | |  || ||| ||||| ||| |||   ||| ||| |||||
Db        227 TATACACATATACACATATACACATATACACATATACACATATACACATATACACATATA 286

Qy        377 CACATATATACACACATATATACACATATATACACACATATATACACATATATACACATA 436
              |||||||| ||| | | | ||| ||| ||||||||| ||| | | | | | ||| ||| |
Db        287 CACATATACACATATACACATATACACATATACACATATACACATATACACATATACACA 346

Qy        437 TATACACATATACACATATATACACATATATACATATATACACATATATATAATATACAC 496
              |||||||||||||||||||| ||| ||| | | ||| | | | | | |||| | | | |
Db        347 TATACACATATACACATATACACATATACACATATACACATATACACATAT-ACACATAT 405

Qy        497 ACATATATATACACATATATACACACATATATACACATATATACACATATATATACACAT 556
              ||| ||||| ||| ||| | | | ||| |||||||||||| ||| ||||| ||| | |
```

```
Db        406 ACACATATACACATATACACATATACACATATACACATATACACATATATACATATATAC 465

Qy        557 ATATACACATATATACATATATACACATATATATACATATATACACATATATACATATAT 616
              ||||| |||||||||||||||||||||| | | | ||| ||||| | ||| ||| ||| ||
Db        466 ATATATACATATATACATATATACACACACACACACACATATATATATACATATATACAT 525

Qy        617 ACACATATATACATATATACACACATATATACACATACATATACACACACATAGATATAC 676
              | | |||||||||| | | ||||||||||||||| | | | ||||||| | | ||| ||||||
Db        526 ATATATATATACACACACACACACATATATATATACATATATACATATATATATATATAC 585

Qy        677 ATATATATACACATATATATACGTATATATATGTATATATATAT 720
              | | | | ||||||||||||||  ||||||||| ||||||||||
Db        586 ACACACACACACATATATATATATATATATATATATATATAT 629


RESULT 49
CL626353/c
LOCUS       CL626353                581 bp    DNA     linear   GSS 01-JUL-2004
DEFINITION  OR_BBa0022K14.f OR_BBa Oryza rufipogon genomic clone OR_BBa0022K14
            5', genomic survey sequence.
ACCESSION   CL626353
VERSION     CL626353.1  GI:49128388
KEYWORDS    GSS.
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 581)
  AUTHORS   Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R.,
            Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
  TITLE     OMAP Project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Insert Length: 161   Std Error: 0.00
            Plate: 0022  row: K  column: 14
            Seq primer: TAA TAC GAC TCA CTA TAG GG
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .581
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
                     /clone="OR_BBa0022K14"
                     /tissue_type="young leaves"
                     /lab_host="DH10B-T1 phage resistant"
                     /clone_lib="OR_BBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

  Query Match             6.6%;  Score 306.4;  DB 9;  Length 581;
  Best Local Similarity   77.4%;  Pred. No. 5.8e-42;
  Matches  384;  Conservative    0;  Mismatches  111;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | || |||||||||||| ||| | ||| |||   || ||| ||||| ||| ||| ||| |||
Db        574 TATACATATATATATACATACACATATACATTCATACATATATACATACACATATACATA 515

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              | |||||||||||||| ||||| | ||| | |||||||||||| |||||||||   |||| |
Db        514 CATACATATATACATACATACATACATACATACATATATATATACATACATACATACATA 455

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              ||||||| | | ||  ||||| | ||||| | ||||||||| ||||| ||||| |||||
Db        454 CATATATATATACATACATACATACATATATATATATATACATACATACATACATATATA 395

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | | ||| | ||||| ||| | ||||||| ||||| | ||||| ||| | |||||
Db        394 TATATATACATACATACATACATATATATATATATACATACATACATACATATATATATA 335

Qy        467 TACATATATACACATATATATAATATACACACATATATATACACATATATACACACATAT 526
              |||||| ||||| | |||||| ||||| | ||||| ||| | |||||||||| | | ||||
Db        334 TACATACATACATACATATAT-ATATATATACATACATACATACATATATATATATATAT 276

Qy        527 ATACACATATATACACATATATATACACATATATACACATATATACATATATACACATAT 586
              | | ||||| ||| | ||||||||| ||||| ||||| ||||||| ||||||| | || |
Db        275 ACATACATACATATATATATATATACATACATACATATATATATATATATATATGT 216

Qy        587 ATATACATATATACACATATATACATATATACACATATATACATATATACACACATATAT 646
              ||||| ||||||  | ||||| | | ||| | | || |||| ||||||| |   ||||||
Db        215 ATATATATATATGTATATATACATACATACATATGTATATATATATATATGTATATAT 156

Qy        647 ACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATAT 706
              | | ||| ||||| | | ||| ||||| |||| |||| | ||||||||| |||||||
Db        155 ATATATATATATATGTATATATATATATATATATGTATATATATATATATATATATAT 96

Qy        707 ATGTATATATATATGC 722
              || |||||||  | ||
Db         95 ATATATATATGGACGC 80


RESULT 50
CA255084
LOCUS       CA255084                613 bp    mRNA    linear   EST 26-SEP-2003
DEFINITION  SCEPFL4175C11.g FL4 Saccharum officinarum cDNA clone SCEPFL4175C11
            5', mRNA sequence.
ACCESSION   CA255084
VERSION     CA255084.1  GI:35939612
KEYWORDS    EST.
SOURCE      Saccharum officinarum
  ORGANISM  Saccharum officinarum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Saccharum; Saccharum officinarum
            complex.
REFERENCE   1  (bases 1 to 613)
  AUTHORS   Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.
  TITLE     The libraries that made SUCEST
  JOURNAL   Genet. Mol. Biol. 24 (1-4), 1-7 (2001)
COMMENT     Contact: Arruda P
            Centro de Biologia Molecular e Engenharia Genetica
            Universidade Estadual de Campinas
            Caixa Postal 6010, 13083-970, Campinas SP, Brazil
            Tel: 55 19 3788 1137
            Fax: 55 19 3788 1089
            Email: parruda@unicamp.br
            Clone distribution: clone distribution information can be found
            through the Brazilian Clone Collection Center (BCCC) at
            http://www.bcccenter.fcav.unesp.br
            Plate: 175  row: C  column: 11
            Seq primer: T7 Promoter Primer.
FEATURES             Location/Qualifiers
     source          1. .613
                     /organism="Saccharum officinarum"
                     /mol_type="mRNA"
                     /db_xref="taxon:4547"
                     /clone="SCEPFL4175C11"
                     /lab_host="DH10B"
                     /clone_lib="FL4"
                     /note="Organ: Developed inflorescence and rachis
                     (20cm-long); Vector: pSport1; Site_1: SalI; Site_2: NotI;
                     An unidirectional cDNA library generated from [Developed
                     inflorescence  and rachis (20cm-long)]. cDNA was prepared
                     from polyA+ mRNA using  SuperScript Plasmid System Kit
                     (Invitrogen). The double-strand cDNAs were  fractionated
                     in a sepharose CL-2B 40cm-columns and fragments sizing
                     between 0.8 and 1.5 Kb were directionally cloned into the
```

vector. Details of each source of RNA and library construction can be obtained at http://sucest.lad.ic.unicamp.br/public"

ORIGIN

Query Match 6.6%; Score 306; DB 6; Length 613;
Best Local Similarity 76.9%; Pred. No. 6.8e-42;
Matches 386; Conservative 0; Mismatches 115; Indels 1; Gaps 1;

```
Qy        220 AAGGCATTTTATATATATATATATATATATACACACACACACACATATATATATACAC 279
              | |||  | | |||  |||| | ||| ||| |  |     | | | ||||||| ||  |
Db        109 ATGGCTCTCTGTATGCATATGTCTATCTATGTGGAGCGCTAGATAGATATATACATCTAT 168

Qy        280 ATATATATACACATATATACATATATACACATATATACACATATATATACACATACATAT 339
              |||||||||  | ||||||  ||||||| | ||||||| | |||||||||  | ||| || |
Db        169 ATATATATATATATATATCGATATATATATATATATATATATATATATATATATATGT 228

Qy        340 GTACACATATATATACACATATGTATACACATATATACACATATATACACACATATATAC 399
              || | |||||||||  | |||| |||| | ||||||| | |||||||  | | |||||||
Db        229 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 288

Qy        400 ACATATATACACACATATATACACATATATACACATATATACACATATACACATATATAC 459
              | ||||||| | | ||||||| | ||||||| | ||||||| | ||||| | |||||||
Db        289 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 348

Qy        460 ACATATATACATATATACACATATATATAATATACACACATATATATACACATATATACA 519
              | ||||||| ||||||| | |||||||| ||||| | | ||||||||| | |||||||  |
Db        349 ATATATATATATATATATATATATATAT-ATATATATATATATATATATATATATATATA 407

Qy        520 CACATATATACACATATATACACATATATATACACATATATACACATATATACATATATA 579
              | ||||||| | ||||||| | ||||||||| | ||||||| | ||||||| |||||||
Db        408 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 467

Qy        580 CACATATATATACATATATACACATATATACATATATACACATATATACATATATACACA 639
              | |||||||||  ||||||| | ||||||| ||||||| | ||||||| ||||||| | |
Db        468 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 527

Qy        640 CATATATACACATACATATACACACACATAGATATACATATATATACACATATATATACG 699
              |||||||  | |||  |||||  | | | |||  ||||| |||||||||  | |||||  |||
Db        528 TATATATATATATATATATATATATATATATATATATATATATATATATATATATAGATATA 587

Qy        700 TATATATATGTATATATATATG 721
              |||||  |||  |||||||  |||
Db        588 TATATCTATATATATATCGATG 609
```

RESULT 51
CA255084/c
LOCUS CA255084 613 bp mRNA linear EST 26-SEP-2003
DEFINITION SCEPFL4175C11.g FL4 Saccharum officinarum cDNA clone SCEPFL4175C11 5', mRNA sequence.
ACCESSION CA255084
VERSION CA255084.1 GI:35939612
KEYWORDS EST.
SOURCE Saccharum officinarum
ORGANISM Saccharum officinarum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Saccharum; Saccharum officinarum complex.
REFERENCE 1 (bases 1 to 613)
AUTHORS Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.
TITLE The libraries that made SUCEST
JOURNAL Genet. Mol. Biol. 24 (1-4), 1-7 (2001)
COMMENT Contact: Arruda P
Centro de Biologia Molecular e Engenharia Genetica
Universidade Estadual de Campinas
Caixa Postal 6010, 13083-970, Campinas SP, Brazil
Tel: 55 19 3788 1137
Fax: 55 19 3788 1089
Email: parruda@unicamp.br
Clone distribution: clone distribution information can be found through the Brazilian Clone Collection Center (BCCC) at http://www.bcccenter.fcav.unesp.br
Plate: 175 row: C column: 11
Seq primer: T7 Promoter Primer.
FEATURES Location/Qualifiers
source 1. .613
/organism="Saccharum officinarum"
/mol_type="mRNA"
/db_xref="taxon:4547"
/clone="SCEPFL4175C11"
/lab_host="DH10B"
/clone_lib="FL4"
/note="Organ: Developed inflorescence and rachis (20cm-long); Vector: pSport1; Site_1: SalI; Site_2: NotI; An unidirectional cDNA library generated from [Developed inflorescence and rachis (20cm-long)]. cDNA was prepared from polyA+ mRNA using SuperScript Plasmid System Kit (Invitrogen). The double-strand cDNAs were fractionated in a sepharose CL-2B 40cm-columns and fragments sizing between 0.8 and 1.5 Kb were directionally cloned into the vector. Details of each source of RNA and library construction can be obtained at http://sucest.lad.ic.unicamp.br/public"

ORIGIN

Query Match 6.6%; Score 304.4; DB 6; Length 613;
Best Local Similarity 77.3%; Pred. No. 1.3e-41;
Matches 382; Conservative 0; Mismatches 111; Indels 1; Gaps 1;

```
Qy        227 TTTATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | |||||| |||||||||| |||||| | | | | | ||||||||||| | |||||||
Db        601 TATATATAGATATATATATCTATATATATATATATATATATATATATATATATATATA 542

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | |||||||||  | ||| |||| || | |
Db        541 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 482

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        481 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 422

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db        421 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 362

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db        361 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 302

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db        301 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 242

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||||| ||||||| ||||||| | ||||||| ||||||| |    |||||
Db        241 TATATATATATATACATATATATATATATATATATATATATATATATATATATCGATATA 182

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | |   ||| || ||  ||     | |||||| ||| || ||| |
Db        181 TATATATATATATATAGATGTATATATCTATCTAGCGCTCCACATAGATAGACATATGCA 122

Qy        706 TATGTATATATATA 719
              ||   | |   |||
Db        121 TACAGAGAGCCATA 108
```

RESULT 52
CL830207
LOCUS CL830207 491 bp DNA linear GSS 09-AUG-2004
DEFINITION OR_CBa0052C06.f OR_CBa Oryza rufipogon genomic clone OR_CBa0052C06 5', genomic survey sequence.
ACCESSION CL830207
VERSION CL830207.1 GI:51075808
KEYWORDS GSS.

```
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 491)
  AUTHORS   Kim,H., Yu,Y., Wissotski,M., Yost,D., Stum,D., Rao,K., Luo,M.,
            Jetty,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and
            Wing,R.
  TITLE     OMAP project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0052  row: C  column: 06
            Seq primer: TAA TAC GAC TCA CTA TAG GG
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .491
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
                     /clone="OR_CBa0052C06"
                     /tissue_type="young leaves"
                     /dev_stage="2 week old seedlings"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OR_CBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII;
                     drk treated 36 hrs before harvest"
ORIGIN

  Query Match             6.5%;  Score 302;  DB 9;  Length 491;
  Best Local Similarity   77.3%;  Pred. No. 3.3e-41;
  Matches  379;  Conservative    0;  Mismatches  110;  Indels    1;  Gaps    1;

Qy        230 ATATATATATATATATATATATATACACACACACACACATATATATATACACATATATATAC 289
              ||||||||||||||||||||||||| | | | | | |||||||||||| | |||||||||
Db          2 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 61

Qy        290 ACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATAT 349
              | ||||||| ||||||| | ||||||    |||||| || | ||  |||| || | ||||
Db         62 ATATATATATATATATATATATATATATGTGTATATATGTATATATGTATATATATATATAT 121

Qy        350 ATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATAC 409
              ||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||||
Db        122 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 181

Qy        410 ACACATATATACACATATATACACATATATACACATATACACATATATACACATATATAC 469
              | | ||||||| | ||||||| | || |||| | ||||| | ||||||| | |||||||
Db        182 ATATATATATATATATATATATATATATCTATATATATATATATATATATATATATATAT 241

Qy        470 ATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATATAT 528
              ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | ||||||
Db        242 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 301

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACACATATAT 588
                  ||||||| | |||||| || | ||||||| | ||||||| ||||||| | ||||||
Db        302 CTCTATATATATATATATATCTATATATATATATAGATATATATATATATATATATATAT 361

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
              |||  |||||| | || ||||| |||||| | | ||| | || |||    | |||||||
Db        362 ATATGTATATATATATGTATACCTATATATATACATACATATCTATCTGTATATATATAT 421

Qy        649 ACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATAT 708
              | ||| ||||    | |||||||| || ||||||||| | |||||||||  |||||||||
Db        422 ATATATATATGTGTATACATAGATCTATATATATATATATATATATATATATATATATAT 481

Qy        709 GTATATATAT 718
                 ||| |||||
Db        482 ATATGTATAT 491


RESULT 53
CL858220
LOCUS       CL858220                546 bp    DNA     linear   GSS 16-AUG-2004
DEFINITION  OR_CBa0089A17.f OR_CBa Oryza rufipogon genomic clone OR_CBa0089A17
            5', genomic survey sequence.
ACCESSION   CL858220
VERSION     CL858220.1  GI:51269336
KEYWORDS    GSS.
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 546)
  AUTHORS   Kim,H., Yu,Y., Wissotski,M., Yost,D., Stum,D., Rao,K., Luo,M.,
            Jetty,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and
            Wing,R.
  TITLE     OMAP project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0089  row: A  column: 17
            Seq primer: TAA TAC GAC TCA CTA TAG GG
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .546
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
                     /clone="OR_CBa0089A17"
                     /tissue_type="young leaves"
                     /dev_stage="2 week old seedlings"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OR_CBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII;
                     drk treated 36 hrs before harvest"
ORIGIN

  Query Match             6.5%;  Score 300.2;  DB 9;  Length 546;
  Best Local Similarity   77.5%;  Pred. No. 6.7e-41;
  Matches  376;  Conservative    0;  Mismatches  108;  Indels    1;  Gaps    1;

Qy        224 CATTTTATATATATATATATATATATATACACACACACACACATATATATATACACATAT 283
              ||  | |||||||||||||||||||||||| | | | | | ||||||||||| | ||||
Db         63 CACCTCATATATATATATATATATATATATATATATATATATATATATATATATATATAT 122

Qy        284 ATATACACATATATACATATATACACATATATACACATATATATACACATACATATGTAC 343
              ||||| | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| ||
Db        123 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 182

Qy        344 ACATATATATACACATATGTATACACATATATACACATATATACACACATATATACACAT 403
              | ||||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | ||
Db        183 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 242

Qy        404 ATATACACACATATATACACATATATACACATATATACACATATACACATATATACACAT 463
              ||||| | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | ||
Db        243 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 302

Qy        464 ATATACATATATACACATATATATAATATACACACATATATATACACATATATACACACA 523
```

```
Db        303 ATATATATATATATATATATATAT-ATATATATATATATATATATATATATATATATA 361

Qy        524 TATATACACATATATACACATATATATACACATATATACACATATATACATATATACACA 583
              |||||| | ||||||| | |||||||||    |||||||   ||||||| ||||||| | |
Db        362 TATATATATATATATATAGATATATATATGTATATATATGTATATATATATATATATATA 421

Qy        584 TATATATACATATATACACATATATACATATATACACATATATACATATATACACACATA 643
              |||||||| ||||||| | ||||||| ||| ||| | ||||||| ||||||| | | |||
Db        422 TATATATATATATATATATATATATATATAGATATATATATATATATATATATATATATA 481

Qy        644 TATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATA 703
              |||| | ||| ||||  | | | ||           |||||| | |||||||||  ||||
Db        482 TATATATATATATATGTATATATATGATATATATATTATATATAGATATATATATATATA 541

Qy        704 TATAT 708
              |||||
Db        542 TATAT 546


RESULT 54
BX180548/c
LOCUS       BX180548                592 bp    DNA     linear   GSS 28-JAN-2003
DEFINITION  Danio rerio genomic clone DKEY-180C22, genomic survey sequence.
ACCESSION   BX180548
VERSION     BX180548.1  GI:28012355
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 592)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the T7 end of BAC 180C22. 180C22
            is part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .592
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-180C22"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             6.5%;  Score 298.6;  DB 9;  Length 592;
  Best Local Similarity   78.0%;  Pred. No. 1.2e-40;
  Matches  372;  Conservative    0;  Mismatches  104;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | |||||||||||||||||||||||| | | | | | | |||||||||| | |||||||
Db        590 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 531

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db        530 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 471

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        470 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 411

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | |||||  | ||||||| | |||||
Db        410 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 351

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | |||||  | | ||||||| | |||||||
Db        350 TATATATATATATATATATATATATATATATACATACACACACATACACACACACACA 291

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              | ||||||| | ||||| | | | |||||||||| ||||| ||||||| ||||| |||
Db        290 CACACACATACACACACACACACACACACATATATATACATACATACATACATACATATA 231

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db        230 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 171

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTAT 702
              || | ||| ||||| | | | ||| |||||| | | |        || | | |  |||
Db        170 TATATATATATATATATATATATATATATATACTTTTTTTCTGTGCTAAACAAATTTAT 114


RESULT 55
BX134913
LOCUS       BX134913                696 bp    DNA     linear   GSS 28-JAN-2003
DEFINITION  Danio rerio genomic clone DKEY-89F23, genomic survey sequence.
ACCESSION   BX134913
VERSION     BX134913.1  GI:27966206
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 696)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the T7 end of BAC 89F23. 89F23 is
            part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .696
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-89F23"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             6.4%;  Score 297.4;  DB 9;  Length 696;
  Best Local Similarity   79.8%;  Pred. No. 1.9e-40;
  Matches  363;  Conservative    0;  Mismatches   91;  Indels    1;  Gaps    1;

Qy        231 TATATATATATATATATATATACACACACACACATATATATATACACATATATATACA 290
              ||||||| |||||||||||||| | | | | | |||||||||| | |||||||| |
Db        243 TATATATTTATATATATATATATATATATATATATATATATATATATATATATATA 302

Qy        291 CATATATACATATATACACATATATACACATATATATACACATACATATGTACACATATA 350
              ||||||| ||||||| | |||||||     |||||| | | ||| |||| || | |||||
Db        303 TATATATATATATATATATATATATAGTAGTATATAGATAGATATATATATATATATATA 362

Qy        351 TATACACATATGTATACACATATATACACATATATACACACATATATACACATATATACA 410
              |||| | ||||||||| | ||||||| | ||||||| | | ||||||| | ||||||| |
Db        363 TATATATATATGTATATATATATATATATATATATATATATATATATATATATATATA 422

Qy        411 CACATATATACACATATATACACATATATACACATATACACATATATACACATATATACA 470
              | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | ||||||| |
Db        423 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 482

Qy        471 TATATACACATATATATAATATACACACATATATATACACATATATACACACATATATAC 530
              |||||| | |||||||| ||||| | | ||||||||| | ||||||| | | ||||||||
Db        483 TATATATATATATATAT-ATATATATATATATATATATATATATATATATATATATAT 541

Qy        531 ACATATATACACATATATATACACATATATACACATATATACATATATACACATATATAT 590
```

```
Db        542 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 601

Qy        591 ACATATATACACATATATACATATATACACATATATACATATATACACACATATATACAC 650
Db        602 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 661

Qy        651 ATACATATACACACACATAGATATACATATATATA 685
Db        662 ATATATATATATATATATATATATATATATATATA 696


RESULT 56
BX163144
LOCUS       BX163144                700 bp    DNA     linear   GSS 13-MAR-2003
DEFINITION  Danio rerio genomic clone DKEY-154G8, genomic survey sequence.
ACCESSION   BX163144
VERSION     BX163144.1  GI:27994657
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 700)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAR-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the T7 end of BAC 154G8. 154G8 is
            part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .700
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-154G8"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             6.4%;  Score 297.4;  DB 9;  Length 700;
  Best Local Similarity   79.0%;  Pred. No. 1.9e-40;
  Matches  391;  Conservative    0;  Mismatches  101;  Indels    3;  Gaps    3;

Qy        227 TTTATATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
Db        198 TATATATATATATATATATATATATATACACACACACACACAT-TATATATATATATATATA 256

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
Db        257 TATATATATATATATATATATATATACACACATACATACATACACACATACATACATACATA 316

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
Db        317 CATACACACACACACACACACACACACACATACATACATACATACATACATACATACATACA 376

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
Db        377 TACATACATACATACATACATACATACATACATACATACATACATACATACATACATACA 436

Qy        467 TACATATATACACATATATATAATATACACACATATATATACACATATATACACACATAT 526
Db        437 TACATACATACATACATACAT-ACATACATACATACATATATATATATATATATATATATAT 495

Qy        527 ATACACATATATACACATATATATACACATATATACACATATATACATATATACACATAT 586
Db        496 ATAATATCATATATATACATATATACATACATACATACATATATATATATATATATATAT 555

Qy        587 ATATACATATATACACATATATACATATATACACATATATACATATATACACACATATAT 646

Db        556 ATATATATATATATATATATATATATATATACACACACACACACATACATACATACATACAT 615

Qy        647 ACACATACATATACACAC-ACATAGATATACATATATATACACATATATATACGTATATA 705
Db        616 ACACACACACACACACACGTGTTATATATATATATATATATATATATATATATATATATA 675

Qy        706 TATGTATATATATAT 720
Db        676 TATATATATATATAT 690


RESULT 57
BX180768/c
LOCUS       BX180768                701 bp    DNA     linear   GSS 28-JAN-2003
DEFINITION  Danio rerio genomic clone DKEY-176A10, genomic survey sequence.
ACCESSION   BX180768
VERSION     BX180768.1  GI:28012577
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 701)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the SP6 end of BAC 176A10. 176A10
            is part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .701
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-176A10"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             6.4%;  Score 297;  DB 9;  Length 701;
  Best Local Similarity   70.1%;  Pred. No. 2.3e-40;
  Matches  413;  Conservative    0;  Mismatches  175;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
Db        701 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 642

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
Db        641 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 582

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
Db        581 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 522

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
Db        521 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 462

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
Db        461 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 402

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
Db        401 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 342

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
```

```
Db        341 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 282

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705

Db        281 TATATATATATATATTCTCGAAAACCTCTGATCAAACTCTTACCTTCATTTAAAACATTT 222

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765

Db        221 TCATTTAAGCCATACACAGCTTAACATAAATGCGCCATCATTAAGTCATCGCCAAGGAAA 162

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGG 814

Db        161 GTAAATAAATACATATCAGCGCATACCTTTCTAGAAAACTGAGTTCAGG 113


RESULT 58
CL763005
LOCUS       CL763005                617 bp    DNA     linear   GSS 27-JUL-2004
DEFINITION  OR_BBa0133F16.f OR_BBa Oryza rufipogon genomic clone OR_BBa0133F16
            5', genomic survey sequence.
ACCESSION   CL763005
VERSION     CL763005.1  GI:50720952
KEYWORDS    GSS.
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 617)
  AUTHORS   Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R.,
            Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
  TITLE     OMAP Project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Insert Length: 161   Std Error: 0.00
            Plate: 0133  row: F  column: 16
            Seq primer: TAA TAC GAC TCA CTA TAG GG
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .617
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
                     /clone="OR_BBa0133F16"
                     /tissue_type="young leaves"
                     /lab_host="DH10B-T1 phage resistant"
                     /clone_lib="OR_BBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

  Query Match             6.4%;  Score 295.4;  DB 9;  Length 617;
  Best Local Similarity   76.3%;  Pred. No. 4.3e-40;
  Matches  376;  Conservative    0;  Mismatches  116;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286

Db         85 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 144

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346

Db        145 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 204

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406

Db        205 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 264

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466

Db        265 TATATATATATATATATATATATATATATATATATATATATATATATATATATAAATATATA 324

Qy        467 TACATATATACACATATATATAATATACACACATATATATACACATATATACACACATAT 526

Db        325 TAAATATATATATATATATATATTATATA-ATATATATATTATTTTATTTTTTTTAATAT 383

Qy        527 ATACACATATATACACATATATATACACATATATACACATATATACATATATACACATAT 586

Db        384 ATATATATATATATTTATATATATATATATATATATATATATATAAAAAAATATATTTTT 443

Qy        587 ATATACATATATACACATATATACATATATACACATATATACATATATACACACATATAT 646

Db        444 ATATATATAGATACATATATATAAAAATATAGAGAGAGATATATATAAATATAAAAAAAG 503

Qy        647 ACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATAT 706

Db        504 ATATATATATATATATATGCAGATACATAGATATATATACATATAGATATATATATAGTT 563

Qy        707 ATGTATATATATA 719

Db        564 ATGTATATATATA 576


RESULT 59
BX134913/c
LOCUS       BX134913                696 bp    DNA     linear   GSS 28-JAN-2003
DEFINITION  Danio rerio genomic clone DKEY-89F23, genomic survey sequence.
ACCESSION   BX134913
VERSION     BX134913.1  GI:27966206
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 696)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the T7 end of BAC 89F23. 89F23 is
            part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .696
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-89F23"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             6.4%;  Score 294.4;  DB 9;  Length 696;
  Best Local Similarity   79.6%;  Pred. No. 6.3e-40;
  Matches  360;  Conservative    0;  Mismatches   91;  Indels    1;  Gaps    1;

Qy        269 TATATATACACATATATATACACATATATACATATATACACATATATACACATATATATA 328

Db        696 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 637

Qy        329 CACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACA 388

Db        636 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 577

Qy        389 CACATATATACACATATATACACACATATATACACATATATACACATATATACACATATA 448
```

```
Db        576 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 517

Qy        449 CACATATATACACATATATACATATATACACATATATATAATATACACACATATATATAC 508
              | ||||||| | ||||||| | |||||||| ||||| | | |||||||||
Db        516 TATATATATATATATATATATATATATATATATATATAT-ATATATATATATATATATAT 458

Qy        509 ACATATATACACACATATATACACATATATACACATATATATACACATATATACACATAT 568
              | ||||||| | | ||||||| | ||||||| | |||||||| | ||||||| | ||||
Db        457 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 398

Qy        569 ATACATATATACACATATATATACATATATACACATATATACATATATACACATATATAC 628
              ||| ||||||| | |||||||||||||||||| | ||||||| ||||||| | || |||
Db        397 ATATATATATATATATATATATATATACATATATATATATATATATATATATATATCTATCT 338

Qy        629 ATATATACACACATATATACACATACATATACACACACATAGATATACATATATATACAC 688
              |||||      | ||||||| | ||| ||||| | | | ||| ||||| |||||||||| |
Db        337 ATATACTACTATATATATATATATATATATATATATATATATATATATATATATATATATAT 278

Qy        689 ATATATATACGTATATATATGTATATATATAT 720
              |||||||||  |||||||||| |||||| ||||
Db        277 ATATATATATATATATATATATATATAAATAT 246


RESULT 60
CF924712/c
LOCUS       CF924712                760 bp    mRNA    linear   EST 10-NOV-2003
DEFINITION  DD005E11.ab1 Codonopsis lanceolata cDNA library Codonopsis
            lanceolata cDNA, mRNA sequence.
ACCESSION   CF924712
VERSION     CF924712.1  GI:38243228
KEYWORDS    EST.
SOURCE      Codonopsis lanceolata
  ORGANISM  Codonopsis lanceolata
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            asterids; campanulids; Asterales; Campanulaceae; Codonopsis.
REFERENCE   1  (bases 1 to 760)
  AUTHORS   Yang,D.C., In,J.G. and Lee,B.S.
  TITLE     Codonopsis lanceolata
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Deok Chun Yang
            BioPia
            1 Seochonri, Kihung, Yongin, Kyungki-do, 449-701, Korea
            Tel: 82 31 201 2688
            Fax: 82 31 205 6323
            Email: dcyang@ibiopia.com.
FEATURES             Location/Qualifiers
     source          1. .760
                     /organism="Codonopsis lanceolata"
                     /mol_type="mRNA"
                     /db_xref="taxon:103999"
                     /clone_lib="Codonopsis lanceolata cDNA library"
                     /note="Organ: root; Vector: pTriplEx2; Site_1: Sfi IA;
                     Site_2: Sfi IB"
ORIGIN

  Query Match             6.4%;  Score 294.2;  DB 7;  Length 760;
  Best Local Similarity   79.8%;  Pred. No. 6.7e-40;
  Matches  347;  Conservative    0;  Mismatches   88;  Indels    0;  Gaps    0;

Qy        227 TTTATATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | | |||||||| |||||||||||||| | | | | | |||||||||||| | |||||||
Db        748 TATGTATATATAGATATATATATATATATAGATATATATATATATATATATATATATATA 689

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | |||| || | ||||||||| | ||| |||| || | |
Db        688 TATATATATATATATATATATATATATATCTATATATATATATATATATATATATATATATA 629

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        628 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 569

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db        568 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 509

Qy        467 TACATATATACACATATATATAATATACACACATATATATACACATATATACACACATAT 526
              || ||||||| | |||||||| ||||| | | ||||||||| | ||||||| | | ||||
Db        508 TATATATATATATATATATATCATATATATATATATATATATATATATATATATATATAT 449

Qy        527 ATACACATATATACACATATATATACACATATATACACATATATACATATATACACATAT 586
              ||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | ||||
Db        448 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 389

Qy        587 ATATACATATATACACATATATACATATATACACATATATACATATATACACACATATAT 646
              ||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| |  |||||||
Db        388 ATATATATATATATATATATATATATATATATATATATATATATATATATATATGCATATAT 329

Qy        647 ACACATACATATACA 661
              | | ||| ||||| |
Db        328 ATATATATATATATA 314


RESULT 61
BX197183
LOCUS       BX197183                771 bp    DNA     linear   GSS 28-JAN-2003
DEFINITION  Danio rerio genomic clone DKEY-207A10, genomic survey sequence.
ACCESSION   BX197183
VERSION     BX197183.1  GI:28029069
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 771)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the SP6 end of BAC 207A10. 207A10
            is part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .771
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-207A10"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             6.4%;  Score 293.6;  DB 9;  Length 771;
  Best Local Similarity   77.1%;  Pred. No. 8.5e-40;
  Matches  370;  Conservative    0;  Mismatches  109;  Indels    1;  Gaps    1;

Qy        168 ATTGCACTCCAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAGGCATT 227
              ||| | || |  ||  |  || | | ||  |  | |  |||||||||      |
Db        292 ATTCAATTCAAAAGCATAAATAAATTGGGGAAAAAAAAATAAATAAAAAAAAAAAAATAT 351

Qy        228 TTATATATATATATATATATATATACACACACACACATATATATATACACATATATAT 287
              |||||||||||||||||||||||| | | | | | |||||||||| | ||||||||
Db        352 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 411

Qy        288 ACACATATATACACATATATACACATATATACACATATATATACACATACATATGTACACAT 347
              | | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | ||
Db        412 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 471

Qy        348 ATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATAT 407
              ||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | ||||||
Db        472 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 531
```

```
Qy        408 ACACACATATATACACATATATACACATATATACACATATACACATATATACACATATAT 467
              | | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | ||||||
Db        532 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 591

Qy        468 ACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATAT 526
              | ||||||| | |||||||||| ||||| | | ||||||||| | ||||||| | | ||||
Db        592 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 651

Qy        527 ATACACATATATACACATATATATACACATATATACACATATATACATATATACACACATAT 586
              ||| | ||||||| | |||||||||| | ||||||| | ||||||| ||||||| | ||||
Db        652 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 711

Qy        587 ATATACATATATACACATATATACATATATACACATATATACATATATACACACATATAT 646
              ||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | ||||||||
Db        712 ATATATATATATATATATATATATATATATATATATATATATATATATATATATACATATAT 771


RESULT 62
BZ215061
LOCUS       BZ215061                643 bp    DNA     linear   GSS 11-OCT-2002
DEFINITION  CH230-280D4.TV CHORI-230 Segment 2 Rattus norvegicus genomic clone
            CH230-280D4, genomic survey sequence.
ACCESSION   BZ215061
VERSION     BZ215061.1  GI:23873314
KEYWORDS    GSS.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 643)
  AUTHORS   Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K.,
            Shvartsbeyn,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D.,
            Riggs,F., de Jong,P. and Fraser,C.M.
  TITLE     Rat BAC End Sequences from Library CHORI-230 MboI segment
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: CH230-280D4.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the rat BAC library CHORI-230
            (http://www.chori.org/bacpac/rat230.htm). For BAC library
            availability, please contact Pieter de Jong (pdejong@mail.cho.org).
            Clones may be purchased from BACPAC Resources
            (http://www.chori.org/bacpac/or ering_information.htm). BAC end
            page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
            Plate: 280  row: D  column: 4
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .643
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /strain="BN/SsNHsd/MCW"
                     /db_xref="taxon:10116"
                     /clone="CH230-280D4"
                     /sex="Female"
                     /cell_type="Brain"
                     /clone_lib="CHORI-230 Segment 2"
                     /note="Vector: pTARBAC1.3; Site_1: MboI; Site_2: MboI;
                     CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
                     Pieter de Jong"
ORIGIN

  Query Match             6.3%;  Score 293;  DB 8;  Length 643;
  Best Local Similarity   75.7%;  Pred. No. 1.1e-39;
  Matches  376;  Conservative    0;  Mismatches  120;  Indels    1;  Gaps    1;

Qy        229 TATATATATATATATATATATATACACACACACACATATATATATACACATATATATA 288
              || | | | | ||| | | |||||||||| |||||||||| ||||||||| ||| | |
Db        118 TACAAACACACATACACACATATACACACATACACACATATACATATACACACATACACA 177

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              |||||||| ||||| ||||||| ||||| ||| | |||||||||||||| || ||||   |
Db        178 CACATATACACATACATACACACATATAAACACACATATACACATACACATATACATGCA 237

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              | ||||||| || || ||||| ||| ||    ||||||||||||||| ||| | ||| |
Db        238 CACATACACACATATACACACACATACACTTCCATATACACACATATACACAGACATACA 297

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              |||| ||| | | ||||| | ||| ||| | | ||||| ||||||||| ||||| ||| |
Db        298 CACATATACACAGACATACACACATATACACAGACATACACACATATAGACACACATACA 357

Qy        469 CATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATATA 527
              | ||  || ||| | ||||| | | ||| ||| | |||||||| ||||||||||||||||
Db        358 CTTACTTATACACACATATACACAGACATACACACATATACACAGATATACACACATATA 417

Qy        528 TACACATATATACACATATATATACACATATATACACATATATACATATATACACATATA 587
              ||| | ||| |||||||||| | ||||||| |    ||||||| ||| |||||||| | |
Db        418 CACAGACATACACACATATAGACACACATACACTTACATATACACACATATACACAGACA 477

Qy        588 TATACATATATACACATATATACATATATACACATATATACATATATACACACATATATA 647
              || ||| |||||||||| ||||||| ||| ||||| ||| ||| |||||||| | | | |
Db        478 TACACACATATACACACATATACACATACACACACATACACACATATACACATACACACA 537

Qy        648 CACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATA 707
              || | ||| ||| ||||| | | ||| ||||| | | | ||| | ||| ||  ||| | |
Db        538 CATACACACATATACACATACATATACACATACACACAAACACACATACACACATACACA 597

Qy        708 TGTATATATATATGCTC 724
              ||   | | |  | |
Db        598 CATACGAACACACACAC 614


RESULT 63
BM906116
LOCUS       BM906116                1156 bp    mRNA    linear   EST 12-MAR-2002
DEFINITION  AGENCOURT_6621231 NIH_MGC_125 Homo sapiens cDNA clone IMAGE:5589966
            5', mRNA sequence.
ACCESSION   BM906116
VERSION     BM906116.1  GI:19356495
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1156)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Invitrogen
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12363  row: i  column: 07
            High quality sequence start: 323
            High quality sequence stop: 613.
FEATURES             Location/Qualifiers
     source          1. .1156
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5589966"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_125"
                     /note="Organ: ovary (pool of 3); Vector: pCMV-SPORT6;
```

Site_1: EcoRV (destroyed); Site_2: NotI; RNA source pool of three ovaries, from females ranging in age from 38 to 49 yo. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 2.1 kb, insert size range 1-3.5 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 036."

ORIGIN

Query Match 6.3%; Score 292.6; DB 5; Length 1156;
Best Local Similarity 76.5%; Pred. No. 1.2e-39;
Matches 385; Conservative 0; Mismatches 114; Indels 4; Gaps 2;

```
Qy      210 GAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACACATAT 269
            || | | |   |||    |  | || | |||||| |||||||| | | | ||| | ||||
Db      223 GACATAGATGGAGGAGGATCCTCTAGACATATATGTATATATAAATAGATACATATATAT 282

Qy      270 ATATATACACATATATATACACATATATACATATATACACATATATACACATATATATAC 329
            | |||||   | ||  ||| | ||||||||||||||| | ||||||||||||||||||
Db      283 AGATATATCTACATGCATATATATATATACATATATATAGATATATACACATATATATAT 342

Qy      330 ACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACAC 389
            | ||| |||| || | |||||||| | |||| |||| | ||||||| | ||||||| |
Db      343 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 402

Qy      390 ACATATATACACATATATACACACATATATACACATATATACACATATATACACATATAC 449
            | ||||||| | ||||||| | | ||||||| | ||||||| | ||||||| | |||||
Db      403 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 462

Qy      450 ACATATATACACATATATACATATATACACATATATATA-ATATACACACATATATATAC 508
            | ||||||| | ||||||| ||||||| | ||||||||| ||||| | | |||||||||
Db      463 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 522

Qy      509 ACATATATACACACATATATACACATATATACACATATATATACACATATATACACATAT 568
            | ||||||| | | ||||||| | ||||||| | ||||||||| | ||||||| | ||||
Db      523 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 582

Qy      569 ATACATATATACACATATATATACATATATACACATATATACATATATACACATATATAC 628
            ||| ||||||| | ||||||||| ||||||| | ||||||| ||||||| ||||||||||
Db      583 ATATATATATATATATATATATATATATATATATATATATATATATATATATATACATATAC 642

Qy      629 ATATATACACACATATATACACATACATATACACACACATAGATATACATATATATACAC 688
            ||||||| | | ||||||| | ||| ||||| | | | ||| |||| |    | ||||
Db      643 ATATATATATATATATATATATATATATATATATATATATACATATGGA---CTGTACAT 699

Qy      689 ATATATATACGTATATATATGTA 711
            ||||||| ||  | ||   ||||
Db      700 ATATATACACCCACATGACTGTA 722
```

RESULT 64
CL907900/c
LOCUS CL907900 714 bp DNA linear GSS 14-SEP-2004
DEFINITION OA_ABa0007A21.r OA_ABa Oryza australiensis genomic clone OA_ABa0007A21 3', genomic survey sequence.
ACCESSION CL907900
VERSION CL907900.1 GI:52016779
KEYWORDS GSS.
SOURCE Oryza australiensis
ORGANISM Oryza australiensis
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 714)
AUTHORS Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
TITLE OMAP Project
JOURNAL Unpublished (2004)
COMMENT Contact: Rod A. Wing
Arizona Genomics Institute
University of Arizona
Forbes Building Room 303, Tucson, AZ 85721-0036, USA
Tel: 520 626 9595
Fax: 520 621 1259
Email: http://genome.arizona.edu
PCR PRimers
FORWARD: TAA TAC GAC TCA CTA TAG GG
BACKWARD: CAC TCA TTA GGC ACC CCA
Plate: 0007 row: A column: 21
Seq primer: CAC TCA TTA GGC ACC CCA
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .714
/organism="Oryza australiensis"
/mol_type="genomic DNA"
/db_xref="taxon:4532"
/clone="OA_ABa0007A21"
/tissue_type="young leaves"
/lab_host="DH10B T1 phage resistant"
/clone_lib="OA_ABa"
/note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

Query Match 6.3%; Score 290.2; DB 9; Length 714;
Best Local Similarity 77.7%; Pred. No. 3.2e-39;
Matches 363; Conservative 0; Mismatches 103; Indels 1; Gaps 1;

```
Qy      286 ATACACATATATACATATATACACATATATACACATATATATACACATACATATGTACAC 345
            ||| | ||||||| ||||||| | ||||||  | |||| |||| | ||| |||| || |
Db      714 ATATATATATATATATATATATATATATTTATATTTATATATATATATATATATAT 655

Qy      346 ATATATATACACATATGTATACACATATATACACATATATACACACATATATACACATAT 405
            ||||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | ||||
Db      654 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 595

Qy      406 ATACACACATATATACACATATATACACATATATACACATATACACATATATACACATAT 465
            ||| | | |||||||| | |||||||| | ||||||| | ||||| | ||||||| | ||||
Db      594 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 535

Qy      466 ATACATATATACACATATATATAATATACACACATATATATACACATATATACACACATA 525
            ||| ||||||| | |||||||| ||||| | | ||||||||| | ||||||| | | |||
Db      534 ATATATATATATATATATATAT-ATATATATATATATATATATATATATATATATATATA 476

Qy      526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
            |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db      475 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 416

Qy      586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
            |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db      415 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 356

Qy      646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
            || | ||||||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db      355 TATATATACATATATATATATATATATATATATATATATATATATATATATATATATATA 296

Qy      706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATT 752
            ||| |||||||| | || |  | || |  | |   | ||  |  |||
Db      295 TATATATATATACACGCGCGCGTGTGCGCACGTTATTGCGTTGAATT 249
```

RESULT 65
BX180768
LOCUS BX180768 701 bp DNA linear GSS 28-JAN-2003
DEFINITION Danio rerio genomic clone DKEY-176A10, genomic survey sequence.
ACCESSION BX180768
VERSION BX180768.1 GI:28012577
KEYWORDS GSS.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 701)
AUTHORS Humphray,S.J., Huckle,E. and Durham,J.L.
TITLE Direct Submission

```
  JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the SP6 end of BAC 176A10. 176A10
            is part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .701
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-176A10"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             6.3%;  Score 290;  DB 9;  Length 701;
  Best Local Similarity   76.1%;  Pred. No. 3.5e-39;
  Matches  370;  Conservative    0;  Mismatches  115;  Indels    1;  Gaps    1;

Qy        213 AAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACATATATA 272
              ||| |||| |  ||  ||  |  ||    | |    | |      |    | |||||||
Db        216 AAATGAAAATGTTTTAAATGAAGGTAAGAGTTTGATCAGAGGTTTTCGAGAATATATATA 275

Qy        273 TATACACATATATATACACATATATACATATATACACATATATACACATATATATACACA 332
              ||||  | |||||||||  | |||||||  |||||||  | |||||||  | |||||||||  | |
Db        276 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 335

Qy        333 TACATATGTACACATATATATACACATATGTATACACATATATACACATATATACACACA 392
              ||  ||||  ||  |  |||||||||  |  ||||  ||||  |  |||||||  |  |||||||  |  | |
Db        336 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 395

Qy        393 TATATACACATATATACACACATATATACACATATATACACATATATACACATATACACA 452
              ||||||  |  |||||||  |  |  |||||||  |  |||||||  |  |||||||  |  |||||  |  |
Db        396 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 455

Qy        453 TATATACACATATATACATATATACACATATATATA-ATATACACACATATATATACACA 511
              ||||||  |  |||||||  |||||||  |  |||||||||  |||||  |  |  |||||||||  |  |
Db        456 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 515

Qy        512 TATATACACACATATATACACATATATACACATATATATACACATATATACACATATATA 571
              ||||||  |  |||||||  |  |||||||  |  |||||||||  |  |||||||  |  |||||||
Db        516 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 575

Qy        572 CATATATACACATATATATACATATATACACATATATACATATATACACATATATACATA 631
              |  |||||||  |  |||||||||  |||||||  |  |||||||  |||||||  |  |||||||  |||
Db        576 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 635

Qy        632 TATACACACATATATACACATACATATACACACACATAGATATACATATATATACACATA 691
              ||||  |  |  |||||||  |  |||  |||||  |  |  |||  |||||  |||||||||  |  |||
Db        636 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 695

Qy        692 TATATA 697
              ||||||
Db        696 TATATA 701


RESULT 66
CL907900
LOCUS       CL907900                 714 bp    DNA     linear   GSS 14-SEP-2004
DEFINITION  OA_ABa0007A21.r OA_ABa Oryza australiensis genomic clone
            OA_ABa0007A21 3', genomic survey sequence.
ACCESSION   CL907900
VERSION     CL907900.1  GI:52016779
KEYWORDS    GSS.
SOURCE      Oryza australiensis
  ORGANISM  Oryza australiensis
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 714)
  AUTHORS   Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R.,
            Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
  TITLE     OMAP Project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0007  row: A  column: 21
            Seq primer: CAC TCA TTA GGC ACC CCA
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .714
                     /organism="Oryza australiensis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4532"
                     /clone="OA_ABa0007A21"
                     /tissue_type="young leaves"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OA_ABa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

  Query Match             6.3%;  Score 289.4;  DB 9;  Length 714;
  Best Local Similarity   79.4%;  Pred. No. 4.4e-39;
  Matches  355;  Conservative    0;  Mismatches   91;  Indels    1;  Gaps    1;

Qy        251 TACACACACACACACATATATATATACACATATATATACACATATATACATATATACACA 310
              | | |||||  | | | |  ||||||||  |  |||||||||  |  |||||||  |||||||  | |
Db        266 TGCGCACACGCGCGCGTGTATATATATATATATATATATATATATATATATATATATATA 325

Qy        311 TATATACACATATATATACACATACATATGTACACATATATATACACATATGTATACACA 370
              ||||||  |  |||||||||  |   ||  ||||  ||  |  |||||||||  |  ||||  ||||  |  |
Db        326 TATATATATATATATATATATGTATATATATATATATATATATATATATATATATATATA 385

Qy        371 TATATACACATATATACACACATATATACACATATATACACACATATATACACATATATA 430
              ||||||  |  |||||||  |  |  |||||||  |  |||||||  |  |  |||||||  |  |||||||
Db        386 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 445

Qy        431 CACATATATACACATATACACATATATACACATATATACATATATACACATATATAAT 490
              |  |||||||  |  |||||  |  |||||||  |  |||||||  |||||||  |  |||||||  ||
Db        446 TATATATATATATATATATATATATATATATATATATATATATATATATATATATAT-AT 504

Qy        491 ATACACACATATATATACACATATATACACACATATATACACATATATACACATATATAT 550
              |||  |  |  |||||||||  |  |||||||  |  |  |||||||  |  |||||||  |  |||||||||
Db        505 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 564

Qy        551 ACACATATATACACATATATACATATATACACATATATATACATATATACACATATATAC 610
              |  |  |||||||  |  |||||||  |||||||  |  |||||||||  |||||||  |  |||||||
Db        565 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 624

Qy        611 ATATATACACATATATACATATATACACACATATATACACATACATATACACACACATAG 670
              |||||||  |  |||||||  |||||||  |  |  |||||||  |  |||  |||||  |  |  |||
Db        625 ATATATATATATATATATATATATATATATATATATATATATATATATATATATAAATATATAA 684

Qy        671 ATATACATATATATACACATATATATA 697
              |||||  |||||||||  |  |||||||||
Db        685 ATATATATATATATATATATATATATA 711


RESULT 67
DR42I22S
LOCUS       DR42I22S                 673 bp    DNA     linear   GSS 22-NOV-2002
DEFINITION  Danio rerio genomic clone DKEY-42I22, genomic survey sequence.
ACCESSION   AL983929
VERSION     AL983929.1  GI:25183962
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
```

```
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 673)
  AUTHORS   Humphray,S.J., Huckle,E. and Hunt,S.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-NOV-2002) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail contact:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the SP6 end of BAC 42I22. 42I22 is
            part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene.
            Further details: http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .673
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-42I22"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             6.2%;  Score 288.8;  DB 9;  Length 673;
  Best Local Similarity   80.0%;  Pred. No. 5.6e-39;
  Matches  352;  Conservative    0;  Mismatches   87;  Indels    1;  Gaps    1;

Qy        281 TATATATACACATATATACATATATACACATATATACACATATATATACACATACATATG 340
              | | | || | ||||||| ||||||| | ||||||| | |||||||||| | ||| ||||
Db        234 TGTGTGTATATATATATATATATATATATATATATATATATATATATATATATATATA 293

Qy        341 TACACATATATATACACATATGTATACACATATATACACATATATACACACATATATACA 400
              || | |||||||||| | |||| |||||||| | | ||||| | | ||||| ||| | |||
Db        294 TATATATATATATATATATATATATACACACACACACACACACACACACATATACACACA 353

Qy        401 CATATATACACACATATATACACATATATACACATATATACACATATACACATATATACA 460
              |||||||| | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| |
Db        354 CATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 413

Qy        461 CATATATACATATATACACATATATATA-ATATACACACATATATATACACATATATACA 519
              ||||||| ||||||| | |||||||||| ||||| | | |||||||||| | ||||||| |
Db        414 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 473

Qy        520 CACATATATACACATATATACACATATATATACACATATATACACATATATACATATATA 579
              | ||||||| | ||||||| | ||||||||| | ||||||| | ||||||| |||||||
Db        474 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 533

Qy        580 CACATATATATACATATATACACATATATACATATATACACATATATACATATATACACA 639
              | |||||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | |
Db        534 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 593

Qy        640 CATATATACACATACATATACACACACATAGATATACATATATATACACATATATATACG 699
              ||||||| | ||| ||||| | | | ||| ||||| |||||||||| | ||||||||||
Db        594 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 653

Qy        700 TATATATATGTATATATATA 719
              |||||||||| ||||||||||
Db        654 TATATATATATATATATATA 673


RESULT 68
CW007564
LOCUS       CW007564                569 bp    DNA     linear   GSS 23-SEP-2004
DEFINITION  ZMMBLa0009D17.r ZMMBLa Zea mays genomic clone ZMMBLa0009D17 3',
            genomic survey sequence.
ACCESSION   CW007564
VERSION     CW007564.1  GI:52588327
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 569)
  AUTHORS   Wing,R., Luo,M., Soderlund,C. and Haller,K.
  TITLE     ZMMBL sequences
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            Plate: 0009  row: D  column: 17
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .569
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /cultivar="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBLa0009D17"
                     /tissue_type="immature ears"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="ZMMBLa"
                     /note="Vector: pAGIBAC1; Site_1: SalI; Site_2: SalI"
ORIGIN

  Query Match             6.2%;  Score 288.4;  DB 9;  Length 569;
  Best Local Similarity   74.5%;  Pred. No. 6.7e-39;
  Matches  389;  Conservative    0;  Mismatches  131;  Indels    2;  Gaps    2;

Qy        200 TGTCATCTCAGAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACAC 259
              ||| |  || ||||| ||  | |||  ||   |||||  |   |  |    ||
Db         48 TGTAAAGTCTGAAAACAACCACGGCCATTGTAATATAATTTCCCAACTGCTCACGATGCT 107

Qy        260 ACA-CACATATATATATACACATATATATACACATATATACATATATACACATATATACA 318
              ||| | || |  || | |  |   |  |||| ||||||| ||||||| | ||||||| |
Db        108 ACATCTCAAAACTACAAAACCTCGTTGCTACAAATATATATATATATATATATATATATA 167

Qy        319 CATATATATACACATACATATGTACACATATATATACACATATGTATACACATATATACA 378
              ||||||||| | ||| |||| || | ||||||||| | |||| |||| | ||||||| |
Db        168 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 227

Qy        379 CATATATACACACATATATACACATATATACACACATATATACACATATATACACATATA 438
              ||||||| | | ||||||| | ||||||| | | ||||||| | ||||||| | |||||
Db        228 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 287

Qy        439 TACACATATACACATATATACACATATATACATATATACACATATATATA-ATATACACA 497
              || | ||||| | ||||||| | ||||||| ||||||| | ||||||||| ||||| | |
Db        288 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 347

Qy        498 CATATATATACACATATATACACACATATATACACATATATACACATATATATACACATA 557
              ||||||||| | ||||||| | | ||||||| | ||||||| | ||||||||| | |||
Db        348 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 407

Qy        558 TATACACATATATACATATATACACATATATATACATATATACACATATATACATATATA 617
              |||| | ||||||| ||||||| | |||||||| ||||||| | ||||||| |||||||
Db        408 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 467

Qy        618 CACATATATACATATATACACACATATATACACATACATATACACACACATAGATATACA 677
              | ||||||| ||||||| | | ||||||| | ||| ||||| | | | ||| ||||| |
Db        468 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 527

Qy        678 TATATATACACATATATATACGTATATATATGTATATATATA 719
              |||||||| | |||||||||  |||| |||| |||| | |||
Db        528 TATATATATATATATATATATATATAAATATATATAGAAATA 569


RESULT 69
DR42I22S/c
LOCUS       DR42I22S                673 bp    DNA     linear   GSS 22-NOV-2002
DEFINITION  Danio rerio genomic clone DKEY-42I22, genomic survey sequence.
ACCESSION   AL983929
```

```
VERSION     AL983929.1  GI:25183962
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 673)
  AUTHORS   Humphray,S.J., Huckle,E. and Hunt,S.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-NOV-2002) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail contact:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the SP6 end of BAC 42I22. 42I22 is
            part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene.
            Further details: http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .673
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-42I22"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             6.2%;  Score 288;  DB 9;  Length 673;
  Best Local Similarity   75.4%;  Pred. No. 7.7e-39;
  Matches  371;  Conservative    0;  Mismatches  120;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | |||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db        673 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 614

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db        613 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 554

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        553 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 494

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db        493 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 434

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db        433 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 374

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | || | | |  | || | |    | | |  | | |  | |||
Db        373 TATATATATATATATATATATGTGTGTGTATATGTGTGTGTGTGTGTGTGTGTGTATATA 314

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db        313 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 254

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||||||||| | | | ||| | | | ||||| | | | ||  | | |
Db        253 TATATATATATATACACACACACACACACACACACACACACACACACACACACACACACA 194

Qy        706 TATGTATATATA 717
                |  || ||||
Db        193 GTTTCATTTATA 182


RESULT 70
CL917654
LOCUS       CL917654                720 bp    DNA     linear   GSS 14-SEP-2004
DEFINITION  OA_ABa0020N13.r OA_ABa Oryza australiensis genomic clone
            OA_ABa0020N13 3', genomic survey sequence.

ACCESSION   CL917654
VERSION     CL917654.1  GI:52029479
KEYWORDS    GSS.
SOURCE      Oryza australiensis
  ORGANISM  Oryza australiensis
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 720)
  AUTHORS   Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R.,
            Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
  TITLE     OMAP Project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0020  row: N  column: 13
            Seq primer: CAC TCA TTA GGC ACC CCA
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .720
                     /organism="Oryza australiensis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4532"
                     /clone="OA_ABa0020N13"
                     /tissue_type="young leaves"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OA_ABa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

  Query Match             6.2%;  Score 288;  DB 9;  Length 720;
  Best Local Similarity   75.2%;  Pred. No. 7.6e-39;
  Matches  385;  Conservative    0;  Mismatches  125;  Indels    2;  Gaps    2;

Qy        196 GAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACAC 255
              || || ||||  || |||||  ||  | | ||  || |  ||  ||| |
Db        200 GACATATCATAGAAGGAAAAAGTTAAATTTTGTAATTAGTAAAACAATTGGAATACATTT 259

Qy        256 ACACACACACATATATATACACATATATATACACATATATACATATATACACATATAT 315
              ||        |||||||||| | ||||||||| | ||||||| ||||||| | ||||||
Db        260 ACTGTGTGGAATATATATATATATATATATATATATATATATATATATATATATATAT 319

Qy        316 ACACATATATATACACATACATATGTACACATATATATACACATATGTATACACATATAT 375
              | | ||||||||| | ||| |||| || | ||||||||| | |||| |||| | ||||||
Db        320 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 379

Qy        376 ACACATATATACACACATATATACACATATATACACACATATATACACATATATACACAT 435
              | | ||||||| | | ||||||| | ||||||| | | ||||||| | ||||||| | ||
Db        380 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 439

Qy        436 ATATACACATATACACATATATACACATATATACATATATACACATATATATAATATACA 495
              ||||| | ||||| | ||||||| | ||||||| ||||||| | |||||||| ||||| |
Db        440 ATATATATATATATATATATATATATATATATATATATATATATATATATATAT-ATATATA 498

Qy        496 CACATATATATACACATATATACACACATATATACACATATATACACATATATATACACA 555
              | |||||||| | ||||||| | | ||||||| | ||||||| | |||||||| | |
Db        499 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 558

Qy        556 TATATACACATATATACATATATACACATATATATACATATATACACATATATACATATA 615
              |||||| | ||||||| ||||||| | |||||||| ||||||| | ||||||| |||||
Db        559 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 618

Qy        616 TACACATATATACATATATACACACATATATACACATACATATACACACACATAGATATA 675
              || | ||||||| |||||||    | ||||||| | ||| ||||| | |   ||| |||||
Db        619 TATATATATATATATATATATTTATATATATATATATATATATATATATTTATATATATA 678
```

```
Qy        676 CATATATATACACATATATATACGTATATATA 707
              |||| ||||   |||||||| ||||||||
Db        679 TATATGTATA-GTATATATATATGTATATATA 709


RESULT 71
BX157670
LOCUS       BX157670                668 bp    DNA     linear   GSS 28-JAN-2003
DEFINITION  Danio rerio genomic clone DKEY-125M16, genomic survey sequence.
ACCESSION   BX157670
VERSION     BX157670.1  GI:27989269
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 668)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the T7 end of BAC 125M16. 125M16
            is part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .668
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-125M16"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             6.2%;  Score 287.4;  DB 9;  Length 668;
  Best Local Similarity   77.5%;  Pred. No. 9.8e-39;
  Matches  373;  Conservative    0;  Mismatches  106;  Indels    2;  Gaps    2;

Qy        242 ATATATATATACACACACACACACATATATATATACACATATATATACACATATATACAT 301
              | | | ||  ||| | | | | | ||| ||| ||| |  ||||| || |  |||||| ||
Db        185 AAAAAAATGAACATAAATAAAGAAATAAATAAATAAATTTATATCTATATCTATATATAT 244

Qy        302 ATATACACATATATACACATATATATACACATACATATGTACACA-TATATATACACATA 360
              |||||| | |||||||| | ||||||||| | ||| |||| || ||| |||||| | ||| |
Db        245 ATATATATATATATATATATATATATATATATATATATATATACAGTATATACATACACA 304

Qy        361 TGTATACACATATATACACATATATACACACATATATACACATATATACACACATATATA 420
                || | ||||||||| ||||| ||| | | ||||||| ||||| ||||| ||||| |||
Db        305 CATACATACATATATATACATACATATATATATATATATACATACATACATACATACATA 364

Qy        421 CACATATATACACATATATACACATATACACATATATACACATATATACATATATACACA 480
              || | ||| | ||||| ||||| | ||||| | ||| | ||||| ||||||| ||||| |
Db        365 CATACATACATACATACATACATACATACATACATACATACATACATACATACATACATA 424

Qy        481 TATATATA-ATATACACACATATATATACACATATATACACACATATATACACATATATA 539
              |||||||| ||||| | | ||||||||| | ||||||| | | ||||||| | |||||||
Db        425 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 484

Qy        540 CACATATATATACACATATATACACATATATACATATATACACATATATATACATATATA 599
              | ||||||||| | ||||||| | ||||||| ||||||| | | ||||||| |||||||
Db        485 TATATATATATATATATATATATATATATATATATATATATATAGATATATATATATATA 544

Qy        600 CACATATATACATATATACACATATATACATATATACACACATATATACACATACATATA 659
              | ||||||| ||||||| | ||||||| ||||||| | | ||||||| | ||| |||||
Db        545 TATATATATATATATATATAGATATATATATATATATATATATATATATATATATATA 604

Qy        660 CACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATA 719
              | | | ||| ||||| ||||||||| | |||||||||  ||||| ||  ||||||||||
Db        605 TATATATATATATATATATATATATATATATATATATATATATATATGTAGATATATATATA 664
```

```
Qy        720 T 720
              |
Db        665 T 665


RESULT 72
BX149919
LOCUS       BX149919                737 bp    DNA     linear   GSS 13-MAR-2003
DEFINITION  Danio rerio genomic clone DKEY-112P5, genomic survey sequence.
ACCESSION   BX149919
VERSION     BX149919.1  GI:27981385
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 737)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAR-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the T7 end of BAC 112P5. 112P5 is
            part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .737
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-112P5"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             6.2%;  Score 287.4;  DB 9;  Length 737;
  Best Local Similarity   76.5%;  Pred. No. 9.6e-39;
  Matches  365;  Conservative    0;  Mismatches  111;  Indels    1;  Gaps    1;

Qy        242 ATATATATATACACACACACACACATATATATATACACATATATATACACATATATACAT 301
              | || | ||||||       ||  | | | | | |     | | | |    | |  || ||
Db        262 AGATTTCTATACAGTTTGGCAACCTTTTCTGTTTTTTGGTTTTTTTTTTAAAAGCTATAT 321

Qy        302 ATATACACATATATACACATATATATACACATACATATGTACACATATATATACACATAT 361
              ||||| | |||||||| | ||||||||| | ||| |||| || | ||||||||| | ||||
Db        322 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 381

Qy        362 GTATACACATATATACACATATATACACACATATATACACATATATACACACATATATAC 421
              |||| | ||||||| | ||||||| | | ||||||| | ||||||| | | |||||||
Db        382 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 441

Qy        422 ACATATATACACATATATACACATATACACATATATACACATATATACATATATACACAT 481
              | ||||||| | ||||||| | ||||| | ||||||| | ||||||| ||||||| | ||
Db        442 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 501

Qy        482 ATATATAATATACACACATATATATACACATATATACACACATATATACACATATATACA 541
              |||||| |||||||| |||||||| | ||||||| | | ||||||| | ||||||| |
Db        502 ATATAT-ATATACACATATATATATATATATATATATATATATATATATATATATATA 560

Qy        542 CATATATATACACATATATACACATATATACATATATACACATATATATACATATATACA 601
              ||||||||| | ||||||| | ||||||| ||||||| | ||||||||| ||||||| |
Db        561 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 620

Qy        602 CATATATACATATATACACATATATACATATATACACACATATATACACATACATATACA 661
              ||||||| |||||||| | ||||||| ||||||| | | ||||||| | ||| |||||| |
Db        621 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 680

Qy        662 CACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATAT 718
              | | ||| ||||||| ||||||| | |||||||||  ||||||||| |||||||||
Db        681 TATATATATATATACACATATATATATATATATATATATATATATATATATATATAT 737
```

```
RESULT 73
BX197183/c
LOCUS       BX197183                771 bp    DNA     linear   GSS 28-JAN-2003
DEFINITION  Danio rerio genomic clone DKEY-207A10, genomic survey sequence.
ACCESSION   BX197183
VERSION     BX197183.1  GI:28029069
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 771)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the SP6 end of BAC 207A10. 207A10
            is part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .771
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-207A10"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             6.2%;  Score 286.6;  DB 9;  Length 771;
  Best Local Similarity   78.1%;  Pred. No. 1.3e-38;
  Matches  357;  Conservative    0;  Mismatches   99;  Indels    1;  Gaps    1;

Qy        270 ATATATACACATATATATACACATATATACATATATACACATATATACACATATATATAC 329
              ||||||  | |||||||||| | ||||||| ||||||| | ||||||| | |||||||||
Db        771 ATATATGTATATATATATATATATATATATATATATATATATATATATATATATATAT 712

Qy        330 ACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACAC 389
              | ||| |||| || | |||||||||| | |||| |||| | ||||||| | ||||||| |
Db        711 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 652

Qy        390 ACATATATACACATATATACACACATATATACACATATATACACATATATACACATATAC 449
              | ||||||| | ||||||| | | ||||||| | ||||||| | ||||||| | |||||
Db        651 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 592

Qy        450 ACATATATACACATATATACATATATACACATATATATAATATACACACATATATATACA 509
              | ||||||| | ||||||| |||||||| | |||||||| ||||| | | |||||||||| |
Db        591 ATATATATATATATATATATATATATATATATATATAT-ATATATATATATATATATATA 533

Qy        510 CATATATACACACATATATACACATATATACACATATATATACACATATATACACATATA 569
              |||||||  | |  ||||||| | ||||||| | ||||||||| | ||||||| | |||||
Db        532 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 473

Qy        570 TACATATATACACATATATATACATATATACACATATATACATATATACACATATATACA 629
              ||  ||||||| | ||||||||| |||||||  | |||||||  ||||||| | ||||||| |
Db        472 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 413

Qy        630 TATATACACACATATATACACATACATATACACACACATAGATATACATATATATACACA 689
              ||||||  | |  |||||||  | |||  |||||  | |  | |||  |||||  ||||||||| | |
Db        412 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 353

Qy        690 TATATATACGTATATATATGTATATATATATGCTCCA 726
              |||||  |    | |  | |||  |||  | |  | |  | | |
Db        352 TATATTTTTTTTTTTTTATTTATTTTTTTTTCCCCAA 316


RESULT 74
CW007564/c
LOCUS       CW007564                569 bp    DNA     linear   GSS 23-SEP-2004
DEFINITION  ZMMBLa0009D17.r ZMMBLa Zea mays genomic clone ZMMBLa0009D17 3',
            genomic survey sequence.
ACCESSION   CW007564
VERSION     CW007564.1  GI:52588327
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 569)
  AUTHORS   Wing,R., Luo,M., Soderlund,C. and Haller,K.
  TITLE     ZMMBL sequences
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            Plate: 0009  row: D  column: 17
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .569
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /cultivar="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBLa0009D17"
                     /tissue_type="immature ears"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="ZMMBLa"
                     /note="Vector: pAGIBAC1; Site_1: SalI; Site_2: SalI"
ORIGIN

  Query Match             6.2%;  Score 286;  DB 9;  Length 569;
  Best Local Similarity   77.3%;  Pred. No. 1.7e-38;
  Matches  360;  Conservative    0;  Mismatches  105;  Indels    1;  Gaps    1;

Qy        281 TATATATACACATATATACATATATACACATATATACACATATATATACACATACATATG 340
              ||| | || | |||| || |||||||| | ||||||| | |||||||||| | ||| ||||
Db        569 TATTTCTATATATATTTATATATATATATATATATATATATATATATATATATATATA 510

Qy        341 TACACATATATATACACATATGTATACACATATATACACATATATACACACATATATACA 400
              || | |||||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| |
Db        509 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 450

Qy        401 CATATATACACACATATATACACATATATACACATATATACACATATACACATATATACA 460
              |||||||  | |  |||||||| | |||||||| | ||||||| | |||||  | |||||||  |
Db        449 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 390

Qy        461 CATATATACATATATACACATATATATA-ATATACACACATATATATACACATATATACA 519
              |||||||  |||||||  | |||||||||  |||||  | |  |||||||||  | |||||||  |
Db        389 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 330

Qy        520 CACATATATACACATATATACACATATATATACACATATATACACATATATACATATATA 579
              | |||||||  | |||||||  | |||||||||  | |||||||  | |||||||  |||||||
Db        329 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 270

Qy        580 CACATATATATACATATATACACATATATACATATATACACATATATACATATATACACA 639
              | |||||||||  |||||||  | |||||||  |||||||  | |||||||  |||||||  | |
Db        269 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 210

Qy        640 CATATATACACATACATATACACACACATAGATATACATATATATACACATATATATACG 699
              |||||||  | |||  |||||  | |  | |||  |||||  |||||||||  | |||||||||
Db        209 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 150

Qy        700 TATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCA 745
              |||||||||  | ||   |   | |   |   |  | |||  ||||||
Db        149 TATATATATTTGTAGCAACGAGGTTTTGTAGTTTTGAGATGTAGCA 104
```

```
RESULT 75
CL917654/c
LOCUS       CL917654               720 bp    DNA     linear   GSS 14-SEP-2004
DEFINITION  OA_ABa0020N13.r OA_ABa Oryza australiensis genomic clone
            OA_ABa0020N13 3', genomic survey sequence.
ACCESSION   CL917654
VERSION     CL917654.1  GI:52029479
KEYWORDS    GSS.
SOURCE      Oryza australiensis
  ORGANISM  Oryza australiensis
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 720)
  AUTHORS   Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R.,
            Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
  TITLE     OMAP Project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0020  row: N  column: 13
            Seq primer: CAC TCA TTA GGC ACC CCA
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .720
                     /organism="Oryza australiensis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4532"
                     /clone="OA_ABa0020N13"
                     /tissue_type="young leaves"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OA_ABa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

  Query Match             6.2%;  Score 286;  DB 9;  Length 720;
  Best Local Similarity   79.7%;  Pred. No. 1.7e-38;
  Matches  362;  Conservative    0;  Mismatches   90;  Indels    2;  Gaps    2;

Qy        283 TATATACACATATATACATA-TATACACATATATACACATATATATACACATACATATGT 341
              |||||| ||||||||| ||| |||||| ||||||| | ||| ||||| | ||| |||| |
Db        709 TATATATACATATATATATACTATACATATATATATATATAAATATATATATATATATAT 650

Qy        342 ACACATATATATACACATATGTATACACATATATACACATATATACACACATATATACAC 401
              | | ||||| ||| | |||| |||| | ||||||| | ||||||| | | ||||||| |
Db        649 ATATATATAAATATATATATATATATATATATATATATATATATATATATATATATATAT 590

Qy        402 ATATATACACACATATATACACATATATACACATATATACACATATACACATATATACAC 461
              ||||||| | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| |
Db        589 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 530

Qy        462 ATATATACATATATACACATATATATA-ATATACACACATATATATACACATATATACAC 520
              ||||||| ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| |
Db        529 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 470

Qy        521 ACATATATACACATATATACACATATATATACACATATATACACATATATACATATATAC 580
              | ||||||| | ||||||| | ||||||||| | ||||||| | ||||||| |||||||
Db        469 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 410

Qy        581 ACATATATATACATATATACACATATATACATATATACACATATATACATATATACACAC 640
              | ||||||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | |
Db        409 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 350

Qy        641 ATATATACACATACATATACACACACATAGATATACATATATATACACATATATATACGT 700
              ||||||| | ||| ||||| | | | ||| ||||| ||||||||| | ||||||||  |
Db        349 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 290

Qy        701 ATATATATGTATATATATATGCTCCAGAGTTCAT 734
              |||||||| |||||||||||| |  || ||| ||
Db        289 ATATATATATATATATATATTCCACACAGTAAAT 256


RESULT 76
CF924712
LOCUS       CF924712               760 bp    mRNA    linear   EST 10-NOV-2003
DEFINITION  DD005E11.ab1 Codonopsis lanceolata cDNA library Codonopsis
            lanceolata cDNA, mRNA sequence.
ACCESSION   CF924712
VERSION     CF924712.1  GI:38243228
KEYWORDS    EST.
SOURCE      Codonopsis lanceolata
  ORGANISM  Codonopsis lanceolata
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            asterids; campanulids; Asterales; Campanulaceae; Codonopsis.
REFERENCE   1  (bases 1 to 760)
  AUTHORS   Yang,D.C., In,J.G. and Lee,B.S.
  TITLE     Codonopsis lanceolata
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Deok Chun Yang
            BioPia
            1 Seochonri, Kihung, Yongin, Kyungki-do, 449-701, Korea
            Tel: 82 31 201 2688
            Fax: 82 31 205 6323
            Email: dcyang@ibiopia.com.
FEATURES             Location/Qualifiers
     source          1. .760
                     /organism="Codonopsis lanceolata"
                     /mol_type="mRNA"
                     /db_xref="taxon:103999"
                     /clone_lib="Codonopsis lanceolata cDNA library"
                     /note="Organ: root; Vector: pTriplEx2; Site_1: Sfi IA;
                     Site_2: Sfi IB"
ORIGIN

  Query Match             6.2%;  Score 285.8;  DB 7;  Length 760;
  Best Local Similarity   80.0%;  Pred. No. 1.8e-38;
  Matches  360;  Conservative    0;  Mismatches   88;  Indels    2;  Gaps    2;

Qy        268 ATATATATACACATATATATACACATATATACATATATACACATATATACACATATATAT 327
              ||||||||| | ||||||||| |||||||| ||||||| | ||||||| | ||||||||
Db        313 ATATATATATATATATATATATGCATATATATATATATATATATATATATATATATAT 372

Qy        328 ACACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATAC 387
              | | ||| |||| || | ||||||||| | |||| |||| | ||||||| | |||||||
Db        373 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 432

Qy        388 ACACATATATACACATATATACACACATATATACACATATATACACATATATACACATAT 447
              | | ||||||| | ||||||| | | ||||||| | ||||||| | ||||||| ||||
Db        433 ATATATATATATATATATATATATATATATATATATATATATATATATATATATA-TGATAT 491

Qy        448 ACACATATATACACATATATACATATATACACATATATATAATATACACACATATATATA 507
              | | ||||||| | ||||||| ||||||| | |||||||| ||||| | | |||||||||
Db        492 ATATATATATATATATATATATATATATATATATATATAT-ATATATATATATATATA 550

Qy        508 CACATATATACACACATATATACACATATATACACATATATATACACATATATACACATA 567
              | ||||||| | | ||||||| | ||||||| | ||||||||| | ||||||| | |||
Db        551 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 610

Qy        568 TATACATATATACACATATATATACATATATACACATATATACATATATACACATATATA 627
              |||| ||||||| | ||||||||| ||||||| | ||||||| ||||||| | |||||||
Db        611 TATATATATATATATATATATATATATATATATATATATATATATATAGATATATATA 670

Qy        628 CATATATACACACATATATACACATACATATACACACACATAGATATACATATATATACA 687
              ||||||| | | ||||||| | ||| ||||| | | | ||| ||||  ||||||||| |
Db        671 TATATATATATATATATATATATATATATATATATATATATATATATATCTATATATATA 730
```

```
Qy        688 CATATATATACGTATATATATGTATATATA 717
              |||| ||||  || ||||||||||||||||
Db        731 TATATCTATATATACATATATGTATATATA 760


RESULT 77
BX145980/c
LOCUS       BX145980               709 bp    DNA     linear   GSS 28-JAN-2003
DEFINITION  Danio rerio genomic clone DKEY-104E13, genomic survey sequence.
ACCESSION   BX145980
VERSION     BX145980.1  GI:27977335
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 709)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the T7 end of BAC 104E13. 104E13
            is part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .709
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-104E13"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             6.2%;  Score 285.4;  DB 9;  Length 709;
  Best Local Similarity   74.0%;  Pred. No. 2.1e-38;
  Matches  375;  Conservative    0;  Mismatches  131;  Indels    1;  Gaps    1;

Qy        230 ATATATATATATATATATATATACACACACACACACATATATATATACACATATATATAC 289
              |||||||||||||||||||| | | | | | |||||||||||| | ||| |||||
Db        709 ATATATATATATATATATATATATATATATATATATATATATATATATATATAGATATAT 650

Qy        290 ACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATAT 349
              | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | || |
Db        649 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATGT 590

Qy        350 ATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATAC 409
              ||||| | |||| |||    ||||||| | ||||||| | | ||||||| | |||||||
Db        589 ATATATATATATATATGTGTATATATATATATATATATATATATATATATATATATATAT 530

Qy        410 ACACATATATACACATATATACACATATATACACATATACACATATATACACATATATAC 469
              | | ||||||| | ||||||| | ||||||| | ||||| |   ||| || | |||||||
Db        529 ATATATATATATATATATATATATATATATATATATATATATGTATGTATATATATATAT 470

Qy        470 ATATATACACATATATATAATATACACACATATATATACACATATATACACACATATATA 529
              |||||||   |||||| | ||||| | | ||||||||| | || |||| | | |||||||
Db        469 ATATATATGTATATATGT-ATATATATATATATATATATATGTATATATATATATATA 411

Qy        530 CACATATATACACATATATATACACATATATACACATATATACATATATACACATATATA 589
              | ||||||| | ||||||||| | ||||||| | ||||||| |||||||     | ||||
Db        410 TATATATATATATATATATATATATATATATATATATATATATATATATATGTGTGTATA 351

Qy        590 TACATATATACACATATATACATATATACACATATATACATATATACACACATATATACA 649
              || ||||||| | |||| |   ||||||| | ||||||   |||||| | | ||||||| |
Db        350 TATATATATATATATATGTGTATATATATATATATATGTGTATATATATATATATATA 291

Qy        650 CATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATATG 709
              ||| ||||| | | ||| ||||| |||||||||   |||| || |   | |
Db        290 TATATATATATATATATATATATATATATATATATGCATAAATTTGACTTTTGCATGC 231
```

```
Qy        710 TATATATATATGCTCCAGAGTTCATAA 736
              || |  ||  || |  ||| ||  | |
Db        230 TAAAATTACTTGATTAAGATTTTGTCA 204


RESULT 78
BU507480/c
LOCUS       BU507480               1836 bp    mRNA    linear   EST 12-SEP-2002
DEFINITION  AGENCOURT_10309824 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:6501420
            5', mRNA sequence.
ACCESSION   BU507480
VERSION     BU507480.1  GI:22813713
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1836)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM14055  row: n  column: 13
            High quality sequence start: 74
            High quality sequence stop: 447.
FEATURES             Location/Qualifiers
     source          1. .1836
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6501420"
                     /tissue_type="leiomyosarcoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_71"
                     /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 2.1 kb.  "
ORIGIN

  Query Match             6.2%;  Score 284.6;  DB 5;  Length 1836;
  Best Local Similarity   75.5%;  Pred. No. 2.6e-38;
  Matches  361;  Conservative    0;  Mismatches  116;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | | | | | | | | | | | |||||||||||||||| |       | | | | |
Db        548 TCTNTTTCTCTNTTTCTTTTTTTGTACACACACACACACACTCTCTCTCTCTTTTTTC 489

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | | | ||| | | ||| |||||||| | |||||||||| ||| |||| || | |
Db        488 TATATACACACACACACACATATATATANATATATATATACATATATATATATATATA 429

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | |||||||| | |||||
Db        428 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 369

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db        368 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 309

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db        308 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 249

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
```

```
Db        248 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 189

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db        188 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 129

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATA 703
              || | ||| ||||| | | | ||| ||||| |||||||||| | ||| || | ||   |
Db        128 TATATATATATATATATATATATATATATATATATATATATATATATANATNTCCGNGNA 71


RESULT 79
CA364719/c
LOCUS       CA364719                683 bp    mRNA    linear   EST 06-NOV-2002
DEFINITION  639681 NCCCWA 1RT Oncorhynchus mykiss cDNA clone 1RT70B09_B_A05 5',
            mRNA sequence.
ACCESSION   CA364719
VERSION     CA364719.1  GI:24674188
KEYWORDS    EST.
SOURCE      Oncorhynchus mykiss (rainbow trout)
  ORGANISM  Oncorhynchus mykiss
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei;
            Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
REFERENCE   1  (bases 1 to 683)
  AUTHORS   Rexroad,C.E. 3rd, Lee,Y., Keele,J.W., Karamycheva,S., Brown,G.,
            Koop,B., Gahr,S.A., Palti,Y. and Quackenbush,J.
  TITLE     Sequence analysis of a rainbow trout cDNA library and creation of a
            gene index
  JOURNAL   Cytogenet. Genome Res. 102 (1-4), 347-354 (2003)
COMMENT     Contact: Rexroad CE
            USDA, ARS, National Center for Cool and Cold Water Aquaculture
            11876 Leetown Road, Kearneysville, WV 25430, USA
            Tel: 304 724 8340 x2129
            Fax: 304 725 0351
            Email: crexroad@ncccwa.ars.usda.gov
            Single pass sequencing. Bases called with phred v0.020425.c and
            trimmed with the aid of the trim_alt option. Vector identified by
            cross_match v0.990329.
            Seq primer: AGCGGATAACAATTTCACACAGGA.
FEATURES             Location/Qualifiers
     source          1. .683
                     /organism="Oncorhynchus mykiss"
                     /mol_type="mRNA"
                     /db_xref="taxon:8022"
                     /clone="1RT70B09_B_A05"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /clone_lib="NCCCWA 1RT"
                     /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
                     Library made from pooled tissue from brain, gill, liver,
                     spleen, muscle, and kidney."
ORIGIN

  Query Match             6.1%;  Score 282.2;  DB 6;  Length 683;
  Best Local Similarity   74.5%;  Pred. No. 7.5e-38;
  Matches  369;  Conservative    0;  Mismatches  123;  Indels    3;  Gaps    1;

Qy        229 TATATATATATATATATATATATACACACACACACACATATATATATACACATATATATA 288
              |||||||||||||| |||||||||| ||| ||| | |||||||| ||    | ||||||
Db        674 TATATATATATATACATATATATACATACATACATATATATATATGTATGTGTGTATATA 615

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              || | || |||| ||  | |||||| ||| || ||||||||| || |   | |||||
Db        614 TGTATGTGTATATATTTATGCGTATATATACACATGTATACACATGCACACACATACATA 555

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              | |||  | ||||||||| |||||||| | | | |||   |  |||||||| |||||||
Db        554 TGTATGTATATATGTATATACATATATATATACACATATGTATGTATATACATATATATA 495

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              ||||||  |||  | ||| ||| | ||||||| ||| | |||||  |||| | |||||||
Db        494 CACACACGTATGTATATACATATATATATATACACACACACGTATATATATATATA 435

Qy        469 CATATATACACATATATATA---ATATACACACATATATATACACATATATACACACATA 525
              ||| |   ||||||| |||   ||||| | | ||| | |||||||  |||| | | |||
Db        434 TATACACGTACATATACATACACATATATATATATACACATACACGTATATATATATA 375

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||||||||| ||| | ||||||||||   | |||    ||| ||| ||||||| ||| |
Db        374 TATACACATACATATATATATATATACGTGTGTATGTGTATACATATATATATATACACA 315

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              || | || |||| | | ||||||| ||||| ||  |||||||  ||||| | | | |||||
Db        314 TACACACGTATACATATATATATATATATACACGTATATATATATATACATATATATATA 255

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              |||||||||| |  |||||||||| ||||| |||||||||   || | ||||  ||||||
Db        254 TACACATACACACGCACACACATATATATATATATATATATGTATGTGTATATATATATA 195

Qy        706 TATGTATATATATAT 720
              |||  ||||   |||
Db        194 TATACATATGCGTAT 180


RESULT 80
CW518121
LOCUS       CW518121                663 bp    DNA     linear   GSS 08-OCT-2004
DEFINITION  OP__Ba0025C13.f OP__Ba Oryza punctata genomic clone OP__Ba0025C13
            5',genomic survey sequence.
ACCESSION   CW518121
VERSION     CW518121.1  GI:53992343
KEYWORDS    GSS.
SOURCE      Oryza punctata
  ORGANISM  Oryza punctata
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 663)
  AUTHORS   SanMiguel,P., Westerman,R., Kim,H., Yu,Y., Wissotski,M., Yost,D.,
            Stum,D., Rao,K., Luo,M., Jetty,R., Kudrna,D., Muller,C.,
            Hatfield,J., Soderlund,C., Wing,R. and Jackson,S.A.
  TITLE     OMAP Project - Purdue University
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Scott A. Jackson
            Jackson Laboratory
            Purdue University
            915 W. State St., West Lafayette, IN 47907, USA
            Tel: 7654963621
            Fax: 7654967255
            Email: sjackson@purdue.edu
            Basecalling by phred version 0.020425.c. This sequence was derived
            from the raw sequence read by clipping with lucy version 1.19s.
            Bases 50-712 of the raw sequence (length 1364) were retained after
            clipping.
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Insert Length: 161000   Std Error: 0.00
            Plate: 0025  row: C  column: 13
            Seq primer: TAA TAC GAC TCA CTA TAG GG
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .663
                     /organism="Oryza punctata"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4537"
                     /clone="OP__Ba0025C13"
                     /tissue_type="young leaves"
                     /lab_host="DH10B-T1 phage resistant"
                     /clone_lib="OP__Ba"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

  Query Match             6.1%;  Score 280.4;  DB 9;  Length 663;
  Best Local Similarity   77.2%;  Pred. No. 1.5e-37;
```

```
  Matches  366;  Conservative    0;  Mismatches  106;  Indels    2;  Gaps    2;

Qy        208 CAGAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACACAT 267
              || ||||||||||     |||     || |  ||  | | | |||||  | | | | | ||
Db        191 CAAAAAAAAAAAATCCAATTCACCATGTCAATGCAAAAAGATATATATATATATATATAT 250

Qy        268 ATATATATACACATATATATACACATATATACATATATACACATATATACACATATATAT 327
              |||||||||  | |||||||||  | |||||||  |||||||  | |||||||  | ||||||||
Db        251 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 310

Qy        328 ACACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATAC 387
              |  | |||  ||||  ||  | |||||||||  | ||||  ||||  | |||||||  | |||||||
Db        311 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 370

Qy        388 ACACATATATACACATATATACACACATATATACACATATATACACATATATACACATAT 447
              |  | |||||||  | |||||||  |||||||  | |||||||  | |||||||  | ||||
Db        371 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 430

Qy        448 ACACATATATACACATATATACATATATACACATATATATA-ATATACACACATATATAT 506
              |  | |||||||  | |||||||  |||||||  | |||||||||  |||||  |  | ||||||||
Db        431 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 490

Qy        507 ACACATATATACACACATATATACACATATATACACATATATATACACATATATACACAT 566
              |  | |||||||  |  |  ||||||||||||||||  |  |  |||||  |||  |  |||||  |  |  |
Db        491 ATATATATATATATATATATATATACACATATAAATAAATATAGATAGATATATA-ATATAA 549

Qy        567 ATATACATATATACACATATATATACATATATACACATATATACATATATACACATATAT 626
              |||||  |||||||  |  |||||||||  |||||||  |  |||||||  |||||||  |||||||
Db        550 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 609

Qy        627 ACATATATACACACATATATACACATACATATACACACACATAGATATACATAT 680
              |  |||||||  |  |  |||||||  |  |||  |||||  |  |  |  |||  |||||  ||||
Db        610 ATATATATATATATATATATATATATATATATATATATATATATATATATATATAAATAT 663


RESULT 81
BX149919/c
LOCUS       BX149919                737 bp    DNA     linear   GSS 13-MAR-2003
DEFINITION  Danio rerio genomic clone DKEY-112P5, genomic survey sequence.
ACCESSION   BX149919
VERSION     BX149919.1  GI:27981385
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 737)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAR-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the T7 end of BAC 112P5. 112P5 is
            part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .737
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-112P5"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             6.1%;  Score 280;  DB 9;  Length 737;
  Best Local Similarity   75.6%;  Pred. No. 1.7e-37;
  Matches  360;  Conservative    0;  Mismatches  115;  Indels    1;  Gaps    1;

Qy        230 ATATATATATATATATATATATACACACACACACACATATATATATACACATATATATAC 289
              |||||||||||||||||||||||| | | | | | |||| | |||| | |||||||||

Db        737 ATATATATATATATATATATATATATATATATATATATATATATGTGTATATATATATATATAT 678

Qy        290 ACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATAT 349
              | |||||||  |||||||  | |||||||  | |||||||||  | |||  ||||  ||  | ||||
Db        677 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 618

Qy        350 ATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATAC 409
              |||||  | ||||  ||||  | |||||||  | |||||||  |  | |||||||  | |||||||
Db        617 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 558

Qy        410 ACACATATATACACATATATACACATATATACACATATACACATATATACACATATATAC 469
              |  | |||||||  | |||||||  | |||||||  | |||||  |  | ||||  | |||||||
Db        557 ATATATATATATATATATATATATATATATATATATATATATATATGTGTATATATATATATAT 498

Qy        470 ATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATATAT 528
              |||||||  | |||||||||  |||||  | |  |||||||||  |||||||  |  | |||||||
Db        497 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 438

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACACATATAT 588
              |  | |||||||  | |||||||||  | |||||||  | |||||||  |||||||  | ||||||
Db        437 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 378

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
              |||  |||||||  | |||||||  |||||||  | |||||||  |||||||  |  | |||||||
Db        377 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 318

Qy        649 ACATACATATACACACACATAGATATACATATATATACACATATATATACGTATAT 704
              |  |  |  | | | | ||  | | | | |  |  |  |  |  |||  ||||
Db        317 AGCTTTTAAAAAAAAAACCAAAAAACAGAAAAGGTTGCCAAACTGTATAGAAATCT 262


RESULT 82
CL044834
LOCUS       CL044834                1031 bp    DNA     linear   GSS 31-DEC-2003
DEFINITION  CH216-61O18_RM1.1 CH216 Xenopus tropicalis genomic clone
            CH216-61O18, genomic survey sequence.
ACCESSION   CL044834
VERSION     CL044834.1  GI:40500747
KEYWORDS    GSS.
SOURCE      Xenopus tropicalis (western clawed frog)
  ORGANISM  Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 1031)
  AUTHORS   Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
            Mardis,E. and Wilson,R.
  TITLE     A physical map of the xenopus tropicalis genome
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 175000   Std Error: 0.00
            Seq primer: RM1 TACGACTCACTATAGGGAGA
            Class: BAC ends
            High quality sequence start: 47
            High quality sequence stop: 682.
FEATURES             Location/Qualifiers
     source          1. .1031
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /strain="Nigerian frog"
                     /db_xref="taxon:8364"
                     /clone="CH216-61O18"
                     /sex="male"
                     /cell_line="Stock 248 F7A2, inbred N7"
                     /clone_lib="CH216"
                     /note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis
                     BAC library"
ORIGIN

  Query Match             6.1%;  Score 280;  DB 9;  Length 1031;
```

```
  Best Local Similarity   73.4%;  Pred. No. 1.7e-37;
  Matches  373;  Conservative    0;  Mismatches  130;  Indels    5;  Gaps    1;

Qy        215 AAAAAAAGGCATTTTATATATATATATATATATATACACACACACACACATATATATA 274
              |||| |       |  ||| | | | ||| ||| ||||||||| ||| ||||| ||| |
Db        203 AAAATACATACATACATACACACACACATACATACATACACACATACATACATACATACA 262

Qy        275 TACACATATATATACACATATATACATATATACACATATATACACATATATATACACATA 334
              |||| | ||| |||||  | ||| | | ||| ||||| ||| ||||||| ||| | ||| |
Db        263 TACATACATACATACATACATAAAGACATACACACACATACACACATACATACATACACA 322

Qy        335 CATATGTACACATATATATACACATATGTATACACATATATACACATATATACACACATA 394
              ||||  |||||||| ||| | |||||  || | ||| | ||||||| ||| | ||| | |
Db        323 CATACATACACATACATACATACATACATACATACACACACATACACACATACATACATACA 382

Qy        395 TATACACATATATACACACATATATACACATATATACACATATATACACATATACACATA 454
              | | ||||| ||||| ||||| ||||| | | | | ||||| ||||||| ||||| | |
Db        383 CACATACATACATACAGACATACATACATACACACATACATACATACACACATACATACA 442

Qy        455 TATACACATATATACATATATACACATATATATAATATACACACATATATATACACATAT 514
              || ||||||| ||||||| | ||||| ||     | | ||| ||||| ||| |||||||
Db        443 TACACACATACATACATACACACACAT-----ACACACATACATACATACACACATAC 497

Qy        515 ATACACACATATATACACATATATACACATATATATACACATATATACACATATATACAT 574
              |||||||| | ||||| | ||| | ||||| ||| | ||||| ||||||| ||| | |
Db        498 ATACACACACACATACATACATACATACATACATACATACATACACACACATACATAC 557

Qy        575 ATATACACATATATATACATATATACACATATATACATATATACACATATATACATATAT 634
              ||| ||||||| ||| | | ||| ||||||| ||||||| | ||| | ||| | | |||
Db        558 ATACACACATACATACATACATACACACATACATACATACAGACATACATACATACATAC 617

Qy        635 ACACACATATATACACATACATATACACACACATAGATATACATATATATACACATATAT 694
              | ||||| ||| |||||||||||| | ||| | | | ||| ||| ||| ||||||| |||
Db        618 ATACACACATACACACATACATATACATACATAAAGACATACACACATACATACACACATAC 677

Qy        695 ATACGTATATATATGTATATATATATGC 722
              |||| || ||  ||  ||| | | || |
Db        678 ATACATACATTCATACATACACACATAC 705


RESULT 83
AQ506817/c
LOCUS       AQ506817                537 bp    DNA     linear   GSS 29-APR-1999
DEFINITION  RPCI-11-281J17.TV RPCI-11 Homo sapiens genomic clone
            RPCI-11-281J17, genomic survey sequence.
ACCESSION   AQ506817
VERSION     AQ506817.1  GI:4711564
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 537)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: RPCI-11-281J17.TJ
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genet cs (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: T7
            Class: BAC ends.
FEATURES            Location/Qualifiers
     source         1. .537
                    /organism="Homo sapiens"
                    /mol_type="genomic DNA"
                    /db_xref="GDB:7607752"
                    /db_xref="taxon:9606"
                    /clone="RPCI-11-281J17"
                    /sex="Male"
                    /cell_type="Lymphocytes"
                    /clone_lib="RPCI-11"
                    /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                    RPCI11 Human Male BAC Library"
ORIGIN

  Query Match             6.0%;  Score 278.8;  DB 8;  Length 537;
  Best Local Similarity   76.0%;  Pred. No. 2.9e-37;
  Matches  357;  Conservative    0;  Mismatches  112;  Indels    1;  Gaps    1;

Qy        230 ATATATATATATATATATATATACACACACACACATATATATATACACATATATATAC 289
              |||||||||||||||||| ||| | | ||| | | | ||||||||| | ||||| |||
Db        530 ATATATATATATATATATACATATATATACATATATACATATATATAAATATATACATAT 471

Qy        290 ACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATAT 349
              | ||||| | ||||||| | ||||| | | ||||| ||| | |||||||| || | ||||
Db        470 ATATATAAATATATATATATATATAAATATATATAAATATATATACATATATAAATATAT 411

Qy        350 ATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATAC 409
              ||||| | |||| || | | ||| ||| | ||||||| | | | ||||| | |||||||
Db        410 ATATAAATATATATAAATATATACATATATATATATATATATAAATATATAAATATATAT 351

Qy        410 ACACATATATACACATATATACACATATATACACATATACACATATATACACATATATAC 469
              | | |||||||| ||||||| || ||||||| | ||||||| ||||||| | |||||||
Db        350 ATAAATATATACATATATATCCATATATATAAATATATACATATATATAAATATATATAA 291

Qy        470 ATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATATAT 528
              ||||||| | ||||||| | ||||| ||| ||||  ||| | ||  ||| | | | ||||
Db        290 ATATATAAATATATATAAATATATATACATATATCAATATATATCAATATATATAAATAT 231

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACATATAT 588
              | |   |||||| | ||||||||| | ||||||| | ||||||| ||||||| | ||||||
Db        230 ATATCAATATATAAATATATATAAATATATATAAATATATATAAATATATAAATATATAT 171

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
              | | ||||||| | ||||||| ||||||| | ||||||| | ||||| | | ||||  |
Db        170 ACATATATATAAATATATATAAATATATAAATATATATATAAATATAAATATATATCAAT 111

Qy        649 ACATACATATACACACACATAGATATACATATATATACACATATATATAC 698
              | ||||||||| | | | ||| ||||| |||     | ||  |||||  | |
Db        110 ATATACATATATATATAAATATATATATATACTGTTTCATTTATATCTGC 61


RESULT 84
BX159146
LOCUS       BX159146                789 bp    DNA     linear   GSS 28-JAN-2003
DEFINITION  Danio rerio genomic clone DKEY-152P14, genomic survey sequence.
ACCESSION   BX159146
VERSION     BX159146.1  GI:27990759
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 789)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the SP6 end of BAC 152P14. 152P14
            is part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
```

```
          http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .789
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-152P14"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             6.0%;  Score 278.8;  DB 9;  Length 789;
  Best Local Similarity   75.5%;  Pred. No. 2.8e-37;
  Matches  373;  Conservative    0;  Mismatches  117;  Indels    4;  Gaps    2;

Qy        213 AAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACACATATATA 272
              ||| || |   |||||||||||||||| | | | ||| |   |||  |||||||| |||||
Db        300 AAATAAGATACCATTTTATATATATACACACACATACA-GTACATGCACACACACATATA 358

Qy        273 TATACACATATATATACACATATATACATATATACACATATATACACATATATATACACA 332
              |||||||| | | | |||||||||||| |||| |||  ||||| | | |||||| |
Db        359 TATACACACACACAAACACATATATATATTTACGTATATACATATATATATGTGTGTG 418

Qy        333 TACATATGTACACATATATATACACATATGTATACACATATATACACATATATACACACA 392
              || || ||||    ||| |||| | |||| |||| | |||||||| | ||||| |||||||
Db        419 TATATGTGTATGTGTATGTATATATATATATATATATATATATATATATATACACACACA 478

Qy        393 TATATACACATATATACACACATATATACACATATATACACATATATACACATATACACA 452
              | ||| | ||||||| | | ||||||| | ||||||| | ||||||| ||| | |||||
Db        479 CACATATATATATATATATATATATATATATATATATATATATATATATACACACACACA 538

Qy        453 TATATACACATATATACATATATACACATATATATAATATACACACATATATATACACAT 512
              |||||| | ||||||| |||| || | ||||||   ||||| | | |||||||||| | ||
Db        539 TATATATATATATATATATATGTATATATATAT---ATATATATATATATATATATATAT 595

Qy        513 ATATACACACATATATACACATATATACACATATATATACACATATATACACATATATAC 572
              ||||| | | |||||||||||| | | ||| |||||||||| | ||||||| | |||||||
Db        596 ATATATATATATATATACACACACACACATATATATATATATATATATATATATATAT 655

Qy        573 ATATATACACATATATATACATATATACACATATATACATATATACACATATATACATAT 632
              ||||||| | |||||||||| ||||| ||||| | | ||| | | ||||| | | ||| |
Db        656 ATATATATATATATATATATATATATACACACACACACACACACACACACACACACAC 715

Qy        633 ATACACACATATATACACATACATATACACACACATAGATATACATATATATACACATAT 692
              ||| | | |||||||| | ||| ||||| | | |   | ||||||||||| ||| | ||||
Db        716 ATATATATATATATATATATATATATATATATATATGTGTATATACATATACATATATATAT 775

Qy        693 ATATACGTATATAT 706
              |||||  |||||||
Db        776 ATATATATATATAT 789


RESULT 85
BX240321
LOCUS       BX240321                446 bp    DNA     linear   GSS 29-JAN-2003
DEFINITION  Danio rerio genomic clone DKEY-276D22, genomic survey sequence.
ACCESSION   BX240321
VERSION     BX240321.1  GI:28162655
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 446)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the SP6 end of BAC 276D22. 276D22
            is part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
```

```
          http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .446
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-276D22"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             6.0%;  Score 277.6;  DB 9;  Length 446;
  Best Local Similarity   79.5%;  Pred. No. 4.8e-37;
  Matches  353;  Conservative    0;  Mismatches   89;  Indels    2;  Gaps    2;

Qy        268 ATATATATACACATATATATACACATATATACATATATACACATATATACACATATATAT 327
              |||||||| | ||||||||| | ||||||| ||||||| | ||||||| | ||||||||
Db          4 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 63

Qy        328 ACACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATAC 387
              | | ||| | || |||| | ||| ||| | |||  |||| | ||||||| |||||||||
Db         64 ATATATATAAATATACATA-ATAAATAGAAATAGTTATATATATATATAGACATATATAT 122

Qy        388 ACACATATATACACATATATACACACATATATACACATATATACACATATATACACATAT 447
              | | ||||||| | | ||||| | | ||||||| | ||||| | | | ||||| | |||
Db        123 ATATATATATATATAGATATATATATATATATATATATATAGATAGACATATATATATAG 182

Qy        448 ACACATATATACACATATATACATATATACACATATATATA-ATATACACACATATATAT 506
              | | ||||||| | ||||||| ||||||| | ||||||||| ||||| | | |||||||||
Db        183 ATATATATATATATATATATATATATATATATATATATATAGATATATATATATATATAT 242

Qy        507 ACACATATATACACACATATATACACATATATACACATATATATACACATATATACACAT 566
              | | ||||||| | | ||||||| | ||||||| | ||||||||| | ||||||| | ||
Db        243 ATATATATATATATATATATATATATATATATATAGATATATATATATATATATATATAT 302

Qy        567 ATATACATATATACACATATATATACATATATACACATATATACATATATACACATATAT 626
              ||||| ||||||| | ||||||||| ||||||| ||||||| ||||||| | ||||||
Db        303 ATATATATATATATATATATATATATATATATATACATATATATATATATATATATATAT 362

Qy        627 ACATATATACACACATATATACACATACATATACACACACATAGATATACATATATATAC 686
              | ||||||| | | ||||||| | ||| ||||| | | |||||  ||||| |||||||||
Db        363 ATATATATATATATATATATATATATATATATATATATACATATATATATATATATAT 422

Qy        687 ACATATATATACGTATATATATGT 710
              | |||||||||  |||||||| |
Db        423 ATATATATATATATATATATATAT 446


RESULT 86
BX241290
LOCUS       BX241290                722 bp    DNA     linear   GSS 29-JAN-2003
DEFINITION  Danio rerio genomic clone DKEY-282A22, genomic survey sequence.
ACCESSION   BX241290
VERSION     BX241290.1  GI:28163624
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 722)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the T7 end of BAC 282A22. 282A22
            is part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .722
                     /organism="Danio rerio"
```

```
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-282A22"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             6.0%;  Score 276.6;  DB 9;  Length 722;
  Best Local Similarity   77.6%;  Pred. No. 6.6e-37;
  Matches  350;  Conservative    0;  Mismatches   94;  Indels    7;  Gaps    1;

Qy        250 ATACACACACACACACATATATATATACACATATATATACACATATATACATATATACAC 309
              ||||| |     |   |||| |||||| | ||||||||| | ||||||| ||||||| |
Db         13 ATACATATTATTATTTATATTTATATATATATATATATATATATATATATATATATATAT 72

Qy        310 ATATATACACATATATATACACATACATATGTACACATATATATACACATATGTATACAC 369
              ||||||| ||||||| ||||||| | | || |||| ||||||| | ||| || |||||||
Db         73 ATATATATACATATACATACACACATACATATACATATATATACATACACATATATACAC 132

Qy        370 ATATATACACATATATACACACATATATACACATATATACACACATATATACACATATAT 429
              ||| ||| ||||| ||| | | | ||| | ||||| ||||| ||||| ||||| | ||
Db        133 ATACATATACATACATATATATACATACATACATACATACATACATACATACATACATGC 192

Qy        430 ACACATATATACACATATACACATATATACACATATATACATATATACACATATATATA- 488
              | ||||  ||||| ||||| | ||||||||||| | ||||| ||| ||||||||||
Db        193 ATACATGCATACATATATATATATATATACACATACACACATACATATACATATATATAC 252

Qy        489 ------ATATACACACATATATATACACATATATACACACATATATACACATATATACAC 542
                   ||||| ||||||| ||||||| | ||||| | | ||||| | ||||| |||||
Db        253 ATACACATATATACACATACATATACATACATATATATATATATACATACATACATACAT 312

Qy        543 ATATATATACACATATATACACATATATACATATATACACATATATATACATATATACAC 602
              | ||| ||||| | ||| | ||||| ||||||| ||||| | ||| ||||||| ||| |
Db        313 ACATACATACATACATACATACATACATACATACATACATACATACATACATACATATAT 372

Qy        603 ATATATACATATATACACATATATACATATATACACACATATATACACATACATATACAC 662
              |||||||||||||||||||| ||||||||| | | | ||||||||||||||||| | ||
Db        373 ATATATACATATATACACATACATACATATACATATATATATATACACATACATACATAC 432

Qy        663 ACACATAGATATACATATATATACACATATA 693
              | ||||| ||| | | ||| ||||| | |||
Db        433 ATACATACATACATACATACATACATACATA 463


RESULT 87
CL763005/c
LOCUS       CL763005                617 bp    DNA     linear   GSS 27-JUL-2004
DEFINITION  OR_BBa0133F16.f OR_BBa Oryza rufipogon genomic clone OR_BBa0133F16
            5', genomic survey sequence.
ACCESSION   CL763005
VERSION     CL763005.1  GI:50720952
KEYWORDS    GSS.
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 617)
  AUTHORS   Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R.,
            Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
  TITLE     OMAP Project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Insert Length: 161   Std Error: 0.00
            Plate: 0133  row: F  column: 16
            Seq primer: TAA TAC GAC TCA CTA TAG GG
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .617
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
                     /clone="OR_BBa0133F16"
                     /tissue_type="young leaves"
                     /lab_host="DH10B-T1 phage resistant"
                     /clone_lib="OR_BBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

  Query Match             6.0%;  Score 276;  DB 9;  Length 617;
  Best Local Similarity   73.6%;  Pred. No. 8.6e-37;
  Matches  365;  Conservative    0;  Mismatches  130;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | |||| |||  |||||||||||| || |   | | | |  ||| ||| | || |||||||
Db        572 TATATACATAACTATATATATATCTATATGTATATATATCTATGTATCTGCATATATATA 513

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | |||| |   | ||||   |||||||    | | |||    || |||| |  |
Db        512 TATATATATCTTTTTTTATATTTATATATATCTCTCTCTATATTTTTATATATATGTATC 453

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | | || |||      |||||| | ||||||| | | ||||||| | |||||
Db        452 TATATATATAAAAATATATTTTTTTTATATATATATATATATATATATATATAAATATA 393

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||  | | | ||| |   ||||| | ||       ||||||| | |||||
Db        392 TATATATATATATTAAAAAAAATAAAATAATATATATATTATATAATATATATATATATA 333

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || || |||| | || ||||||  ||||| | | |||||||||  | |||||||  | | |||
Db        332 TATATTTATATATATTTATATATATATATATATATATATATATATATATATATATATATA 273

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | |||||||  | |||||||||  | |||||||  | |||||||  |||||||  | |||
Db        272 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 213

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              ||||||  |||||||  | |||||||  |||||||  | |||||||  |||||||  | |||||
Db        212 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 153

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| |||||||||  | |||||||||   ||||||
Db        152 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 93

Qy        706 TATGTATATATATATG 721
              ||| ||||||||||||
Db         92 TATATATATATATATG 77


RESULT 88
CL800151
LOCUS       CL800151                715 bp    DNA     linear   GSS 06-AUG-2004
DEFINITION  OR_CBa0011K05.r OR_CBa Oryza rufipogon genomic clone OR_CBa0011K05
            3', genomic survey sequence.
ACCESSION   CL800151
VERSION     CL800151.1  GI:51024771
KEYWORDS    GSS.
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 715)
  AUTHORS   Kim,H., Yu,Y., Wissotski,M., Yost,D., Stum,D., Rao,K., Luo,M.,
            Jetty,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and
            Wing,R.
```

```
  TITLE     OMAP project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0011  row: K  column: 05
            Seq primer: CAC TCA TTA GGC ACC CCA
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .715
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
                     /clone="OR_CBa0011K05"
                     /tissue_type="young leaves"
                     /dev_stage="2 week old seedlings"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OR_CBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII;
                     drk treated 36 hrs before harvest"
ORIGIN

  Query Match             6.0%;  Score 275.4;  DB 9;  Length 715;
  Best Local Similarity   80.3%;  Pred. No. 1.1e-36;
  Matches  335;  Conservative    0;  Mismatches   81;  Indels    1;  Gaps    1;

Qy        303 TATACACATATATACACATATATATACACATACATATGTACACATATATATACACATATG 362
              ||||  |||||||| | ||||||||| | ||| |||| || | ||||||||| | ||||
Db        300 TATAATCATATATATATATATATATATATATATATATATATATATATATATATATATATA 359

Qy        363 TATACACATATATACACATATATACACACATATATACACATATATACACACATATATACA 422
              |||| | ||||||| | ||||||| | | ||||||| | ||||||| | | ||||||| |
Db        360 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 419

Qy        423 CATATATACACATATATACACATATACACATATATACACATATATACATATATACACATA 482
              ||||||| | ||||||| | ||||| | ||||||| | ||||||| ||||||| | |||
Db        420 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 479

Qy        483 TATATAATATACACACATATATATACACATATATACACACATATATACACATATATACAC 542
              ||||| ||||| | | ||||||||| | ||||||| | | ||||||| | ||||||| |
Db        480 TATAT-ATATATATATATATATATATATATATATATATATATATATATATATATATATAT 538

Qy        543 ATATATATACACATATATACACATATATACATATATACACATATATATACATATATACAC 602
              ||||||||| | ||||||| | ||||||| ||||||| | ||||||||| ||||||| |
Db        539 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 598

Qy        603 ATATATACATATATACACATATATACATATATACACACATATATACACATACATATACAC 662
              ||||||| ||||||| | ||||||| ||||||| | | ||||||| | ||| ||||| |
Db        599 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 658

Qy        663 ACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATA 719
              | | ||| ||||| ||||||||| | |||||||||  ||||||||| ||||||||||
Db        659 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 715


RESULT 89
CL800151/c
LOCUS       CL800151                715 bp    DNA     linear   GSS 06-AUG-2004
DEFINITION  OR_CBa0011K05.r OR_CBa Oryza rufipogon genomic clone OR_CBa0011K05
            3', genomic survey sequence.
ACCESSION   CL800151
VERSION     CL800151.1  GI:51024771
KEYWORDS    GSS.
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 715)
  AUTHORS   Kim,H., Yu,Y., Wissotski,M., Yost,D., Stum,D., Rao,K., Luo,M.,
            Jetty,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and
            Wing,R.
  TITLE     OMAP project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0011  row: K  column: 05
            Seq primer: CAC TCA TTA GGC ACC CCA
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .715
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
                     /clone="OR_CBa0011K05"
                     /tissue_type="young leaves"
                     /dev_stage="2 week old seedlings"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OR_CBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII;
                     drk treated 36 hrs before harvest"
ORIGIN

  Query Match             6.0%;  Score 275.4;  DB 9;  Length 715;
  Best Local Similarity   80.3%;  Pred. No. 1.1e-36;
  Matches  335;  Conservative    0;  Mismatches   81;  Indels    1;  Gaps    1;

Qy        229 TATATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 288
              ||||||||||||||||||||||||| | | | | | |||||||||||| | |||||||||
Db        715 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 656

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |||
Db        655 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 596

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              |||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||||
Db        595 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 536

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||||
Db        535 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 476

Qy        469 CATATATACACATATATATAATATACACACATATATATACACATATATACACACATATAT 528
              ||||||| | |||||||| ||||| | | ||||||||| | ||||||| | | ||||||
Db        475 TATATATATATATATATAT-ATATATATATATATATATATATATATATATATATATAT 417

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACATATAT 588
              | | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | ||||||
Db        416 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 357

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              ||| ||||||| | ||||||| ||||||| | ||||||| ||||||| |    ||||
Db        356 ATATATATATATATATATATATATATATATATATATATATATATATATATATGATTATA 300


RESULT 90
BX158149
LOCUS       BX158149                705 bp    DNA     linear   GSS 28-JAN-2003
DEFINITION  Danio rerio genomic clone DKEY-151O11, genomic survey sequence.
ACCESSION   BX158149
```

```
VERSION     BX158149.1  GI:27989752
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 705)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the SP6 end of BAC 151O11. 151O11
            is part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .705
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-151O11"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             6.0%;  Score 275;  DB 9;  Length 705;
  Best Local Similarity   78.9%;  Pred. No. 1.2e-36;
  Matches  343;  Conservative    0;  Mismatches   85;  Indels    7;  Gaps    1;

Qy        266 ATATATATATACACATATATATACACATATATACATATATACACATATATACACATATAT 325
              |||| |||||| | ||||||||| | ||||||| ||||||| | ||||||| |||||||
Db         12 ATATTTATATATATATATATATATATATATATATATATATATATATATATACATATAC 71

Qy        326 ATACACATACATATGTACACATATATATACACATATGTATACACATATATACACATATAT 385
              ||||||| | | || |||| ||||||| | ||| || |||||||||| ||| ||||| ||
Db         72 ATACACACATACATATACATATATATACATACACATATATACACATACATATACATACAT 131

Qy        386 ACACACATATATACACATATATACACACATATATACACATATATACACATATATACACAT 445
              | | | ||| | ||||| ||||| ||||| ||||| | || | |||| ||||| ||
Db        132 ATATATACATACATACATACATACATACATACATACATGCATACATGCATACATAT 191

Qy        446 ATACACATATATACACATATATACATATATACACATATATATA-------ATATACACAC 498
              ||| | |||||||||||||| | ||||| ||| |||||||||||       ||||| ||||
Db        192 ATATATATATATACACATACACACATACATATACATATATATACATACACATATATACAC 251

Qy        499 ATATATATACACATATATACACACATATATACACATATATACACATATATATACACATAT 558
              ||| ||||||| | ||||| | | ||||| | ||||| ||||| | ||| ||||| | ||
Db        252 ATACATATACATACATATATATATATATACATACATACATACATACATACATACATAT 311

Qy        559 ATACACATATATACATATATACACATATATATACATATATACACATATATACATATATAC 618
              | | ||||| ||||||| ||||| | ||| ||||||| ||| | ||||||||||||||||
Db        312 ACATACATACATACATACATACATACATACATACATATATATATATACATATATAC 371

Qy        619 ACATATATACATATATACACACATATATACACATACATATACACACACATAGATATACAT 678
              ||||| ||||||||| | | | |||||||||||||||||||| | ||| ||||| ||| | |
Db        372 ACATACATACATATACATATATATATATACACATACATACATACATACATACATACATAC 431

Qy        679 ATATATACACATATA 693
              ||| |||||| | |||
Db        432 ATACATACATACATA 446


RESULT 91
BM906116/c
LOCUS       BM906116               1156 bp    mRNA    linear   EST 12-MAR-2002
DEFINITION  AGENCOURT_6621231 NIH_MGC_125 Homo sapiens cDNA clone IMAGE:5589966
            5', mRNA sequence.
ACCESSION   BM906116
VERSION     BM906116.1  GI:19356495
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1156)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Invitrogen
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12363  row: i  column: 07
            High quality sequence start: 323
            High quality sequence stop: 613.
FEATURES             Location/Qualifiers
     source          1. .1156
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5589966"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_125"
                     /note="Organ: ovary (pool of 3); Vector: pCMV-SPORT6;
                     Site_1: EcoRV (destroyed); Site_2: NotI; RNA source pool
                     of three ovaries, from females ranging in age from 38 to
                     49 yo. Library is oligo-dT primed and directionally cloned
                     (EcoRV site is destroyed upon cloning). Average insert
                     size 2.1 kb, insert size range 1-3.5 kb. Library is
                     normalized and enriched for full-length clones and was
                     constructed by C. Gruber (Invitrogen). Research Genetics
                     tracking code 036."
ORIGIN

  Query Match             5.9%;  Score 274.2;  DB 5;  Length 1156;
  Best Local Similarity   76.3%;  Pred. No. 1.6e-36;
  Matches  350;  Conservative    0;  Mismatches  108;  Indels    1;  Gaps    1;

Qy        268 ATATATATACACATATATATACACATATATACATATATACACATATATACACATATATAT 327
              |||| |||| | |||||||||| | ||||||| ||||||| | |||| || | || |||||
Db        688 ATATGTATATATATATATATATATATATATATATATATATATATATGTATATGTATAT 629

Qy        328 ACACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATAC 387
              | | ||| |||| || | ||||||||| | |||| |||| | ||||||| | |||||||
Db        628 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 569

Qy        388 ACACATATATACACATATATACACACATATATACACATATATACACATATATACACATAT 447
              | | ||||||| | ||||||| | | ||||||| | ||||||| | ||||||| | ||||
Db        568 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 509

Qy        448 ACACATATATACACATATATACATATATACACATATATATAATATACACACATATATATA 507
              | | ||||||| | ||||||| ||||||| | |||||||| ||||| | | |||||||||
Db        508 ATATATATATATATATATATATATATATATATATATATAT-ATATATATATATATATATA 450

Qy        508 CACATATATACACACATATATACACATATATACACATATATATACACATATATACACATA 567
              | ||||||| | | ||||||| | ||||||| | ||||||||| | ||||||| | |||
Db        449 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 390

Qy        568 TATACATATATACACATATATATACATATATACACATATATACATATATACACATATATA 627
              |||| ||||||| | ||||||||| ||||||| | ||||||| ||||||| | |||| |
Db        389 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATGTG 330

Qy        628 CATATATACACACATATATACACATACATATACACACACATAGATATACATATATATACA 687
              ||||||   | | |||| || | ||| |||| ||    | ||| || || |||||| ||
Db        329 TATATATCTATATATATGTATATATATATATGCATGTAGATATATCTATATATATGTATC 270

Qy        688 CATATATATACGTATATATATGTATATATATATGCTCCA 726
              || ||||||   |||||| | || |       | |||||
Db        269 TATTTATATATACATATATGTCTAGAGGATCCTCCTCCA 231
```

```
RESULT 92
CC075205/c
LOCUS       CC075205               556 bp    DNA     linear   GSS 16-APR-2003
DEFINITION  CSU-K33r.31C19.T7 CSU-K33r Aedes aegypti genomic clone
            CSU-K33r.31C19, genomic survey sequence.
ACCESSION   CC075205
VERSION     CC075205.1  GI:29916666
KEYWORDS    GSS.
SOURCE      Aedes aegypti (yellow fever mosquito)
  ORGANISM  Aedes aegypti
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Aedes;
            Stegomyia.
REFERENCE   1  (bases 1 to 556)
  AUTHORS   Loftus,B., Shetty,J., Severson,D., Brown,S. and Knudson,D.
  TITLE     End sequencing of Aedes aegypti BACs
  JOURNAL   Unpublished (2003)
COMMENT     Other_GSSs: CSU-K33r.31C19.SP6
            Contact: Brendan Loftus
            Department of Eukaryotic Genomics
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-3543
            Fax: 301-838-0208
            Email: enta@tigr.org
            Library was provided by Susan Brown and Dennis Knudson at Colorado
            State University.
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .556
                     /organism="Aedes aegypti"
                     /mol_type="genomic DNA"
                     /strain="Rexville"
                     /db_xref="taxon:7159"
                     /clone="CSU-K33r.31C19"
                     /clone_lib="CSU-K33r"
                     /note="Vector: pBeloBAC11; Site_1: HindIII"
ORIGIN

  Query Match             5.9%;  Score 273.4;  DB 8;  Length 556;
  Best Local Similarity   71.7%;  Pred. No. 2.4e-36;
  Matches  372;  Conservative    0;  Mismatches  146;  Indels    1;  Gaps    1;

Qy        230 ATATATATATATATATATATATACACACACACACATATATATATACACATATATATAC 289
              ||||||||||||||||||||||| | | | | | |  |||||||||| | |||||||||
Db        556 ATATATATATATATATATATATATATATATATATATCTATATATATATATATATATAT 497

Qy        290 ACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATAT 349
              | |||||| ||||||| | ||||||| | |||||||||| | ||| |||| || | ||||
Db        496 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 437

Qy        350 ATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATAC 409
              |||||  | ||||  |||| | ||||||| | ||||||| | | ||||||| | |||||||
Db        436 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 377

Qy        410 ACACATATATACACATATATACACATATATACACATATACACATATATACACATATATAC 469
              | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||||
Db        376 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 317

Qy        470 ATATATACACATATATATAATATACACACATATATATACACATATATACACACATATATA 529
              |||||||| |||||||| ||||| | | ||||||||| | ||||||| | | |||||||
Db        316 ATATATATATATATATAT-ATATATATATATATATATATATATATATATATATATATA 258

Qy        530 CACATATATACACATATATATACACATATATACACATATATACATATATACACATATATA 589
              | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||||||
Db        257 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 198

Qy        590 TACATATATACACATATATACATATATACACATATATACATATATACACACATATATACA 649
              || ||||||| | ||||||| |||||   | |||||  |   ||  ||  | | |  |
Db        197 TATATATATATATATATATATATATAAGTATATATAAATTAATCAACCTAGAAAAGTTTC 138

Qy        650 CATACATATATACACACACATAGATATACATATATATACACATATATATACGTATATATATG 709
              | |  | |        | |   | | |  || |||   |||  || ||  ||   ||
Db        137 AACAGCTGTCGTGGTGAAAAATAAAAAAAGTAGATAATAATAACTACACTAATCCCTAAT 78

Qy        710 TATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTT 748
              |||  ||  ||||   |          ||||  | | |||||
Db         77 TATCTACCTATGTAGCTTGTCCTTTAAGGTGAAACAGTT 39


RESULT 93
BX146163/c
LOCUS       BX146163               648 bp    DNA     linear   GSS 28-JAN-2003
DEFINITION  Danio rerio genomic clone DKEY-104K20, genomic survey sequence.
ACCESSION   BX146163
VERSION     BX146163.1  GI:27977521
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 648)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the SP6 end of BAC 104K20. 104K20
            is part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .648
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-104K20"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             5.9%;  Score 272.2;  DB 9;  Length 648;
  Best Local Similarity   70.4%;  Pred. No. 3.8e-36;
  Matches  392;  Conservative    0;  Mismatches  163;  Indels    2;  Gaps    2;

Qy        238 ATATATATATATATACACACACACACATATATATATACACATATATATACACATATAT 297
              |||||||||||||| | | | | | | |||||||||| | ||||||||| | ||||||
Db        648 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 589

Qy        298 ACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACAC 357
              | ||||||| | ||||||| | |||||||||| | ||| |||| || | |||||||||| |
Db        588 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 529

Qy        358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417
              ||||  | || | || | ||   ||||||| | | ||||||| | ||||||| | | ||||
Db        528 ATATATGTATATATGTGTATGTATATATATATATATATATATATATATATATATATAT 469

Qy        418 ATACACATATATACACATATATACACATATACACATATATACACATATATACATATATAC 477
              ||| | ||||||| | ||||||| | ||||| | ||||||| | ||||||| ||| ||||
Db        468 ATATATATATATATATATATATATATATATATATATATATATATATATATATAAATACATAC 409

Qy        478 ACATATATATAATATATACACACATATATATACACATATATACACACATATATACACATATA 537
              | |||||||| |||||||||| | | | |||| | | ||||||| | | ||||| | |
Db        408 ATATATATAT-ATATACACACACACACACACACACACACACACACACACACACACACACA 350

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATACATATA 597
              |||||  | ||||||| ||||||| | |||||||||||||| | ||||||||| |||||
Db        349 CACACACACATATACATATATATATATATATATACATATATATATATATATATATATA 290

Qy        598 TACACATATATACATATATACACATATATACATATATACACACATATATACACATACATA 657
              || | ||||||| ||||||| | ||||||| ||||||||| |||||| ||  |||   ||
Db        289 TATATATATATATATATATATATATATATATATATATACATACATATTTAATCATTTTTA 230
```

```
Qy        658 TACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATA 717
              |    |  || | || |      ||||  |  |   |  |  || |||| | | ||| ||
Db        229 T-TTTATTCAAATATTTTATATAATATTTATTTTATTTCATTTAAATATTTTTTTATTTA 171

Qy        718 TATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGA 777
               ||  |  |   |  |||    | | | || ||   |   |  ||| |   | | | |
Db        170 AATATTTAATTTGTTTAAATATTTGTATTTCATATGCTAGGTAAATATATATATATATA 111

Qy        778 GTTTGACAGCAGTGTAT 794
                |  | |    | |||
Db        110 TATATATATATATATAT 94


RESULT 94
CW020046/c
LOCUS       CW020046                892 bp    DNA     linear   GSS 23-SEP-2004
DEFINITION  ZMMBLc0013O08.f ZMMBLc Zea mays genomic clone ZMMBLc0013O08 5',
            genomic survey sequence.
ACCESSION   CW020046
VERSION     CW020046.1  GI:52613214
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 892)
  AUTHORS   Wing,R., Luo,M., Soderlund,C. and Haller,K.
  TITLE     ZMMBL sequences
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            Plate: 0013  row: O  column: 08
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .892
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /cultivar="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBLc0013O08"
                     /tissue_type="immature ears"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="ZMMBLc"
                     /note="Vector: pAGIBAC1; Site_1: Hpa II; Site_2: Hpa II"
ORIGIN

  Query Match             5.9%;  Score 271;  DB 9;  Length 892;
  Best Local Similarity   72.7%;  Pred. No. 5.8e-36;
  Matches  363;  Conservative    0;  Mismatches  135;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | |||||||||||||||||||||||| | | | | | | |||||||||||| | |||||||
Db        790 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 731

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | |||||||||                   |
Db        730 TATATATATATATATATATATATATATATATATATATATATAATATATATATATATTA 671

Qy        347 TATATATACACATATGTATACACATATATACACATATA-TACACACATATATACACATAT 405
              |||||||| | |||| |||| | ||||||| | ||||| || | | ||||||| | ||||
Db        670 TATATATATATATATATATATATATATATATATATATACTATATATATATATATATATAT 611

Qy        406 ATACACACATATATACACATATATACACATATATACACATATACACATATATACACATAT 465
              ||| | | ||||||| | ||||||| | ||||||| | ||||  | ||||||| | ||||
Db        610 ATATATATATATATATATATATATATATATATATATATATATTTATATATATATATAT 551
```

```
Qy        466 ATACATATATACACATATATATAATATACACACATATATATACACATATATACACACATA 525
              ||| ||||||| | |||||||||         | |||    || |    |||    | ||
Db        550 ATATATATATATATATATATATATATATATAATATAATATTATAATATTATTATTATATT 491

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
               ||    |||    |    |||||    ||||||| | ||||||| ||||||| | |||
Db        490 ATTAATATTATTATTATTATTATATTATTATATATATATATATATATATATATATATA 431

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db        430 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 371

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db        370 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 311

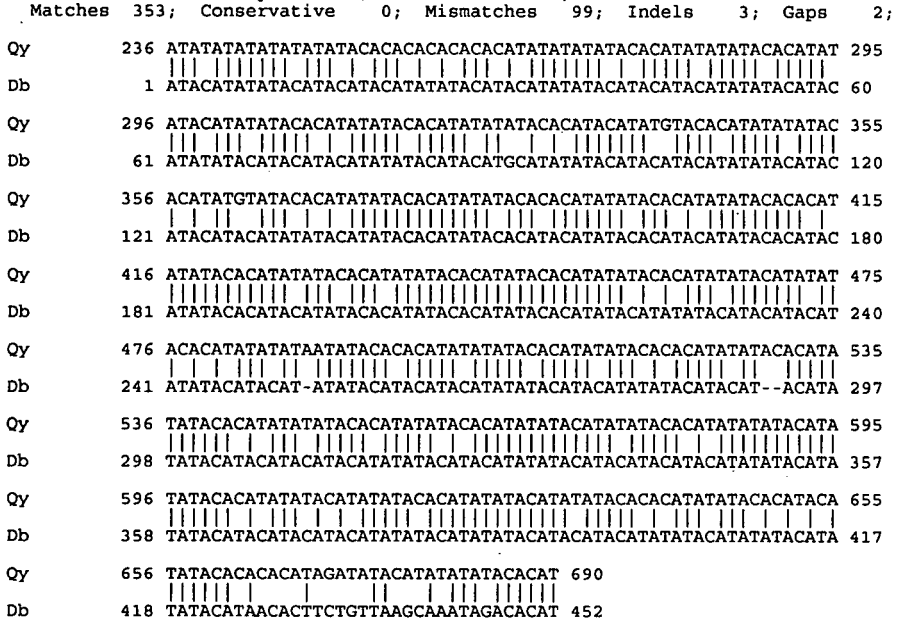
Qy        706 TATGTATATATATATGCTC 724
              ||| |||||||||||||  |
Db        310 TATATATATATATATGCAC 292


RESULT 95
AZ318983
LOCUS       AZ318983                647 bp    DNA     linear   GSS 29-SEP-2000
DEFINITION  1M0038D20F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0038D20 F, genomic survey sequence.
ACCESSION   AZ318983
VERSION     AZ318983.1  GI:10369306
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 647)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,
            Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von
            Niederhausern,A. and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0038  row: D  column: 20
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 647.
FEATURES             Location/Qualifiers
     source          1. .647
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0038D20"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
```

adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

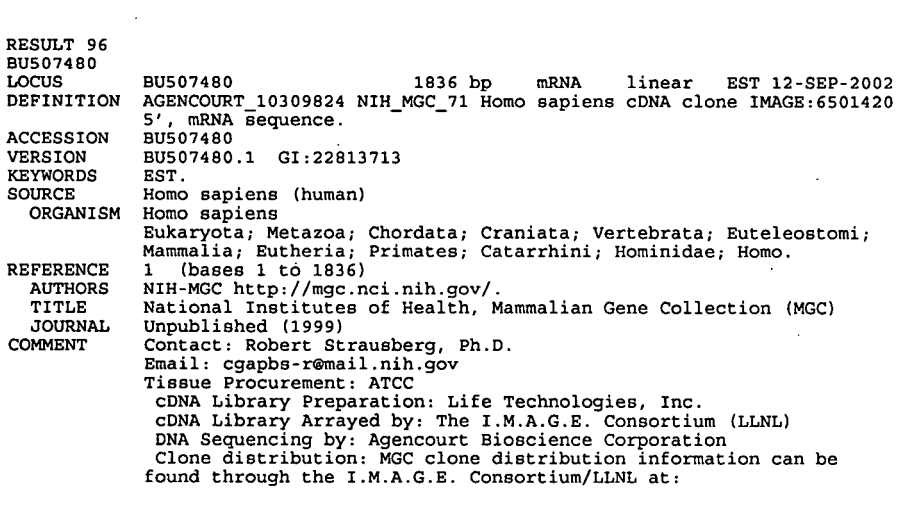
ORIGIN

Query Match 5.9%; Score 270.6; DB 8; Length 647;
Best Local Similarity 77.6%; Pred. No. 7e-36;
Matches 353; Conservative 0; Mismatches 99; Indels 3; Gaps 2;



RESULT 96
BU507480
LOCUS BU507480 1836 bp mRNA linear EST 12-SEP-2002
DEFINITION AGENCOURT_10309824 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:6501420 5', mRNA sequence.
ACCESSION BU507480
VERSION BU507480.1 GI:22813713
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1836)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM14055 row: n column: 13
High quality sequence start: 74
High quality sequence stop: 447.
FEATURES Location/Qualifiers
source 1. .1836
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6501420"
/tissue_type="leiomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_71"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2.1 kb. "
ORIGIN

Query Match 5.9%; Score 270.6; DB 5; Length 1836;
Best Local Similarity 72.5%; Pred. No. 6.2e-36;
Matches 359; Conservative 0; Mismatches 135; Indels 1; Gaps 1;



RESULT 97
BH329576/c
LOCUS BH329576 458 bp DNA linear GSS 03-DEC-2001
DEFINITION CH230-106A4.TV CHORI-230 Segment 1 Rattus norvegicus genomic clone CH230-106A4, genomic survey sequence.
ACCESSION BH329576
VERSION BH329576.1 GI:17260290
KEYWORDS GSS.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 458)

```
  AUTHORS   Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K.,
            Shvartsbeyn,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D.,
            Riggs,F., de Jong,P. and Fraser,C.M.
  TITLE     Rat BAC End Sequences from Library CHORI-230 EcoRI segment
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: CH230-106A4.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the rat BAC library CHORI-230
            (http://www.chori.org/bacpac/rat230.htm). For BAC library
            availability, please contact Pieter de Jong (pdejong@mail.cho.org).
            Clones may be purchased from BACPAC Resources
            (http://www.chori.org/bacpac/or ering_information.htm). BAC end
            page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
            Plate: 106  row: A  column: 4
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .458
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /strain="BN/SsNHsd/MCW"
                     /db_xref="taxon:10116"
                     /clone="CH230-106A4"
                     /sex="Female"
                     /cell_type="Brain"
                     /clone_lib="CHORI-230 Segment 1"
                     /note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;
                     CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
                     Pieter de Jong"
ORIGIN

  Query Match             5.8%;  Score 269.8;  DB 8;  Length 458;
  Best Local Similarity   77.2%;  Pred. No. 1e-35;
  Matches  353;  Conservative    0;  Mismatches  102;  Indels    2;  Gaps    2;

Qy        230 ATATATATATATATATATATATACACACACACACACATATATATATACACATATATATAC 289
              |||||||||||||||||||||| | | | |   |||||||||| | |||||||||
Db        458 ATATATATATATATATATATATATATATATATAGGGATATATATATATATATATATATAT 399

Qy        290 ACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATAT 349
              | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | ||||
Db        398 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 339

Qy        350 ATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATAC 409
              ||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||||
Db        338 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 279

Qy        410 ACACATATATACACATATATACACATATATACACATATACACATATATACACATATATAC 469
              | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||||
Db        278 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 219

Qy        470 ATATATACACATATATATAATATACACACATATATATACACATATATACACACATATATA 529
              ||||||| | |||||||| ||||| | | ||||||||| | ||||||| | | |||||||
Db        218 ATATATATATATATATAT-ATATATATATATATATATATATATATATATATATATATATA 160

Qy        530 CACATATATACACATATATATACACATATATACACATATATACATATATACACATATATA 589
                | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||||||
Db        159 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 100

Qy        590 TACATATATACACATATATACATATATACACATATATACATATAT-ACACACATATATAC 648
              || ||||||| | ||||||| ||||||| | |||||||  | | | ||||    | | ||
Db         99 TATATATATATATATATATATATATATATATATATATATATATTTCTCTCACACTATCACAGAC 40

Qy        649 ACATACATATACACACACATAGATATACATATATATA 685
              ||| | | | | | | || |||  | | | | | | |
Db         39 ACAGAGAGACAGAGAGACGTAGTGACAGACACAGAGA 3
```

```
RESULT 98
CW518121/c
LOCUS       CW518121                663 bp    DNA     linear   GSS 08-OCT-2004
DEFINITION  OP__Ba0025C13.f OP__Ba Oryza punctata genomic clone OP__Ba0025C13
            5', genomic survey sequence.
ACCESSION   CW518121
VERSION     CW518121.1  GI:53992343
KEYWORDS    GSS.
SOURCE      Oryza punctata
  ORGANISM  Oryza punctata
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 663)
  AUTHORS   SanMiguel,P., Westerman,R., Kim,H., Yu,Y., Wissotski,M., Yost,D.,
            Stum,D., Rao,K., Luo,M., Jetty,R., Kudrna,D., Muller,C.,
            Hatfield,J., Soderlund,C., Wing,R. and Jackson,S.A.
  TITLE     OMAP Project - Purdue University
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Scott A. Jackson
            Jackson Laboratory
            Purdue University
            915 W. State St., West Lafayette, IN 47907, USA
            Tel: 7654963621
            Fax: 7654967255
            Email: sjackson@purdue.edu
            Basecalling by phred version 0.020425.c. This sequence was derived
            from the raw sequence read by clipping with lucy version 1.19s.
            Bases 50-712 of the raw sequence (length 1364) were retained after
            clipping.
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Insert Length: 161000   Std Error: 0.00
            Plate: 0025  row: C  column: 13
            Seq primer: TAA TAC GAC TCA CTA TAG GG
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .663
                     /organism="Oryza punctata"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4537"
                     /clone="OP__Ba0025C13"
                     /tissue_type="young leaves"
                     /lab_host="DH10B-T1 phage resistant"
                     /clone_lib="OP__Ba"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

  Query Match             5.8%;  Score 268.8;  DB 9;  Length 663;
  Best Local Similarity   77.9%;  Pred. No. 1.4e-35;
  Matches  349;  Conservative    0;  Mismatches   97;  Indels    2;  Gaps    2;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              |||||||||||||||||||||||||| | | | | | |||||||||| | ||||||||
Db        660 TTTATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 601

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| || | ||   |
Db        600 TATATATATATATATATATATATATATATATATATATATATATATATATATTTATA-TTA 542

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              ||||| ||   |||| |||  | || | || | ||||||| | | ||||||| | |||||
Db        541 TATATCTATCTATATTTATTTATATGTGTATATATATATATATATATATATATATATATA 482

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||| | ||||| | ||||||| | |||||
Db        481 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 422

Qy        467 TACATATATACACATATATATAATATACACACATATATATACACATATATACACACATAT 526
              || ||||||| | |||||||| ||||| | | ||||||||| | ||||||| | | ||||
Db        421 TATATATATATATATATATAT-ATATATATATATATATATATATATATATATATATATAT 363
```

```
Qy        527 ATACACATATATACACATATATATACACATATATACACATATATACATATATACACACATAT 586
              ||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | ||||
Db        362 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 303

Qy        587 ATATACATATATACACATATATACATATATACACATATATACATATATACACACATATAT 646
              ||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | ||||||
Db        302 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 243

Qy        647 ACACATACATATACACACACATAGATAT 674
              | | ||| ||||       ||| || ||
Db        242 ATATATATATATCTTTTTGCATTGACAT 215


RESULT 99
CW019737
LOCUS       CW019737                600 bp    DNA     linear   GSS 23-SEP-2004
DEFINITION  ZMMBLc0013H15.r ZMMBLc Zea mays genomic clone ZMMBLc0013H15 3',
            genomic survey sequence.
ACCESSION   CW019737
VERSION     CW019737.1  GI:52612602
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 600)
  AUTHORS   Wing,R., Luo,M., Soderlund,C. and Haller,K.
  TITLE     ZMMBL sequences
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            Plate: 0013  row: H  column: 15
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .600
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /cultivar="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBLc0013H15"
                     /tissue_type="immature ears"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="ZMMBLc"
                     /note="Vector: pAGIBAC1; Site_1: Hpa II; Site_2: Hpa II"
ORIGIN

  Query Match             5.8%;  Score 267.2;  DB 9;  Length 600;
  Best Local Similarity   78.8%;  Pred. No. 2.7e-35;
  Matches  331;  Conservative    0;  Mismatches   88;  Indels    1;  Gaps    1;

Qy        228 TTATATATATATATATATATATATACACACACACACACATATATATATACACATATATAT 287
              || | ||  | | | |       | |||||||||||||||||||||||| | ||||||||
Db        182 TTCTCTAATTTTTTTTGCTCGAACACACACACACACACATATATATATATATATATATAT 241

Qy        288 ACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACAT 347
              | | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | ||
Db        242 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 301

Qy        348 ATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATAT 407
              ||||||| | |||| |||| | |||||||| | ||||||| | | ||||||| | ||||||
Db        302 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 361

Qy        408 ACACACATATATACACATATATACACATATATACACATATACACATATATACACATATAT 467
              | | | ||||||| | ||||||| | ||||||| | ||||| | |||||||   |||| |
Db        362 ATATATATATATATATATATATATATATATATATATATATATATATATATATATGTATATGT 421
```

```
Qy        468 ACATATATACACATATATATAATATACACACATATATATACACATATATACACACATATA 527
              | ||||||| | |||||||| ||||| | | ||||||||| | ||||||| | | |||||
Db        422 ATATATATATATATATATAT-ATATATATATATATATATATATATATATATATATATATA 480

Qy        528 TACACATATATACACATATATATACACATATATACACATATATACATATATACACATATA 587
              || | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||||
Db        481 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 540

Qy        588 TATACATATATACACATATATACATATATACACATATATACATATATACACACATATATA 647
              |||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||||
Db        541 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 600


RESULT 100
CC075205
LOCUS       CC075205                556 bp    DNA     linear   GSS 16-APR-2003
DEFINITION  CSU-K33r.31C19.T7 CSU-K33r Aedes aegypti genomic clone
            CSU-K33r.31C19, genomic survey sequence.
ACCESSION   CC075205
VERSION     CC075205.1  GI:29916666
KEYWORDS    GSS.
SOURCE      Aedes aegypti (yellow fever mosquito)
  ORGANISM  Aedes aegypti
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Aedes;
            Stegomyia.
REFERENCE   1  (bases 1 to 556)
  AUTHORS   Loftus,B., Shetty,J., Severson,D., Brown,S. and Knudson,D.
  TITLE     End sequencing of Aedes aegypti BACs
  JOURNAL   Unpublished (2003)
COMMENT     Other_GSSs: CSU-K33r.31C19.SP6
            Contact: Brendan Loftus
            Department of Eukaryotic Genomics
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-3543
            Fax: 301-838-0208
            Email: enta@tigr.org
            Library was provided by Susan Brown and Dennis Knudson at Colorado
            State University.
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .556
                     /organism="Aedes aegypti"
                     /mol_type="genomic DNA"
                     /strain="Rexville"
                     /db_xref="taxon:7159"
                     /clone="CSU-K33r.31C19"
                     /clone_lib="CSU-K33r"
                     /note="Vector: pBeloBAC11; Site_1: HindIII"
ORIGIN

  Query Match             5.8%;  Score 267;  DB 8;  Length 556;
  Best Local Similarity   71.4%;  Pred. No. 2.9e-35;
  Matches  365;  Conservative    0;  Mismatches  145;  Indels    1;  Gaps    1;

Qy        186 CAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATAT 245
              || |  |  || | |  |  | | | | || | |  | | ||  |||   ||| |
Db         47 CACCTTAAAGGACAAGCTACATAGGTAGATAATTAGGGATTAGTGTAGTTATTATTATCT 106

Qy        246 ATATATACACACACACACATATATATATACACATATATATACACATATATACATATAT 305
              |  | |          |||    | |  | |  | |  | | ||      ||  |  |||||
Db        107 ACTTTTTTTATTTTCACCACGACAGCTGTTGAAACTTTCTAGGTTGATTAATTTATAT 166

Qy        306 ACACATATATACACATATATATACACATACATATGTACACATATATATACACATATGTAT 365
              | || |||||  | |||||||| | ||| |||| || | |||||||| | |||| |||
Db        167 ATACTTATATATATATATATATATATATATATATATATATATATATATATATATATAT 226

Qy        366 ACACATATATACACATATATACACACATATATACACATATATACACACATATATACACAT 425
              | | |||||||  | |||||||  | | |||||||  | |||||||  | | |||||||  | ||
Db        227 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 286
```

```
Qy        426 ATATACACATATATACACATATACACATATATACACATATATACATATATACACATATAT 485
              ||||| | ||||||| | ||||| | |||||||| | ||||||| ||||||| | ||||||
Db        287 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 346

Qy        486 ATAATATACACACATATATATACACATATATACACACATATATACACATATATACACATA 545
              || ||||| | | ||||||||| | ||||||| | | ||||||| | ||||||| | |||
Db        347 AT-ATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 405

Qy        546 TATATACACATATATACACATATATACATATATACACATATATATACATATATACACATA 605
              |||||| | ||||||| | ||||||| ||||||| | ||||||||| ||||||| | |||
Db        406 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 465

Qy        606 TATACATATATACACATATATACATATATACACACATATATACACATACATATACACACA 665
              |||| ||||||| | ||||||| ||||||| | | ||||||| | ||| ||||| | | |
Db        466 TATATATATATATATATATATATATATATATATATATATATATATATATATATAGATATA 525

Qy        666 CATAGATATACATATATATACACATATATAT 696
                ||| ||||| ||||||||| | ||||||||
Db        526 TATATATATATATATATATATATATATATAT 556
```

Search completed: April 22, 2005, 04:38:54
Job time : 9260.96 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 20, 2005, 22:49:00 ; Search time 1342.5 Seconds
(without alignments)
18171.562 Million cell updates/sec

Title: US-09-939-209A-3_COPY_1_4121
Perfect score: 4121
Sequence: 1 agttcaagaccagcctgagc.....tgtgagggcaaattgaagtc 4121

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_16Dec04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	4121	100.0	20300	6	ABK47337	Abk47337 Genomic n
c	2	356.6	8.7	121600	6	ABT10748	Abt10748 Human bre
c	3	356.6	8.7	121600	12	ADQ18307	Adq18307 Human sof
c	4	327.4	7.9	188971	12	ADL08108	Adl08108 Human gen
	5	323.6	7.9	163382	13	ABD32659	Abd32659 Human can
c	6	322.2	7.8	26997	4	AAS46748	Aas46748 Tumour su
c	7	311.2	7.6	26997	4	AAS46747	Aas46747 Tumour su
	8	308	7.5	1990	10	ADC87572	Adc87572 Human GPC
	9	296	7.2	15418	3	AAA63785	Aaa63785 Nucleotid
	10	296	7.2	15418	6	AAL38601	Aal38601 Human TER
	11	296	7.2	15418	6	ABS54997	Abs54997 Lambda cl
	12	296	7.2	15418	6	AAL38595	Aal38595 DNA of pl
	13	296	7.2	15418	10	ADC21253	Adc21253 Lambda cl
c	14	294.8	7.2	1486	10	ADC87552	Adc87552 Human GPC
	15	293.6	7.1	27067	10	AAD50021	Aad50021 Human sec
c	16	291.6	7.1	1672	10	ADC87570	Adc87570 Human GPC
	17	288	7.0	30393	4	AAK67239	Aak67239 Human imm
c	18	286.6	7.0	44147	6	ABK84481	Abk84481 Human cDN
c	19	286.6	7.0	44147	10	ADD14691	Add14691 Human src
	20	280.8	6.8	32206	4	AAK89374	Aak89374 Human dig
	21	270.4	6.6	20420	4	AAK73165	Aak73165 Human imm
	22	270.4	6.6	20420	6	ABK69933	Abk69933 Human sec
c	23	268.6	6.5	110000	13	ABD32806_4	Continuation (5 of
c	24	268.6	6.5	191127	10	ADH13788	Adh13788 Prostate
	25	267.6	6.5	110000	13	ABD32806_4	Continuation (5 of
	26	267.4	6.5	19965	4	AAK73166	Aak73166 Human imm
	27	267.4	6.5	19965	6	ABK69932	Abk69932 Human sec
c	28	266.4	6.5	6971	6	ABL33237	Abl33237 Human imm
c	29	263.6	6.4	2522	11	ADM01858	Adm01858 Human cDN
c	30	263	6.4	117962	8	AAD54480	Aad54480 Human CIP
	31	263	6.4	165199	6	ABK83460	Abk83460 Human cDN
c	32	263	6.4	227968	6	ABK83497	Abk83497 Human cDN
c	33	263	6.4	227968	12	ADQ18538	Adq18538 Human sof
c	34	261.4	6.3	12069	6	ABK39930	Abk39930 Human che
c	35	260.6	6.3	20420	4	AAK73165	Aak73165 Human imm
c	36	260.6	6.3	20420	6	ABK69933	Abk69933 Human sec
	37	258.8	6.3	1489	10	ADC87554	Adc87554 Human GPC
c	38	258.8	6.3	2761	12	ADO15781	Ado15781 4 synthes
	39	258.8	6.3	201766	10	ADL13771	Adl13771 Osteoarth
c	40	258.6	6.3	163382	13	ABD32659	Abd32659 Human can
	41	258.4	6.3	82938	6	ABV72623	Abv72623 Human tra
c	42	258	6.3	19965	4	AAK73166	Aak73166 Human imm
c	43	258	6.3	19965	6	ABK69932	Abk69932 Human sec
	44	257.6	6.3	12069	6	ABK39930	Abk39930 Human che
	45	256.4	6.2	1486	10	ADC87552	Adc87552 Human GPC

ALIGNMENTS

RESULT 1
ABK47337
ID ABK47337 standard; DNA; 20300 BP.
XX
AC ABK47337;
XX
DT 18-JUN-2002 (first entry)
XX
DE Genomic nucleotide sequence encoding human RGS-4 protein.
XX
KW RGS-4; schizophrenia; human; regulator of G protein signalling 4;
KW neuroleptic; polymorphism; schizotypy; schizoid; chromosome 1q21-22;
KW gene; ds.
XX
OS Homo sapiens.
XX

FH	Key	Location/Qualifiers
FT	variation	replace(4121,T)
FT		/*tag= a
FT		/standard_name= "Single-nucleotide polymorphism"
FT	variation	replace(4123,A)
FT		/*tag= b
FT		/standard_name= "Single-nucleotide polymorphism"
FT	variation	replace(4368,C)
FT		/*tag= c
FT		/standard_name= "Single-nucleotide polymorphism"
FT	variation	replace(4621,C)
FT		/*tag= d
FT		/standard_name= "Single-nucleotide polymorphism"
FT	variation	replace(4790,T)
FT		/*tag= e
FT		/standard_name= "Single-nucleotide polymorphism"
FT	variation	replace(4816,T)
FT		/*tag= f
FT		/standard_name= "Single-nucleotide polymorphism"
FT	variation	replace(4970,T)
FT		/*tag= g
FT		/standard_name= "Single-nucleotide polymorphism"
FT	variation	replace(5055,G)
FT		/*tag= h
FT		/standard_name= "Single-nucleotide polymorphism"
FT	variation	replace(5295,A)
FT		/*tag= i

```
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(5695,A)
FT                   /*tag= j
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(7375,T)
FT                   /*tag= k
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(7759,A)
FT                   /*tag= l
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(8596,A)
FT                   /*tag= m
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(9602. .9610,CA)
FT                   /*tag= n
FT                   /note= "Deletion polymorphism"
FT   variation       replace(9892,A)
FT                   /*tag= o
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(9963,A)
FT                   /*tag= p
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(10132,A)
FT                   /*tag= q
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(11056,C)
FT                   /*tag= r
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(11091,T)
FT                   /*tag= s
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(11106,A)
FT                   /*tag= t
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(11774,T)
FT                   /*tag= u
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(12143,A)
FT                   /*tag= v
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(12145,T)
FT                   /*tag= w
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(14367,G)
FT                   /*tag= x
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(17027. .17029,GA)
FT                   /*tag= y
FT                   /note= "Deletion polymorphism"
FT   variation       replace(17630,T)
FT                   /*tag= Z
FT                   /standard_name= "Single-nucleotide polymorphism"
XX
PN   WO200216653-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US026622.
XX
PR   24-AUG-2000; 2000US-0228021P.
XX
PA   (UYPI-) UNIV PITTSBURGH.
XX
PI   Levitt PR,  Mirnics K,  Kodavali VC,  Nimgaonkar VL;
XX
DR   WPI; 2002-292070/33.
XX
PT   Diagnosing, assessing susceptibility and treating schizophrenia, involves
PT   observing regulator of G-protein signaling 4, RGS4 levels in a subject.
XX
PS   Claim 1; Page 20-33; 112pp; English.
XX
CC   This invention relates to a novel method for diagnosing schizophrenia or
CC   determining susceptibility to schizophrenia in a human. The method
CC   comprises obtaining from a patient a DNA sample and detecting variations
CC   in the regulator of G-protein signalling 4 (RGS4) gene. Alternatively,
CC   the method involves measuring RGS4 mRNA or protein levels in a tissue
CC   sample from the patient and determining if there is a reduced level. The
CC   method of the invention is useful for diagnosing and determining
CC   susceptibility to schizophrenia. The invention also comprises a method
CC   that is useful for treating schizophrenia which includes a prophylactic
CC   treatment. The method of genotyping polymorphic variants in the RGS-4
CC   gene is applied to diagnosing pathologies of the schizophrenic spectrum,
CC   such as in particular schizotypy, schizoid individuals, etc. This method
CC   offers the possibility of diagnosing schizophrenia by a biological test
CC   and no longer exclusively by clinical evaluations. The present sequence
CC   represents the genomic DNA encoding the human regulator of G-protein
CC   signalling 4 (RGS4) protein used in the method of the invention. The gene
CC   for the RGS4 protein is located on human chromosome 1q21-22
XX
SQ   Sequence 20300 BP; 6157 A; 4102 C; 3775 G; 6266 T; 0 U; 0 Other;

  Query Match           100.0%;  Score 4121;  DB 6;  Length 20300;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 4121;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAATTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAATTA 60

Qy         61 GACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAAT 120

Qy        121 CACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAGC 180

Qy        181 CCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATA 240

Qy        241 TATATATATATACACACACACACATATATATATACACATATATATACACATATATACA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TATATATATATACACACACACACATATATATATACACATATATATACACATATATACA 300

Qy        301 TATATACACATATATACACATATATATACACATACATATGTACACATATATATACACATA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TATATACACATATATACACATATATATACACATACATATGTACACATATATATACACATA 360

Qy        361 TGTATACACATATATACACATATATACACACATATATACACATATATACACACATATATA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGTATACACATATATACACATATATACACACATATATACACATATATACACACATATATA 420

Qy        421 CACATATATACACATATATACACATATACACATATATACACATATATACATATATACACA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CACATATATACACATATATACACATATACACATATATACACATATATACATATATACACA 480

Qy        481 TATATATAATATACACACATATATATACACATATATACACACATATATACACATATATAC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TATATATAATATACACACATATATATACACATATATACACACATATATACACATATATAC 540

Qy        541 ACATATATATACACATATATACACATATATACATATATACACATATATATACATATATAC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ACATATATATACACATATATACACATATATACATATATACACATATATATACATATATAC 600

Qy        601 ACATATATACATATATACACATATATACATATATACACACATATATACACATACATATAC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ACATATATACATATATACACATATATACATATATACACACATATATACACATACATATAC 660

Qy        661 ACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATAT 720

Qy        721 GCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        721 GCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTT 780
Qy        781 TGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATAC 840
Db        781 TGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATAC 840
Qy        841 ATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATATATATA 900
Db        841 ATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATATATATA 900
Qy        901 TAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTATAGAAGAATGG 960
Db        901 TAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTATAGAAGAATGG 960
Qy        961 CTAAGATGCCAAAGCTCAGTGTATGTGGCAGAGGCATGGTGTAGGGTGTGTCCAGGTTCA 1020
Db        961 CTAAGATGCCAAAGCTCAGTGTATGTGGCAGAGGCATGGTGTAGGGTGTGTCCAGGTTCA 1020
Qy       1021 TATATTGCATTAAGTGTGAGAACACCCTGGAGTATGAACCAAGAAAATGCAAAAGCCAGA 1080
Db       1021 TATATTGCATTAAGTGTGAGAACACCCTGGAGTATGAACCAAGAAAATGCAAAAGCCAGA 1080
Qy       1081 AGTGATGGAGGAAATGAGACACAATAATGAAGATATTGAGAGGAGGGTGTGGGCCTAGAG 1140
Db       1081 AGTGATGGAGGAAATGAGACACAATAATGAAGATATTGAGAGGAGGGTGTGGGCCTAGAG 1140
Qy       1141 TGAAGCTTTTCGTGCCAGTACTTCTTTTGAAGGCCCAGTTCTCTTCTCTCTCGGGGGCTC 1200
Db       1141 TGAAGCTTTTCGTGCCAGTACTTCTTTTGAAGGCCCAGTTCTCTTCTCTCTCGGGGGCTC 1200
Qy       1201 CTTCATCTCTCATAGAGTCCACAGCTTTTAAGGGCCAACACTTGAGGTCAGCCTGGCTCT 1260
Db       1201 CTTCATCTCTCATAGAGTCCACAGCTTTTAAGGGCCAACACTTGAGGTCAGCCTGGCTCT 1260
Qy       1261 CTCATTTGAGCTGGATAGAACATTTTAGAGCACCATCTATTCTTCAAGAGGAAGTTTAAA 1320
Db       1261 CTCATTTGAGCTGGATAGAACATTTTAGAGCACCATCTATTCTTCAAGAGGAAGTTTAAA 1320
Qy       1321 AATAAAAGAACCTTGAAGAGGAAAAAATGTAGACATTCAATCTAACCTTTTCATTTTACT 1380
Db       1321 AATAAAAGAACCTTGAAGAGGAAAAAATGTAGACATTCAATCTAACCTTTTCATTTTACT 1380
Qy       1381 AGCCAAAGCTAAATAGAATGCAGGTTACCTGTTTTTCAGCCAGGCACCATCATTTCCTAA 1440
Db       1381 AGCCAAAGCTAAATAGAATGCAGGTTACCTGTTTTTCAGCCAGGCACCATCATTTCCTAA 1440
Qy       1441 TTGTTATAAAATTTATTATTATTGTTGTTATTATTATTATTTGCCATAAGAAGTTTCCCA 1500
Db       1441 TTGTTATAAAATTTATTATTATTGTTGTTATTATTATTATTTGCCATAAGAAGTTTCCCA 1500
Qy       1501 TATCCTTTTAGTATAACAAAAACACAATTCACAAGCATTATAAAACCCATGGTGTCTAAC 1560
Db       1501 TATCCTTTTAGTATAACAAAAACACAATTCACAAGCATTATAAAACCCATGGTGTCTAAC 1560
Qy       1561 TATTAAAAAAAATTAAGTGGAACACACTTGTCCCAGCTACTGGGGAGGCTGAGGAGGGAGG 1620
Db       1561 TATTAAAAAAAATTAAGTGGAACACACTTGTCCCAGCTACTGGGGAGGCTGAGGAGGGAGG 1620
Qy       1621 ATCACGTGATCCCAGGGGGTCAAGGTTATGGAGAGCTATGATTGTGCCACTGCACTCCAG 1680
Db       1621 ATCACGTGATCCCAGGGGGTCAAGGTTATGGAGAGCTATGATTGTGCCACTGCACTCCAG 1680
Qy       1681 CCTGGGTGACAGGGAAAGACCCTGTCTCTAAAATTTTTTTTAAAAAAACTAAACTGGTTT 1740
Db       1681 CCTGGGTGACAGGGAAAGACCCTGTCTCTAAAATTTTTTTTAAAAAAACTAAACTGGTTT 1740
Qy       1741 TATTACAGAGATTCTGGAGACAGCTACACATAAAAGGGTGGTATGCCTCATATTAGCTAC 1800
Db       1741 TATTACAGAGATTCTGGAGACAGCTACACATAAAAGGGTGGTATGCCTCATATTAGCTAC 1800
Qy       1801 CCAGGGAGGTGGAATGCCAACTTAGGTGGTGTCACCACTATTAAAAATGCCCCAAAGCAA 1860
Db       1801 CCAGGGAGGTGGAATGCCAACTTAGGTGGTGTCACCACTATTAAAAATGCCCCAAAGCAA 1860

Qy       1861 TCAAAACTGAGAACTTCCTGGGAGCTTAGCATTGTGCAAAAGCAGCACAAAACACTTAAA 1920
Db       1861 TCAAAACTGAGAACTTCCTGGGAGCTTAGCATTGTGCAAAAGCAGCACAAAACACTTAAA 1920
Qy       1921 CAATTCACAGTTGTGTTGGAATGGGAAGGCCTGGAAATATAAACCAAAGAGTATATTGTC 1980
Db       1921 CAATTCACAGTTGTGTTGGAATGGGAAGGCCTGGAAATATAAACCAAAGAGTATATTGTC 1980
Qy       1981 TAAATTGATAGAGATTACAATTGCCTGAAAGAAAAAGTTGACTTTTAACTAGAATGTTCA 2040
Db       1981 TAAATTGATAGAGATTACAATTGCCTGAAAGAAAAAGTTGACTTTTAACTAGAATGTTCA 2040
Qy       2041 GAGTAGGTTTACAGAAGAAGCTCTTAAACTGGGCTCCAGTGGATTTGTCAATGCTTTGGA 2100
Db       2041 GAGTAGGTTTACAGAAGAAGCTCTTAAACTGGGCTCCAGTGGATTTGTCAATGCTTTGGA 2100
Qy       2101 AGCTGGTGGGGTGGGAGGGTTGGAGGGGGCATAAAAAAGTCATGTTGGTATGCTCTGCTCA 2160
Db       2101 AGCTGGTGGGGTGGGAGGGTTGGAGGGGGCATAAAAAGTCATGTTGGTATGCTCTGCTCA 2160
Qy       2161 AGTCTCCATTCTGTTTCCTTTTCCTCTTTTCAATGTCATGTCCCATTATTTCATTATGGG 2220
Db       2161 AGTCTCCATTCTGTTTCCTTTTCCTCTTTTCAATGTCATGTCCCATTATTTCATTATGGG 2220
Qy       2221 CTTCCCTTTATCCAGGATCAATATGCCACCTCTTGGTTGTCTTTTACCTACTTCTCCACC 2280
Db       2221 CTTCCCTTTATCCAGGATCAATATGCCACCTCTTGGTTGTCTTTTACCTACTTCTCCACC 2280
Qy       2281 TCACTATGGAATCGTCCTTGGGTAGCTCCTGTGCTTGGGAACCTGCACGGGCACTTTTCT 2340
Db       2281 TCACTATGGAATCGTCCTTGGGTAGCTCCTGTGCTTGGGAACCTGCACGGGCACTTTTCT 2340
Qy       2341 GATGTCTTGATTCCAGCTTTACTCCTAAAACTTAAATGCTGAGGGGCCAACACCATGGCA 2400
Db       2341 GATGTCTTGATTCCAGCTTTACTCCTAAAACTTAAATGCTGAGGGGCCAACACCATGGCA 2400
Qy       2401 GTGGTAGGGATGGGAATGGGGGTCTTGTAACACACTACATAAACTACACGAAATAAACTA 2460
Db       2401 GTGGTAGGGATGGGAATGGGGGTCTTGTAACACACTACATAAACTACACGAAATAAACTA 2460
Qy       2461 CATGAAACTCAACATGTTTGCAAGACTCAGTTCACATCCATGAGGAGCTCATGCTTCTCC 2520
Db       2461 CATGAAACTCAACATGTTTGCAAGACTCAGTTCACATCCATGAGGAGCTCATGCTTCTCC 2520
Qy       2521 CTCCTGCTCCCCTAGCACACATGATTATCTCTATTTGGAAATGTTTGGCATTTTTGGTGA 2580
Db       2521 CTCCTGCTCCCCTAGCACACATGATTATCTCTATTTGGAAATGTTTGGCATTTTTGGTGA 2580
Qy       2581 AGTGAATGGTTCAATAACTTTCTCCACCATCAGAACAAAAGCTCTTTAAGGTTAGGGATG 2640
Db       2581 AGTGAATGGTTCAATAACTTTCTCCACCATCAGAACAAAAGCTCTTTAAGGTTAGGGATG 2640
Qy       2641 GGATCATACACACTTCCCTTGTCCAAGTCCCCATCACCCCTTATCTAGACAATTGCTACA 2700
Db       2641 GGATCATACACACTTCCCTTGTCCAAGTCCCCATCACCCCTTATCTAGACAATTGCTACA 2700
Qy       2701 GTTTCCTACACACTCTTCTAACCTCTTGCAGTCTATTTTCATAAAACAGCTAGAGAACTT 2760
Db       2701 GTTTCCTACACACTCTTCTAACCTCTTGCAGTCTATTTTCATAAAACAGCTAGAGAACTT 2760
Qy       2761 TGAGATGTAAGTCAAAAAATAGAACATGTCGCTCTTTCCCATTGTTTTTGAAATAAAGTT 2820
Db       2761 TGAGATGTAAGTCAAAAAATAGAACATGTCGCTCTTTCCCATTGTTTTTGAAATAAAGTT 2820
Qy       2821 CAACCCCCTTACCAGGGTCAACAAGGCCCTGCAATGATTTGGTCCTGTTAAAAATTCTTT 2880
Db       2821 CAACCCCCTTACCAGGGTCAACAAGGCCCTGCAATGATTTGGTCCTGTTAAAAATTCTTT 2880
Qy       2881 AGCCTTAACTCATGCTGTTCTTCCTTACACTCACTGCATTCTAGCCATTGAGGTTTCTAT 2940
Db       2881 AGCCTTAACTCATGCTGTTCTTCCTTACACTCACTGCATTCTAGCCATTGAGGTTTCTAT 2940
```

```
Qy       2941 GCATCAAACTTTTTTTGGTCCCAGCACTGTGCACATCCTTCTGGGTAGAATGCCCCTTGA 3000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2941 GCATCAAACTTTTTTTGGTCCCAGCACTGTGCACATCCTTCTGGGTAGAATGCCCCTTGA 3000

Qy       3001 TTTGTATAATTAGCACCTCCTTCATCATTTAGGTCTTAGTATAACTACTACCTTCTTAGA 3060
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3001 TTTGTATAATTAGCACCTCCTTCATCATTTAGGTCTTAGTATAACTACTACCTTCTTAGA 3060

Qy       3061 GAAGCTCTGCTTCTTCATCCTATAAAAAAGTAAAATTCCTTACCCTGTTATTTTTTAAGT 3120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3061 GAAGCTCTGCTTCTTCATCCTATAAAAAAGTAAAATTCCTTACCCTGTTATTTTTTAAGT 3120

Qy       3121 CATCCGTGTTTCATTCTGTTAAAGTTCTTATCACAATTTATCATTATTTTATTTACAGTC 3180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3121 CATCCGTGTTTCATTCTGTTAAAGTTCTTATCACAATTTATCATTATTTTATTTACAGTC 3180

Qy       3181 ATGTGCCACATAACAATGTTTCAGTCAGGGATAGAACACAAATGTATCTGGCCCCATAAT 3240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3181 ATGTGCCACATAACAATGTTTCAGTCAGGGATAGAACACAAATGTATCTGGCCCCATAAT 3240

Qy       3241 ATTATAAGCTGAGAAATTTCTATTAACTAGTGATATCGCAGCCATCATAAGTGTAATGCA 3300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3241 ATTATAAGCTGAGAAATTTCTATTAACTAGTGATATCGCAGCCATCATAAGTGTAATGCA 3300

Qy       3301 GGACATTACCTTTTCTATGTTTAGATATGTTAGATACACAAATATATTTCATTGTGTTAT 3360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3301 GGACATTACCTTTTCTATGTTTAGATATGTTAGATACACAAATATATTTCATTGTGTTAT 3360

Qy       3361 AATTTCCTACAGTATTCAGTACAGTAACATGCTGTACAGGTTTGTAACCTAGGAGTAATA 3420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3361 AATTTCCTACAGTATTCAGTACAGTAACATGCTGTACAGGTTTGTAACCTAGGAGTAATA 3420

Qy       3421 GGCTATACCATACAGCTTAGGTGTGTAGTAGGCTATAACCATCTAGGTTTGTGTAAGTAC 3480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3421 GGCTATACCATACAGCTTAGGTGTGTAGTAGGCTATAACCATCTAGGTTTGTGTAAGTAC 3480

Qy       3481 ATTCTATGATATTCCCACAATGATGAAATCACCTAACTACACATTTCTCAGAATGTTTCA 3540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3481 ATTCTATGATATTCCCACAATGATGAAATCACCTAACTACACATTTCTCAGAATGTTTCA 3540

Qy       3541 CTGTTGTGAAGTGACCCATGACTATATTTTCCTATATACTTGATATTTTTGTGCATCTGC 3600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3541 CTGTTGTGAAGTGACCCATGACTATATTTTCCTATATACTTGATATTTTTGTGCATCTGC 3600

Qy       3601 CCATGAGAATGTAGTGTAAGATCAAAGGATGCAAGAATGGGTTCTATCCAGTATAGTACC 3660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3601 CCATGAGAATGTAGTGTAAGATCAAAGGATGCAAGAATGGGTTCTATCCAGTATAGTACC 3660

Qy       3661 CACTACACTGGTGGATGTCAATATGTATTTGTTAGATTAATATCTCAAGAATGAGCACCT 3720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3661 CACTACACTGGTGGATGTCAATATGTATTTGTTAGATTAATATCTCAAGAATGAGCACCT 3720

Qy       3721 TTCTCAGACACATAAAAGATGCTCAATATAAAAGTTTGTTGAACTGAACGTTATTGGCAA 3780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3721 TTCTCAGACACATAAAAGATGCTCAATATAAAAGTTTGTTGAACTGAACGTTATTGGCAA 3780

Qy       3781 ATGTAACATGATCGGATTTAAAGAGGAGCGAAACAGAGGTCTGGCTCAAACACCATACTT 3840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3781 ATGTAACATGATCGGATTTAAAGAGGAGCGAAACAGAGGTCTGGCTCAAACACCATACTT 3840

Qy       3841 CTAGAGTGCATAAGAGGTAGCAGTTGATTACCACTGGCGACAGGAGAAAAAAGAGCTTGA 3900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3841 CTAGAGTGCATAAGAGGTAGCAGTTGATTACCACTGGCGACAGGAGAAAAAAGAGCTTGA 3900

Qy       3901 CCGCAGGGTACTGTGAAGACATTTCAGGTTGATGGCACAGAACAGGGGAAATACATAAAT 3960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3901 CCGCAGGGTACTGTGAAGACATTTCAGGTTGATGGCACAGAACAGGGGAAATACATAAAT 3960

Qy       3961 GTGTGGGAATATTCAGTGGTCTGGGATGACTACATAGTAGAATATAATGAAGAAAAGAGT 4020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3961 GTGTGGGAATATTCAGTGGTCTGGGATGACTACATAGTAGAATATAATGAAGAAAAGAGT 4020

Qy       4021 GGAAGGGAAAGATGAAAAGTTGGAATGGGGATGAATTATGAAAGTACCAGAATGTTATGC 4080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4021 GGAAGGGAAAGATGAAAAGTTGGAATGGGGATGAATTATGAAAGTACCAGAATGTTATGC 4080

Qy       4081 TAAGGAATCTAGATTTTAAAATGTGAGGGCAAATTGAAGTC 4121
              |||||||||||||||||||||||||||||||||||||||||
Db       4081 TAAGGAATCTAGATTTTAAAATGTGAGGGCAAATTGAAGTC 4121
```

RESULT 2
ABT10748/c
ID ABT10748 standard; cDNA; 121600 BP.
XX
AC ABT10748;
XX
DT 04-DEC-2002 (first entry)
XX
DE Human breast cancer associated coding sequence SEQ ID NO: 882.
XX
KW Human; breast specific gene; breast cancer; differential expression;
KW cytostatic; gene therapy; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200259271-A2.
XX
PD 01-AUG-2002.
XX
PF 25-JAN-2002; 2002WO-US002176.
XX
PR 25-JAN-2001; 2001US-0263757P.
PR 25-APR-2001; 2001US-0286090P.
PR 23-MAY-2001; 2001US-0292517P.
XX
PA (GENE-) GENE LOGIC INC.
XX
PI Orr MS, Nation M, Diggans JC, Zeng W;
XX
DR WPI; 2002-674803/72.
XX
PT Diagnosing breast cancer in a patient comprises detecting the level of
PT gene expression in cell or tissue samples, where a differential gene
PT expression is indicative of breast cancer.
XX
PS Claim 1; SEQ ID NO 882; 260pp + Sequence Listing; English.
XX
CC The present invention relates to methods of diagnosing breast cancer in a
CC patient, which comprise detecting the level of expression in a tissue
CC sample of two or more genes selected from those shown in ABT09867-
CC ABT11112, where a differential expression of the genes indicates breast
CC cancer. The methods are useful in diagnosing, treating, detecting the
CC progression, and in monitoring treatment of breast cancer in patients.
CC The methods are also useful as a screening tool for agents that modulate
CC the onset or progression of breast cancer. The breast cancer genes may be
CC used as diagnostic markers for the prediction or identification of the
CC malignant state of breast tissue, for confirming the type and progression
CC of cancer, and for drug screening and assays. The present sequence is a
CC coding sequence of the invention. Note: The sequence data for this patent
CC did not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub.published_pct_sequences
XX
SQ Sequence 121600 BP; 38372 A; 24652 C; 24048 G; 34528 T; 0 U; 0 Other;

```
  Query Match             8.7%;  Score 356.6;  DB 6;  Length 121600;
  Best Local Similarity   74.7%;  Pred. No. 2.1e-51;
  Matches  543;  Conservative    0;  Mismatches  169;  Indels   15;  Gaps    7;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAA--- 57
              || || |||||| ||||    ||||  |||||||||| |||||||||||||| |||||
Db      44776 AGATCGAGACCATCCTGGCTAACACAGTGAAACCCCGTCTCTACTAAAAATCCAAAAAAA 44717

Qy         58 --TTAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGG 115
                |||| | ||| |||||     | |||||| |||||||||||  |||||| ||||||||
```

```
Db      44716 AGTTAGCCGGGCGTGGTGGCGGGCACCTGTAGTCCCAGCTACTCTGGAGGCTGAGGCAGG 44657

Qy        116 AGAATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCAC 174
              ||||| | ||||| ||  || || |||| |||| | ||| | ||||    |||| |||||
Db      44656 AGAATGGCGTGAACTTG--GGAGGCGGAGCTTGCAGTGAGCCGAGATAGCGCCACTGCAC 44599

Qy        175 TCCAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATA 234
              ||||| |  ||| || ||||||| |   ||||||| |||||||| |       | ||||||
Db      44598 TCCAGTCTGGGCGAC-AGAGCGAGACTCCATCTCAAAAAAAAAATA----TATATATATA 44544

Qy        235 TATATATATATATATATACACACACACACATATATATATACACATATATATACACATA 294
              || ||||||||| ||||||| | | | | | ||||||| ||| | | ||||||| | |||
Db      44543 TAGATATATATAGATATACAGATATATAGATATATATAGATATATAGATATATATAGATA 44484

Qy        295 TATACATATATACACATATATACACATATATATACACATACATATGTACACATATATATA 354
              |||| ||||||| | ||||||| | ||||||| | | ||| |||| || | |||||||
Db      44483 TATAGATATATATAGATATATAGATATATATAGATATATAGATATATATAGATATATAGA 44424

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              | ||||  ||| | | ||||| | | ||||| | | ||||| | | ||||| | | | |
Db      44423 TATATATAGATATATAGATATATATAGATATATAGATATATAGATATATATAGATATATA 44364

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              |||||  | | |||||  | |||||||  | |||||  | |||||||  | |||||||  |||||
Db      44363 GATATATATAGATATATAGATATATATAGATATATAGATATATAGATATATATAGATATA 44304

Qy        475 TACACATATATATAATATACACACATATATATACACATATATACACACATATATACACAT 534
              || |  |||| || ||||| | | ||||| ||| | ||| ||| | | |||||||  | ||
Db      44303 TAGATGTATAGAT-ATATATAGATATATAGATATATATAGATATATAGATATATAGATAT 44245

Qy        535 ATATACACATATATATACACATATATACACATATATACATATATACACATATATATACAT 594
              |||||  | |||||  |||  | |||  |||  | |||||  | |||||  | |  |||  ||||| |
Db      44244 ATATAGATATATAGATATATATAGATATATATATAGATATATAGATATATAGATATATAG 44185

Qy        595 ATATACACATATATACATATATACA-CATATATACATATATACACACATATATACACATA 653
              |||||  | |||||||  |||||||  |  |||||||  |||||||  | |  |||||  | |  |||
Db      44184 ATATATAGATATATAGATATATAGATGATATATAGATATATATAGATATATAGATATATA 44125

Qy        654 CATATACACACACATAGATATACATATATATACACATATATATACGTATATATATATGTATA 713
              |||||  | | |  |||||||||  |||  |||||  | |||||||  |  ||||  ||||  ||||
Db      44124 GATATATAGATATATAGATATATATAGATATATAGATATATAGATATATAGATATATATA 44065

Qy        714 TATATAT 720
              |||||||
Db      44064 GATATAT 44058


RESULT 3
ADQ18307/c
ID   ADQ18307 standard; DNA; 121600 BP.
XX
AC   ADQ18307;
XX
DT   26-AUG-2004  (first entry)
XX
DE   Human soft tissue sarcoma-upregulated DNA - SEQ ID 1125.
XX
KW   soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO2004048938-A2.
XX
PD   10-JUN-2004.
XX
PF   26-NOV-2003; 2003WO-US038193.
XX
PR   26-NOV-2002; 2002US-0429739P.
XX
PA   (PROT-) PROTEIN DESIGN LABS INC.
XX
PI   Aziz N,  Ginsburg WM,  Zlotnik A;
XX
DR   WPI; 2004-441208/41.
XX
PT   Early detection of soft tissue sarcoma comprises determining expression
PT   of a gene in a first soft tissue sample and a normal soft tissue sample
PT   and comparing the gene expression, also useful in treating soft tissue
PT   sarcoma.
XX
PS   Example 2; SEQ ID NO 1125; 210pp; English.
XX
CC   The invention relates to a novel method for detecting soft tissue sarcoma
CC   which comprises obtaining a first soft tissue sample from an individual
CC   and a normal soft tissue sample from the same or different individual,
CC   determining the expression of a gene in both samples and comparing the
CC   expression of the gene in both soft tissue samples, where a higher level
CC   of protein expression in the first soft tissue sample indicates the
CC   presence of soft tissue sarcoma. The method of the invention has
CC   cytostatic applications and may be useful for detecting soft tissue
CC   sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC   acid sequences may be useful in diagnostic and screening applications.
CC   The current sequence is that of a human soft tissue sarcoma-upregulated
CC   DNA of the invention. The current sequence is not shown within the
CC   specification per se but was submitted in CD format by the inventor.
XX
SQ   Sequence 121600 BP; 38372 A; 24652 C; 24048 G; 34528 T; 0 U; 0 Other;

  Query Match             8.7%;  Score 356.6;  DB 12;  Length 121600;
  Best Local Similarity   74.7%;  Pred. No. 2.1e-51;
  Matches  543;  Conservative    0;  Mismatches  169;  Indels   15;  Gaps    7;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAA--- 57
              || || |||||| ||||    ||||  |||||||||| |||||||||||||| |||||
Db      44776 AGATCGAGACCATCCTGGCTAACACAGTGAAACCCCGTCTCTACTAAAAATCCAAAAAAA 44717

Qy         58 --TTAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGG 115
                |||| | ||| |||||      | |||||| |||||||||||  |||||| ||||||||
Db      44716 AGTTAGCCGGGCGTGGTGGCGGGCACCTGTAGTCCCAGCTACTCTGGAGGCTGAGGCAGG 44657

Qy        116 AGAATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCAC 174
              ||||| | ||||| ||  || || |||| |||| | ||| | ||||    |||| |||||
Db      44656 AGAATGGCGTGAACTTG--GGAGGCGGAGCTTGCAGTGAGCCGAGATAGCGCCACTGCAC 44599

Qy        175 TCCAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATA 234
              ||||| |  ||| || ||||||| |   ||||||| |||||||| |       | ||||||
Db      44598 TCCAGTCTGGGCGAC-AGAGCGAGACTCCATCTCAAAAAAAAAATA----TATATATATA 44544

Qy        235 TATATATATATATATATACACACACACACATATATATATACACATATATATACACATA 294
              || ||||||||| ||||||| | | | | | ||||||| ||| | | ||||||| | |||
Db      44543 TAGATATATATAGATATACAGATATATAGATATATATAGATATATAGATATATATAGATA 44484

Qy        295 TATACATATATACACATATATACACATATATATACACATACATATGTACACATATATATA 354
              |||| ||||||| | ||||||| | ||||||| | | ||| |||| || | |||||||
Db      44483 TATAGATATATATAGATATATAGATATATATAGATATATAGATATATATAGATATATAGA 44424

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              | ||||  ||| | | ||||| | | ||||| | | ||||| | | ||||| | | | |
Db      44423 TATATATAGATATATAGATATATATAGATATATAGATATATAGATATATATAGATATATA 44364

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              |||||  | | |||||  | |||||||  | |||||  | |||||||  | |||||||  |||||
Db      44363 GATATATATAGATATATAGATATATATAGATATATAGATATATAGATATATATAGATATA 44304

Qy        475 TACACATATATATAATATACACACATATATATACACATATATACACACATATATACACAT 534
              || |  |||| || ||||| | | ||||| ||| | ||| ||| | | |||||||  | ||
Db      44303 TAGATGTATAGAT-ATATATAGATATATAGATATATATAGATATATAGATATATAGATAT 44245

Qy        535 ATATACACATATATATACACATATATACACATATATACATATATACACATATATATACAT 594
              |||||  | |||||  |||  | |||  |||  | |||||  | |||||  | |  |||  ||||| |
Db      44244 ATATAGATATATAGATATATATAGATATATATATAGATATATAGATATATAGATATATAG 44185

Qy        595 ATATACACATATATACATATATACA-CATATATACATATATACACACATATATACACATA 653
              |||||  | |||||||  |||||||  |  |||||||  |||||||  | |  |||||  | |  |||
Db      44184 ATATATAGATATATAGATATATAGATGATATATAGATATATATAGATATATAGATATATA 44125
```

```
Qy        654 CATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATA 713
              ||||| | | | ||||||||| ||| ||||| | ||||||| |  |||| |||| ||||
Db      44124 GATATATAGATATATAGATATATATAGATATATAGATATATAGATATATAGATATATATA 44065

Qy        714 TATATAT 720
              |||||||
Db      44064 GATATAT 44058
```

RESULT 4
ADL08108/c
ID ADL08108 standard; DNA; 188971 BP.
XX
AC ADL08108;
XX
DT 20-MAY-2004 (first entry)
XX
DE Human gene associated with low HDL-C APOA1.
XX
KW Human; ds; SNP; single nucleotide polymorphism;
KW high density lipoprotein-C; HDL-C; vascular disease; metabolic disease;
KW coronary artery disease; gene.
XX
OS Homo sapiens.
XX
FH Key   Location/Qualifiers
FT variation   replace(123408,T)
FT   /*tag= a
FT   /standard_name= "Single nucleotide polymorphism"
XX
PN US2004043389-A1.
XX
PD 04-MAR-2004.
XX
PF 04-SEP-2002; 2002US-00235192.
XX
PR 04-SEP-2002; 2002US-00235192.
XX
PA (VITI-) VITIVITY INC.
XX
PI Mccarthy J;
XX
DR WPI; 2004-214170/20.
XX
PT Determining whether a subject has, or is at risk of developing, an
PT abnormally low high density lipoprotein-C (HDL-C) level comprises
PT detecting an allelic variant of a polymorphic region from any of a set of
PT 27 genes.
XX
PS Disclosure; SEQ ID NO 27; 37pp; English.
XX
CC The invention relates to determining whether a subject has, or is at risk
CC of developing, an abnormally low high density lipoprotein-C (HDL-C) level
CC comprises determining whether the subject has an allelic variant of a
CC polymorphic region from any of 27 genes (alleles listed in Table 5 of the
CC specification). Also included are determining whether a male subject has,
CC or is at risk of developing, an abnormally low HDL-C level, comprising
CC determining whether the male subject has an allelic variant of a
CC polymorphic region listed in Table 5 which is associated with abnormally
CC low HDL-C levels in males, and determining whether a female subject has,
CC or is at risk of developing, an abnormally low HDL-C level, comprising
CC determining whether the female subject has an allelic variant of a
CC polymorphic region listed in Table 5 which is associated with abnormally
CC low HDL-C levels in females. The allelic variant in determining whether a
CC subject has, or is at risk of developing, an abnormally low HDL-C level
CC is APOA 1 CC, CD14 1 CT, COL5A2 1 GG, EDNRB 1 AG or AA, FABP3 1 CT, GBE1
CC 1 AG or GG, LIPC 5 AA, MTHFR 1 CC, VWF 2 GG, or their complements. The
CC allelic variant in determining whether a male subject has, or is at risk
CC of developing, an abnormally low HDL-C level, LRP1 3 CC or CT, PAI2 4 GG,
CC or PPARG 1 CG, or their complements. The allelic variants are also COL5A2
CC 1 GG, CD14 1 CT or CC, and FABP3 1 CT, in combination, or their
CC complements. The methods are useful for diagnosing (a predisposition to)
CC abnormally low levels of low high density lipoprotein-C (HDL-C) in a
CC subject, The methods are useful in diagnosing (a predisposition to) or
CC prognosticating diseases and disorders associated with abnormal lipid
CC levels such as vascular and metabolic diseases, e.g., coronary artery
CC disease. The present sequence is a human gene containing a SNP (single
CC nucleotide polymorphism associated with low high density lipoprotein-C
CC (HDL-C) levels.
XX
SQ Sequence 188971 BP; 53091 A; 44581 C; 42653 G; 48646 T; 0 U; 0 Other;

Query Match 7.9%; Score 327.4; DB 12; Length 188971;
Best Local Similarity 67.6%; Pred. No. 2e-46;
Matches 601; Conservative 0; Mismatches 261; Indels 27; Gaps 9;

```
Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC--AAAAT 58
              || ||||||||| ||||  |||||||||||||| || ||||||||||||||||  |||||
Db     104585 AGATCAAGACCATCCTGGCCAACATGGTGAAACTCCGTCTCTACTAAAAATACAAAAAAT 104526

Qy         59 TAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGA 118
              |||   ||| ||||| |   | ||||||||||||||||||| |||||| || ||||||||
Db     104525 TAGCTGGGCGTGGTGGTGGGCTCCTGTAATCCCAGCTACTTGGGAGGCTGAAGCAGGAGA 104466

Qy        119 ATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTG---CAC 174
              ||| ||||||||  | || || ||||||||| | ||| |||||||  |||||||   |||
Db     104465 ATCGCTTGAACC--CAGGAGGCGGAGGTTGCAGTGAGCCAAGATCGCGCCATTGCACCAC 104408

Qy        175 TCCAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATA 234
              |||||||  ||||| ||||| ||||     ||||| ||||||||||   |   | | |
Db     104407 TCCAGCCTGGGCAAAAAGAGAGAAACTCTGTCTCAAAAAAAAAAAAAAAAAAAAAAAAAA 104348

Qy        235 TATATATATATATATATACACACACACACATATATATATACACATATATATACACATA 294
              ||||||||||||||||| | | | | ||||| | |||||||  ||||||||       |
Db     104347 AATATATATATATATATATATATATACACACACATATATATTCATATATA-------A 104295

Qy        295 TATACATATATACACATATATACACATATATATACACATACATATGTACACATATATATA 354
              |  | |||||||  ||||||||   ||  ||||| | || ||||| ||  ||||||||||
Db     104294 TTCATATATATATTCATATATA---ATTCATATATATATTCATATATATTCATATATATA 104238

Qy        355 CACATATGT-ATACACATATATACACATATATACACACATATATACACATATATACACAC 413
              |||||  | || || |||||||  |||||||  || |||||||  |||||||| |  |
Db     104237 TTCATATATAATTCATATATATATTCATATATATTCATATATATATTCATATATATATTC 104178

Qy        414 ATATATACACATATATACACATATATACACATATACACATATATACACATATATACATAT 473
              |||||||     ||  | | |||||| || |||||  ||||||| |   ||||| ||||
Db     104177 ATATATA-----ATTCATATATATATTCATATATATTCATATATATATTCATATATATAT 104123

Qy        474 ATACACATATATATAATATACACACATATATATACACATATATACACACATATATACACA 533
              | | |||  |   |||||  | ||||||||| || |||||||  || |||||| || |
Db     104122 TCATATATAATTCATATATATATTCATATATATTCATATATATATTCATATATATTCATA 104063

Qy        534 TATATACACATATATATACACATATATACACATATATACATATATACACATATATATACA 593
              ||||||  |||||||||  ||||||||  |||||||| ||  |||  | ||||  ||| |
Db     104062 TATATATTCATATATATATTCATATATATTCATATATATATTCATATATATATTCATATA 104003

Qy        594 TATATACACATATATACATATATACACATATATACATATATACACACATATATACACATA 653
              ||||| || |||||||   ||||| | ||  ||| |||||| || | ||||  | | |||
Db     104002 TATATTCATATATATATTCATATATATATTCATATATATATTCATATATATTCATATATA 103943

Qy        654 CATATACACACACATAGATATACA---TATATATACACATATATATACGTATATATATGT 710
              ||  | | | | ||  ||||| |     ||||||| |   |||||||  |  ||||||
Db     103942 TATTCATATATATATTCATATATAATTCATATATATATTCATATATATATTCATATATAA 103883

Qy        711 ATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGA 770
              ||  ||||||  | || |  |       |  |||   || ||    | |   |||||
Db     103882 ATTCATATATAATTCATATATATATTCATATAGATATTCATATATAATTCATATAGATAT 103823

Qy        771 AAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTA 830
              | |||  || | |    |  ||  || |  | || |      |     || ||| |
Db     103822 TCATAGATATTCATATATATTCATATAGATATTCATATATTCATATATATTCATATATTC 103763

Qy        831 GGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGA 879
              | | || |||| |  | |    | |   || |  |  |   || | |
Db     103762 ATGTATATTCATATAGATATTCATATATAGATATTCATAGATATTCATA 103714
```

```
RESULT 5
ABD32659
ID   ABD32659 standard; DNA; 163382 BP.
XX
AC   ABD32659;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Human cancer-associated genomic DNA HD13-065.
XX
KW   Human; ds; cancer-associated protein; gene; cytostatic; cancer;
KW   leukaemia; lymphoma; CAP.
XX
OS   Homo sapiens.
XX
PN   WO2004074320-A2.
XX
PD   02-SEP-2004.
XX
PF   17-FEB-2004; 2004WO-US004730.
XX
PR   14-FEB-2003; 2003US-00367094.
PR   14-MAR-2003; 2003US-00388838.
PR   15-APR-2003; 2003US-00417375.
PR   13-JUN-2003; 2003US-00461862.
PR   15-SEP-2003; 2003US-00663431.
PR   15-DEC-2003; 2003US-00737318.
XX
PA   (SAGR-) SAGRES DISCOVERY INC.
XX
PI   Morris DW,  Morris DW,  Malandro MS;
XX
DR   WPI; 2004-652914/63.
XX
PT   New isolated cancer-associated polynucleotides and polypeptides useful
PT   for diagnosing, preventing or treating cancers, especially lymphoma and
PT   leukemia, or in screening for agents that modulate cancer.
XX
PS   claim 16; seqid 199; 310pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising at least 10
CC   contiguous nucleotides of any of the 233 polynucleotide sequences given
CC   in the specification, or its complement. The nucleic acids encode cancer-
CC   associated proteins. Also included are an expression vector comprising
CC   the isolated nucleic acid cited above, a host cell comprising the above
CC   recombinant nucleic acid or expression vector, a microarray for detecting
CC   a cancer-associated (CA) nucleic acid comprising at least one probe
CC   comprising at least 10 contiguous nucleotides of any of the above-
CC   mentioned nucleotide sequences, an isolated polypeptide (encoded within
CC   an open reading frame of a CA sequence selected from any of the 95
CC   polynucleotide sequences as mentioned in the specification, or its
CC   complement), an isolated antibody, (or its antigen binding fragment) that
CC   binds to the above polypeptide, a hybridoma that produces the above
CC   monoclonal antibody, a pharmaceutical composition comprising the above
CC   antibody and a pharmaceutical excipient, a kit for detecting cancer
CC   cells(comprising the antibody cited above, methods for diagnosing cancer
CC   or for detecting the presence or absence of cancer cells in an
CC   individual, a method for inhibiting growth of cancer cells in an
CC   individual, a method for delivering a therapeutic agent to cancer cells
CC   in an individual, an electronic library comprising the above
CC   polynucleotide or polypeptide (or their fragments), methods of screening
CC   for anticancer activity or for a bioactive agent capable of modulating
CC   the activity of a CA protein (CAP), methods for detecting cancer
CC   associated with expression of a polypeptide in a test cell sample, a
CC   method for treating cancers and a method for inhibiting the expression of
CC   CA gene in a cell. The composition and methods are useful for detecting,
CC   diagnosing, preventing and treating cancers, especially lymphoma and
CC   leukaemia. These may also be used in screening for agents that modulate
CC   cancer. The present sequence is a human CAP genomic sequence. Note: The
CC   sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 163382 BP; 50073 A; 25805 C; 27837 G; 53184 T; 0 U; 6483 Other;

  Query Match             7.9%;  Score 323.6;  DB 13;  Length 163382;
  Best Local Similarity   68.2%;  Pred. No. 8.7e-46;
  Matches  495;  Conservative    0;  Mismatches  224;  Indels    7;  Gaps    3;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              |||||||||||||||||  ||||||||||||||||||||||||||||||||| ||||||
Db     144405 AGTTCAAGACCAGCCTGGCCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAAATT 144464

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              || ||||| |||||   ||  |||||||||||||||||||   |||||| ||||||||||||
Db     144465 AGCCAGGCGTGGTGGCACGTGCCTGTAATCCCAGCTACCCAGGAGGCTGAGGCAGGAGAA 144524

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              |  |||||| |    | ||  | || |   | |   | ||||||||||||||||||||
Db     144525 TTGCTTGAAGC-CAAGAGGCAGAAGTTGCAGTGAGCCGAGATCATGCCATTGCACTCCAG 144583

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  ||  |||||| ||| |  |||| ||| ||||||||||        |   |   |   |
Db     144584 CCTGGGGGACAAGAACGAGACTTCATGTCAAAAAAAAAAAATACACAGTCTTCCTTGGCT 144643

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
                      | | | | | | || ||| |||| | |||| |||| | || |||
Db     144644 TAGCTGTATGGGAAATATATATATATATTTATTTATATATATATTTATATATATTTATTT 144703

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACAT 359
              |||||||   || |||| | || ||| || | |||  ||| || | || ||| || | |
Db     144704 ATATATATTTATTTATATATATTTATTTATATATATTTATTTATATATTTATTTATATTT 144763

Qy        360 ATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATAT 419
              || | || | || |||  | ||||||     | |||| ||    ||||||  |   ||||||
Db     144764 ATTTTTATATATTTATTTATATATATTTGTATATATTTATTTATATATTTATTTATATAT 144823

Qy        420 ACACATATATACACATATATACACATATACACATATATACACATATATACATA-----TA 474
                |  ||||||   |||||||     |  | | |||| || | || |||| |||     ||
Db     144824 TTATTTATATATTTATATATATTAGTTCATATATATTTATATATTTATATATATTAGTTA 144883

Qy        475 TACACATATATATAATATACACACATATATATACACATATATACACACATATATACACAT 534
              || | || |||||| ||   | | |||| |||| | || | || | |  ||| || | ||
Db     144884 TATATATTTATATATTAGTTATATATATTTATATATATTTTTATATATTTATTTATATAT 144943

Qy        535 ATATACACATATATATACACATATATACACATATATACATATATACACATATATATACAT 594
              || || | ||  | |||   |||||||  |||||||  ||||||| | |||||||  ||
Db     144944 ATGTATATATTAACATATTTATATATATTCATATATTTATATATAAATATATATATTAAT 145003

Qy        595 ATATACACATATATACATATATACACATATATACATATATACACACATATATACACATAC 654
              ||||  | ||||||  ||||||  | |||| |  ||||||  | ||||||  ||   |||
Db     145004 ATATTTATATATATTTATATATTAATATATGTTTATATATTTATACATATTTATGTATAT 145063

Qy        655 ATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATAT 714
                ||| | | |  || |||||  |||||||  |  ||||||||  |||||||  | ||| |
Db     145064 TAATATATATATTTATATATATTTATATATGTATTTATATATATTTATATATTTATATTT 145123

Qy        715 ATATAT 720
              || |||
Db     145124 ATTTAT 145129


RESULT 6
AAS46748/c
ID   AAS46748 standard; DNA; 26997 BP.
XX
AC   AAS46748;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Tumour suppressor gene derived chemically modified sequence #472.
XX
KW   Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer;
KW   tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
```

```
KW   cytosine methylation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200168912-A2.
XX
PD   20-SEP-2001.
XX
PF   15-MAR-2001; 2001WO-EP002955.
XX
PR   15-MAR-2000; 2000DE-01013847.
PR   06-APR-2000; 2000DE-01019058.
PR   07-APR-2000; 2000DE-01019173.
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2001-602752/68.
XX
PT   Fragments of chemically modified genes associated with tumor suppressor
PT   genes and oncogenes, useful in designing primers and probes for analyzing
PT   diseases associated with cytosine methylation state e.g. cancer.
XX
PS   Claim 1; SEQ ID NO 472; 27pp; English.
XX
CC   The invention relates to a nucleic acid comprising a sequence of 18
CC   bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC   bisulphite, of genes associated with tumour suppression and oncogenes
CC   having a sequence taken from 536 (actually 533 since numbers 408, 458 and
CC   500 are missing from the sequence listing) sequences (Ss) and sequences
CC   complementary to (Ss). The nucleic acid may be a peptide nucleic acid-
CC   oligomer (PNA) of at least 9 nucleotides and may form part of a set of
CC   probes for detecting the cytosine methylation state and/or single
CC   nucleotide polymorphisms and also to be used in an array for analysing
CC   diseases associated with CpG dinucleotides e.g. cancers and tumours. The
CC   probes can also be used in a method for ascertaining genetic and/or
CC   epigenetic parameters for the diagnosis and/or therapy of existing
CC   diseases or the predisposition to specific diseases, by analysing
CC   cytosine methylations. The parameters may be compared to another set of
CC   genetic and/or epigenetic parameters, the differences serving as basis
CC   for diagnosis and/or prognosis events which are disadvantageous to
CC   patients. The present sequence is one of the 533 genomic sequences
CC   derived from tumour suppressor genes and oncogenes. Sequences with even
CC   numbered Seq ID numbers are the complementary sequence of the
CC   corresponding odd numbered sequence (e.g. ID 2 and ID1, ID 536 and ID
CC   535, except for those whose partner sequence is missing). Note: The
CC   sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 26997 BP; 7650 A; 354 C; 5924 G; 13069 T; 0 U; 0 Other;

  Query Match             7.8%;  Score 322.2;  DB 4;  Length 26997;
  Best Local Similarity   66.0%;  Pred. No. 1.4e-45;
  Matches  498;  Conservative    0;  Mismatches  253;  Indels    4;  Gaps    2;

Qy        207 TCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATATACACACACACACACA 266
              | |  |||||||| |     | ||||||||||||||||||||||||||||||| | | | |
Db       6221 TAAATAAAAAAAATATAATATATATATATATATATATATATATACACACATATATATA 6162

Qy        267 TATATATATACACATATATATACACATATATACATATATACACATATATACACATATATA 326
              | |||||||||| | |||||| ||||| ||| ||||| ||||||||||||||||||||
Db       6161 AACATATATACACACACATATACATATATACACACATATATACATATATACACATATATA 6102

Qy        327 TACACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATA 386
              || | ||| | || || | | ||||||| | | |  |||| | ||||||||| |||||||
Db       6101 TATATATATACATATATATACATATATATATACACATATATATATATATACATATATA 6042

Qy        387 CACACATATATACACATATATACACACATATATACACATATATACACATATATACACATA 446
              || | | |||||||||||||||| | | ||| ||| | ||| ||| ||||||| ||| |||
Db       6041 CATATACATATACACATATATATATATATACATATATATACATATACATATACACATATA 5982

Qy        447 TACACATATATACACATATATACATATATACACATATATATA-ATATACACACATATATA 505
              || | ||||| | | | ||| | |||||||||||||||||||| ||||||| | | ||| |
Db       5981 TATATATATACATATACATATATATACACATATATATATATATACATATACATACA 5922

Qy        506 TACACATATATACACACATATATACACATATATACACATATATATACACATATATACACA 565
              || | ||| | | | | ||||||| |||||||||||| |||||| | | ||| | ||| |
Db       5921 TATATATACACATATATATATATATACATATATACACGTATATACATATATACACACATA 5862

Qy        566 TATATACATATATACACATATATATACATATATACACATATATACATATATACACA---T 622
              | |||| |||  || | | ||||||| ||||||||| ||||||| ||||||| |||   |
Db       5861 TTTATATATACGTATATACATATATATATATATACATATATATATATATATACATATT 5802

Qy        623 ATATACATATATACACACATATATACACATACATATACACACACATAGATATACATATAT 682
              ||||| ||||||| | ||||||||| | ||||||||| | | | ||| | ||| ||| ||
Db       5801 ATATATATATATATACATATATAAATATACATATATATATATATATACATATATACAT 5742

Qy        683 ATACACATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTA 742
              ||| | ||||||| |  |||||||| |||| |||||          || |     ||
Db       5741 ATATATATATATAAATATATATATATATATACGTATATATATATACACACACACACCATA 5682

Qy        743 GCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAA 802
                |    ||   | |     | | |  |||| |   |||   |  | |  | |   | |
Db       5681 AAATACTATACACCCATAAAAAAAAAAAAAAATAACATTTACAACAACTTAAATAAAATTA 5622

Qy        803 GGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTAT 862
                |  | |  | | | ||  | |     | | |   |  | | |||  |  ||
Db       5621 AAAACCATTATTCTAACCGAAATAACTCAAAAATAAAAAACCAAATATTATATATTCTCA 5562

Qy        863 CTGTAAGGTAGTTAAGAAGGTAACACTATATATATATATAGTGAAAGCAGTGTAAACCTA 922
              ||   |  ||   |  ||  ||  |  | | | | |||   | |  |   ||||| |
Db       5561 CTTATAACTAAAAACTAAACTATAAAAACACAAAAACATAAAAATAATACAATAAACTTT 5502

Qy        923 AAGGATGGGCCAAGGATTTAAATGTTATAGAAGAA 957
              ||  |      ||  |    ||   || | || ||
Db       5501 AAAAACTCTAAAAAAAACATAAAAATAAATAAAAA 5467


RESULT 7
AAS46747/c
ID   AAS46747 standard; DNA; 26997 BP.
XX
AC   AAS46747;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Tumour suppressor gene derived chemically modified sequence #471.
XX
KW   Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer;
KW   tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW   cytosine methylation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200168912-A2.
XX
PD   20-SEP-2001.
XX
PF   15-MAR-2001; 2001WO-EP002955.
XX
PR   15-MAR-2000; 2000DE-01013847.
PR   06-APR-2000; 2000DE-01019058.
PR   07-APR-2000; 2000DE-01019173.
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2001-602752/68.
XX
```

```
PT   Fragments of chemically modified genes associated with tumor suppressor
PT   genes and oncogenes, useful in designing primers and probes for analyzing
PT   diseases associated with cytosine methylation state e.g. cancer.
XX
PS   Claim 1; SEQ ID NO 471; 27pp; English.
XX
CC   The invention relates to a nucleic acid comprising a sequence of 18
CC   bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC   bisulphite, of genes associated with tumour suppression and oncogenes
CC   having a sequence taken from 536 (actually 533 since numbers 408, 458 and
CC   500 are missing from the sequence listing) sequences (Ss) and sequences
CC   complementary to (Ss). The nucleic acid may be a peptide nucleic acid-
CC   oligomer (PNA) of at least 9 nucleotides and may form part of a set of
CC   probes for detecting the cytosine methylation state and/or single
CC   nucleotide polymorphisms and also to be used in an array for analysing
CC   diseases associated with CpG dinucleotides e.g. cancers and tumours. The
CC   probes can also be used in a method for ascertaining genetic and/or
CC   epigenetic parameters for the diagnosis and/or therapy of existing
CC   diseases or the predisposition to specific diseases, by analysing
CC   cytosine methylations. The parameters may be compared to another set of
CC   genetic and/or epigenetic parameters, the differences serving as basis
CC   for diagnosis and/or prognosis events which are disadvantageous to
CC   patients. The present sequence is one of the 533 genomic sequences
CC   derived from tumour suppressor genes and oncogenes. Note: The sequence
CC   data for this patent did not form part of the printed specification, but
CC   was obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 26997 BP; 7233 A; 354 C; 6190 G; 13220 T; 0 U; 0 Other;

  Query Match          7.6%;  Score 311.2;  DB 4;  Length 26997;
  Best Local Similarity  76.9%;  Pred. No. 1e-43;
  Matches  409;  Conservative    0;  Mismatches  113;  Indels   10;  Gaps    2;

Qy        203 CATCTCAGAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACA 262
              | | ||| ||| | | ||       | ||||||||||||||||||| | | || || | | |
Db      21339 CTTTTCATAAATACATAATATTCCATAATATATATATATATATACACACGCATATATA 21280

Qy        263 CACATATATATATACACATATATATACACATATATACATATATACACATATATACACATA 322
              ||||||||    | ||| ||| ||||| | |||||||||||||| ||| ||||||| | ||
Db      21279 CACATATACCCACACATATACATATATATATATATACATATACACATATATATATATCTA 21220

Qy        323 TATATACACATACATATGT-----ACACATATATATACACATATGTATACACATATATAC 377
              |||||||||  | |||  |     | | ||| | ||| |||||| |||| ||| ||||||
Db      21219 CATATACACACATATACATATATAATATATACACATATACATATATATATACACATATAC 21160

Qy        378 ACATATATACACACATATATACACATATATACACACATATATACACATATATACACATAT 437
              ||||||| | || || ||||| ||||||||| | |||||||||||| |||| ||| |||
Db      21159 ACATATACATACGCACATATAAACATATATATATACATATATACACGTATACACATATAC 21100

Qy        438 ATACACATATACACATATATACACATATATACATATATACACATATATATA-----ATAT 492
              ||| ||| ||| | ||||||||| ||||| ||||||| | | ||| |||||     |||
Db      21099 ATATACACATATATATATATACATATATACACATATACATATATACATATATATATAC 21040

Qy        493 ACACACATATATATACACATATATACACACATATATACACATATATACACATATATATAC 552
              | | | | | ||||||| ||||| | | | ||||||| |||||||||||||||||| |||
Db      21039 ATATATACACATATACATATATACATATATATATATATACATATATACACATATACATAT 20980

Qy        553 ACATATATACACATATATACATATATACACATATATATACATATATACACATATATACAT 612
              | | ||||| | |||||||||||||||||||||||| ||| ||| ||| | |||||||
Db      20979 ATACATATATATATATATACATATATACACATATACATATATACATATATATATATAC 20920

Qy        613 ATATACACATATATACATATATACACACATATATACACATACATATACACACACATAGAT 672
              ||||| ||| |||||||||||||| | ||||||| | ||| ||||| | | | ||| ||
Db      20919 ATATATACACATATACATATATACATATATATATATATATATATATATATATATACAT 20860

Qy        673 ATACATATATATACACATATATATACGTATATATATGTATATATATATGCTC 724
              ||| |||||||||| |||||| ||  ||||||||| |||||||||| | |
Db      20859 ATATATATATATACATATATATCTATATATATATATATATATATATATAC 20808


RESULT 8
ADC87572
ID   ADC87572 standard; DNA; 1990 BP.
XX
AC   ADC87572;
XX
DT   01-JAN-2004  (first entry)
XX
DE   Human GPCR gene SEQ ID NO:2025.
XX
KW   ds; gene; human; GPCR;
KW   guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1270724-A2.
XX
PD   02-JAN-2003.
XX
PF   18-JUN-2002; 2002EP-00013517.
XX
PR   18-JUN-2001; 2001JP-00246789.
XX
PA   (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA   (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI   Suwa M,  Asai K,  Akiyama Y,  Aburatani H;
XX
DR   WPI; 2003-315783/31.
DR   P-PSDB; ADC87573.
XX
PT   New polynucleotide, useful for preparing a composition for treating a
PT   patient in need of increased or suppressed activity or expression of the
PT   guanosine triphosphate-binding protein coupled receptor.
XX
PS   Claim 1; SEQ ID NO 2025; 28pp; English.
XX
CC   The invention relates to a novel polynucleotide encoding a guanosine
CC   triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC   the invention may have a use in gene therapy. The polynucleotide and
CC   polypeptide are useful for preparing a composition for treating a patient
CC   in need of increased or suppressed activity or expression of the
CC   guanosine triphosphate-binding protein coupled receptor. The
CC   polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC   invention.
XX
SQ   Sequence 1990 BP; 833 A; 366 C; 180 G; 611 T; 0 U; 0 Other;

  Query Match          7.5%;  Score 308;  DB 10;  Length 1990;
  Best Local Similarity  80.2%;  Pred. No. 3.3e-43;
  Matches  401;  Conservative    0;  Mismatches   90;  Indels    9;  Gaps    3;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | || | | |||| |||||||| | | | | ||||||| || ||||||||||||||||||
Db        826 TATACACACATATGTATATATACACATATATACACACATATGTATATATACACATATATA 885

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              |||||||| || ||||| ||| ||||| |||||||| |||| | | |||||| ||||||
Db        886 CACACATATGTATATATACACATATATACACACATATGTATATATACACATATATACACA 945

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||| || | |||   ||| | | |  || |  |||||| ||||||||||||||||| |
Db        946 CATATGTATATATACACATATATACACGTATATGTATATATACACATATATACACATACA 1005

Qy        407 TACACA----CATATATACACATATATACACATATATACACATATACACATATATACACA 462
              ||||||     ||||| |||||||||||||||| |||||||||||  ||||||||||
Db       1006 TACACATATGTATATACACACATATATACACATACATACACATATACGTATATATACACA 1065

Qy        463 TATATACATATATACACATATATATAATATACACACATATATATACACATATATACACAC 522
              |||| ||||| ||||||||||   | ||||| |||||||||| ||| ||    || | ||
Db       1066 TATACACATACATACACATATACGT-ATATATACACATATATACACATATGCGTATATAC 1124

Qy        523 ATATATACACATATATACACATATATATACACATATATACACATATATACATATATACAC 582
              ||||||||||||||||||||    ||||||  |||||||||||||||||||| ||||||
Db       1125 ATATATACACATATATACAC----ATATACGTATATATACACATATATACACATATACGT 1180
```

```
Qy        583 ATATATATACATATATACACATATATACATATATACACATATATACATATATACACACAT 642
              ||||||| |||||||||||||||||||||| |||||   ||||| |||||||||||||| ||
Db       1181 ATATATACACATATATACACATATATACACATATATGTATATACACATATATACACATAT 1240

Qy        643 ATATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTAT 702
              |||| | | ||||||| ||||| |||  |||| ||| ||||| |||  ||| ||  |||
Db       1241 GTATATATACACATATATACACATATATGTATATATACATATATACACGTATGTATATAT 1300

Qy        703 ATATATGTATATATATATGC 722
              | | || ||| ||||||| |
Db       1301 ACACATATATGTATATATAC 1320


RESULT 9
AAA63785
ID   AAA63785 standard; DNA; 15418 BP.
XX
AC   AAA63785;
XX
DT   04-DEC-2000  (first entry)
XX
DE   Nucleotide sequence of the human TERT promoter and gene from pGRN142.
XX
KW   Telomerase reverse transcriptase; TERT; promoter; telomerase complex;
KW   telomere length; hyperplasic disease; cancer; oncolytic virus;
KW   cis-acting transcriptional control sequence; viral replication;
KW   cell proliferation; aging; immunological disorder; infertility; ss.
XX
OS   Homo sapiens.
XX
PN   WO200046355-A2.
XX
PD   10-AUG-2000.
XX
PF   04-FEB-2000; 2000WO-US003104.
XX
PR   04-FEB-1999;   99US-00244438.
XX
PA   (GERO-) GERON CORP.
XX
PI   Morin GB,  Lichtsteiner S,  Vasserot A,  Adams R,  Cardoza LM;
PI   Lebkowski JS;
XX
DR   WPI; 2000-532898/48.
XX
PT   New polynucleotides comprising cis-acting transcriptional control
PT   sequences, e.g. promoter sequence, of telomerase reverse transcriptase
PT   genes, useful in the treatment of cancer.
XX
PS   Claim 9; Page 58-61; 63pp; English.
XX
CC   The present sequence represents the human telomerase reverse
CC   transcriptase (TERT) promoter and gene. TERT is part of the telomerase
CC   complex responsible for maintaining telomere length and increasing
CC   replicative capacity of progenitor cells. Telomerase activity is turned
CC   off in mature differentiated cells, but is turned back on again in
CC   hyperplasic diseases, including many cancers. The polynucleotide
CC   comprises cis-acting transcriptional control sequences, e.g. promoter
CC   sequence. These promoter sequences are used to produce oncolytic viruses,
CC   in which a toxin or genetic element essential for viral replication is
CC   placed under the control of a TERT promoter. As a result, the virus
CC   replicates preferentially in cells expressing TERT, and selectively lyses
CC   cancer cells. The oncolytic viruses are useful for treating cancer in
CC   humans or animals. The TERT promoter sequences are useful in the
CC   treatment of cancer and other diseases of cell proliferation such as
CC   degenerative and aging processes and diseases of aging, immunological
CC   disorders, or infertility
XX
SQ   Sequence 15418 BP; 4518 A; 3797 C; 3765 G; 3338 T; 0 U; 0 Other;

  Query Match             7.2%;  Score 296;  DB 3;  Length 15418;
  Best Local Similarity   69.2%;  Pred. No. 3.9e-41;
  Matches  479;  Conservative    0;  Mismatches  200;  Indels   13;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||| |||||||||||  ||| || | |||||||||||| |||||||| || ||||||
Db        646 AGTTCGAGACCAGCCTGGCCAATATAGCGAAACCCCATCTGTACTAAAAACACAAAAATT 705

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   ||  |||||| ||| |||||||||||||| ||||  ||||||  ||||||||||
Db        706 AGCTGGGTGTGGTGACACATGCCTGTAATCCCAGGTACTCAGGAGGCTAAGGCAGGAGAA 765

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              |  |||||||  |  ||  | || |||     |   ||||||    ||| ||||||||||
Db        766 TTGCTTGAACTGGGAGGCAGAGGTGGT-----GAGCCAAGATTGCACCAGTGCACTCCAG 820

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  ||  || |||| ||||    ||||||| ||| ||| |||   | |  ||||| |||
Db        821 CCTTGGTGAC-AGAGTGAAACTCCATCTCAAAAACAAACAAACAAAATACATATACATAA 879

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              |||||||  | | | | ||| | ||| ||||||| ||| |||||||  |  |||||||
Db        880 ATATATATGCACATATATATACATATATAAATATATATACACATATATAAATCTATATAC 939

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACAT-----ATATATA 354
              ||||||||| ||||| ||| |||||| ||  | |||||||| || ||||     ||||||
Db        940 ATATATACATATATACACATATATAAATCTATATACATATATATACATATATAATATATT 999

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              |||||| || | | ||| ||| | | ||||||| | ||| ||||| ||||| | | |||
Db       1000 TACATATATAAATATATACATATATAAATATACATATATAAATACATATATAAATATACATA 1059

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACACATATA 474
              |||||| | ||| ||| | | ||||||| ||||| | ||||| | | ||| ||| | |||
Db       1060 TATATAAATATACATATATAAATATACATATATAAATATATACATATATAAATATACATA 1119

Qy        475 TACACATATATATA-ATATACACACATATATATACACATATATACACACATATATACACA 533
              || | ||||||||| ||||| | | ||||| ||| |||   |||| ||| |||||||| |
Db       1120 TATAAATATATATACATATATAAATATATAAATATACAAGTATATACAAATATATACATA 1179

Qy        534 TATATACACATATATATACACATATATACACATATATACATATATACACATATATATACA 593
              |||| |   |||||  || | | ||||| | ||| ||| ||| | | || | |   |
Db       1180 TATAAATGTATATACGTATATACATATATATATAAATATATAAAAAAACTTTTGGCTGGG 1239

Qy        594 TATATACACATATATACATATATACACATATATACATATATACACACATATATACACATA 653
              | |   || || |       | | | |  | |      ||  | || ||| | |  | |||||
Db       1240 CACCTTTCCAAATCTCATGGCACATATAAGTCTCATGGTAACCTCAAATAAAAAAACATA 1299

Qy        654 CATATACACACACATAGATATACATATATATA 685
              |           || | ||| | |   ||| |
Db       1300 TAACAGATACACCAAAAATAAAAACCAATAAA 1331


RESULT 10
AAL38601
ID   AAL38601 standard; DNA; 15418 BP.
XX
AC   AAL38601;
XX
DT   16-AUG-2002  (first entry)
XX
DE   Human TERT promoter and upstream sequence.
XX
KW   Cytostatic; glycosyltransferase; tumour; cell-surface carbohydrate;
KW   tissue specific; transcriptional control element; cancer cell;
KW   gene therapy; human TERT; telomerase reverse transcriptase; ds.
XX
OS   Homo sapiens.
XX
PN   WO200242468-A2.
XX
PD   30-MAY-2002.
XX
PF   26-NOV-2001; 2001WO-US044306.
XX
```

PR 27-NOV-2000; 2000US-0253395P.
XX
PA (GERO-) GERON CORP.
XX
PI Schiff MJ;
XX
DR WPI; 2002-479954/51.
XX
PT New polynucleotide encoding glycosyltransferase enzymes including histo
PT blood group transferase useful for treating conditions associated with
PT hyperproliferation, such as cancers and other neoplasias.
XX
PS Claim 9; Page 17-20; 49pp; English.
XX
CC The invention relates to a polynucleotide comprising an encoding sequence
CC for a glycosyltransferase, under control of a heterologous tumour
CC specific or tissue specific transcriptional control element, where
CC expression of the polynucleotide in a human cell causes the cell to
CC express a cell-surface carbohydrate determinant to which some or all
CC humans have a naturally occurring antibody. The polynucleotide of the
CC invention is useful for killing cancer cells and for preparing a
CC medicament for the treatment of cancer. The polynucleotide sequence of
CC the invention can be used to treat disorders by gene therapy. This
CC polynucleotide sequence represents the human TERT (telomerase reverse
CC transcriptase) promoter and upstream sequence related to the invention
XX
SQ Sequence 15418 BP; 4518 A; 3797 C; 3765 G; 3338 T; 0 U; 0 Other;

Query Match 7.2%; Score 296; DB 6; Length 15418;
Best Local Similarity 69.2%; Pred. No. 3.9e-41;
Matches 479; Conservative 0; Mismatches 200; Indels 13; Gaps 5;

```
Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||| ||||||||||  ||| || | ||||||||||| |||||||| || ||||||
Db        646 AGTTCGAGACCAGCCTGGCCAATATAGCGAAACCCCATCTGTACTAAAAACACAAAAATT 705

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   ||  |||||| ||| ||||||||||||| ||||  ||||||  ||||||||||
Db        706 AGCTGGGTGTGGTGACACATGCCTGTAATCCCAGGTACTCAGGAGGCTAAGGCAGGAGAA 765

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              |  |||||||  |  ||  | || |||     |   ||||||    ||| ||||||||||
Db        766 TTGCTTGAACTGGGAGGCAGAGGTGGT-----GAGCCAAGATTGCACCAGTGCACTCCAG 820

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  ||  || |||| ||||   ||||||| ||| ||| |||   | |  ||||| |||
Db        821 CCTTGGTGAC-AGAGTGAAACTCCATCTCAAAAACAAACAAACAAAATACATATACATAA 879

Qy        240 ATATATATATATACACACACACACACATATATATATACACATATATATACACATATATAC 299
              ||||||||  | | | | ||| | ||| ||||||| ||| ||||||| |  |||||||
Db        880 ATATATATGCACATATATATACATATATAAATATATATACACATATATAAATCTATATAC 939

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACAT-----ATATATA 354
              ||||||||| ||||| ||| ||||| ||  | |||||||| || ||||      ||||||
Db        940 ATATATACATATATACACATATATAAATCTATATACATATATATACATATATAATATATT 999

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              ||||||  || | | ||| ||| | | ||||||| | ||| |||| |||| | | |||
Db       1000 TACATATATAAATATATACATATATAAATATACATATATAAATACATATATAAATATACA 1059

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              |||||| | ||| ||| | | ||||||| ||||| | ||||| | | ||| ||| | |||
Db       1060 TATATAAATATACATATATAAATATACATATATAAATATATACATATATAAATATACATA 1119

Qy        475 TACACATATATATA-ATATACACACATATATATACACATATATACACACATATATACACA 533
              || | |||||||||  ||||| | | ||||| ||| |||  |||| ||| |||||||||  |
Db       1120 TATAAATATATATACATATATAAATATATAAATATACAAGTATATACAAATATATACATA 1179

Qy        534 TATATACACATATATATACACATATATACACATATATACATATATACACATATATATACA 593
              ||||  |   |||||  ||  | | ||||| | ||| ||| ||| | | || | |   |
Db       1180 TATAAATGTATATACGTATATACATATATATATAAATATATAAAAAAACTTTTGGCTGGG 1239

Qy        594 TATATACACATATATACATATATACACATATATACATATATACACACATATATACACATA 653
              |  |   || || |      | | | |  | |     ||  | || ||| | | |||||
Db       1240 CACCTTTCCAAATCTCATGGCACATATAAGTCTCATGGTAACCTCAAATAAAAAAACATA 1299

Qy        654 CATATACACACACATAGATATACATATATATA 685
                |          || | ||| | |   ||| |
Db       1300 TAACAGATACACCAAAAATAAAAACCAATAAA 1331
```

RESULT 11
ABS54997
ID ABS54997 standard; DNA; 15418 BP.
XX
AC ABS54997;
XX
DT 10-DEC-2002 (first entry)
XX
DE Lambda clone containing human TERT genomic insert.
XX
KW Telomerase reverse transcriptase; TERT; replication-conditional virus;
KW adenovirus replication gene; cancer cell; lung; pancreatic cancer;
KW medulloblastoma; cervical carcinoma; fibrosarcoma; osteosarcoma;
KW cytolysis; replication defective adenovirus vector; congenital defect;
KW proinflammatory; antiinflammatory; heterologous effector gene;
KW cancer therapy; cytostatic; gene therapy; lambda clone; human; ds.
XX
OS Bacteriophage lambda.
OS Homo sapiens.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT misc_feature 1. .43
FT /*tag= a
FT /note= "From Lambda clone"
FT misc_feature 44. .15375
FT /*tag= b
FT /note= "Human telomerase reverse transcriptase (hTERT)
FT sequence including the hTERT promoter and upstream
FT sequence"
FT misc_feature 15376. .15418
FT /*tag= c
FT /note= "From Lambda clone"
XX
PN WO200253760-A2.
XX
PD 11-JUL-2002.
XX
PF 17-DEC-2001; 2001WO-US048785.
XX
PR 18-DEC-2000; 2000US-0256418P.
XX
PA (GERO-) GERON CORP.
XX
PI Irving JM, Lebkowski JS;
XX
DR WPI; 2002-723123/78.
XX
PT Novel replication-conditional virus useful for cytolysis of target cells
PT e.g. cancer cells and preparing a medicament for treating cancer,
PT comprises heterologous replication element in an adenovirus-based
PT construct.
XX
PS Claim 11; Page 26-29; 32pp; English.
XX
CC The present invention relates to a new replication-conditional virus with
CC a genome comprising adenovirus replication genes and one or more
CC heterologous gene(s) that functionally replaces one or more adenovirus
CC gene(s) required for replication or assembly of the virus. The invention
CC is useful for killing a cancer cell (such as lung, pancreatic cancer,
CC medulloblastoma, cervical carcinoma, fibrosarcoma or osteosarcoma),
CC killing a cell expressing TERT (telomerase reverse transcriptase), and in
CC preparing a medicament for treating cancer and a condition associated
CC with increased expression of TERT in affected cells, in a subject. The
CC invention is also useful for cytolysis of specific target cells. The

CC invention is further useful for producing replication defective
CC adenovirus vector which is useful for transient expression of a
CC heterologous therapeutic gene to correct a congenital defect, introducing
CC proinflammatory or antiinflammatory activity, enhancing telomerase
CC function, and delivering heterologous effector genes that induce killing
CC of the transduced cells. The invention is more safe for use in cancer
CC therapy. The present nucleic acid sequence represents the human TERT
CC sequence contained within a lambda clone sequence of the invention
XX
SQ Sequence 15418 BP; 4518 A; 3797 C; 3765 G; 3338 T; 0 U; 0 Other;

Query Match 7.2%; Score 296; DB 6; Length 15418;
Best Local Similarity 69.2%; Pred. No. 3.9e-41;
Matches 479; Conservative 0; Mismatches 200; Indels 13; Gaps 5;

```
Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||| |||||||||||  ||| || | ||||||||||||| |||||||||  || ||||||
Db        646 AGTTCGAGACCAGCCTGGCCAATATAGCGAAACCCCATCTGTACTAAAAACACAAAAATT 705

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   ||  |||||| ||| ||||||||||||||| ||||  ||||||  ||||||||||
Db        706 AGCTGGGTGTGGTGACACATGCCTGTAATCCCAGGTACTCAGGAGGCTAAGGCAGGAGAA 765

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              |  |||||||  |  ||  | || |||     |   ||||||    ||| ||||||||||
Db        766 TTGCTTGAACTGGGAGGCAGAGGTGGT-----GAGCCAAGATTGCACCAGTGCACTCCAG 820

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  ||  || |||| ||||   ||||||| ||| ||| |||   | |  ||||| |||
Db        821 CCTTGGTGAC-AGAGTGAAACTCCATCTCAAAAACAAACAAACAAAATACATATACATAA 879

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              ||||||||  | | | | | ||| | ||| ||||||| ||| ||||||| |  |||||||
Db        880 ATATATATGCACATATATATACATATATAAATATATATACACATATATAAATCTATATAC 939

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACAT-----ATATATA 354
              ||||||||| ||||| ||| ||||| ||  | |||||||| || ||||     ||||||
Db        940 ATATATACATATATACACATATATAAATCTATATACATATATATACATATATAATATATT 999

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              |||||| || | | ||| ||| | | ||||||| | ||| ||||| ||||| | | |||
Db       1000 TACATATATAAATATATACATATATAAATATACATATATAAATACATATATAAATATACA 1059

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              |||||| | ||| ||| | | ||||||| ||||| | ||||| | | ||| ||| | |||
Db       1060 TATATAAATATACATATATAAATATACATATATAAATATATACATATATAAATATACATA 1119

Qy        475 TACACATATATATA-ATATACACACATATATATACACATATATACACACATATATACACA 533
              || | |||||||| ||||| | | ||||| ||| |||  |||| ||| |||||||| |
Db       1120 TATAAATATATATACATATATAAATATATAAATATACAAGTATATACAAATATATACATA 1179

Qy        534 TATATACACATATATATACACATATATACACATATATACATATATACACATATATATACA 593
              |||| |   |||||  || | | ||||| | ||| ||| ||| | | || | |   |
Db       1180 TATAAATGTATATACGTATATACATATATATATAAATATATAAAAAAACTTTTGGCTGGG 1239

Qy        594 TATATACACATATATACATATATACACATATATACATATATACACACATATATACACATA 653
              |  |   || || |      | | | | | |      ||  | || ||| | | |||||
Db       1240 CACCTTTCCAAATCTCATGGCACATATAAGTCTCATGGTAACCTCAAATAAAAAAACATA 1299

Qy        654 CATATACACACACATAGATATACATATATATA 685
              |       || | ||| | |   |||  |
Db       1300 TAACAGATACACCAAAAATAAAAACCAATAAA 1331
```

RESULT 12
AAL38595
ID AAL38595 standard; DNA; 15418 BP.
XX
AC AAL38595;
XX
DT 16-AUG-2002 (first entry)
XX
DE DNA of plasmid pGRN144 containing human hTERT gene.
XX
KW Anticonvulsant; cerebroprotective; vasotropic; nootropic; stem cell;
KW neuroprotective; antiparkinsonian; antiinflammatory; undifferentiated;
KW cardiant; transcriptional control element; human embryonic; tissue graft;
KW regenerative medicine; tissue reconstitution; neural progenitor cell;
KW nerve system; epilepsy; stroke; ischaemia; Huntington's disease;
KW Parkinson's disease; multiple sclerosis; leukodystrophy; neuritis;
KW hepatocyte; liver damage; cardiomyocyte; cryoinjury; cardiac muscle;
KW pGRN144; human hTERT gene; chimeric; ds.
XX
OS Homo sapiens.
OS Unidentified.
OS Chimeric.
XX
PN WO200242445-A2.
XX
PD 30-MAY-2002.
XX
PF 26-NOV-2001; 2001WO-US044309.
XX
PR 27-NOV-2000; 2000US-0253357P.
PR 27-NOV-2000; 2000US-0253443P.
PR 13-FEB-2001; 2001US-00783203.
XX
PA (GERO-) GERON CORP.
XX
PI Gold JD, Lebrowski JS;
XX
DR WPI; 2002-479952/51.
XX
PT Depleting a cell e.g., human embryonic stem cell population of
PT undifferentiated stem cells (UC) for use in regenerative medicine
PT comprises genetically altering UC in a population to express nucleic acid
PT encoding a lethal product.
XX
PS Example 9; Page 62-66; 67pp; English.
XX
CC The invention relates to a system for depleting a cell population of
CC undifferentiated stem cells, by introducing nucleic acids of structure P-
CC X, where X is a nucleic acid encoding product that is lethal to a cell in
CC which it is expressed; or renders a cell in which it is expressed
CC susceptible to the lethal effect of an external agent, and P is a
CC transcriptional control element causing X to be preferentially expressed
CC in the undifferentiated stem cells. The system is used for depleting a
CC cell population (preferably, human embryonic stem cells) of
CC undifferentiated stem cells. A population of differentiated cells is
CC useful in regenerative medicine, and for preparing antibodies and cDNA
CC libraries that are specific for a differentiated phenotype. The cell
CC populations are also useful for drug screening and therapeutic
CC applications. The differentiated cells are useful for tissue
CC reconstitution or regeneration in a human patient in need of treatment.
CC The cells are administered in a manner that permits to graft to the
CC intended tissue side and reconstitute or regenerate the functionally
CC deficient area. The neural progenitor cells are useful for treating acute
CC or chronic damage to the nerve system e.g. epilepsy, stroke, ischaemia,
CC Huntington's disease, Parkinson's disease, multiple sclerosis,
CC leukodystrophies, neuritis, etc. The hepatocytes and hepatocyte
CC precursors are useful for assessing animal models for ability to repair
CC liver damage. The cardiomyocyte population is useful for assessing animal
CC models for cryoinjury, regenerating cardiac muscle and to treat
CC insufficient cardiac function. This polynucleotide sequence represents
CC the DNA of plasmid pGRN144 containing the human hTERT gene relating to
CC the invention
XX
SQ Sequence 15418 BP; 4519 A; 3797 C; 3764 G; 3338 T; 0 U; 0 Other;

Query Match 7.2%; Score 296; DB 6; Length 15418;
Best Local Similarity 69.2%; Pred. No. 3.9e-41;
Matches 479; Conservative 0; Mismatches 200; Indels 13; Gaps 5;

```
Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||| |||||||||||  ||| || | ||||||||||||| |||||||||  || ||||||
Db        646 AGTTCGAGACCAGCCTGGCCAATATAGCGAAACCCCATCTGTACTAAAAACACAAAAATT 705
```

```
Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   ||  |||||| ||| ||||||||||||||| ||||  ||||||  ||||||||||
Db        706 AGCTGGGTGTGGTGACACATGCCTGTAATCCCAGGTACTCAGGAGGCTAAGGCAGGAGAA 765

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              |  |||||||  |  ||  | || |||     |   ||||||    ||| ||||||||||
Db        766 TTGCTTGAACTGGGAGGCAGAGGTGGT-----GAGCCAAGATTGCACCAGTGCACTCCAG 820

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  ||  || |||| ||||   ||||||| ||| ||| |||   | |  ||||| |||
Db        821 CCTTGGTGAC-AGAGTGAAACTCCATCTCAAAAACAAACAAACAAAATACATATACATAA 879

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              ||||||||  | | | | | ||| | ||| ||||||| ||| ||||||| |  |||||||
Db        880 ATATATATGCACATATATATACATATATAAATATATATACACATATATAAATCTATATAC 939

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACAT-----ATATATA 354
              |||||||| ||||| ||| ||||| ||  | |||||||| || ||||     ||||||
Db        940 ATATATACATATATACACATATATAAATCTATATACATATATATACATATATAATATATT 999

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              ||||||  || | | ||| ||| | | ||||||| | ||| ||||| ||||| | | |||
Db       1000 TACATATATAAATATATACATATATAAATATACATATATAAATACATATATAAATATACA 1059

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              |||||| | ||| ||| | | ||||||| ||||| | ||||| | | ||| ||| | |||
Db       1060 TATATAAATATACATATATAAATATACATATATAAATATATACATATATAAATATACATA 1119

Qy        475 TACACATATATATA-ATATACACACATATATATACACATATATACACACATATATACACA 533
              || | ||||||||| ||||| | | ||||| ||| |||  |||| ||| |||||||| |
Db       1120 TATAAATATATATACATATATAAATATATAAATATACAAGTATATACAAATATATACATA 1179

Qy        534 TATATACACATATATATACACATATATACACATATATACATATATACACATATATATACA 593
              |||| |   |||||  || | | ||||| | ||| ||| ||| | | || | |   |
Db       1180 TATAAATGTATATACGTATATACATATATATATAAATATATAAAAAAACTTTGGCTGGG 1239

Qy        594 TATATACACATATATACATATATACACATATATACATATATACACACATATATACACATA 653
              |  |   || || |     | | | |  | |     ||  | || ||| | | |||||
Db       1240 CACCTTTCCAAATCTCATGGCACATATAAGTCTCATGGTAACCTCAAATAAAAAAACATA 1299

Qy        654 CATATACACACACATAGATATACATATATATA 685
              |          || | ||| | |   ||| |
Db       1300 TAACAGATACACCAAAAATAAAAACCAATAAA 1331
```

RESULT 13
ADC21253
ID ADC21253 standard; DNA; 15418 BP.
XX
AC ADC21253;
XX
DT 18-DEC-2003 (first entry)
XX
DE Lambda clone (lambdaGphi5) containing human TERT DNA.
XX
KW Viral vector; heterologous control element; gene expression;
KW human telomerase reverse transcriptase; hTERT; tumour specific gene;
KW cell death; transcriptional control element;
KW human telomerase RNA component; hTR; cancer cell; liver cancer;
KW prostate cancer; muscle cancer; neural cell cancer; lung cancer;
KW pancreatic cancer; medulloblastoma; cervical carcinoma; fibrosarcoma;
KW osteosarcoma; lambda clone; lambdaGphi5; human; ds.
XX
OS Synthetic.
OS Homo sapiens.
OS Bacteriophage lambda.
XX
PN US2003099616-A1.
XX
PD 29-MAY-2003.
XX
PF 25-JUL-2002; 2002US-00206447.
XX
PR 25-JUL-2001; 2001US-0308029P.
XX
PA (IRVI/) IRVING J M.
PA (KARP/) KARPF D B.
PA (SCHI/) SCHIFF J M.
XX
PI Irving JM, Karpf DB, Schiff JM;
XX
DR WPI; 2003-730140/69.
XX
PT New dual specificity vectors driven by the telomerase promoter, useful
PT for killing or slowing the growth of tumor cells, or for treating cancer,
PT e.g. liver cancer, prostate cancer, lung cancer, or pancreatic cancer.
XX
PS Example 1; Page 13-20; 25pp; English.
XX
CC The present invention relates to a viral vector comprising first and
CC second genes controlled by heterologous control elements, where the first
CC gene is preferentially expressed in cells expressing human telomerase
CC reverse transcriptase (hTERT), and another gene under the control of a
CC heterologous transcriptional control element for a tissue or tumour
CC specific gene other than TERT, and where transduction of the vector into
CC a mammalian cell expressing TERT causes the death of the cell or its
CC progeny. In particular, the second gene may be under the control of a
CC transcriptional control element for a tissue specific gene selected from
CC albumin, alpha-fetoprotein, prostate-specific antigen, mitochondrial
CC creatine kinase, myelin basic protein, glial fibrillary acidic protein,
CC and neuron-specific enolase. The second gene may be under the control of
CC a transcriptional control element for a human telomerase RNA component
CC (hTR). The vector of the invention is useful for killing a cancer cell,
CC or treating a subject for a condition associated with increased
CC expression of TERT in affected cells. The vector is also useful in the
CC preparation of a medicament for treatment of a condition associated with
CC increased expression of TERT, particularly cancer, and especially liver
CC cancer, prostate cancer, muscle cancer, neural cell cancer, lung cancer,
CC pancreatic cancer, medulloblastoma, cervical carcinoma, fibrosarcoma, and
CC osteosarcoma. The present sequence represents a lambda clone
CC (lambdaGphi5) containing human TERT DNA.
XX
SQ Sequence 15418 BP; 4518 A; 3797 C; 3765 G; 3338 T; 0 U; 0 Other;

Query Match 7.2%; Score 296; DB 10; Length 15418;
Best Local Similarity 69.2%; Pred. No. 3.9e-41;
Matches 479; Conservative 0; Mismatches 200; Indels 13; Gaps 5;

```
Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              |||||  ||||||||||||   ||| || | |||||||||||| |||||||| || ||||||
Db        646 AGTTCGAGACCAGCCTGGCCAATATAGCGAAACCCCATCTGTACTAAAAACACAAAAATT 705

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   ||  |||||| ||| ||||||||||||||| ||||  ||||||  ||||||||||
Db        706 AGCTGGGTGTGGTGACACATGCCTGTAATCCCAGGTACTCAGGAGGCTAAGGCAGGAGAA 765

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              |  |||||||  |  ||  | || |||     |   ||||||    ||| ||||||||||
Db        766 TTGCTTGAACTGGGAGGCAGAGGTGGT-----GAGCCAAGATTGCACCAGTGCACTCCAG 820

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  ||  || |||| ||||   ||||||| ||| ||| |||   | |  ||||| |||
Db        821 CCTTGGTGAC-AGAGTGAAACTCCATCTCAAAAACAAACAAACAAAATACATATACATAA 879

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              ||||||||  | | | | | ||| | ||| ||||||| ||| ||||||| |  |||||||
Db        880 ATATATATGCACATATATATACATATATAAATATATATACACATATATAAATCTATATAC 939

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACAT-----ATATATA 354
              |||||||| ||||| ||| ||||| ||  | |||||||| || ||||     ||||||
Db        940 ATATATACATATATACACATATATAAATCTATATACATATATATACATATATAATATATT 999

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              ||||||  || | | ||| ||| | | ||||||| | ||| ||||| ||||| | | |||
Db       1000 TACATATATAAATATATACATATATAAATATACATATATAAATACATATATAAATATACA 1059
```

```
Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              |||||| | ||| ||| | | ||||||| ||||| | ||||| | | ||| ||| | |||
Db       1060 TATATAAATATACATATATAAATATACATATATAAATATATACATATATAAATATACATA 1119

Qy        475 TACACATATATATA-ATATACACACATATATATACACATATATACACACATATATACACA 533
              || | ||||||||| ||||| | | ||||| ||| |||  |||| ||| |||||||| |
Db       1120 TATAAATATATATACATATATAAATATATAAATATACAAGTATATACAAATATATACATA 1179

Qy        534 TATATACACATATATATACACATATATACACATATATACATATATACACATATATATACA 593
              |||| |   |||||  || | | ||||| | ||| ||| ||| | | || | |    |
Db       1180 TATAAATGTATATACGTATATACATATATATATAAATATATAAAAAAACTTTTGGCTGGG 1239

Qy        594 TATATACACATATATACATATATACACATATATACATATATACACACATATATACACATA 653
               |   |   || || |       | | | |  | |      ||  | || ||| | | |||||
Db       1240 CACCTTTCCAAATCTCATGGCACATATAAGTCTCATGGTAACCTCAAATAAAAAAACATA 1299

Qy        654 CATATACACACACATAGATATACATATATATA 685
               |          || | ||| | |   ||| |
Db       1300 TAACAGATACACCAAAAATAAAAACCAATAAA 1331
```

RESULT 14
ADC87552/c
ID ADC87552 standard; DNA; 1486 BP.
XX
AC ADC87552;
XX
DT 01-JAN-2004 (first entry)
XX
DE Human GPCR gene SEQ ID NO:2005.
XX
KW ds; gene; human; GPCR;
KW guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1270724-A2.
XX
PD 02-JAN-2003.
XX
PF 18-JUN-2002; 2002EP-00013517.
XX
PR 18-JUN-2001; 2001JP-00246789.
XX
PA (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI Suwa M, Asai K, Akiyama Y, Aburatani H;
XX
DR WPI; 2003-315783/31.
DR P-PSDB; ADC87553.
XX
PT New polynucleotide, useful for preparing a composition for treating a
PT patient in need of increased or suppressed activity or expression of the
PT guanosine triphosphate-binding protein coupled receptor.
XX
PS Claim 1; SEQ ID NO 2005; 28pp; English.
XX
CC The invention relates to a novel polynucleotide encoding a guanosine
CC triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC the invention may have a use in gene therapy. The polynucleotide and
CC polypeptide are useful for preparing a composition for treating a patient
CC in need of increased or suppressed activity or expression of the
CC guanosine triphosphate-binding protein coupled receptor. The
CC polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC invention.
XX
SQ Sequence 1486 BP; 517 A; 149 C; 242 G; 578 T; 0 U; 0 Other;

Query Match 7.2%; Score 294.8; DB 10; Length 1486;
Best Local Similarity 74.9%; Pred. No. 5.8e-41;
Matches 397; Conservative 0; Mismatches 127; Indels 6; Gaps 2;

```
Qy        197 AAATGTCATCTCAGAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACA 256
              | || |   | || | || | | |      || || | | ||  | ||||||| |||||||
Db        999 ATATATACACACATATAATATATACATATTTATACACACATGCACATATATAAATACACA 940

Qy        257 CACACACACATATATATATACACATATATATACACATATATACATATATACACATATATA 316
              | | ||| ||||| | | | | || |||||| |||||||||||||||||||||||| ||
Db        939 TATATACATATATACACACATATATGTATATATACATATATACACATATGTA 880

Qy        317 CACATATATATACACATACATATGTACACATATATATACACATATGTATACACATATATA 376
              | ||| ||||| ||||| |||   | |  ||||||||||||||||||||| | | | |||
Db        879 TATATACATATATACATATATACACATATGTATATATACACATATGTATATATACACATA 820

Qy        377 CACATATATACACACATATATACACATATATACACACATATATACACATATATACACATA 436
              ||||||| || | | | | |||||||||| || | | | | |||||||||| || | |||
Db        819 CACATATGTATATATACACATACACATATGTATATATACACATACACATATGTATATATA 760

Qy        437 TATACACATATACACATATATACACATATATACATATATACACATATATATA-----ATA 491
              | | ||| |||   |||||||||||||| ||  |||||||||||||||| ||      | |
Db        759 CACATACACATATGTATATATACACATACATGTATATATACACATATATGTATATATACA 700

Qy        492 TACACACATATATATACACATATATACACACATATATACACATATATACACATATATATA 551
              | |||||| |||| ||| | | ||| |   |||||||||||||| | | | ||| | |||
Db        699 CATACACATGTATACACATACACATATATGTATATATACACATACACATATATACACATA 640

Qy        552 CACATATATACACATATATACATATATACACATATATATACATATATACACATATATACA 611
              ||| ||||||||||||||||| ||||||||||||||   |||||||||||||| |||
Db        639 TACACATATACACATATATACACATATACACATATATAAGTATATATACACATATACACA 580

Qy        612 TATATACACATATATACATATATACACACATATATACACATACATAT-ACACACACATAG 670
              ||||   | ||||| |||||||||||||| |||  || | ||||| || | ||||| |||
Db        579 TATACGTATATATACACATATATACACATATACGTATATATACACATAATACACATATAC 520

Qy        671 ATATACATATATATACACATATATATACGTATATATATGTATATATATAT 720
              |||| ||| | ||| | | | |||||||||||||||| ||||||| | ||
Db        519 GTATAAATACACATATATACACATATACGTATATATACGTATATACACAT 470
```

RESULT 15
AAD50021
ID AAD50021 standard; DNA; 27067 BP.
XX
AC AAD50021;
XX
DT 24-MAR-2003 (first entry)
XX
DE Human secreted protein encoding gene.
XX
KW Human; secreted protein; cancer; gene therapy; cytostatic; gene; ds.
XX
OS Homo sapiens.
XX
FH Key   Location/Qualifiers
FT variation   replace(921,T)
FT   /*tag= a
FT   /standard_name= "Single nucleotide polymorphism"
FT variation   replace(1022,T)
FT   /*tag= ba
FT   /standard_name= "Single nucleotide polymorphism"
FT variation   replace(1781,T)
FT   /*tag= b
FT   /standard_name= "Single nucleotide polymorphism"
FT variation   replace(1850,A)
FT   /*tag= c
FT   /standard_name= "Single nucleotide polymorphism"
FT variation   replace(1882,T)
FT   /*tag= bb
FT   /standard_name= "Single nucleotide polymorphism"
FT variation   replace(1951,A)
FT   /*tag= bc
FT   /standard_name= "Single nucleotide polymorphism"
FT CDS   2031. .25941
FT   /*tag= d

```
FT                   /product= "Human secreted protein"
FT   exon            2031. .2094
FT                   /*tag= e
FT   intron          2095. .5569
FT                   /*tag= f
FT   exon            2132. .2195
FT                   /*tag= bd
FT   intron          2196. .5670
FT                   /*tag= be
FT   variation       replace(2839,G)
FT                   /*tag= g
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(2940,G)
FT                   /*tag= bf
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(3730,A)
FT                   /*tag= h
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(3831,A)
FT                   /*tag= bg
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            5570. .5762
FT                   /*tag= i
FT   exon            5671. .5863
FT                   /*tag= bh
FT   intron          5763. .7571
FT                   /*tag= j
FT   intron          5864. .7672
FT                   /*tag= bi
FT   variation       replace(6631,A)
FT                   /*tag= k
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(6732,A)
FT                   /*tag= bj
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(6945,A)
FT                   /*tag= l
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(6952,T)
FT                   /*tag= m
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(7457,A)
FT                   /*tag= n
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(7558,A)
FT                   /*tag= bk
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            7572. .7741
FT                   /*tag= o
FT   exon            7673. .7761
FT                   /*tag= bl
FT   intron          7742. .10000
FT                   /*tag= p
FT   intron          7762. .9149
FT                   /*tag= bm
FT   variation       replace(7830,A)
FT                   /*tag= q
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(7931,A)
FT                   /*tag= bn
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(8089,C)
FT                   /*tag= r
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(8190,C)
FT                   /*tag= bo
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(8551,T)
FT                   /*tag= s
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(8652,T)
FT                   /*tag= bp
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            9150. .9302
FT                   /*tag= bq
FT   variation       replace(9269,C)
FT                   /*tag= t
FT                   /standard_name= "Single nucleotide polymorphism"
FT   intron          9303. .10101
FT                   /*tag= br
FT   variation       replace(9362,T)
FT                   /*tag= u
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(9370,C)
FT                   /*tag= bs
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(9463,T)
FT                   /*tag= bt
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(9782,T)
FT                   /*tag= v
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(9883,T)
FT                   /*tag= bu
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            10001. .10173
FT                   /*tag= w
FT   exon            10102. .10274
FT                   /*tag= bv
FT   intron          10174. .10298
FT                   /*tag= x
FT   intron          10275. .10399
FT                   /*tag= bw
FT   exon            10299. .10485
FT                   /*tag= y
FT   exon            10400. .10586
FT                   /*tag= bx
FT   intron          10486. .12027
FT                   /*tag= z
FT   intron          10587. .12128
FT                   /*tag= by
FT   variation       replace(11493,W)
FT                   /*tag= aa
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(11594,W)
FT                   /*tag= bz
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            12028. .12193
FT                   /*tag= ab
FT   exon            12129. .12294
FT                   /*tag= ca
FT   intron          12194. .25821
FT                   /*tag= ac
FT   variation       replace(12260,G)
FT                   /*tag= ad
FT                   /standard_name= "Single nucleotide polymorphism"
FT   intron          12295. .25922
FT                   /*tag= cb
FT   variation       replace(12361,G)
FT                   /*tag= cc
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(13086,C)
FT                   /*tag= ae
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(13183,C)
FT                   /*tag= af
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(13187,C)
FT                   /*tag= cd
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(13284,C)
FT                   /*tag= ce
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(21240,G)
FT                   /*tag= ag
FT                   /standard_name= "Single nucleotide polymorphism"
```

```
FT   variation       replace(21341,G)
FT                   /*tag=  cf
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(21695,G)
FT                   /*tag=  ah
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(21796,G)
FT                   /*tag=  cg
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(22058,T)
FT                   /*tag=  ai
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(22159,T)
FT                   /*tag=  ch
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(22233,A)
FT                   /*tag=  aj
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(22244. .22245,A)
FT                   /*tag=  ak
FT                   /standard_name= "Single nucleotide polymorphism"

  Query Match             7.1%;  Score 293.6;  DB 10;  Length 27067;
  Best Local Similarity   75.8%;  Pred. No. 1e-40;
  Matches  391;  Conservative    0;  Mismatches  119;  Indels    6;  Gaps    2;

Qy        230 ATATATATATATATATATATATACACACACACACACATATAT-ATATACACATATATATA 288
              || ||||  |  | |  |  ||| |     | ||| |||||| | ||| | ||||| | |
Db      26509 ATGTATAACTTAAAAATTTCATATATGTGTATACATATATATGAAATATATATATACAGA 26568

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              |||| ||||| || |||||| ||||||||||||||||||  | | | ||||  |||||||
Db      26569 CACACATATATATGTATACATATATATACACATATATATGTATACATATATACACACATA 26628

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              ||| || |||||| |||||||| |||||||||||||||||||| ||| | | ||| | |
Db      26629 TATGTATACATATATATACACACATATACACATATATACACACACATACATATATACACA 26688

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              |||| ||||| |||||||||||||| ||  |||||||| |  |||| | | ||| | |
Db      26689 CACATATATACACACATATATACACACATGCACACATATATATGTATACATATATACACA 26748

Qy        469 CATATATACACATATATATA-----ATATACACACATATATATACACATATATACACACA 523
              ||| |||||  ||||||| |     |||||||||||||||||||||||| ||||||||||
Db      26749 CATGTATACGTATATATACACACATATATACACACATATATATACACACATATACACACA 26808

Qy        524 TATATACACATATATACACATATATATACACATATATACACATATATACATATATACACA 583
              || | |||||||||||||||| |||||||||| ||||| ||| | ||| |||| || |||
Db      26809 TACACACACATATATACACACATATATACACACATATATACACACATATATATGTATACA 26868

Qy        584 TATATATACATATATACACATATATACATATATACACATATATACATATATACACACATA 643
              |||||||||| | ||| | | | ||||| | ||||| ||||| ||||||| ||| | | |
Db      26869 TATATATACACACATATATACACATACACACATACATATATACACATATACACATATACA 26928

Qy        644 TATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATA 703
              | ||| ||||| | |   | ||||| ||||||| | || || |||||| |||||  | |
Db      26929 CACACATATACACACATGTATACATATATATACACACATGTATACATATGTATACACACA 26988

Qy        704 TATATGTATATATATATGCTCCAGAGTTCATAAGAG 739
              |||| | |||| ||||    |     | |||| | |
Db      26989 CATATATGTATACATATATACACACATACATATGTG 27024


RESULT 16
ADC87570/c
ID   ADC87570 standard; DNA; 1672 BP.
XX
AC   ADC87570;
XX
DT   01-JAN-2004  (first entry)
XX
DE   Human GPCR gene SEQ ID NO:2023.
XX
KW   ds; gene; human; GPCR;
KW   guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1270724-A2.
XX
PD   02-JAN-2003.
XX
PF   18-JUN-2002; 2002EP-00013517.
XX
PR   18-JUN-2001; 2001JP-00246789.
XX
PA   (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA   (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI   Suwa M,  Asai K,  Akiyama Y,  Aburatani H;
XX
DR   WPI; 2003-315783/31.
DR   P-PSDB; ADC87571.
XX
PT   New polynucleotide, useful for preparing a composition for treating a
PT   patient in need of increased or suppressed activity or expression of the
PT   guanosine triphosphate-binding protein coupled receptor.
XX
PS   Claim 1; SEQ ID NO 2023; 28pp; English.
XX
CC   The invention relates to a novel polynucleotide encoding a guanosine
CC   triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC   the invention may have a use in gene therapy. The polynucleotide and
CC   polypeptide are useful for preparing a composition for treating a patient
CC   in need of increased or suppressed activity or expression of the
CC   guanosine triphosphate-binding protein coupled receptor. The
CC   polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC   invention.
XX
SQ   Sequence 1672 BP; 499 A; 171 C; 300 G; 602 T; 0 U; 100 Other;

  Query Match             7.1%;  Score 291.6;  DB 10;  Length 1672;
  Best Local Similarity   75.5%;  Pred. No. 2e-40;
  Matches  376;  Conservative    0;  Mismatches  119;  Indels    3;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACATATATATATACACATATATA 286
              | |||||||||||||||||||||||| | | ||||||||| | ||||||||||||||||
Db      1390 TATATATATATATATATATATATATATATACACACACACATATATACACATATATA 1331

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              |||||| | | | ||||| ||| ||||| |||||||||||||||||||| |||   | | |
Db      1330 TACACACACATATATACACATATATACACACATATATATACACATATATACACATATA 1271

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||| | | | ||||  | | | |||| |||||||||||||||||| |||| || ||||
Db      1270 TATACACATATATATACACATATATATGTACACATATATACACACATGTATATACGTATA 1211

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              |    |||||||||||||| | ||| |  ||| |  | |||||||||||||  | |||||
Db      1210 TGTGTACATATATACACACACATATATGTATGTGTATATATACACATATATGTATATATA 1151

Qy        467 TACATATATACACATATATATA---ATATACACACATATATATACACATATATACACACA 523
              |||| | | | |||| ||||||    |||| | ||| | ||| | | || |||| |
Db      1150 TACACACACAGACATGTATATATGTGTATATATACACACATAGATATATGTATATATGTG 1091

Qy        524 TATATACACATATATACACATATATATACACATATATACACATATATACATATATACACA 583
              |||||||||| | ||| |  | |||||||||| |||||    |||||||| | | | | |
Db      1090 TATATACACACACATATATGTGTATATACACACATATATGTGTATATACACACACATATA 1031

Qy        584 TATATATACATATATACACATATATACATATATACACATATATACATATATACACACATA 643
              | | |||| | | | ||| |||| |  ||||| ||||| |||||  | |||| ||||| |
Db      1030 TGTGTATATACACACACATATATGTGTATATACACACACATATATGTGTATATACACACA 971

Qy        644 TATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATA 703
              |||  |   |||||||||||||||  |||||  | |||||||||||||||| ||  |||
Db        970 CATATATGTGTATATACACACACATGTATATATGTGTATATACACATATATGTATATATG 911
```

```
Qy        704 TATATGTATATATATATG 721
              | ||| || | |||||||
Db        910 TGTATATACACATATATG 893


RESULT 17
AAK67239
ID   AAK67239 standard; DNA; 30393 BP.
XX
AC   AAK67239;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:22051.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001354.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0235484P.
PR   27-SEP-2000; 2000US-0235834P.
PR   27-SEP-2000; 2000US-0235836P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   13-OCT-2000; 2000US-0239937P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241221P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   20-OCT-2000; 2000US-0241826P.
PR   01-NOV-2000; 2000US-0244617P.
PR   08-NOV-2000; 2000US-0246474P.
PR   08-NOV-2000; 2000US-0246475P.
PR   08-NOV-2000; 2000US-0246476P.
PR   08-NOV-2000; 2000US-0246477P.
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246526P.
PR   08-NOV-2000; 2000US-0246527P.
PR   08-NOV-2000; 2000US-0246528P.
PR   08-NOV-2000; 2000US-0246532P.
PR   08-NOV-2000; 2000US-0246609P.
PR   08-NOV-2000; 2000US-0246610P.
PR   08-NOV-2000; 2000US-0246611P.
PR   08-NOV-2000; 2000US-0246613P.
PR   17-NOV-2000; 2000US-0249207P.
PR   17-NOV-2000; 2000US-0249208P.
PR   17-NOV-2000; 2000US-0249209P.
PR   17-NOV-2000; 2000US-0249210P.
PR   17-NOV-2000; 2000US-0249211P.
PR   17-NOV-2000; 2000US-0249212P.
PR   17-NOV-2000; 2000US-0249213P.
PR   17-NOV-2000; 2000US-0249214P.
PR   17-NOV-2000; 2000US-0249215P.
PR   17-NOV-2000; 2000US-0249216P.
PR   17-NOV-2000; 2000US-0249217P.
PR   17-NOV-2000; 2000US-0249218P.
PR   17-NOV-2000; 2000US-0249244P.
PR   17-NOV-2000; 2000US-0249245P.
PR   17-NOV-2000; 2000US-0249264P.
```

PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS Disclosure; SEQ ID NO 22051; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention
XX
SQ Sequence 30393 BP; 8271 A; 7315 C; 8076 G; 6731 T; 0 U; 0 Other;

Query Match 7.0%; Score 288; DB 4; Length 30393;
Best Local Similarity 76.1%; Pred. No. 9.3e-40;
Matches 382; Conservative 0; Mismatches 115; Indels 5; Gaps 2;

```
Qy      227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
            | ||||||||| | ||||| ||||||| | | ||||||||| | | | | ||| |||||||
Db    14755 TATATATATACACATATACATATATATAGACACACACACACACACATATATATA 14814

Qy      287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
            ||||| | ||||||||||||||| ||||||| ||||| | ||| |||||||| |||| |
Db    14815 CACACACACATACATATATACACACATATACATATATACACACATATACATATATACATA 14874

Qy      347 TATA----TATACACATATGTATACACATATATACACATATATACACACATATATACACA 402
            ||||    | |||| ||||    | | | ||||| ||| ||||||| | ||||| |||||
Db    14875 TATACACGTGTACATATATACGCATATACATATATACACATATACATATATATACACACA 14934

Qy      403 TATATACACACATATATACACATATATACACATATATACACATATACACATATATACACA 462
            |||| | | | ||||| | | | ||  |||||||||| |||||||| ||||||||||
Db    14935 TATACATATATACATATATATACACATGTACATATATACGCATATACATATATATACACA 14994

Qy      463 TATATACATATATACACATATATATA-ATATACACACATATATATACACATATATACACA 521
            |||| | ||||||||||  |||| | ||||||| | ||| | ||  |||||||||| ||
Db    14995 TATACATATATATACACACCTATACATATATACATATATACACATGTACATATATACGCA 15054
```

Qy 522 CATATATACACATATATACACATATATATACACATATATACACATATATACATATATACA 581
|||| ||| | | | | | | || |||||| ||||| ||||||||| | |||| | |
Db 15055 TATACATATATACACACATATACATGTATACATATATACACACATATACATGTATACATA 15114

Qy 582 CATATATATACATATATACACATATATACATATATACACATATATACATATATACACACA 641
||| | || | ||||| || |||||||||||||||||||| | ||||| ||||| |
Db 15115 TATACACATGTACATATATACGCATATACATATATACACATATACATATATACACACATA 15174

Qy 642 TATATACACATACATATATACACACACATAGATATACATATATATACACATATATATACGTA 701
||||| ||||| ||| ||| ||||| ||| |||||| ||| | | | |||||| ||
Db 15175 CATATATACATATATACACATGTACATATATACGCATATACATATATACACATATACATA 15234

Qy 702 TATATATGTATATATATATGCT 723
|||| | | || |||||| ||
Db 15235 TATACACATGTACATATATACT 15256

RESULT 18
ABK84481/c
ID ABK84481 standard; cDNA; 44147 BP.
XX
AC ABK84481;
XX
DT 14-AUG-2002 (first entry)
XX
DE Human cDNA differentially expressed in granulocytic cells #1052.
XX
KW Human; ss; granulocytic cell; DNA chip; bacterial infection;
KW viral infection; parasitic infection; protozoal infection;
KW fungal infection; sterile inflammatory disease; psoriasis;
KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KW cardiac reperfusion injury; renal reperfusion injury; ARDS;
KW adult respiratory distress syndrome; inflammatory bowel disease;
KW Crohn's disease; ulcerative colitis; periodontal disease;
KW granulocyte activation; chronic inflammation; allergy.
XX
OS Homo sapiens.
XX
PN WO200228999-A2.
XX
PD 11-APR-2002.
XX
PF 03-OCT-2001; 2001WO-US030821.
XX
PR 03-OCT-2000; 2000US-0237189P.
XX
PA (GENE-) GENE LOGIC INC.
XX
PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
XX
DR WPI; 2002-435328/46.
XX
PT Detecting granulocyte activation by detecting differential expression of
PT genes associated with granulocyte activation, which serves as diagnostic
PT markers that is useful for monitoring disease states and drug toxicity.
XX
PS Claim 1; SEQ ID NO 1052; 114pp; English.
XX
CC The invention relates to detecting (M1) granulocyte (GC) activation
CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
CC DNA chip analysis as given in the specification, and comparing the
CC expression level to an expression level in an unactivated GC, where
CC differential expression of Gs is indicative of GCA. Also included are
CC modulating (M2) GA by contacting GC with an agent that alters the
CC expression of at least one gene in Gs; (2) screening (M3) for an agent
CC capable of modulating GCA or an inflammation (especially chronic) in a
CC tissue, an allergic response in a subject, exposure of a subject to a
CC pathogen or sterile inflammatory disease using the gene expression
CC profile; (3) detecting (M4) an inflammation (especially chronic) in a
CC tissue, an allergic response in a subject, exposure of a subject to a
CC pathogen or sterile inflammatory disease, by detecting the level of
CC expression in a sample of the tissue of gene(s) from Gs, where the level
CC of expression of the gene is indicative of inflammation; (4) treating

CC (M5) an inflammation (especially chronic) or in a tissue, an allergic
CC response in a subject, exposure of a subject to a pathogen or sterile
CC inflammatory disease, by contacting a tissue having inflammation with an
CC agent that modulates the expression of gene(s) from Gs in the tissue. M1
CC is useful for detecting GCA; M2 is useful for modulating GA; M3 is useful
CC for screening an agent capable of modulating GCA preferably in an
CC inflammation in a tissue; M4 is useful for detecting an inflammation
CC (especially chronic) in a tissue, an allergic response in a subject,
CC exposure of a subject to a pathogen or sterile inflammatory disease (e.g.
CC psoriasis, rheumatoid arthritis, glomerulonephritis, asthma, thrombosis,
CC cardiac reperfusion injury, renal reperfusion injury, ARDS, adult
CC respiratory distress syndrome, inflammatory bowel disease, Crohn's
CC disease, ulcerative colitis, periodontal disease; also bacterial
CC infection, viral infection, parasitic infection, protozoal infection,
CC fungal infection and M5 is useful for treating one of the above
CC conditions. The present sequence represents a gene differentially
CC expressed in granulocytes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 44147 BP; 9813 A; 11709 C; 11789 G; 10833 T; 0 U; 3 Other;

Query Match 7.0%; Score 286.6; DB 6; Length 44147;
Best Local Similarity 72.9%; Pred. No. 1.6e-39;
Matches 453; Conservative 0; Mismatches 154; Indels 14; Gaps 6;

```
Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC---AAAA 57
              ||||  |||||||||||| |||||||| |||||||| |||||| |||||||||   ||||
Db      10942 AGTTTGAGACCAGCCTGACCAACATGGAGAAACCCCGTCTCTATTAAAAATACAAAAAAA 10883

Qy         58 TTAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAG 117
              |||| ||||| |||||  ||| ||||||||||||||||||||| | ||||| || |||||||
Db      10882 TTAGCCAGGCGTGGTGGCACATGCCTGTAATCCCAGCTACTTGGAAGGCTGAAGCAGGAG 10823

Qy        118 AATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTC 176
              || | ||||||||  | || || ||||||| ||| ||| | ||||||||||||||||||
Db      10822 AAGCGCTTGAACC--CGGGAGGCGGAGGTTTCGGTGAGCCCAGATCATGCCATTGCACTG 10765

Qy        177 CAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATA 236
              |||||  |||||||||||| ||||      || |    || ||| |   |||| |||||||
Db      10764 CAGCCTGGGCAACAAGAGCAAAAT------TCCGTTTAATAAATAATTATTTAATATATA 10711

Qy        237 TATATATATATATATACACACACACACATATATATATACACATATATATACACATATA 296
              ||||||| | | | || | | | | |||| |||||||||||  ||||||||||||   ||
Db      10710 TATATATGTGTGTGTATATATATATACACGTATATATATACGTATATATATACACGCGTA 10651

Qy        297 TACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACA 356
              || |||||    | ||||||||||   |||||| || || |||| ||| | ||||||| |
Db      10650 TATATATACGTATATATATACACGCGTATATATACGTATATATATACGCGTATATATATA 10591

Qy        357 CATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATA 416
              | ||| |||| |||  || | | |  |||| | | |||  || | |||||·||  |||||
Db      10590 CGTATATATATACACGTACATATACGTATATATATATACGTATATATATACACGTACATA 10531

Qy        417 TATACACATATATACACATATATACACATATACACATATATACACATATATACATATATA 476
              ||     ||||||| ||| ||||||  ||||| | | | ||| | |  |||| ||| | |
Db      10530 TACGTGTATATATATACACATATACGTATATATATATACACATATATACGTATATATACACA 10471

Qy        477 CACATATATATAATATACACACATATATATACACATATATACACACATATATACACATAT 536
              | ||||| |  ||||  | |||| |||||||| |||||||  |    ||    |  || |||
Db      10470 TATATATACA-CATATGTATACATCTATATACATATATATAAAATAATTATTTTACGTAT 10412

Qy        537 ATACACATATATATACACATATATA-CACATATATACATATATACACATATATATACATA 595
              |||      |   | |||  ||  | |  ||||| | |||| | |||||||||  |||
Db      10411 ATAATTTACGTAAATAATATACGTATAATAATTATACGTGTATATATATATATATATATA 10352

Qy        596 TATACACATATATACATATAT 616
              | ||||| ||||| ||||||
Db      10351 CACACACACATATATATATAT 10331
```

RESULT 19

ADD14691/c
ID ADD14691 standard; cDNA; 44147 BP.
XX
AC ADD14691;
XX
DT 01-JAN-2004 (first entry)
XX
DE Human src biomarker polynucleotide SEQ ID NO:85.
XX
KW predictor set; protein tyrosine kinase activity modulator;
KW protein tyrosine kinase pathway; protein tyrosine kinase; cytostatic;
KW gene therapy; drug sensitivity; genetic profile; cancer; human; gene; ss.
XX
OS Homo sapiens.
XX
PN WO2003062395-A2.
XX
PD 31-JUL-2003.
XX
PF 17-JAN-2003; 2003WO-US001981.
XX
PR 18-JAN-2002; 2002US-0350061P.
XX
PA (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
PI Huang F, Fairchild CR, Lee FY, Shaw P;
XX
DR WPI; 2003-636735/60.
DR P-PSDB; ADD14095.
XX
PT New polynucleotides and polypeptides for predicting the activity of
PT compounds that interact with protein tyrosine kinases and/or protein
PT tyrosine kinase pathways.
XX
PS Claim 2; SEQ ID NO 85; 139pp; English.
XX
CC The present invention describes a predictor set comprising a plurality of
CC polynucleotides or polypeptides whose expression pattern is predictive of
CC the response of cells to treatment with a compound that modulates protein
CC tyrosine kinase activity or members of the protein tyrosine kinase
CC pathway. Also described: (1) predicting whether a compound is capable of
CC modulating the activity of cells, comprising obtaining a sample of cells,
CC determining whether the cells express a plurality of markers, and
CC correlating the expression of the markers to the compound's ability to
CC modulate the activity of the cells; (2) a plurality of cell lines for
CC identifying polynucleotides and polypeptides whose expression levels
CC correlate with compound sensitivity or resistance of cells associated
CC with a disease state; and (3) identifying polynucleotides and
CC polypeptides that predict compound sensitivity or resistance of cells
CC associated with a disease state, comprising subjecting the plurality of
CC cell lines to one or more compounds, analysing the expression pattern of
CC a microarray of polynucleotides or polypeptides, and selecting
CC polynucleotides or polypeptides that predict the sensitivity or
CC resistance of cells associated with a disease state by using the
CC expression pattern of the microarray. The polynucleotides and
CC polypeptides have cytostatic activities, and can be used in gene therapy.
CC The polynucleotides and polypeptides are useful in predicting the
CC activity of compounds that interact with protein tyrosine kinases and/or
CC protein tyrosine kinase pathways. These may be used in determining drug
CC sensitivity in patients to allow the development of individualized
CC genetic profiles which aid in treating diseases and disorders (e.g.
CC cancer) based on patient response at a molecular level. The present
CC sequence is used in the exemplification of the present invention.
XX
SQ Sequence 44147 BP; 9813 A; 11709 C; 11789 G; 10833 T; 0 U; 3 Other;

Query Match 7.0%; Score 286.6; DB 10; Length 44147;
Best Local Similarity 72.9%; Pred. No. 1.6e-39;
Matches 453; Conservative 0; Mismatches 154; Indels 14; Gaps 6;

```
Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC---AAAA 57
              ||||  |||||||||||| |||||||| |||||||| |||||| |||||||||   ||||
Db      10942 AGTTTGAGACCAGCCTGACCAACATGGAGAAACCCCGTCTCTATTAAAAATACAAAAAAA 10883
```

```
Qy         58 TTAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAG 117
              |||| ||||| |||||  ||| ||||||||||||||||||| | |||| || |||||||
Db      10882 TTAGCCAGGCGTGGTGGCACATGCCTGTAATCCCAGCTACTTGGAAGGCTGAAGCAGGAG 10823

Qy        118 AATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTC 176
              || | ||||||||  | || || ||||||| ||| ||| | ||||||||||||||||||
Db      10822 AAGCGCTTGAACC--CGGGAGGCGGAGGTTTCGGTGAGCCCAGATCATGCCATTGCACTG 10765

Qy        177 CAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATA 236
              |||||  |||||||||||| ||||      || |   || ||| |   |||| |||||||
Db      10764 CAGCCTGGGCAACAAGAGCAAAAT------TCCGTTTAATAAATAATTATTTAATATATA 10711

Qy        237 TATATATATATATATACACACACACACACATATATATATACACATATATATACACATATA 296
              ||||||| | | | || | | | | |||| ||||||||||  ||||||||||||   ||
Db      10710 TATATATGTGTGTGTATATATATATACACGTATATATATACGTATATATATACACGCGTA 10651

Qy        297 TACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACA 356
              || |||||   | ||||||||||   |||||| || || |||| ||| | |||||||| |
Db      10650 TATATATACGTATATATATACACGCGTATATATACGTATATATATATACGCGTATATATATA 10591

Qy        357 CATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATA 416
              | ||| |||| |||  || | | |  |||| | | |||  || | ||||| ||  |||||
Db      10590 CGTATATATATACACGTACATATACGTATATATATATACGTATATATATACACGTACATA 10531

Qy        417 TATACACATATATACACATATATACACATATACACATATATACACATATATACATATATA 476
              ||      ||||||| ||| ||||||  ||||| | | | ||| | |  |||| ||| | |
Db      10530 TACGTGTATATATATACACATATACGTATATATATACACATATATACGTATATATACACA 10471

Qy        477 CACATATATATAATATACACACATATATATACACATATATACACACATATATACACATAT 536
              | ||||| |  ||||  | |||| |||||||| |||||||  |   ||   |  || |||
Db      10470 TATATATACA-CATATGTATACATCTATATACATATATATAAAATAATTATTTTACGTAT 10412

Qy        537 ATACACATATATATACACATATATA-CACATATATACATATATACACATATATATACATA 595
              |||          |   | |||  ||  | |  ||||| | |||| | |||||||| |||
Db      10411 ATAATTTACGTAAATAATATACGTATAATAATTATACGTGTATATATATATATATATATA 10352

Qy        596 TATACACATATATACATATAT 616
              | ||||| ||||| ||||||
Db      10351 CACACACACATATATATATAT 10331


RESULT 20
AAK89374
ID   AAK89374 standard; DNA; 32206 BP.
XX
AC   AAK89374;
XX
DT   05-NOV-2001  (first entry)
XX
DE   Human digestive system antigen genomic sequence SEQ ID NO: 2950.
XX
KW   Human; digestive system antigen; gene therapy; cancer; appendicitis;
KW   ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
KW   digestive system disorder; Meckel's diverticulum; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155314-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001324.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0235484P.
PR   27-SEP-2000; 2000US-0235834P.
PR   27-SEP-2000; 2000US-0235836P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   13-OCT-2000; 2000US-0239937P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241221P.
PR   20-OCT-2000; 2000US-0241785P.
```

PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-502630/55.
XX
PT Polynucleotides encoding digestive system antigens, useful for
PT diagnosing, treating, preventing and/or prognozing disorders of the
PT digestive system, particularly cancer and cancer metastases.
XX
PS Disclosure; SEQ ID NO 2950; 986pp; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human digestive system antigens. These can be used in the
CC diagnosis, treatment and prevention of digestive system disorders,
CC including cancer, Meckel's diverticulum, bacterial or parasitic
CC infections, appendicitis, Hirschsprung's disease, chronic colitis or
CC ulcerative colitis. The present sequence is a genomic DNA fragment
CC encoding a digestive system antigen of the invention
XX
SQ Sequence 32206 BP; 9503 A; 5480 C; 7017 G; 10206 T; 0 U; 0 Other;

Query Match 6.8%; Score 280.8; DB 4; Length 32206;
Best Local Similarity 74.4%; Pred. No. 1.6e-38;
Matches 369; Conservative 0; Mismatches 122; Indels 5; Gaps 1;

```
Qy      229 TATATATATATATATATATATACACACACACACATATATATATACACATATATATA 288
            ||||| |||||||| | ||| | | | | ||| | |  |||||||| ||| | | |||||
Db    21803 TATATGTATATATACACATACACATATATACATATATGTATATATACACATACACATATA 21862

Qy      289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
            ||||||||  |||||||||||||| | | |  ||| |||||| | | ||||  ||| ||
Db    21863 TACATATATGTATATATACACATACACATATGTATGTATACATACATATATACACATATG 21922

Qy      349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
            |||||| |||||  |||| ||| || |  | |||||  |  || ||||||||||| | | ||
Db    21923 TATATATACATACATATATACACATGTGTATATATACATGCAGATATATACACGTGTGTA 21982

Qy      409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
            | ||| ||||  | ||| || |||||||| |||| | || | ||| | | |||  ||||
Db    21983 TATACACATATGTATATACATGCACATATACACACGTGTATATATACACATACACGTATA 22042

Qy      469 CATATATACACATATATATA-----ATATACACACATATATATACACATATATACACACA 523
            ||| ||| ||| |||||||     | ||| |||||| |||||| ||||| | | | |||
Db    22043 TATACATATACACATATATACATACACATATACACATGTATATATACATACACATATACA 22102

Qy      524 TATATACACATATATACACATATATATACACATATATACACATATATACATATATACACA 583
            || || | ||| ||||| ||||| |||||  |||||| |||| ||| ||| |||||
Db    22103 CATGTATATATACATACATATATACATACATGTATATATACATACATATATACATACATG 22162

Qy      584 TATATATACATATATACACATATATACATATATACACATATATACATATATACACACATA 643
            |||||||||||| ||| | | ||| |  |||||| ||||| |||||||||||||  |||
Db    22163 TATATATACATACATATATACATACATGTATATATACATACATACATATATACACGTATA 22222

Qy      644 TATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATA 703
            |||||| | | ||||||||||||||| ||| | | |||||| || | || ||||  || |
Db    22223 TATACATACATATATACACACACATATATACATACATATATGCATACATGTATATATACA 22282

Qy      704 TATATGTATATATATA 719
            || |  |||||| |||
Db    22283 TACAGATATATACATA 22298
```

RESULT 21
AAK73165
ID AAK73165 standard; DNA; 20420 BP.
XX
AC AAK73165;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:27977.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001354.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.

PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS Disclosure; SEQ ID NO 27977; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For

CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention
XX
SQ Sequence 20420 BP; 4834 A; 5403 C; 5630 G; 4553 T; 0 U; 0 Other;

Query Match 6.6%; Score 270.4; DB 4; Length 20420;
Best Local Similarity 61.4%; Pred. No. 9e-37;
Matches 467; Conservative 0; Mismatches 291; Indels 2; Gaps 2;

```
Qy       144 GGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAGCCCAGGCAACAAGAGCGAAATGTC 203
            | | | | | || |  || ||  ||   |||| |||||       |         |
Db      4055 GCTGGAGTGCAGTAGCATGATCACAGCTCACTGCAGCCATATATATGTATATATATATAT 4114

Qy       204 ATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACAC 263
            || | | |||| | | |       | |||||||  | | |||||||||| | | | | |
Db      4115 ATATGAAAAAATATATATATATATATATATATGAAAAAATATATATAAATATATATATGA 4174

Qy       264 ACATATATATATACACATATATATACACATATATACATATATACACAT-ATATACACATA 322
            | |||||||||||| | ||||||||  | ||||||| | |||||  | || |||||  | | |
Db      4175 AAATATATATATAAATATATATATGAAAATATATATAAATATATATATAAATATATATAAA 4234

Qy       323 TATATACACATACATATGTACACATATATATACACATATGTATACACATATATACACATA 382
            ||||||  | |||  | ||  || |  ||  | |||  | |||  ||||  | |||  |||  | |||
Db      4235 TATATATATATATAAATATATATATGAAAATATATATAAATATATATATAAATATATATA 4294

Qy       383 TATACACACATATATACACATATATACACACATATATACACATATATACACATATATACA 442
            |  | | |  ||||||  | |||||||    | |||||||  | |  |||||  | ||||  | |
Db      4295 TGAAAATATATATATGAAAATATATATGAAAATATATATATAAATATATATATATGAAAA 4354

Qy       443 CATATACACATATATACACATATATACATATATACACATATATATAATATACACACATAT 502
              |||||  | | |||||  | ||  | | |||||||  | ||||||||  ||||  | |  ||||
Db      4355 TATATATATAAATATATATATGAAAATATATATATAAATATATAT-ATATGAAAATATAT 4413

Qy       503 ATATACACATATATACACACATATATACACATATATACACATATATATACACATATATAC 562
            |||||  | |||||||  |    | |||||  | |||  |||  | ||||||   | |  |||||||
Db      4414 ATATAAATATATATATATGAAAATATATATATAAATATATATATATGAAAATATATATAT 4473

Qy       563 ACATATATACATATATACACATATATATACATATATACACATATATACATATATACACAT 622
            | |||||||  ||||   | |  |||||||||  |||||||  | ||   | |  |||||||  | ||
Db      4474 AAATATATATATATGAAAATATATATATAAATATATATATATGAAAATATATATATAAAT 4533

Qy       623 ATATACATATATACACACATATATACACATACATATACACACACATAGATATACATATAT 682
            |||||  ||||   | | |  |||||||  | |||  |||||    | |  |||  |||||  ||||||
Db      4534 ATATATATATGAAAATATATATATAAATATATATATATGAAAATATATATATAAATATAT 4593

Qy       683 ATACACATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTA 742
            |||   | |||||||  ||  ||||||  |||  | ||||       | |  |||      |
Db      4594 ATATGAAAATATATATATAAATATATATATGAAAATATATATATAAATATATATATATGA 4653

Qy       743 GCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAA 802
              |   |   ||  | |    ||| | ||||| | |  |  | |  |  | |      | |
Db      4654 AAATATATATATAAATATATATATATGAAAATATATATAAATATATATATGAAAATATAT 4713

Qy       803 GGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTAT 862
              | ||  |     |||    | |||     | |||||  ||| ||   ||   | |   |
Db      4714 ATATATATATATATGAAAATATATATATATGAAAATATATATATAAATATATATATGAAA 4773

Qy       863 CTGTAAGGTAGTTAAGAAGGTAACACTATATATATATATA 902
              |   |  ||   |   |  |||   |||||| | |||||
Db      4774 ATAATATATATAAATATATATAAATATATATAAAAATATA 4813
```

RESULT 22

ABK69933
ID ABK69933 standard; DNA; 20420 BP.
XX
AC ABK69933;
XX
DT 15-JUL-2002 (first entry)
XX
DE Human secreted protein gene 68 genomic DNA fragment #21.
XX
KW Human; ds; secreted protein; gene therapy; immunosuppressive;
KW antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;
KW vasotropic; cerebroprotective; nootropic; neuroprotective; antibacterial;
KW virucide; fungicide; opthalmalogical; autoimmune disease; neoplasm;
KW rheumatoid arthritis; hyperproliferative disorder; cardiac arrest;
KW cardiovascular disorder; cerebrovascular disorder; cerebral ischaemia;
KW angiogenesis; nervous system disorder; Alzheimer's disease; infection;
KW ocular disorder; corneal infection; wound healing; skin aging;
KW epithelial cell proliferation; food additive.
XX
OS Homo sapiens.
XX
PN WO200226931-A2.
XX
PD 04-APR-2002.
XX
PF 24-SEP-2001; 2001WO-US029871.
XX
PR 25-SEP-2000; 2000US-0234925P.
PR 12-JAN-2001; 2001WO-US000911.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Komatsoulis G, Duan DR, Rosen CA, Moore PA, Shi Y;
PI Lafleur DW, Olsen H, Brewer LA, Florence KA, Young PE, Soppet DR;
PI Endress GA, Mucenski M, Ebner R;
XX
DR WPI; 2002-362489/39.
XX
PT Novel 71 isolated secreted polypeptides and polynucleotides encoding the
PT polypeptides, useful for treating Huntington's disease, sepsis,
PT meningitis, thrombocytopenia, hemolytic anemia, rheumatoid arthritis,
PT asthma.
XX
PS Example 2; Page 1440-1445; 1478pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (or its
CC fragment, homologue complement or allelic variant) encoding a human
CC secreted protein (and its fragment, domain, epitope, variant, secreted
CC form and species variant). Also included are a recombinant vector
CC comprising the nucleic acid, a recombinant host cell comprising the
CC vector, an antibody against the secreted protein, a recombinant host cell
CC that expresses the secreted protein and a method of identifying a binding
CC partner of the secreted protein. The nucleic acid and protein are used to
CC prevent, diagnose, treat or ameliorate a medical condition in e.g.
CC humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep for
CC example autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC angiogenesis, nervous system disorders e.g. Alzheimer's disease,
CC infections caused by bacteria, viruses and fungi and ocular disorders
CC e.g. corneal infection. Many other diseases and disorders are listed in
CC the specification. The polypeptides can also be used to aid wound healing
CC an epithelial cell proliferation, to prevent skin aging due to sunburn,
CC to maintain organs before transplantation, for supporting cell culture of
CC primary tissues, to regenerate tissues and in chemotaxis. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities. The present sequence
CC represents a ds DNA fragment of the gene for a novel human secreted
CC protein of the invention
XX
SQ Sequence 20420 BP; 4834 A; 5403 C; 5630 G; 4553 T; 0 U; 0 Other;

Query Match 6.6%; Score 270.4; DB 6; Length 20420;

```
  Best Local Similarity   61.4%;  Pred. No. 9e-37;
  Matches  467;  Conservative    0;  Mismatches  291;  Indels    2;  Gaps    2;

Qy        144 GGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAGCCCAGGCAACAAGAGCGAAATGTC 203
              | | | | || |  || ||  ||   |||| |||||      |         |
Db       4055 GCTGGAGTGCAGTAGCATGATCACAGCTCACTGCAGCCATATATATGTATATATATATAT 4114

Qy        204 ATCTCAGAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACAC 263
              || | |||| | | |       | |||||||  | | |||||||||| | | | | |
Db       4115 ATATGAAAAAATATATATATATATATATATATGAAAAAATATATATAAATATATATATGA 4174

Qy        264 ACATATATATATACACATATATATACACATATATACATATATACACAT-ATATACACATA 322
              | |||||||||| | ||||||||  | |||||||| | ||||| | || ||||| | | |
Db       4175 AAATATATATATAAATATATATATGAAAATATATATAAATATATATATAATATATATAAA 4234

Qy        323 TATATACACATACATATGTACACATATATATACACATATGTATACACATATATACACATA 382
              |||||| | ||| | || || | ||  | ||| | |||  |||| | ||| ||| | |||
Db       4235 TATATATATATATAAATATATATATGAAAATATATATAAATATATATATAAATATATATA 4294

Qy        383 TATACACACATATATACACATATATACACACATATATACACATATATACACATATATACA 442
              |  | | | ||||||  | |||||||    | |||||||  | | |||||  | ||||  | |
Db       4295 TGAAAATATATATATGAAAATATATATGAAAATATATATATAAATATATATATATGAAAA 4354

Qy        443 CATATACACATATATACACATATATACATATATACACATATATATAATATACACACATAT 502
              |||||  | | |||||  | ||  | | |||||||  | ||||||||  ||||   | | ||||
Db       4355 TATATATATAAATATATATATGAAAATATATATATAAATATATAT-ATATGAAAATATAT 4413

Qy        503 ATATACACATATATACACACATATATACACATATATACACATATATATACACATATATAC 562
              |||||  | |||||||  |    | |||||  | |||  ||| | ||||||   | | |||||||
Db       4414 ATATAAATATATATATATGAAAATATATATATAAATATATATATATGAAAATATATATAT 4473

Qy        563 ACATATATACATATATACACATATATATACATATATACACATATATACATATATACACAT 622
              | |||||||  ||||   | | |||||||||  |||||||  | ||   | |  |||||||  | ||
Db       4474 AAATATATATATATGAAAATATATATATAAATATATATATATGAAAATATATATATAAAT 4533

Qy        623 ATATACATATATACACACATATATACACATACATATACACACACATAGATATACATATAT 682
              |||||  ||||   | | |  |||||||  | |||  |||||    | |  |||  |||||  ||||||
Db       4534 ATATATATATGAAAATATATATATAAATATATATATATGAAAATATATATATAAATATAT 4593

Qy        683 ATACACATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTA 742
              |||    |  |||||||  ||  ||||||  |||   |  ||||        | |   |||       |
Db       4594 ATATGAAAATATATATATAAATATATATATGAAAATATATATATAAATATATATATATGA 4653

Qy        743 GCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAA 802
              |   |   ||  | |    ||| | ||||| | |   |  | |    | |       | |
Db       4654 AAATATATATATAAATATATATATATGAAAATATATATAAATATATATATGAAAATATAT 4713

Qy        803 GGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTAT 862
              | || |     |||    | |||     | ||||| ||| ||   ||   | |   |
Db       4714 ATATATATATATATGAAAATATATATATATGAAAATATATATATAAATATATATATGAAA 4773

Qy        863 CTGTAAGGTAGTTAAGAAGGTAACACTATATATATATATA 902
              |   |  ||    |   |  |||   |||||| | |||||
Db       4774 ATAATATATATAAATATATATAAATATATATAAAAATATA 4813


RESULT 23
ABD32806_4/c
Continuation (5 of 7) of ABD32806 from base 400001 (Human cancer-associated genomic DNA
WP Sequence split into 7 fragments  LOCUS ABD32806 Accession Abd32806
WP      Fragment Name      Begin     End
WP      ABD32806_0             1    110000
WP      ABD32806_1        100001    210000
WP      ABD32806_2        200001    310000
WP      ABD32806_3        300001    410000
WP      ABD32806_4        400001    510000
WP      ABD32806_5        500001    610000
WP      ABD32806_6        600001    653458

  Query Match             6.5%;  Score 268.6;  DB 13;  Length 110000;
  Best Local Similarity   69.2%;  Pred. No. 1.9e-36;
  Matches  441;  Conservative    0;  Mismatches  184;  Indels   12;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC--AAAAT 58
              ||||  |||||||||||  |||||||| |||||||||||||| | |||| |   |||||
Db      95587 AGTTTGAGACCAGCCTGGCCAACATGGTAAAACCCCATCTCTATTTAAAAAAACAAAAAT 95528

Qy         59 TAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGA 118
              ||  ||||  || || |||||  |||||  |||| ||||||  ||||||  |||||  ||||
Db      95527 TAACCAGGTGTGATGGTACACATCTGTACTCCCGGCTACTCAGGAGGCTGAGGCACGAGA 95468

Qy        119 ATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCA-AGATCATGCCATTGCACTCC 177
              ||   |||||||||  |   ||||||||||  | |||||  ||||     || || |||||
Db      95467 ATTGTTTGAACCTG--GAAAGTGGAGGTTGTAGTGAGCAGAGATCGCAACACTGTACTCC 95410

Qy        178 AGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAGGCATTTTATATATAT 237
              ||||  ||  |||         | ||||          ||||||||||    |    ||| | |
Db      95409 AGCCTGGGTGACAGAGCAAGACTGTC------TGAAAAAAAAAATCAACAAAATAAACAA 95356

Qy        238 ATATATATATATATACACACACACACATATATATATACACATATATATACACATATAT 297
              | |  ||    |   ||| | | | | | | |||||||||  |  |||||| ||| ||||
Db      95355 AGAGGTAATGACTGACATATATATATATACATATATATTTCCCATATATATACACATAT 95296

Qy        298 ACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACAC 357
              || |||||| || |||||||||| |||||| ||  |||| | | || || | |||| ||
Db      95295 ACGTATATATACGTATATACACGTATATACACGTATACGTGTATATACGTGTATACACGT 95236

Qy        358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417
              |||  |    | || | || | |||  || | |||| |||| || |||||| ||| ||||
Db      95235 GTATACACGTATATGTGTATATATACGTATATACATGTATATACGTATATATACATATAT 95176

Qy        418 ATACACATATATACACATATATACACATATACACATATATACACATATATACATATATAC 477
              | | |   ||||||  ||||||||  |||||||  ||  |||  |  ||||||  |  | |||
Db      95175 ACATATTCATATACGTATATATACGTATATACATATTCATATATGTATATATACGTGTAC 95116

Qy        478 ACATATATATAATATACACACATATATATACACATATATACACACATATATACACATATA 537
              | |||||| | ||||| ||| |||||  |||||  ||||||    | |||||  | |  |||
Db      95115 ATATATATGT-ATATATACATATATACGTACACGTATATATGTATATATACATATATACG 95057

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATACATATA 597
              |||||  |||||||  | |||||  | |  |||  ||||  ||||  |  |||||||  | | |
Db      95056 TACACGTATATATGTATATATACATATATACGTACACGTATATATGTATATATATACACA 94997

Qy        598 TACACATATATACATATATACACATATATACATATAT 634
              |||  ||  |  |   ||||||  | |||||||  ||  | |
Db      94996 CACATATGTGTGTGTATATATATATATATATTTTT 94960


RESULT 24
ADH13788/c
ID   ADH13788 standard; DNA; 191127 BP.
XX
AC   ADH13788;
XX
DT   11-MAR-2004  (first entry)
XX
DE   Prostate leucine zipper regulatory element/PrLZ gene SEQ ID NO:2.
XX
KW   prostate leucine zipper; PrLZ; prostate; cytostatic; cancer; dysplasia;
KW   ds; human; gene.
XX
OS   Chimeric.
OS   Unidentified.
OS   Homo sapiens.
XX
PN   WO2003087311-A2.
XX
PD   23-OCT-2003.
XX
PF   07-APR-2003; 2003WO-US010536.
XX
PR   05-APR-2002; 2002US-0370557P.
XX
PA   (UYEM-) UNIV EMORY.
XX
PI   Wang R,  Chung LWK;
```

XX
DR WPI; 2003-845313/78.
DR P-PSDB; ADH13786.
XX
PT New isolated prostate leucine zipper (PrLZ) nucleic acid molecule, useful
PT in diagnosing, prognosticating and/or treating cancers or high grade
PT dysplasias affecting PrLZ-positive tissues, such as those in the
PT prostate.
XX
PS Claim 23; SEQ ID NO 2; 103pp; English.
XX
CC The invention relates to a novel isolated nucleic acid molecule that
CC regulats the expression of prostate-expressed genes, specifically the
CC prostate leucine zipper (PrLZ) gene. A polypeptide of the invention has
CC cytostatic activity, and may act as a leucine zipper antagonist. The
CC agent or vector that inhibits the expression or activity of PrLZ is
CC useful for the treatment of cancer or high grade dysplasia, and for the
CC preparation of a medicament for the treatment of cancer or high grade
CC dysplasia, where the cancer or dysplasia affects prostate tissue, breast
CC tissue, renal tissue, tissue within the gastrointestinal tract, or any
CC other PrLZ-positive tissue. The methods and compositions of the present
CC invention are also useful for the diagnosis and prognosis of such
CC cancers. The present sequence represents a genomic sequence that
CC regulates the expression of PrLZ, an additional 5' flanking sequence, the
CC PrLZ gene, and 3' flanking sequence.
XX
SQ Sequence 191127 BP; 51194 A; 37977 C; 39502 G; 62454 T; 0 U; 0 Other;

Query Match 6.5%; Score 268.6; DB 10; Length 191127;
Best Local Similarity 64.4%; Pred. No. 2e-36;
Matches 467; Conservative 0; Mismatches 249; Indels 9; Gaps 4;

```
Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAATTA 60
              || || |||||| |||| ||||||||||||||||          ||||   || ||||
Db     122784 AGATCGAGACCATCCTGGCCAACATGGTGAAACCCGTCTCAACTAAAAACACAAAGATTA 122725

Qy         61 GACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAAT 120
              |   ||| |||||     | |||||| ||||| |||||   ||||| ||||||||||||
Db     122724 GCTGGGCGTGGTGGCGTGCACCTGTAGTCCCAACTACTCATGAGGCTGAGGCAGGAGAAT 122665

Qy        121 CACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGC-AAGATCATGCCATTGCACTCCAG 179
              | | |||||   | || ||  |||||||| | ||||  ||||   ||||   ||||||||
Db     122664 CTCCTGAAC--TCGGGAGGCAGAGGTTGCAGTGAGCTGAGATTGCGCCACCACACTCCAG 122607

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              || ||   |||| ||     ||      | |||| | | | |    | | | | | | |
Db     122606 CCTGGCGACAGAGAGAGA---CTCCATCTAAAAAACATATATGTGTGTGTGTGTGTGTGT 122550

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              | | | | | ||      | | |||| || | ||| | ||   | || | | || |||
Db     122549 GTGTGTGTGTGTATGTGTGTATATACATGTACACATAGATATGCGTGTATATACATGTAC 122490

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACAT 359
              | ||| | |    | || | | || || ||| | | ||  ||   | ||| || |||||||
Db     122489 ACATAGATATGCGTGTATATACATGTACACATAGATATGCGTGTATATACATGTACACAT 122430

Qy        360 ATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATAT 419
              |  ||| |   |||||||  || | | | | ||    |  | ||| ||  ||||||| ||
Db     122429 AGATATGCGTGTATATACATGTACACATAGATATGCGTGTATATACATGTACACATAGAT 122370

Qy        420 ACACATATATACACATATATACACATATACACATATATACACATATATACATATATACAC 479
              |  | | |||| |||| || ||| | |||  | | |||| |||| || ||||| |||  |
Db     122369 ATGCGTGTATATACATGTACACATAGATATGCGTGTATATACATGTACACATAGATATGC 122310

Qy        480 ATATATATA---ATATACACACATATATATACACATATATACACACATATATACACATAT 536
              | ||||||    || ||| | ||||   |  | ||| ||  ||||||| |||  | | |
Db     122309 GTGTATATACATGTACACATAGATATGCGTGTATATACATGTACACATAGATATGCGTGT 122250

Qy        537 ATACACATATATATACACATATATACACATATATACATATATACACATATATATACATAT 596
              ||| |||| || | | || | |  | ||| ||  | | ||| | || | | | || |||
Db     122249 ATATACATGTACACAGATATGTGTGTATATACATGTACACATAGATATGTGTGTATATAC 122190

Qy        597 ATACACATATATACATATATACACATATATACATATATACACACATATATACACATACAT 656
```

```
              | ||| ||||| |  ||||  |||| ||||| | ||||| | ||| | | | | ||||||
Db     122189 ACACATATATACATCTATACGCACACATATATACATATATATACACACACATATATACAT 122130

Qy        657 ATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATAT 716
              ||| ||||||| | ||||| ||| | ||||||| | | | || ||||||||| ||||||
Db     122129 ATATACACACACACATATATATACACATACACACACACACACATATATATATATATAA 122070

Qy        717 ATATG 721
              |  ||
Db     122069 ACTTG 122065
```

RESULT 25
ABD32806_4
Continuation (5 of 7) of ABD32806 from base 400001 (Human cancer-associated genomic DNA F
WP Sequence split into 7 fragments LOCUS ABD32806 Accession Abd32806

	Fragment Name	Begin	End
WP	ABD32806_0	1	110000
WP	ABD32806_1	100001	210000
WP	ABD32806_2	200001	310000
WP	ABD32806_3	300001	410000
WP	ABD32806_4	400001	510000
WP	ABD32806_5	500001	610000
WP	ABD32806_6	600001	653458

Query Match 6.5%; Score 267.6; DB 13; Length 110000;
Best Local Similarity 69.9%; Pred. No. 2.9e-36;
Matches 432; Conservative 0; Mismatches 179; Indels 7; Gaps 5;

```
Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||| |||||||||| ||||||| |||||||||||||||| ||||| || ||||||
Db      94745 AGTTCGAGACCAGCCTGGCCAACATGGCGAAACCCCATCTCTACAAAAAACACAAAAATT 94804

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              || ||||| |||||  | |  ||| || |||||| ||||| |||||| ||||| ||||||
Db      94805 AGCCAGGCGTGGTGGCATATACCTATAGTCCCAG-TACTTGGGAGGCTGAGGCGGGAGAA 94863

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              ||||||||||      || |       |   |||   | ||||||||||| || |||||||
Db      94864 TCACTTGAACCCAGGAGGTGAAAGATTGCAGGGAGCCGAGATCATGCCACTGAACTCCAG 94923

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  ||  || ||||| | |      ||||| |||||||||||      | ||||||||||
Db      94924 CCTGGGTGAC-AGAGCCAGACTCTGTCTCAAAAAAAAAAAAA---TATATATATATATAT 94979

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              ||| | | | |||       | | ||||| ||||| || | ||  || | || ||||
Db      94980 ATACACACACATATGTGTGTGTATATATATACATATATACGTGTACGTATATATGTATAT 95039

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACAT 359
              ||| ||| | |  | |||  |||||| || | |||||||| |||   ||  |||| | |
Db      95040 ATACATATATACGTGTACGTATATATGTATATATACATATATACGTGTACGTATATATGT 95099

Qy        360 ATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATAT 419
              || |||||| ||||||  |||  |||| | ||||||||  | || |||| ||  ||||||
Db      95100 ATATATACATATATATGTACACGTATATATACATATATGAATATGTATATACGTATATAT 95159

Qy        420 ACACATATATACACATATATACACATATACACATATATACACATATATACATATATACAC 479
              ||  |||| | |  ||||||   |||||  || ||||||||  |||||||  |||||| ||
Db      95160 ACGTATATGAATATGTATATATGTATATATACGTATATACATGTATATACGTATATATAC 95219

Qy        480 ATATATA-TAATATACACACATATATATACACATATATACACACATATATACACATATAT 538
              | |||||   |||||||  |||| |   | ||| |   | || | |||| || | |||
Db      95220 ACATATACGTGTATACACGTGTATACACGTATATACACGTATACGTGTATATACGTGTAT 95279

Qy        539 ACACATATATATACACATATATACACATATATACATATATACACATATATATACATATAT 598
              | || |||||||||  |||| |  | |||||  |||| | || |||||| || |||||||
Db      95280 ATACGTATATATACGTATATGTGTATATATATGGGAATATATATATGTATATATATATAT 95339

Qy        599 ACACATATATACATATAT 616
                |   ||| | | |
Db      95340 GTCAGTCATTACCTCTTT 95357
```

```
RESULT 26
AAK73166
ID   AAK73166 standard; DNA; 19965 BP.
XX
AC   AAK73166;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:27978.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001354.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR.  21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0235484P.
PR   27-SEP-2000; 2000US-0235834P.
PR   27-SEP-2000; 2000US-0235836P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   13-OCT-2000; 2000US-0239937P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241221P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   20-OCT-2000; 2000US-0241826P.
PR   01-NOV-2000; 2000US-0244617P.
PR   08-NOV-2000; 2000US-0246474P.
PR   08-NOV-2000; 2000US-0246475P.
PR   08-NOV-2000; 2000US-0246476P.
PR   08-NOV-2000; 2000US-0246477P.
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246526P.
PR   08-NOV-2000; 2000US-0246527P.
PR   08-NOV-2000; 2000US-0246528P.
PR   08-NOV-2000; 2000US-0246532P.
PR   08-NOV-2000; 2000US-0246609P.
PR   08-NOV-2000; 2000US-0246610P.
PR   08-NOV-2000; 2000US-0246611P.
PR   08-NOV-2000; 2000US-0246613P.
PR   17-NOV-2000; 2000US-0249207P.
PR   17-NOV-2000; 2000US-0249208P.
PR   17-NOV-2000; 2000US-0249209P.
PR   17-NOV-2000; 2000US-0249210P.
PR   17-NOV-2000; 2000US-0249211P.
PR   17-NOV-2000; 2000US-0249212P.
PR   17-NOV-2000; 2000US-0249213P.
PR   17-NOV-2000; 2000US-0249214P.
PR   17-NOV-2000; 2000US-0249215P.
PR   17-NOV-2000; 2000US-0249216P.
PR   17-NOV-2000; 2000US-0249217P.
PR   17-NOV-2000; 2000US-0249218P.
PR   17-NOV-2000; 2000US-0249244P.
PR   17-NOV-2000; 2000US-0249245P.
PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250160P.
```

```
PR   01-DEC-2000; 2000US-0250391P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS   Disclosure; SEQ ID NO 27978; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting the
CC   nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention
XX
SQ   Sequence 19965 BP; 4699 A; 5296 C; 5505 G; 4465 T; 0 U; 0 Other;

  Query Match             6.5%;  Score 267.4;  DB 4;  Length 19965;
  Best Local Similarity   63.9%;  Pred. No. 2.9e-36;
  Matches  453;  Conservative    0;  Mismatches  251;  Indels    5;  Gaps    3;

Qy        215 AAAAAAAGGCATTTTATATATATATATATATATATACACACACACACACATATATATA 274
              ||||||      |  ||||||||||  | |||||||||| | | | | | |   | |||||
Db       4148 AAAAAATATATATAAATATATATATGAAAATATATATATAAATATATATATGAAAATATA 4207

Qy        275 TACACATATATATACACATATATACATATATACACATATATACACATATATATACACATA 334
              || | |||||||||   | ||||| ||||  | | ||||||| | ||||||||  | |||
Db       4208 TATAAATATATATATGAAAATATATATATGAAAATATATATAAATATATATATGAAAATA 4267

Qy        335 CATATGTACACATATATATACACATATGTATACACATATATACACATATATACACACATA 394
               ||||| | | ||||||||  | ||||  |||  | |||||||  | |  ||||  | | ||
Db       4268 TATATGAAAATATATATATGAAAATATATATGAAAATATATATATAAATATATATATATG 4327

Qy        395 TATACACATATATACACACATATATACACATATATACACATATATACACATATACACATA 454
               | | | ||||||| | | ||||||  | |||||||  | |  |||||  | ||  | | |||
Db       4328 AAAATATATATATAAATATATATATGAAAATATATATATAAATATATATATGAAAATATA 4387

Qy        455 TATACACATATATACATATATACACATATATAT-AATATACACACATATATATACACATA 513
              ||||  | |||||||  ||||  | | ||||||||  ||||||  | |  ||  | |||  | |||
Db       4388 TATATAAATATATATATATGAAAATATATATATAAATATATATATATGAAAATATATATA 4447

Qy        514 TATACACACATATATACACATATATACACATATATATACACATATATACACATATATACA 573
              ||  | |  | ||||||  | |||||||  | |  | |||||||  | ||   | | |  | |||||||  |
Db       4448 TAAATATATATATATGAAAATATATATATAAATATATATATATGAAAATATATATATAAA 4507

Qy        574 TATATACACATATATATACATATATACACATATATACATATATACACATATATACATATA 633
              ||||||  |   | |||||  |||||  | |  |||||||   | |||  | |||||  | |||||
Db       4508 TATATATATGAAAATATATATATAAATATATATATGAAAATATATATATAAATATATA 4567

Qy        634 TACACACATATATACACATACATATACACACACATAGATATACATATATATACACATATA 693
              || |   | |||||  | |  | |||||  |   | |||  |||||  ||||||   | |  |||||
Db       4568 TATATGAAAATATATATAAATATATATATGAATATATATATATATATATGAAAATATATA 4627

Qy        694 TATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTA 753
              ||||  | |  ||||||  ||||  ||||||      | |  |  |||  |       |  |||  ||
Db       4628 TATATGAAAATATATATATAAATATATATATGAAAATATATATAAATATATATATGAATA 4687

Qy        754 CCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAG 813
                   |    |||  |  |||||  |  |   |  |      |  |  |||     | |   |  || | |
Db       4688 TATATATATATATATGAAAATATATATATATGAAAATATATATATAAATATATATATGAA 4747

Qy        814 GTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAG 873
                | ||   | |||  |    | |||  |||  || |  ||  | |  |||  | |||   |
Db       4748 AATAAT---ATATATAAATATATATAAATATATATAAAAATATAAATATATATAA-AAAT 4803

Qy        874 TTAAGAAGGTAACACTATATATATATATAGTGAAAGCAGTGTAAACCTA 922
                ||  ||  ||    ||||  | |||||||   |||    | ||  |  ||
Db       4804 ATATAAATATATATATATAAACATATATATATAAATATATATATATATA 4852


RESULT 27
ABK69932
ID   ABK69932 standard; DNA; 19965 BP.
XX
AC   ABK69932;
XX
DT   15-JUL-2002  (first entry)
XX
DE   Human secreted protein gene 68 genomic DNA fragment #20.
XX
KW   Human; ds; secreted protein; gene therapy; immunosuppressive;
KW   antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;
KW   vasotropic; cerebroprotective; nootropic; neuroprotective; antibacterial;
KW   virucide; fungicide; opthalmalogical; autoimmune disease; neoplasm;
KW   rheumatoid arthritis; hyperproliferative disorder; cardiac arrest;
KW   cardiovascular disorder; cerebrovascular disorder; cerebral ischaemia;
KW   angiogenesis; nervous system disorder; Alzheimer's disease; infection;
KW   ocular disorder; corneal infection; wound healing; skin aging;
KW   epithelial cell proliferation; food additive.
XX
OS   Homo sapiens.
XX
PN   WO200226931-A2.
XX
PD   04-APR-2002.
XX
PF   24-SEP-2001; 2001WO-US029871.
XX
PR   25-SEP-2000; 2000US-0234925P.
PR   12-JAN-2001; 2001WO-US000911.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Komatsoulis G,  Duan DR,  Rosen CA,  Moore PA,  Shi Y;
PI   Lafleur DW,  Olsen H,  Brewer LA,  Florence KA,  Young PE,  Soppet DR;
PI   Endress GA,  Mucenski M,  Ebner R;
XX
DR   WPI; 2002-362489/39.
XX
PT   Novel 71 isolated secreted polypeptides and polynucleotides encoding the
PT   polypeptides, useful for treating Huntington's disease, sepsis,
PT   meningitis, thrombocytopenia, hemolytic anemia, rheumatoid arthritis,
PT   asthma.
XX
PS   Example 2; Page 1434-1440; 1478pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule (or its
CC   fragment, homologue complement or allelic variant) encoding a human
CC   secreted protein (and its fragment, domain, epitope, variant, secreted
```

```
CC   form and species variant). Also included are a recombinant vector
CC   comprising the nucleic acid, a recombinant host cell comprising the
CC   vector, an antibody against the secreted protein, a recombinant host cell
CC   that expresses the secreted protein and a method of identifying a binding
CC   partner of the secreted protein. The nucleic acid and protein are used to
CC   prevent, diagnose, treat or ameliorate a medical condition in e.g.
CC   humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep for
CC   example autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
CC   disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
CC   e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC   angiogenesis, nervous system disorders e.g. Alzheimer's disease,
CC   infections caused by bacteria, viruses and fungi and ocular disorders
CC   e.g. corneal infection. Many other diseases and disorders are listed in
CC   the specification. The polypeptides can also be used to aid wound healing
CC   an epithelial cell proliferation, to prevent skin aging due to sunburn,
CC   to maintain organs before transplantation, for supporting cell culture of
CC   primary tissues, to regenerate tissues and in chemotaxis. The
CC   polypeptides can also be used as a food additive or preservative to
CC   increase or decrease storage capabilities. The present sequence
CC   represents a ds DNA fragment of the gene for a novel human secreted
CC   protein of the invention
XX
SQ   Sequence 19965 BP; 4699 A; 5296 C; 5505 G; 4465 T; 0 U; 0 Other;

  Query Match             6.5%;  Score 267.4;  DB 6;  Length 19965;
  Best Local Similarity   63.9%;  Pred. No. 2.9e-36;
  Matches  453;  Conservative    0;  Mismatches  251;  Indels    5;  Gaps    3;

Qy        215 AAAAAAAGGCATTTTATATATATATATATATATACACACACACACACATATATATA 274
              ||||||      |  ||||||||||  | |||||||||  | | | | | |    | |||||
Db       4148 AAAAAATATATATAAATATATATATGAAAATATATATATAAATATATATATGAAAATATA 4207

Qy        275 TACACATATATATACACATATATACATATATACACATATATACACATATATATACACATA 334
              || | |||||||||   | |||||  ||||  | | |||||||  | ||||||||  | |||
Db       4208 TATAAATATATATATGAAAATATATATATGAAAATATATATAAATATATATATGAAAATA 4267

Qy        335 CATATGTACACATATATATACACATATGTATACACATATATACACATATATACACACATA 394
              |||||  | |  |||||||||  | ||||  |||  | |||||||  | |  |||||  | |  ||
Db       4268 TATATGAAAATATATATATGAAAATATATATGAAAATATATATATAAATATATATATATG 4327

Qy        395 TATACACATATATACACACATATATACACATATATACACATATATACACATATACACATA 454
              | | | |||||||  | |  ||||||  | |||||||  | |  |||||  | ||  | |  |||
Db       4328 AAAATATATATATAAATATATATATGAAAATATATATATAAATATATATATGAAAATATA 4387

Qy        455 TATACACATATATACATATATACACATATATAT-AATATACACACATATATATACACATA 513
              ||||  | |||||||  ||||  | |  ||||||||  ||||||  | |  ||  | |||  | |||
Db       4388 TATATAAATATATATATATGAAAATATATATATAAATATATATATATGAAAATATATATA 4447

Qy        514 TATACACACATATATACACATATATACACATATATATACACATATATACACATATATACA 573
              ||  | |  |  ||||||  |  |||||||  |  |  |||||||  | ||  |  | |  |||||||  |
Db       4448 TAAATATATATATATGAAAATATATATATAAATATATATATATGAAAATATATATATAAA 4507

Qy        574 TATATACACATATATATACATATATACACATATATACATATATACACATATATACATATA 633
              ||||||  |    |  |||||  |||||  |  |  |||||||    |  |||  |  |||||  |  |||||
Db       4508 TATATATATGAAAATATATATATAAATATATATATATGAAAATATATATATAAATATATA 4567

Qy        634 TACACACATATATACACATACATATACACACACATAGATATACATATATATACACATATA 693
              ||  |   |  |||||  |  |  |  |||||  |    |  |||  |||||  ||||||   |  |  |||||
Db       4568 TATATGAAAATATATATAAATATATATATGAATATATATATATATATATGAAAATATATA 4627

Qy        694 TATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTA 753
              ||||  |  |  ||||||  ||||  ||||||      |  |  |  |||  |        |   |||  ||
Db       4628 TATATGAAAATATATATATAAATATATATATGAAAATATATATAAATATATATATGAATA 4687

Qy        754 CCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAG 813
                  |    |||  |  |||||  |  |   |   |     |  |  |||     |  |    |  ||  |  |
Db       4688 TATATATATATATATGAAAATATATATATATGAAAATATATATATAAATATATATATGAA 4747

Qy        814 GTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAG 873
              |  ||    | |||   |     | |||  |||  || |  ||   |  |   |||  |  |||    |
Db       4748 AATAAT---ATATATAAATATATATAAATATATATAAAAATATAAATATATATAA-AAAT 4803

Qy        874 TTAAGAAGGTAACACTATATATATATATAGTGAAAGCAGTGTAAACCTA 922
```

```
                ||  ||  ||    ||||  |  |||||||   |||    | ||  |  ||
Db       4804 ATATAAATATATATATATAAACATATATATATAAATATATATATATATA 4852


RESULT 28
ABL33237/c
ID   ABL33237 standard; DNA; 6971 BP.
XX
AC   ABL33237;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 1210.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP007537.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful for
PT   diagnosis and treatment of diseases associated with abnormal cytosine
PT   methylation.
XX
PS   Claim 1; SEQ ID NO 1210; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention
XX
SQ   Sequence 6971 BP; 1665 A; 242 C; 1831 G; 3232 T; 0 U; 1 Other;

  Query Match             6.5%;  Score 266.4;  DB 6;  Length 6971;
  Best Local Similarity   71.1%;  Pred. No. 4.1e-36;
  Matches  367;  Conservative    0;  Mismatches  146;  Indels    3;  Gaps    1;

Qy        212 AAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACATATAT 271
              | | | | |    |  | ||||||||  ||||||||| |||||| | | | | ||||||
Db       6472 ATACATATATACCTATATATATATACCTATATATATAAATACACATATATATATATAT 6413

Qy        272 ATATACACATATATATACACATATATACATATATACACATATATACACATATATATACAC 331
              |||   | |||||||| |||  |||| ||||||||||  |||| |||||||||    |
Db       6412 ATACCAATATATATATATACACGTATATATATATACACACTTATATACATATATACCTAT 6353

Qy        332 ATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACACAC 391
              ||| |||  |||||||||||||| | |||  |||| | ||| |||||||| |||| | |
Db       6352 ATATATAAATACACATATATATATATATACCTATATATATAAATACACATTTATATATAT 6293
```

```
Qy        392 ATATATACACATATATACACACATATATACACATATATACACATATATACACATATACAC 451
              | ||| |||  |||||| | | |    || | | ||| |||  |||||| | ||||||
Db       6292 AAATACACATTTATATATATATACGCCTATATAAATACACATTTATATATATATATACCT 6233

Qy        452 ATATATACACATATATACATATATACACATATATATAATATACACACATATATATACACA 511
              ||||||| || |||||| ||| ||||  |||||||||| | |||   ||||||||| | |
Db       6232 ATATATATACCTATATATATAAATACCTATATATATAAAACACATTTATATATATATATA 6173

Qy        512 TATATACACACATATATACACATATATACACATATATATACACATATATACACATATATA 571
              |||| | | ||| | |   ||||||| || | |||||| | |  |||| | | | |||
Db       6172 CCTATATATAAATACACATTTATATATATACCTCTATATATATACCTATATATAAAAATA 6113

Qy        572 CATATATACACATATATATACATATATACACATATATACATATATACACATATATACATA 631
              || || || | |||    || ||||| | | | | ||  |||||||| | |||||| |||
Db       6112 CACATTTATATATACGCCTATATATAAAAATACACATTTATATATACGCCTATATATATA 6053

Qy        632 TATACACACATATATACACATACATATACACACACATAGATATACATATATAT---ACAC 688
              |||||||  |||||| | ||||   || | | | | | ||  |||||||   ||||
Db       6052 AATACACATTTATATATATATACGCCTATATATATAAATACCTATATATATATAAAACAC 5993

Qy        689 ATATATATACGTATATATATGTATATATATATGCTC 724
              || ||||||  ||||||||  |||||||| || | |
Db       5992 ATTTATATATATATATATACCTATATATAAATACAC 5957
```

RESULT 29
ADM01858/c
ID ADM01858 standard; cDNA; 2522 BP.
XX
AC ADM01858;
XX
DT 20-MAY-2004 (first entry)
XX
DE Human cDNA of the invention SEQ ID NO:543.
XX
KW ss; gene; human; gene therapy; diagnostic marker; pharmaceutical.
XX
OS Homo sapiens.
XX
PN EP1347046-A1.
XX
PD 24-SEP-2003.
XX
PF 12-APR-2002; 2002EP-00008400.
XX
PR 22-MAR-2002; 2002JP-00137785.
XX
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX
DR WPI; 2003-723558/69.
DR P-PSDB; ADM04301.
XX
PT New polynucleotides and polypeptides are useful in gene therapy, for
PT developing a diagnostic marker or medicines for regulating their
PT expression and activity, or as a target of gene therapy.
XX
PS Claim 1; SEQ ID NO 543; 305pp; English.
XX
CC The invention relates to a novel human polynucleotide and the encoded
CC polypeptide. A polynucleotide of the invention may have a use in gene
CC therapy. An oligonucleotide of the invention ADM06202-ADM06773 is useful
CC as a primer for synthesizing the polynucleotide or as a probe for
CC detecting the polynucleotide. The polynucleotides ADM01316-ADM03758 are
CC useful in gene therapy, for developing a diagnostic marker or medicines
CC for regulating their expression and activity, or as a target of gene
CC therapy. The proteins ADM03759-ADM06201 encoded by the polynucleotides
CC are useful as pharmaceutical agents. The present sequence represents a
CC cDNA sequence of the invention.
XX

SQ Sequence 2522 BP; 789 A; 399 C; 547 G; 787 T; 0 U; 0 Other;

Query Match 6.4%; Score 263.6; DB 11; Length 2522;
Best Local Similarity 68.9%; Pred. No. 1.2e-35;
Matches 376; Conservative 0; Mismatches 169; Indels 1; Gaps 1;

```
Qy        236 ATATATATATATATATACACACACACACATATATATATACACATATATATACACATAT 295
              ||  |  | | |||||| | | ||||||| |  |||||||    |||||||||||||  |
Db       2217 ATTCAACTGTGTATATATATATACACACATACGTATATATGTGTATATATATACACACGT 2158

Qy        296 ATACATATATACACATATATACACATATATATACACATACATATGTACACATATATATAC 355
              ||| ||| | | | || |||||| | | |   | | ||||| || ||   ||| ||| ||
Db       2157 ATATATACACATATATGTATACATACACACGCATATATACACATATATGTATACATACAC 2098

Qy        356 ACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACAT 415
              || ||| ||||||||||||||  | | ||| |||| ||||||||||||||||||  | ||||
Db       2097 ACGTATATATACACATATATGTATACATACACACGCATATATACACATATATGTATACAT 2038

Qy        416 ATATACACATATATACACATATATACACATATACACATATATACACATATATACATATAT 475
              | | ||  |||||||||||||||| || | |||||||   ||| | | | ||| ||  ||
Db       2037 ACACACGTATATATACACATATATGCATACATACACACGCATATATACACATATATGCAT 1978

Qy        476 ACACATATATATAATATACACACATATATATACACATATATACACACATATATACACATA 535
              ||| | |||  | ||||| |||||||||| || |||||  |||| | | | ||| ||| |||
Db       1977 ACATACATACGT-ATATATACACATATATGTATACATACATACGTATACATACACATATA 1919

Qy        536 TATACACATATATATACACATATATACACATATATACATATATACACATATATATACATA 595
              | || ||||| ||   || ||||| ||| |||| || | ||| | || ||||||||||| |
Db       1918 TGTATACATACATGCGCATATATACACATATATGTATACATACATACGTATATATACACA 1859

Qy        596 TATACACATATATACATATATACACATATATACATATATACACACATATATACACATACA 655
              ||||   ||| |||||||  || ||||| | | || |||||| |||||  || | |||||
Db       1858 TATATGTATACATACATACGTATACATACACATATGTATACATACATACGTATATATACA 1799

Qy        656 TATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATA 715
              ||| |   | | | | || | |||||||||||||||||| ||   ||| ||  |||||||
Db       1798 CATATATGTATACATACATGCGTATATATACACATATATGTATACATACATGCGTATATA 1739

Qy        716 TATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGA 775
              || |       |  | ||||      | ||  |   ||    || | ||| |   | | |
Db       1738 TACACATATATGCATACATACATATGAACACATATATATGTGTGTGTATATATACATATA 1679

Qy        776 GAGTTT 781
                |||
Db       1678 AGATTT 1673
```

RESULT 30
AAD54480/c
ID AAD54480 standard; DNA; 117962 BP.
XX
AC AAD54480;
XX
DT 26-JUN-2003 (first entry)
XX
DE Human CIP DNA #1.
XX
KW Human; p53 pathway; chloride cotransporter interactor protein; CIP;
KW angiogenic disorder; cell proliferation disorder; apoptotic disorder;
KW breast cancer; gene therapy; ds.
XX
OS Homo sapiens.
XX
PN WO200299055-A2.
XX
PD 12-DEC-2002.
XX
PF 03-JUN-2002; 2002WO-US017473.
XX
PR 05-JUN-2001; 2001US-0296076P.
PR 10-OCT-2001; 2001US-0328605P.
PR 15-FEB-2002; 2002US-0357253P.
XX

```
PA   (EXEL-) EXELIXIS INC.
XX
PI   Friedman L,  Plowman GD,  Belvin M,  Francis-Lang H,  Li D,  Funke RP;
XX
DR   WPI; 2003-175140/17.
XX
PT   Identifying p53 pathway modulators for treating or diagnosing disorders
PT   with defective p53 function e.g. breast cancer, by providing an assay
PT   system having a purified cotransporter interactor protein (CIP)
PT   polypeptide or nucleic acid.
XX
PS   Disclosure; Page 38-101; 123pp; English.
XX
CC   The invention relates to a method of Identifying p53 pathway modulating
CC   agent. The method involves contacting a test agent with an assay system
CC   comprising a purified cation Cl- cotransporter interactor protein (CIP)
CC   polypeptide or polynucleotide, or their functionally active fragment or
CC   derivative. The method is useful for identifying modulators of the p53
CC   pathway particularly for identifying agents for treating disorders (e.g.
CC   breast cancer) associated with defective p53 function. Modulators of the
CC   invention are useful as targets for novel therapeutics. CIP sequences are
CC   useful as modifiers of the p53 pathway, and as therapeutic targets for
CC   disorders associated with defective p53 function e.g. angiogenic,
CC   apoptotic or cell proliferation disorders. The invention is useful in
CC   gene therapy. The present sequence is human CIP DNA
XX
SQ   Sequence 117962 BP; 27840 A; 32096 C; 30624 G; 27402 T; 0 U; 0 Other;

  Query Match             6.4%;  Score 263;  DB 8;  Length 117962;
  Best Local Similarity   71.6%;  Pred. No. 1.7e-35;
  Matches  373;  Conservative    0;  Mismatches  145;  Indels    3;  Gaps    2;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||  ||||||||||||  ||||||||||||||||||||||||||||||||| ||||||
Db     107030 AGTTTGAGACCAGCCTGGCCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAAATT 106971

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              || ||||||||||  ||| |||||||||||||||||| | |||||| || | |||||||
Db     106970 AGCCAGGCATGGTGGCACATGCCTGTAATCCCAGCTACCTGGGAGGCTGAAGTAGGAGAA 106911

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              || || ||||| |  || ||  |||| ||| | |||| |         | | |   ||
Db     106910 TCTCTGGAACCAG--GGAGGCAGAGGCTGCAGTGAGCCAAGATCATTCCTGGGTGTCAGA 106853

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
                || |  ||      |      |  | || ||||||||||   |    | | | | |
Db     106852 GAGAGACTCCATCCCTGCCCGTCCCCCGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 106793

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              | | | ||||||| | | | | | ||||||||||| | ||||| | ||||| | |||
Db     106792 AAAAAAATATATATATATATATATATATATATATATATATATACACACACACACATAT 106733

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACAT 359
              ||||| | | ||||| ||||||||||||||| ||| |||  || | ||| |||||  | ||
Db     106732 ATATACATATATATACACACATATATATACATATATATACATATATATACATATATATAT 106673

Qy        360 ATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATAT 419
              | ||||| | ||| ||| | ||||| ||| | ||||| ||| |||||||| | ||||||
Db     106672 ACGTATATATATACATATATATATACACATATATATACACATATATATACATATATATAT 106613

Qy        420 ACACATATATACACATATATACACATATACACATATATACACATATATACATATATACAC 479
              |||||||||| ||| ||||| | | ||| | ||||| ||| ||||||||||||||| |
Db     106612 ACACATATATATACACATATATATACATATATATATACACATATATATACATATATATAT 106553

Qy        480 ATATATATAATATACACACATATATATACACATATATACAC 520
              || |||||  |||  | |       || |    || || |||
Db     106552 ATGTATATTTTATTGAGATCACATTAAATGTGTACATTCAC 106512


RESULT 31
ABK83460
ID   ABK83460 standard; cDNA; 165199 BP.
XX
AC   ABK83460;
XX
DT   14-AUG-2002  (first entry)
XX
DE   Human cDNA differentially expressed in granulocytic cells #31.
XX
KW   Human; ss; granulocytic cell; DNA chip; bacterial infection;
KW   viral infection; parasitic infection; protozoal infection;
KW   fungal infection; sterile inflammatory disease; psoriasis;
KW   rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KW   cardiac reperfusion injury; renal reperfusion injury; ARDS;
KW   adult respiratory distress syndrome; inflammatory bowel disease;
KW   Crohn's disease; ulcerative colitis; periodontal disease;
KW   granulocyte activation; chronic inflammation; allergy.
XX
OS   Homo sapiens.
XX
PN   WO200228999-A2.
XX
PD   11-APR-2002.
XX
PF   03-OCT-2001; 2001WO-US030821.
XX
PR   03-OCT-2000; 2000US-0237189P.
XX
PA   (GENE-) GENE LOGIC INC.
XX
PI   Beazer-Barclay Y,  Weissman SM,  Yamaga S,  Vockley J;
XX
DR   WPI; 2002-435328/46.
XX
PT   Detecting granulocyte activation by detecting differential expression of
PT   genes associated with granulocyte activation, which serves as diagnostic
PT   markers that is useful for monitoring disease states and drug toxicity.
XX
PS   Claim 1; SEQ ID NO 31; 114pp; English.
XX
CC   The invention relates to detecting (M1) granulocyte (GC) activation
CC   (GCA), by detecting the level of expression of gene(s) (Gs) identified by
CC   DNA chip analysis as given in the specification, and comparing the
CC   expression level to an expression level in an unactivated GC, where
CC   differential expression of Gs is indicative of GCA. Also included are
CC   modulating (M2) GA by contacting GC with an agent that alters the
CC   expression of at least one gene in Gs; (2) screening (M3) for an agent
CC   capable of modulating GCA or an inflammation (especially chronic) in a
CC   tissue, an allergic response in a subject, exposure of a subject to a
CC   pathogen or sterile inflammatory disease using the gene expression
CC   profile; (3) detecting (M4) an inflammation (especially chronic) in a
CC   tissue, an allergic response in a subject, exposure of a subject to a
CC   pathogen or sterile inflammatory disease, by detecting the level of
CC   expression in a sample of the tissue of gene(s) from Gs, where the level
CC   of expression of the gene is indicative of inflammation; (4) treating
CC   (M5) an inflammation (especially chronic) or in a tissue, an allergic
CC   response in a subject, exposure of a subject to a pathogen or sterile
CC   inflammatory disease, by contacting a tissue having inflammation with an
CC   agent that modulates the expression of gene(s) from Gs in the tissue. M1
CC   is useful for detecting GCA; M2 is useful for modulating GA; M3 is useful
CC   for screening an agent capable of modulating GCA preferably in an
CC   inflammation in a tissue; M4 is useful for detecting an inflammation
CC   (especially chronic) in a tissue, an allergic response in a subject,
CC   exposure of a subject to a pathogen or sterile inflammatory disease (e.g.
CC   psoriasis, rheumatoid arthritis, glomerulonephritis, asthma, thrombosis,
CC   cardiac reperfusion injury, renal reperfusion injury, ARDS, adult
CC   respiratory distress syndrome, inflammatory bowel disease, Crohn's
CC   disease, ulcerative colitis, periodontal disease; also bacterial
CC   infection, viral infection, parasitic infection, protozoal infection,
CC   fungal infection and M5 is useful for treating one of the above
CC   conditions. The present sequence represents a gene differentially
CC   expressed in granulocytes. Note: The sequence data for this patent did
CC   not form part of the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
```

```
SQ   Sequence 165199 BP; 48510 A; 33223 C; 34406 G; 49060 T; 0 U; 0 Other;

  Query Match             6.4%;  Score 263;  DB 6;  Length 165199;
  Best Local Similarity   69.3%;  Pred. No. 1.8e-35;
  Matches  472;  Conservative    0;  Mismatches  200;  Indels    9;  Gaps    8;

Qy         46 AAAAATACAAAATTAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGG 105
              | ||| | | |   |   | |||||||| | ||| |||||||||||||||||   |||||
Db     138669 AGAAAGAAAGATGGATGGATGCATGGTGGTGCACACCTGTAATCCCAGCTACACAGGAGG 138728

Qy        106 CCGAGGCAGGAGAATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGA-GCAAGATCAT 164
              | | ||||  | ||||||||||||||  || ||| ||| | ||   ||  ||||||  |
Db     138729 CTGGGGCATTAAAATCACTTGAACCTG--GGAGGTAGAGTTAGCAATGACACAAGATTGT 138786

Qy        165 GCCATTGCACTCCAGCCCAGGCAACAAGAGCGAAATGTC-ATCTCAGAAAAAAAAAAAGG 223
              ||   ||||||||||||  ||   ||  |    | ||||  |||| | |   |   |
Db     138787 GCTGCTGCACTCCAGCCTGGGTGGCAGAATGAGACTGTCTCTCTCTGTATCTAGCTATCT 138846

Qy        224 CATTTTATATATATATATATATATATACACACACA-CACACATATATATATACACATA 282
                 | |||||||||| |||||| | || | | | || | ||||  ||||| | |||
Db     138847 ATCCATCCATATATATATCCATATATGTCCATATATATCATATATATCCATATATAGATA 138906

Qy        283 TATATACACATATATACATATATACACATATATACACATATATATACACATACATATGT- 341
              ||||||   |||||||  |||||| | ||  ||| | |||| |||| | ||  |||| |
Db     138907 TATATATCTATATATATCTATATATA-ATCCATATATATATCTATATATATCTATATATA 138965

Qy        342 ACACATATATATACACATATGTATACACATATATACACATATATACACACATATATACAC 401
              |  ||||||||||  ||||| ||| || ||||| | | |||||| || | ||||  | |
Db     138966 ATCCATATATATATCCATATATATCCATATATAGATATATATATCCATATATATCCATAT 139025

Qy        402 ATATATACACACATATATACACATATATACACATATATACA-CATATACACATATATACA 460
              ||| ||| | | ||  ||| | |||| || | || |||| |   ||  | | ||||||  ||
Db     139026 ATAGATATATATATCCATATATATATCTATATATCTATATATAATCAATATATATATCCA 139085

Qy        461 CATATATACATATATACACATATATA-TAATATACACACATATATATACACATATATACA 519
              |||||  |  ||||||  |||||||| | ||||| |  ||||||||||  |||||||| |
Db     139086 TATATAGATGTATATATCCATATATAGTTATATATATCCATATATATATCCATATATAGA 139145

Qy        520 CACATATATACACATATATACACATATATATACACATATATACACATATATACATATATA 579
              | |||||| || |||||||  ||||||||  | | ||||||  || |||||||   |||||
Db     139146 TATATATATCCATATATATATCCATATATAGATATATATATCCATATATATATCCATATA 139205

Qy        580 CACATATATATACATATATACACATATATACATATATACACATATATACATATATACACA 639
              | |||||||||   |||||  | ||  |||  ||| |||  | ||||  | |||||||  ||
Db     139206 TAGATATATATATCCATATATATATCCATATATAGATATATATATCCATATATATATCCA 139265

Qy        640 CATATATACACATACATATACACACACATAGATATACATATATATACACATATATATACG 699
              |||||  | | |||   |||  | |  |  | |||||   |||||||  | ||||||||  |
Db     139266 TATATAGATATATATCCATATATATATCCATATATATCCATATATAGATATATATATCCA 139325

Qy        700 TATATATATGTATATATATAT 720
              |||||||||  |||||||  ||
Db     139326 TATATATATCCATATATAGAT 139346


RESULT 32
ABK83497/c
ID   ABK83497 standard; cDNA; 227968 BP.
XX
AC   ABK83497;
XX
DT   14-AUG-2002  (first entry)
XX
DE   Human cDNA differentially expressed in granulocytic cells #68.
XX
KW   Human; ss; granulocytic cell; DNA chip; bacterial infection;
KW   viral infection; parasitic infection; protozoal infection;
KW   fungal infection; sterile inflammatory disease; psoriasis;
KW   rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KW   cardiac reperfusion injury; renal reperfusion injury; ARDS;
KW   adult respiratory distress syndrome; inflammatory bowel disease;
KW   Crohn's disease; ulcerative colitis; periodontal disease;
KW   granulocyte activation; chronic inflammation; allergy.
XX
OS   Homo sapiens.
XX
PN   WO200228999-A2.
XX
PD   11-APR-2002.
XX
PF   03-OCT-2001; 2001WO-US030821.
XX
PR   03-OCT-2000; 2000US-0237189P.
XX
PA   (GENE-) GENE LOGIC INC.
XX
PI   Beazer-Barclay Y,  Weissman SM,  Yamaga S,  Vockley J;
XX
DR   WPI; 2002-435328/46.
XX
PT   Detecting granulocyte activation by detecting differential expression of
PT   genes associated with granulocyte activation, which serves as diagnostic
PT   markers that is useful for monitoring disease states and drug toxicity.
XX
PS   Claim 1; SEQ ID NO 68; 114pp; English.
XX
CC   The invention relates to detecting (M1) granulocyte (GC) activation
CC   (GCA), by detecting the level of expression of gene(s) (Gs) identified by
CC   DNA chip analysis as given in the specification, and comparing the
CC   expression level to an expression level in an unactivated GC, where
CC   differential expression of Gs is indicative of GCA. Also included are
CC   modulating (M2) GA by contacting GC with an agent that alters the
CC   expression of at least one gene in Gs; (2) screening (M3) for an agent
CC   capable of modulating GCA or an inflammation (especially chronic) in a
CC   tissue, an allergic response in a subject, exposure of a subject to a
CC   pathogen or sterile inflammatory disease using the gene expression
CC   profile; (3) detecting (M4) an inflammation (especially chronic) in a
CC   tissue, an allergic response in a subject, exposure of a subject to a
CC   pathogen or sterile inflammatory disease, by detecting the level of
CC   expression in a sample of the tissue of gene(s) from Gs, where the level
CC   of expression of the gene is indicative of inflammation; (4) treating
CC   (M5) an inflammation (especially chronic) or in a tissue, an allergic
CC   response in a subject, exposure of a subject to a pathogen or sterile
CC   inflammatory disease, by contacting a tissue having inflammation with an
CC   agent that modulates the expression of gene(s) from Gs in the tissue. M1
CC   is useful for detecting GCA; M2 is useful for modulating GA; M3 is useful
CC   for screening an agent capable of modulating GCA preferably in an
CC   inflammation in a tissue; M4 is useful for detecting an inflammation
CC   (especially chronic) in a tissue, an allergic response in a subject,
CC   exposure of a subject to a pathogen or sterile inflammatory disease (e.g.
CC   psoriasis, rheumatoid arthritis, glomerulonephritis, asthma, thrombosis,
CC   cardiac reperfusion injury, renal reperfusion injury, ARDS, adult
CC   respiratory distress syndrome, inflammatory bowel disease, Crohn's
CC   disease, ulcerative colitis, periodontal disease; also bacterial
CC   infection, viral infection, parasitic infection, protozoal infection,
CC   fungal infection and M5 is useful for treating one of the above
CC   conditions. The present sequence represents a gene differentially
CC   expressed in granulocytes. Note: The sequence data for this patent did
CC   not form part of the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 227968 BP; 53185 A; 60440 C; 61985 G; 52358 T; 0 U; 0 Other;

  Query Match             6.4%;  Score 263;  DB 6;  Length 227968;
  Best Local Similarity   71.6%;  Pred. No. 1.8e-35;
  Matches  373;  Conservative    0;  Mismatches  145;  Indels    3;  Gaps    2;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||  ||||||||||  |||||||||||||||||||||||||||||||||| ||||||
Db       5103 AGTTTGAGACCAGCCTGGCCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAAATT 5044

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              || ||||||||||  ||| |||||||||||||||| | |||||| || | |||||||
Db       5043 AGCCAGGCATGGTGGCACATGCCTGTAATCCCAGCTACCTGGGAGGCTGAAGTAGGAGAA 4984
```

```
Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              || || ||||| |  || ||  |||| ||| | |||| |        | | |   ||
Db       4983 TCTCTGGAACCAG--GGAGGCAGAGGCTGCAGTGAGCCAAGATCATTCCTGGGTGTCAGA 4926

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
                | |  ||     |       |  | || ||||||||||    |    | | | | |
Db       4925 GAGAGACTCCATCCCTGCCCGTCCCCCGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4866

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              | | | ||||||| | | | | | ||||||||||| | ||||| | ||||| | |||
Db       4865 AAAAAAATATATATATATATATATATATATATATATATATATATACACACACACATAT 4806

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACAT 359
              ||||| | | |||||| ||||||||||||||| ||| |||  || | ||| |||||  | ||
Db       4805 ATATACATATATATACACACATATATATACATATATATACATATATATACATATATATAT 4746

Qy        360 ATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATAT 419
              | ||||| | ||| ||| | ||||| ||| | ||||| ||| |||||||||  | ||||||
Db       4745 ACGTATATATATACATATATATATACACATATATATACACATATATATACATATATATAT 4686

Qy        420 ACACATATATACACATATATACACATATACACATATATACACATATATACATATATACAC 479
              |||||||||||| ||| ||||| | | ||| | ||||| ||| |||||||||||||| |
Db       4685 ACACATATATATACACATATATATACATATATATATACACATATATATACATATATATAT 4626

Qy        480 ATATATATAATATACACACATATATATACACATATATACAC 520
              || |||||  |||  | |       || |    || || |||
Db       4625 ATGTATATTTTATTGAGATCACATTAAATGTGTACATTCAC 4585


RESULT 33
ADQ18538/c
ID   ADQ18538 standard; DNA; 227968 BP.
XX
AC   ADQ18538;
XX
DT   26-AUG-2004  (first entry)
XX
DE   Human soft tissue sarcoma-upregulated DNA - SEQ ID 1357.
XX
KW   soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO2004048938-A2.
XX
PD   10-JUN-2004.
XX
PF   26-NOV-2003; 2003WO-US038193.
XX
PR   26-NOV-2002; 2002US-0429739P.
XX
PA   (PROT-) PROTEIN DESIGN LABS INC.
XX
PI   Aziz N,  Ginsburg WM,  Zlotnik A;
XX
DR   WPI; 2004-441208/41.
XX
PT   Early detection of soft tissue sarcoma comprises determining expression
PT   of a gene in a first soft tissue sample and a normal soft tissue sample
PT   and comparing the gene expression, also useful in treating soft tissue
PT   sarcoma.
XX
PS   Example 2; SEQ ID NO 1357; 210pp; English.
XX
CC   The invention relates to a novel method for detecting soft tissue sarcoma
CC   which comprises obtaining a first soft tissue sample from an individual
CC   and a normal soft tissue sample from the same or different individual,
CC   determining the expression of a gene in both samples and comparing the
CC   expression of the gene in both soft tissue samples, where a higher level
CC   of protein expression in the first soft tissue sample indicates the
CC   presence of soft tissue sarcoma. The method of the invention has
CC   cytostatic applications and may be useful for detecting soft tissue
CC   sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC   acid sequences may be useful in diagnostic and screening applications.
CC   The current sequence is that of a human soft tissue sarcoma-upregulated
CC   DNA of the invention. The current sequence is not shown within the
CC   specification per se but was submitted in CD format by the inventor.
XX
SQ   Sequence 227968 BP; 53185 A; 60440 C; 61985 G; 52358 T; 0 U; 0 Other;

  Query Match             6.4%;  Score 263;  DB 12;  Length 227968;
  Best Local Similarity   71.6%;  Pred. No. 1.8e-35;
  Matches  373;  Conservative    0;  Mismatches  145;  Indels    3;  Gaps    2;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||  ||||||||||  |||||||||||||||||||||||||||||||||| ||||||
Db       5103 AGTTTGAGACCAGCCTGGCCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAAATT 5044

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              || ||||||||||  ||| |||||||||||||||||| | |||||| || | |||||||
Db       5043 AGCCAGGCATGGTGGCACATGCCTGTAATCCCAGCTACCTGGGAGGCTGAAGTAGGAGAA 4984

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              || || |||||  |  || ||  |||| ||| | |||| |        | | |   ||
Db       4983 TCTCTGGAACCAG--GGAGGCAGAGGCTGCAGTGAGCCAAGATCATTCCTGGGTGTCAGA 4926

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
                | |  ||     |       |  | || ||||||||||    |    | | | | |
Db       4925 GAGAGACTCCATCCCTGCCCGTCCCCCGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4866

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              | | | ||||||| | | | | | ||||||||||| | ||||| | ||||| | |||
Db       4865 AAAAAAATATATATATATATATATATATATATATATATATATATACACACACACATAT 4806

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACAT 359
              ||||| | | ||||| |||||||||||||| ||| |||  || | ||| |||||  | ||
Db       4805 ATATACATATATATACACACATATATATACATATATATACATATATATACATATATATAT 4746

Qy        360 ATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATAT 419
              | ||||| | ||| ||| | ||||| ||| | ||||| ||| |||||||||  | ||||||
Db       4745 ACGTATATATATACATATATATATACACATATATATACACATATATATACATATATATAT 4686

Qy        420 ACACATATATACACATATATACACATATACACATATATACACATATATACATATATACAC 479
              |||||||||||| ||| ||||| | | ||| | ||||| ||| |||||||||||||| |
Db       4685 ACACATATATATACACATATATATACATATATATATACACATATATATACATATATATAT 4626

Qy        480 ATATATATAATATACACACATATATATACACATATATACAC 520
              || |||||  |||  | |       || |    || || |||
Db       4625 ATGTATATTTTATTGAGATCACATTAAATGTGTACATTCAC 4585


RESULT 34
ABK39930/c
ID   ABK39930 standard; DNA; 12069 BP.
XX
AC   ABK39930;
XX
DT   21-MAY-2002  (first entry)
XX
DE   Human chemically pretreated gene sequence #6 strand 1.
XX
KW   Human; ds; bisulphite treatment; CpG; DNA methylation; cancer; tumour;
KW   cytostatic; ALDH6; CYP11A; CYP11B1; CYP3A3; DPYD; EPHX2; OCLN; TXNRD1;
KW   UGT8; MRP; pharmacogenomics; SNP; single nucleotide polymorphism.
XX
OS   Homo sapiens.
XX
PN   WO200202806-A2.
XX
PD   10-JAN-2002.
XX
PF   29-JUN-2001; 2001WO-EP007470.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
```

```
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-154757/20.
XX
PT   New nucleic acid, oligonucleotides and peptide nucleic acid-oligomers,
PT   useful for detecting cytosine methylation state of genes associated with
PT   pharmacogenomics and for therapy of diseases e.g. cancer.
XX
PS   Claim 1; SEQ ID NO 11; 24pp; English.
XX
CC   The invention relates to a nucleic acid comprising a sequence at least 18
CC   bases in length of a segment of the chemically pretreated DNA of genes
CC   associated with pharmacogenomics according to one of the sequences of the
CC   genes ALDH6 (NM_000693), CYP11A (NM_000781), CYP11B1 (NM_000497), CYP3A3
CC   (NM_000776 and NM_017460), DPYD (NM_000110), EPHX2 (NM_001979), OCLN
CC   (NM_002538), TXNRD1 (NM_003330), UGT8 (NM_003360), MRP (NM_004996,
CC   NM_019900, NM_019901, NM_019902, NM_019862, NM_019898, NM_019899) and
CC   their complementary sequences, or a sequence (S1) chosen from 87
CC   sequences and their complements. The chemical pretreatment is bisulphite
CC   treatment to convert cytosines (but not methyl-cytosines) into uracils.
CC   Also included are an oligomer (II) in particular an oligonucleotide or a
CC   peptide nucleic acid (PNA)-oligomer, comprising in each case at least one
CC   base sequence having a length of 9 nucleotides which hybridises to or is
CC   identical to a chemically pretreated DNA of genes associated with
CC   pharmacogenomics and their complements, arranged in an array for
CC   analysing diseases associated with the methylation state (CpG) and/or
CC   detecting SNPs (single nucleotide polymorphisms) of the 87 sequences. The
CC   oligomers may also be used as PCR primers. The set of 87 nucleic acids
CC   and their complements is useful for diagnosis and therapy of solid
CC   tumours and cancer. The present sequence represents one the 87 DNA
CC   sequences or its complement. Note: The sequence data for this patent did
CC   not form part of the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 12069 BP; 3210 A; 185 C; 2849 G; 5823 T; 0 U; 2 Other;

  Query Match             6.3%;  Score 261.4;  DB 6;  Length 12069;
  Best Local Similarity   70.5%;  Pred. No. 3e-35;
  Matches  363;  Conservative    0;  Mismatches  151;  Indels    1;  Gaps    1;

Qy        204 ATCTCAGAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACAC 263
              || ||   | | | | ||     | ||||||||||||| |||| |||||| | | | | |
Db       9846 ATATCTATATATATATAAATATATCTATATATATATAAATATCTATATATATATAAATAT 9787

Qy        264 ACATATATATATACACATATATATACACATATATACATATATACACATATATACACATAT 323
               |||||  |||| | || |||||| | ||| |||  ||||| ||| |||||  | ||||
Db       9786 CTATATACGTATAAATATCTATATATATATAAATATCTATATCCACGTATATCTATATAT 9727

Qy        324 ATATACACATACATATGTACACATATATATACACATATGTATACACATATATACACATAT 383
              ||||| | ||  |||| || | | ||||||| |   | |||||   ||||||| | | ||
Db       9726 ATATAAATATCTATATATATATAAATATATATATCCACGTATATCTATATATATATAAAT 9667

Qy        384 ATACACACATATATACACATATATACACACATATATACACATATATACACATATATACAC 443
              ||  | | ||||||| | |||||||  ||| |||||  | ||||||| | |||| |||
Db       9666 ATCTATATATATATAAATATATATATCCACGTATATCTATATATATATAAATATCTACGT 9607

Qy        444 ATATACACATATATACACATATATACATATATACACATATATATAATATACACACATATA 503
              ||| | | |||||| ||| |||||  ||||||| | | |||| | |   || | | |||||
Db       9606 ATAAATATATATATCCACGTATATCTATATATATATAAATATCT-ACGTATAAATATATA 9548

Qy        504 TATACACATATATACACACATATATACACATATATACACATATATATACACATATATACA 563
              ||| ||| |||||  | | ||||||| | || |||| | ||||| ||| | ||||  ||
Db       9547 TATCCACGTATATCTATATATATATAAATATCTATATATATATAAATATATATATCCACG 9488

Qy        564 CATATATACATATATACACATATATATACATATATACACATATATACATATATACACATA 623
              |||| || ||||||| | |||| |||| ||||||| | || | || ||||||| | |||
Db       9487 TATATCTATATATATATAAATATCTATATATATATAAATATCTCTATATATATAAATATA 9428

Qy        624 TATACATATATACACACATATATACACATACATATACACACACATAGATATACATATATA 683
              |||| | ||| |   ||||||| | | | || || | | | ||| ||||  |||||||
Db       9427 TATATCCACATATATCTATATATATATAAATATCTATATATATATAAATATCTATATATA 9368

Qy        684 TACACATATATATACGTATATATATGTATATATAT 718
              || | | |||| ||  |||||||  ||| |||||
Db       9367 TATATAAATATCTATATATATATAAATATCTATAT 9333


RESULT 35
AAK73165/c
ID   AAK73165 standard; DNA; 20420 BP.
XX
AC   AAK73165;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:27977.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001354.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
```

PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS Disclosure; SEQ ID NO 27977; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention
XX
SQ Sequence 20420 BP; 4834 A; 5403 C; 5630 G; 4553 T; 0 U; 0 Other;

Query Match 6.3%; Score 260.6; DB 4; Length 20420;
Best Local Similarity 71.7%; Pred. No. 4.2e-35;
Matches 355; Conservative 0; Mismatches 139; Indels 1; Gaps 1;

```
Qy       227 TTTATATATATATATATATATATACACACACACACATATATATATACACATATATA 286
             |||  ||||||||||||||||||||    || | | | |  |||||||||   ||||||||
Db      4732 TTTTCATATATATATATATATATTTTCATATATATATTTATATATATTTTCATATATA 4673

Qy       287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
             || |  |||||| |||| | || |||||||  |  ||||||||  |  | |||||  || |
Db      4672 TATATTTATATATATATTTTCATATATATATATTTATATATATATTTTCATATATATATT 4613

Qy       347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
             ||||||||  |  |   ||||  | ||  ||||  | ||||  | ||  | |||||||   |||||
Db      4612 TATATATATATTTTCATATATATATTTATATATATATTTTCATATATATATATTTATATA 4553

Qy       407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
             || |      |||||  | ||||  ||  | ||||||   ||||||  | ||||  ||  | |||||
Db      4552 TATATTTTCATATATATATATTTATATATATATTTTCATATATATATATTTATATATATA 4493

Qy       467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
             |    |||||  | ||||  ||||  |||||    ||||||||||  |  ||||||  |  |  |
```

```
Db       4492 TTTTCATATATATATATTTATATATATATTTTCATATATATATATTTATATATATATTTT 4433

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              ||| | ||||||  | ||||||||   |||||||| | || |||| ||||||   ||||
Db       4432 CATATATATATATTTATATATATATTTTCATATATATATATTTATATATATATTTTCATA 4373

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              ||||||  ||||||  | || |  | |||||||  |  ||||||  |||| | || | ||||
Db       4372 TATATATTTATATATATATTTTCATATATATATATTTATATATATATTTTCATATATATT 4313

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              | || ||| ||||   || | ||| || || ||||||||  | ||||||  | | ||||||
Db       4312 TTCATATATATATTTTCATATATATATTTATATATATATTTATATATATATTTTCATATATA 4253

Qy        706 TATGTATATATATAT 720
              ||| ||||||||||||
Db       4252 TATTTATATATATAT 4238
```

RESULT 36
ABK69933/c
ID ABK69933 standard; DNA; 20420 BP.
XX
AC ABK69933;
XX
DT 15-JUL-2002 (first entry)
XX
DE Human secreted protein gene 68 genomic DNA fragment #21.
XX
KW Human; ds; secreted protein; gene therapy; immunosuppressive;
KW antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;
KW vasotropic; cerebroprotective; nootropic; neuroprotective; antibacterial;
KW virucide; fungicide; opthalmalogical; autoimmune disease; neoplasm;
KW rheumatoid arthritis; hyperproliferative disorder; cardiac arrest;
KW cardiovascular disorder; cerebrovascular disorder; cerebral ischaemia;
KW angiogenesis; nervous system disorder; Alzheimer's disease; infection;
KW ocular disorder; corneal infection; wound healing; skin aging;
KW epithelial cell proliferation; food additive.
XX
OS Homo sapiens.
XX
PN WO200226931-A2.
XX
PD 04-APR-2002.
XX
PF 24-SEP-2001; 2001WO-US029871.
XX
PR 25-SEP-2000; 2000US-0234925P.
PR 12-JAN-2001; 2001WO-US000911.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Komatsoulis G, Duan DR, Rosen CA, Moore PA, Shi Y;
PI Lafleur DW, Olsen H, Brewer LA, Florence KA, Young PE, Soppet DR;
PI Endress GA, Mucenski M, Ebner R;
XX
DR WPI; 2002-362489/39.
XX
PT Novel 71 isolated secreted polypeptides and polynucleotides encoding the
PT polypeptides, useful for treating Huntington's disease, sepsis,
PT meningitis, thrombocytopenia, hemolytic anemia, rheumatoid arthritis,
PT asthma.
XX
PS Example 2; Page 1440-1445; 1478pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (or its
CC fragment, homologue complement or allelic variant) encoding a human
CC secreted protein (and its fragment, domain, epitope, variant, secreted
CC form and species variant). Also included are a recombinant vector
CC comprising the nucleic acid, a recombinant host cell comprising the
CC vector, an antibody against the secreted protein, a recombinant host cell
CC that expresses the secreted protein and a method of identifying a binding
CC partner of the secreted protein. The nucleic acid and protein are used to
CC prevent, diagnose, treat or ameliorate a medical condition in e.g.
CC humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep for
CC example autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC angiogenesis, nervous system disorders e.g. Alzheimer's disease,
CC infections caused by bacteria, viruses and fungi and ocular disorders
CC e.g. corneal infection. Many other diseases and disorders are listed in
CC the specification. The polypeptides can also be used to aid wound healing
CC an epithelial cell proliferation, to prevent skin aging due to sunburn,
CC to maintain organs before transplantation, for supporting cell culture of
CC primary tissues, to regenerate tissues and in chemotaxis. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities. The present sequence
CC represents a ds DNA fragment of the gene for a novel human secreted
CC protein of the invention
XX
SQ Sequence 20420 BP; 4834 A; 5403 C; 5630 G; 4553 T; 0 U; 0 Other;

Query Match 6.3%; Score 260.6; DB 6; Length 20420;
Best Local Similarity 71.7%; Pred. No. 4.2e-35;
Matches 355; Conservative 0; Mismatches 139; Indels 1; Gaps 1;

```
Qy        227 TTTATATATATATATATATATATATACACACACACACATATATATATACACATATATA 286
              |||  ||||||||||||||||||||    || | | | |  |||||||||   ||||||||
Db       4732 TTTTCATATATATATATATATATTTTCATATATATATTTATATATATTTTCATATATA 4673

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || |  |||||| |||| | || ||||||| |  |||||||| |  | |||| || |
Db       4672 TATATTTATATATATATTTTCATATATATATTTATATATATATTTTCATATATATATT 4613

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| |  |  |||| | || |||| | |||| | || | |||||||   |||||
Db       4612 TATATATATATTTTCATATATATATTTATATATATATTTTCATATATATATATTTATATA 4553

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || |      ||||| | |||| || | ||||||   |||||| | |||| || | |||||
Db       4552 TATATTTTCATATATATATATTTATATATATATTTTCATATATATATATTTATATATATA 4493

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              |    ||||| | |||| |||| |||||    |||||||||| |  |||||| | |  |
Db       4492 TTTTCATATATATATATTTATATATATATTTTCATATATATATATTTATATATATATTTT 4433

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||  | ||||||||   |||||||| | || |||| ||||||   ||||
Db       4432 CATATATATATATTTATATATATATTTTCATATATATATATTTATATATATATTTTCATA 4373

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              ||||||  ||||||  | || |  | |||||||  |  ||||||  |||| | || | ||||
Db       4372 TATATATTTATATATATATTTTCATATATATATATTTATATATATATTTTCATATATATT 4313

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              | || ||| ||||   || | ||| || || ||||||||  | ||||||  | | ||||||
Db       4312 TTCATATATATATTTTCATATATATATTTATATATATATTTATATATATATTTTCATATATA 4253

Qy        706 TATGTATATATATAT 720
              ||| ||||||||||||
Db       4252 TATTTATATATATAT 4238
```

RESULT 37
ADC87554
ID ADC87554 standard; DNA; 1489 BP.
XX
AC ADC87554;
XX
DT 01-JAN-2004 (first entry)
XX
DE Human GPCR gene SEQ ID NO:2007.
XX
KW ds; gene; human; GPCR;
KW guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX

OS Homo sapiens.
XX
PN EP1270724-A2.
XX
PD 02-JAN-2003.
XX
PF 18-JUN-2002; 2002EP-00013517.
XX
PR 18-JUN-2001; 2001JP-00246789.
XX
PA (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI Suwa M, Asai K, Akiyama Y, Aburatani H;
XX
DR WPI; 2003-315783/31.
DR P-PSDB; ADC87555.
XX
PT New polynucleotide, useful for preparing a composition for treating a
PT patient in need of increased or suppressed activity or expression of the
PT guanosine triphosphate-binding protein coupled receptor.
XX
PS Claim 1; SEQ ID NO 2007; 28pp; English.
XX
CC The invention relates to a novel polynucleotide encoding a guanosine
CC triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC the invention may have a use in gene therapy. The polynucleotide and
CC polypeptide are useful for preparing a composition for treating a patient
CC in need of increased or suppressed activity or expression of the
CC guanosine triphosphate-binding protein coupled receptor. The
CC polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC invention.
XX
SQ Sequence 1489 BP; 553 A; 208 C; 160 G; 568 T; 0 U; 0 Other;

Query Match 6.3%; Score 258.8; DB 10; Length 1489;
Best Local Similarity 72.9%; Pred. No. 7.7e-35;
Matches 360; Conservative 0; Mismatches 132; Indels 2; Gaps 2;

```
Qy       229 TATATATATATATATATATATATACACACACACACACATATATATATACACATATATATA 288
Db       519 TGTATATATATGTATGTATATATATACATATATACACATATGTATATATACGTACC-ATA 577

Qy       289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
Db       578 TATGTATATGTGTATATATACGTATATATACATGTATATATACGTATATATACACATATA 637

Qy       349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
Db       638 CATATATACATTTATATATACGTATATATACGTATACATATATATACACGTATATATACA 697

Qy       409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
Db       698 TACATATATACACGTGTATATATACGTATACATATATACACGTGTATATATACGTATACA 757

Qy       469 CATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATATA 527
Db       758 TATATATACACGTGTATATATACGTATACATATATACACGTGTATATATACGTATACATA 817

Qy       528 TACACATATATACACATATATATACACATATATACACATATATACATATATACACATATA 587
Db       818 TATACACGTGTATATATACGTATACATATATACACGTGTATATATACGTATACATATATA 877

Qy       588 TATACATATATACACATATATACATATATACACATATATACATATATACACACATATATA 647
Db       878 CACGTGTATATACATACATATATACACGTGTATATACATATGTATATACACGTGTATATA 937

Qy       648 CACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATA 707
Db       938 TGTATATGTATACATATACGTATATATACATGTATATACACGTATATATACGTGTATATA 997

Qy       708 TGTATATATATATG 721
Db       998 TGTATACATATACG 1011
```

RESULT 38
ADO15781/c
ID ADO15781 standard; DNA; 2761 BP.
XX
AC ADO15781;
XX
DT 29-JUL-2004 (first entry)
XX
DE 4 synthesis-period of neuroblastoma related DNA, SEQ ID 43.
XX
KW Human; 4 synthesis-period; neuroblastoma; stage 4S; ds.
XX
OS Homo sapiens.
XX
PN WO2004039975-A1.
XX
PD 13-MAY-2004.
XX
PF 30-OCT-2003; 2003WO-JP013932.
XX
PR 30-OCT-2002; 2002JP-00316586.
XX
PA (HISM) HISAMITSU PHARM CO LTD.
PA (CHIB-) CHIBA PREFECTURE.
XX
PI Nakagawara A, Ohira M;
XX
DR WPI; 2004-390323/36.
XX
PT Novel nucleic acid obtained from 4 synthesis-period of neuroblastoma
PT cells useful for prognosing and determining progress stage of
PT neuroblastomas.
XX
PS Claim 1; SEQ ID NO 43; 455pp; Japanese.
XX
CC The present invention relates to human nucleic acid sequences (I;
CC ADO15739-ADO15912) obtained from 4 synthesis-period (stage 4S) of
CC neuroblastoma cell. (I) is useful for prognosing and determining the
CC progress stage of 4 synthesis-period of neuroblastoma.
XX
SQ Sequence 2761 BP; 847 A; 475 C; 598 G; 841 T; 0 U; 0 Other;

Query Match 6.3%; Score 258.8; DB 12; Length 2761;
Best Local Similarity 71.0%; Pred. No. 7.8e-35;
Matches 375; Conservative 0; Mismatches 142; Indels 11; Gaps 2;

```
Qy       229 TATATATATATATATATATATATACACACACACACACATATATATATACACATATATATA 288
Db      2444 TGTGTATATATATATACATATATACGTATATATGTGTATATATATACACACGTATATATA 2385

Qy       289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
Db      2384 CACATATATGTATACATACACACGTATATATACACATATATGTATACATACACACGCATA 2325

Qy       349 TATATACACATATGTATACA--------CATATATACACATATATACACACATATATACA 400
Db      2324 TATACACATATATGTATACATACACACGTATATATACACATATATGTATACATACACACG 2265

Qy       401 CATATATACACACATATATACACATATATACACATATATACACATATACACATATATACA 460
Db      2264 TATATATACACATATATGTATACATACATACGTATATATACACATATATGTATACATACA 2205

Qy       461 CATATATACATATATACACATATATATAATATACACACATATATATACACATATATACAC 520
Db      2204 TACGTATATATACACATATATGTAT---ACATACATACGTATATATACACATATATGTAT 2148

Qy       521 ACATATATACACATATATACACATATATATACACATATATACACATATATACATATATAC 580
Db      2147 ACATACATACGTATATATACACATATATGTATACATACATACGTATATATACACATATAT 2088
```

```
Qy        581 ACATATATATACATATATACACATATATACATATATACACATATATACATATATACACAC 640
              ||| ||| |    |||| | | | ||| || |||||| | ||    ||||||||||
Db       2087 GTATACATACATGCGTATATATACACATATATGTATACATACATGCGTATATATACACAT 2028

Qy        641 ATATATACACATACATATACACACACATAGATATACATATATATACACATATATATACGT 700
              |||| || ||||| ||||  ||||| ||| || |   | |||| | | |||||      |
Db       2027 ATATGTATACATATATATGAACACATATATATGTGTGTGTATACATATATATAAGATTTT 1968

Qy        701 ATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTT 748
              |     | ||    ||| |   |||| ||  || | ||| | ||
Db       1967 TCTTGACCATGAATCACCATGTTTGGGAGTACAGTAGGGTTAGGAATT 1920
```

RESULT 39
ADL13771
ID ADL13771 standard; DNA; 201766 BP.
XX
AC ADL13771;
XX
DT 06-MAY-2004 (first entry)
XX
DE Osteoarthritis-associated polymorphic nucleotide #303.
XX
KW ds; gene; osteopathic; antiinflammatory; antiarthritic; gene therapy;
KW joint space narrowing; osteophyte development; joint pain;
KW osteoarthritis; SNP; single nucleotide polymorphism.
XX
OS Homo sapiens.
XX
PN WO2003054166-A2.
XX
PD 03-JUL-2003.
XX
PF 19-DEC-2002; 2002WO-US041225.
XX
PR 20-DEC-2001; 2001US-0342603P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Jones KA, Schafer A;
XX
DR WPI; 2003-559141/52.
XX
PT Determining susceptibility of an individual to joint space narrowing,
PT osteophyte development and/or joint pain comprises identifying whether
PT the individual has at least one polymorphism in a polynucleotide encoding
PT a protein.
XX
PS Disclosure; SEQ ID NO 303; 297pp; English.
XX
CC The invention relates to a method of determining susceptibility of an
CC individual to joint space narrowing and/or osteophyte development and/or
CC joint pain comprising identifying whether the individual has at least one
CC polymorphism in a polynucleotide encoding at least one of the protein
CC listed in the specification. The methods, composition and agent are
CC useful for modulating the susceptibility of an individual to joint space
CC narrowing and/or osteophyte development and/or joint pain that is
CC associated with a disease, preferably osteoarthritis. The cell line and
CC the non-human animal are useful for screening for an agent for diagnosing
CC an individual having susceptibility to joint space narrowing and/or
CC osteophyte development and/or joint pain. This sequence corresponds to
CC the polynucleotide encoding a protein listed in the specification. (Note:
CC The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences).
XX
SQ Sequence 201766 BP; 57593 A; 35575 C; 36821 G; 54688 T; 0 U; 17089 Other;

Query Match 6.3%; Score 258.8; DB 10; Length 201766;
Best Local Similarity 72.9%; Pred. No. 9.2e-35;
Matches 360; Conservative 0; Mismatches 132; Indels 2; Gaps 2;

```
Qy        229 TATATATATATATATATATATATACACACACACACACATATATATATACACATATATATA 288
              | ||||||||| ||| |||||||| ||| | | |||||||| |||||| || ||   |||
Db     188638 TGTATATATATGTATGTATATATATACATATATACACATATGTATATATACGTACC-ATA 188696

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              | |||||   |||||| || |||||| |||| |||||| || || ||||  ||| |||
Db     188697 TATGTATATGTGTATATATACGTATATATACATGTATATATACGTATATATACACATATA 188756

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              ||||| |||| | |||| || |||||| || |||| | | | ||| |   | |||||
Db     188757 CATATATACATTTATATATACGTATATATACGTATACATATATATACACGTATATATACA 188816

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              ||| ||||| ||   |||||| || |||| | | ||| ||   |||||| || |||| |
Db     188817 TACATATATACACGTGTATATATACGTATACATATATACACGTGTATATATACGTATACA 188876

Qy        469 CATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATATA 527
              |||||||||| | |||||| |  || ||| ||||| |      |||||||  | | |||
Db     188877 TATATATACACGTGTATATATACGTATACATATATACACGTGTATATATACGTATACATA 188936

Qy        528 TACACATATATACACATATATATACACATATATACACATATATACATATATACACATATA 587
              || |||  | || | |||  |||||| ||||| ||   ||||||  |  |||||| |||||
Db     188937 TATACACGTGTATATATACGTATACATATATACACGTGTATATATACGTATACATATATA 188996

Qy        588 TATACATATATACACATATATACATATATACACATATATACATATATACACACATATATA 647
              |    |||||||| | ||||| | |  |  | ||| |||  |||||||||   ||||||
Db     188997 CACGTGTATATACATACATATATACACGTGTATATACATATGTATATACACGTGTATATA 189056

Qy        648 CACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATA 707
              |||  |||||| | || || |||||||| ||||||||| |||||||||| ||||||
Db     189057 TGTATATGTATACATATACGTATATATACATGTATATACACGTATATATACGTGTATATA 189116

Qy        708 TGTATATATATATG 721
              |||||| ||||| |
Db     189117 TGTATACATATACG 189130

RESULT 40
ABD32659/c
ID ABD32659 standard; DNA; 163382 BP.
XX
AC ABD32659;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human cancer-associated genomic DNA HD13-065.
XX
KW Human; ds; cancer-associated protein; gene; cytostatic; cancer;
KW leukaemia; lymphoma; CAP.
XX
OS Homo sapiens.
XX
PN WO2004074320-A2.
XX
PD 02-SEP-2004.
XX
PF 17-FEB-2004; 2004WO-US004730.
XX
PR 14-FEB-2003; 2003US-00367094.
PR 14-MAR-2003; 2003US-00388838.
PR 15-APR-2003; 2003US-00417375.
PR 13-JUN-2003; 2003US-00461862.
PR 15-SEP-2003; 2003US-00663431.
PR 15-DEC-2003; 2003US-00737318.
XX
PA (SAGR-) SAGRES DISCOVERY INC.
XX
PI Morris DW, Morris DW, Malandro MS;
XX
DR WPI; 2004-652914/63.
XX
PT New isolated cancer-associated polynucleotides and polypeptides useful
PT for diagnosing, preventing or treating cancers, especially lymphoma and
PT leukemia, or in screening for agents that modulate cancer.
```

```
XX
PS   claim 16; seqid 199; 310pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising at least 10
CC   contiguous nucleotides of any of the 233 polynucleotide sequences given
CC   in the specification, or its complement. The nucleic acids encode cancer-
CC   associated proteins. Also included are an expression vector comprising
CC   the isolated nucleic acid cited above, a host cell comprising the above
CC   recombinant nucleic acid or expression vector, a microarray for detecting
CC   a cancer-associated (CA) nucleic acid comprising at least one probe
CC   comprising at least 10 contiguous nucleotides of any of the above-
CC   mentioned nucleotide sequences, an isolated polypeptide (encoded within
CC   an open reading frame of a CA sequence selected from any of the 95
CC   polynucleotide sequences as mentioned in the specification, or its
CC   complement), an isolated antibody, (or its antigen binding fragment) that
CC   binds to the above polypeptide, a hybridoma that produces the above
CC   monoclonal antibody, a pharmaceutical composition comprising the above
CC   antibody and a pharmaceutical excipient, a kit for detecting cancer
CC   cells(comprising the antibody cited above, methods for diagnosing cancer
CC   or for detecting the presence or absence of cancer cells in an
CC   individual, a method for inhibiting growth of cancer cells in an
CC   individual, a method for delivering a therapeutic agent to cancer cells
CC   in an individual, an electronic library comprising the above
CC   polynucleotide or polypeptide (or their fragments), methods of screening
CC   for anticancer activity or for a bioactive agent capable of modulating
CC   the activity of a CA protein (CAP), methods for detecting cancer
CC   associated with expression of a polypeptide in a test cell sample, a
CC   method for treating cancers and a method for inhibiting the expression of
CC   CA gene in a cell. The composition and methods are useful for detecting,
CC   diagnosing, preventing and treating cancers, especially lymphoma and
CC   leukaemia. These may also be used in screening for agents that modulate
CC   cancer. The present sequence is a human CAP genomic sequence. Note: The
CC   sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 163382 BP; 50073 A; 25805 C; 27837 G; 53184 T; 0 U; 6483 Other;

  Query Match             6.3%;  Score 258.6;  DB 13;  Length 163382;
  Best Local Similarity   70.0%;  Pred. No. 9.9e-35;
  Matches  362;  Conservative    0;  Mismatches  154;  Indels    1;  Gaps    1;

Qy        204 ATCTCAGAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACAC 263
              || | |   | | | |||        |||| ||||||| |||| |||| | ||| | | |
Db     145185 ATATAAATATATATAAATAAATAAATATAAATATATAAATATGTATAAATACATATATAA 145126

Qy        264 ACATATATATATACACATATATATACACATATATACATATATACACATATATACACATAT 323
              | | ||||| ||| | | |||||||  | ||| ||| ||| ||| | ||| ||| | ||||
Db     145125 ATAAATATAAATATATAAATATATATAAATACATATATAAATATATATAAATATATATAT 145066

Qy        324 ATATACACATACATATGTACACATATATATACACATATGTATACACATATATACACATAT 383
                ||| ||||| ||||||| | |||||||  ||| ||||  ||| | | ||||| | | ||
Db     145065 TAATATACATAAATATGTATAAATATATAAACATATATTAATATATAAATATATATAAAT 145006

Qy        384 ATACACACATATATACACATATATACACACATATATACACATATATACACATATATACAC 443
              |||    | |||||||    |||||||  | | ||  ||| | ||| |||      ||||| ||
Db     145005 ATATTAATATATATATTTATATATAAATATATGAATATATATAAATATGTTAATATATAC 144946

Qy        444 ATATACACATATATACACATATATACATATATACACATATATATAATATACACACATATA 503
              ||||| | | ||| | | ||| | | |||||||  | |||||||     ||| | | |||||
Db     144945 ATATATATAAATAAATATATAAAAATATATATAAATATATATAACTAATATATAAATATA 144886

Qy        504 TATACACATATATACACACATATATACACATATATACACATATATATACACATATATACA 563
              |||| ||  ||||| | | ||||||| | ||||||  ||  |||||||  | ||||||| |
Db     144885 TATA-ACTAATATATATAAATATATAAATATATATGAACTAATATATATAAATATATAAA 144827

Qy        564 CATATATACATATATACACATATATATACATATATACACATATATACATATATACACATA 623
              | ||||| | ||| | | ||| ||| | ||||| | | ||||||| ||| ||| | | |
Db     144826 TAAATATATAAATAAATATATAAATAAATATATACAAATATATATAAATAAATATATAAA 144767

Qy        624 TATACATATATACACACATATATACACATACATATACACACACATAGATATACATATATA 683
              ||| ||||| | | | ||||| | | ||| ||||| | | | ||| ||| | | |||||
Db     144766 AATAAATATAAATAAATATATAAATAAATATATATAAATAAATATATATAAATAAATATA 144707
```

```
Qy        684 TACACATATATATACGTATATATATGTATATATATAT 720
              || | ||| ||||| || ||||||| |||| ||| ||
Db     144706 TATAAATAAATATATATAAATATATATATAAATAAAT 144670


RESULT 41
ABV72623
ID   ABV72623 standard; DNA; 82938 BP.
XX
AC   ABV72623;
XX
DT   27-NOV-2002  (first entry)
XX
DE   Human transporter protein encoding gene.
XX
KW   Human; transporter protein; hTP; gamma-aminobutyric-acid; GABA;
KW   gene therapy; protein therapy; vaccine; gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1120. .81932
FT                   /*tag=  a
FT                   /product= "Transporter protein"
FT                   /note= "Contains 8 introns"
FT   exon            1120. .1223
FT                   /*tag=  b
FT                   /number= 1
FT   intron          1224. .27733
FT                   /*tag=  c
FT                   /number= 1
FT   variation       replace(2419,C)
FT                   /*tag=  d
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(3191,A)
FT                   /*tag=  e
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(4195,T)
FT                   /*tag=  f
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(5910,C)
FT                   /*tag=  g
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(7119,T)
FT                   /*tag=  h
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(7529,G)
FT                   /*tag=  h
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(8855,C)
FT                   /*tag=  i
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(8956,A)
FT                   /*tag=  j
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(10220,G)
FT                   /*tag=  k
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(11628,C)
FT                   /*tag=  l
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(11670,A)
FT                   /*tag=  m
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(11671,G)
FT                   /*tag=  n
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(11931,C)
FT                   /*tag=  o
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(15984,G)
FT                   /*tag=  p
```

```
FT                     /standard_name= "Single nucleotide polymorphism"
FT   variation         replace(17735,A)
FT                     /*tag= q
FT                     /standard_name= "Single nucleotide polymorphism"
FT   variation         replace(17910,A)
FT                     /*tag= r
FT                     /standard_name= "Single nucleotide polymorphism"
FT   variation         replace(18130,C)
FT                     /*tag= s
FT                     /standard_name= "Single nucleotide polymorphism"
FT   variation         replace(19293. .19294,TWA)
FT                     /*tag= t
FT                     /standard_name= "Single nucleotide polymorphism"
FT   variation         replace(19294. .19295,ATT)
FT                     /*tag= u
FT                     /standard_name= "Single nucleotide polymorphism"
FT   variation         replace(19296. .19297,CYT)
FT                     /*tag= v
FT                     /standard_name= "Single nucleotide polymorphism"
FT   variation         replace(19303. .19304,TYC)
FT                     /*tag= w
FT                     /standard_name= "Single nucleotide polymorphism"
FT   variation         replace(19304. .19305,CYC)
FT                     /*tag= x
FT                     /standard_name= "Single nucleotide polymorphism"
FT   variation         replace(19307. .19308,TMC)
FT                     /*tag= y
FT                     /standard_name= "Single nucleotide polymorphism"
FT   variation         replace(19308. .19309,CCC)
FT                     /*tag= z
FT                     /standard_name= "Single nucleotide polymorphism"
FT   variation         replace(19313. .19314,TMA)
FT                     /*tag= aa
FT                     /standard_name= "Single nucleotide polymorphism"
FT   variation         replace(19314. .19315,AWT)
FT                     /*tag= ab
FT                     /standard_name= "Single nucleotide polymorphism"
FT   variation         replace(19315. .19316,TCC)
FT                     /*tag= ac
FT                     /standard_name= "Single nucleotide polymorphism"
FT   variation         replace(19316. .19317,CYT)
FT                     /*tag= ad
FT                     /standard_name= "Single nucleotide polymorphism"
FT   variation         replace(19317. .19318,TWA)
FT                     /*tag= ae
FT                     /standard_name= "Single nucleotide polymorphism"
FT   variation         replace(19318. .19319,ATT)
FT                     /*tag= af
FT                     /standard_name= "Single nucleotide polymorphism"
FT   variation         replace(19860,G)
FT                     /*tag= ag
FT                     /standard_name= "Single nucleotide polymorphism"
FT   variation         replace(20843,G)
FT                     /*tag= ah
FT                     /standard_name= "Single nucleotide polymorphism"
FT   variation         replace(21144,T)
FT                     /*tag= ai
FT                     /standard_name= "Single nucleotide polymorphism"
FT   variation         replace(21371,G)
FT                     /*tag= aj
FT                     /standard_name= "Single nucleotide polymorphism"
FT   variation         replace(22061. .22062,A)
FT                     /*tag= ak
FT                     /standard_name= "Single nucleotide polymorphism"
FT   variation         replace(22321,T)
FT                     /*tag= al
FT                     /standard_name= "Single nucleotide polymorphism"
FT   variation         replace(22470. .22471,A)
FT                     /*tag= am
FT                     /standard_name= "Single nucleotide polymorphism"
FT   variation         replace(22472. .22473,A)
FT                     /*tag= an
FT                     /standard_name= "Single nucleotide polymorphism"
FT   variation         replace(24990,C)
FT                     /*tag= ao
FT                     /standard_name= "Single nucleotide polymorphism"
FT   variation         replace(25284,A)
FT                     /*tag= ap
FT                     /standard_name= "Single nucleotide polymorphism"
FT   variation         replace(25934,G)
FT                     /*tag= aq
FT                     /standard_name= "Single nucleotide polymorphism"
FT   variation         replace(27509. .27510,TAT)
FT                     /*tag= ar
FT                     /standard_name= "Single nucleotide polymorphism"
FT   variation         replace(27513. .27514,TTT)
FT                     /*tag= as
FT                     /standard_name= "Single nucleotide polymorphism"
FT   exon              27734. .27882
FT                     /*tag= at
FT                     /number= 2
FT   intron            27883. .41045
FT                     /*tag= au
FT                     /number= 2
FT   variation         replace(28608,G)
FT                     /*tag= av
FT                     /standard_name= "Single nucleotide polymorphism"
FT   variation         replace(28627,C)
FT                     /*tag= aw
FT                     /standard_name= "Single nucleotide polymorphism"
FT   variation         replace(29709,T)
FT                     /*tag= ax
FT                     /standard_name= "Single nucleotide polymorphism"
FT   variation         replace(29788,T)
FT                     /*tag= ay
FT                     /standard_name= "Single nucleotide polymorphism"
FT   variation         replace(29878,T)
FT                     /*tag= az
FT                     /standard_name= "Single nucleotide polymorphism"
FT   variation         replace(30294,T)
FT                     /*tag= ba
FT                     /standard_name= "Single nucleotide polymorphism"
FT   variation         replace(30896. .30897,T)
FT                     /*tag= bb
FT                     /standard_name= "Single nucleotide polymorphism"
FT   variation         replace(31815,G)
FT                     /*tag= bc
FT                     /standard_name= "Single nucleotide polymorphism"
FT   variation         replace(31999,TTT)
FT                     /*tag= bd
FT                     /standard_name= "Single nucleotide polymorphism"
FT   variation         replace(32026,T)
FT                     /*tag= be
FT                     /standard_name= "Single nucleotide polymorphism"
FT   variation         replace(32146,A)
FT                     /*tag= bf
FT                     /standard_name= "Single nucleotide polymorphism"
FT   variation         replace(32184,S)
FT                     /*tag= bg
FT                     /standard_name= "Single nucleotide polymorphism"
FT   variation         replace(32214. .32215,T)
FT                     /*tag= bh
FT                     /standard_name= "Single nucleotide polymorphism"
FT   variation         replace(32215,C)
FT                     /*tag= bi

  Query Match             6.3%;  Score 258.4;  DB 6;  Length 82938;
  Best Local Similarity   70.5%;  Pred. No. 1e-34;
  Matches  361;  Conservative    0;  Mismatches  146;  Indels    5;  Gaps    1;

Qy       214 AAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACATATATAT 273
             | | | | |    | |||| | | | |||| |||| |||||||||   | |||||| |||
Db      6946 ATACATATGTATATATATACACACACATATGTATACATACACACATGTATACATATGTAT 7005

Qy       274 ATACACATATATATACACATATATACATATATACACATATATACACATATATATACACAT 333
             |||   ||||| |||||||| ||||||||| || | || |||| | ||| | ||  | |
```

```
Db       7006 ATATGTATATAGATACACATGTATACATATGTATATATGTATATAGATACACATGTATAC 7065

Qy        334 ACATATGTACACATATATATACACATATGTATACACATATATACACATATATACACACAT 393
              | || | || |  |||||| | ||| ||||||||| || |||| |  |||||| | |||
Db       7066 ATATGTATATATGTATATAGATACACATGTATACATATGTATATATGTATATAGATACAC 7125

Qy        394 ATATACACATATATACACACATATATACACATATATACACATATATACACATATACACAT 453
              || || |||||| || | |  |||||| | | | ||  | | |||| | ||||  | ||
Db       7126 ATGTATACATATGTATATATGTATATAGATACACATGTATACATATGTATATATGTATAT 7185

Qy        454 ATATACACATATATACATATATACACATATATATA-----ATATACACACATATATATAC 508
              | |||||||| ||||||||| || | || ||||||     | ||  | |||||| ||||
Db       7186 AGATACACATGTATACATATGTATATATGTATATAGATACACATGTATACATATGTATAT 7245

Qy        509 ACATATATACACACATATATACACATATATACACATATATATACACATATATACACACATAT 568
              |  |||||| | ||| || || |||||| || | || ||||||||  ||| || | || |
Db       7246 ATGTATATAGATACACATGTATACATATGTATATATGTATATACATGTATGTATATATGT 7305

Qy        569 ATACATATATACACATATATATACATATATACACATATATACATATATACACATATATAC 628
              |||  ||||||||  ||| ||||  | ||||   ||||||  |||| || | | || ||
Db       7306 ATATGTATATACATGTATGTATATGTGTATATGTATATATGTATATGTATATACATGTAT 7365

Qy        629 ATATATACACACATATATACACATACATATACACACACATAGATATACATATATATACAC 688
              ||||||  |||  |||||||| |  ||||| | |||  || |||   ||||| ||| |
Db       7366 GTATATATACAGCTATATACATGTGTATATATATACAGCTATATACGTATATACATATAT 7425

Qy        689 ATATATATACGTATATATATGTATATATATAT 720
              || |||||| ||||| |||||| |||||||||
Db       7426 ATGTATATATGTATACATATGTGTATATATAT 7457


RESULT 42
AAK73166/c
ID   AAK73166 standard; DNA; 19965 BP.
XX
AC   AAK73166;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:27978.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001354.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0235484P.
PR   27-SEP-2000; 2000US-0235834P.
PR   27-SEP-2000; 2000US-0235836P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   13-OCT-2000; 2000US-0239937P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241221P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   20-OCT-2000; 2000US-0241826P.
PR   01-NOV-2000; 2000US-0244617P.
PR   08-NOV-2000; 2000US-0246474P.
PR   08-NOV-2000; 2000US-0246475P.
PR   08-NOV-2000; 2000US-0246476P.
PR   08-NOV-2000; 2000US-0246477P.
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
```

PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS Disclosure; SEQ ID NO 27978; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention
XX
SQ Sequence 19965 BP; 4699 A; 5296 C; 5505 G; 4465 T; 0 U; 0 Other;

Query Match 6.3%; Score 258; DB 4; Length 19965;
Best Local Similarity 64.7%; Pred. No. 1.2e-34;
Matches 431; Conservative 0; Mismatches 230; Indels 5; Gaps 3;

```
Qy        59 TAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGA 118
             | | ||||   ||||   ||||||| ||||||||||   || ||||||||||||||||
Db      4999 TGGCCAGGAGCGGTGGCTCACGCCTATAATCCCAGCACGTTGGGAGGCCGAGGCAGGGAG 4940

Qy       119 ATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCA 178
             |  |  | |  | | |  ||          |||| |   ||  ||| ||   |
Db      4939 GATCGCTTGAGAT-CAAGAGCTCCAGACCAGCCTAGGCAACACGGTGAAATTCCATCTCT 4881

Qy       179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATA 238
             |  |   || |  |   |||    |  |   | | | | |      | ||||| | ||
Db      4880 ACAAAAAAAATACAAAAAAAAAAAAAAAAATATATATATATATATTTATATATATATGTTTA 4821

Qy       239 TATATATATATATACACACACACACATATATATATACACATATATATACACATATATA 298
             |||||||||||| || | |      | | |||| |||||    ||||||||  | ||||||
Db      4820 TATATATATATTTATATATTTTTATATATATTTATATTTTTATATATATTTATATATATT 4761

Qy       299 CATATATA-CACATATATACACATATATATACACATACATATGTACACATATATATACAC 357
             |||||||  |  |  ||| | |||| |||| | |||  |   || | |||||| | ||
Db      4760 TATATATATTATTTTCATATATATATTTATATATATATTTTCATATATATATTTTCAT 4701

Qy       358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417
             |||| |||| | ||||  | | ||||||  | | |||| | || |||||||   | ||||
Db      4700 ATATATATATATATATTCATATATATATTTATATATATTTTCATATATATATTTATATAT 4641

Qy       418 ATACACATATATACACATATATACACATATACACATATATACACATATATACACATATATAC 477
             |||      ||||| | |||| | || ||||| | |||||||  ||||||||  || ||||
Db      4640 ATATTTTCATATATATATATTTTCATATATATATATATATATTCATATATATATTTATAT 4581

Qy       478 ACATATATATA---ATATACACACATATATATACACATATATACACACATATATACACAT 534
             | || |  |||   |||||   | ||||||||  |||||||| | |  |||||| | ||
Db      4580 ATATTTTCATATATATATATTTATATATATATTTTCATATATATATATTTATATATATAT 4521

Qy       535 ATATACACATATATATACACATATATACACATATATACATATATACACATATATATACAT 594
             |  | | ||||||  || | ||||||   ||||||||  |||| || | ||||||  | |||
Db      4520 TTTCATATATATATTTATATATATATTTTCATATATATATATTTATATATATATTTTCAT 4461

Qy       595 ATATACACATATATACATATATACACATATATACATATATACACACATATATACACATAC 654
             ||||| | || ||||  ||||||   |||||||| |||| || | |  |||| | || |||
Db      4460 ATATATATATTTATATATATATTTTCATATATATATATTTATATATATATTTTCATATAT 4401

Qy       655 ATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATAT 714
             |||||   | | ||| || | ||||||||||   |||||||||  |  |||||| ||| |
Db      4400 ATATATTTATATATATATTTTCATATATATATTTATATATATATTTTCATATATATATTT 4341

Qy       715 ATATAT 720
             ||||||
Db      4340 ATATAT 4335
```

RESULT 43
ABK69932/c
ID ABK69932 standard; DNA; 19965 BP.
XX
AC ABK69932;
XX
DT 15-JUL-2002 (first entry)
XX
DE Human secreted protein gene 68 genomic DNA fragment #20.
XX
KW Human; ds; secreted protein; gene therapy; immunosuppressive;
KW antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;
KW vasotropic; cerebroprotective; nootropic; neuroprotective; antibacterial;
KW virucide; fungicide; opthalmalogical; autoimmune disease; neoplasm;
KW rheumatoid arthritis; hyperproliferative disorder; cardiac arrest;
KW cardiovascular disorder; cerebrovascular disorder; cerebral ischaemia;
KW angiogenesis; nervous system disorder; Alzheimer's disease; infection;
KW ocular disorder; corneal infection; wound healing; skin aging;
KW epithelial cell proliferation; food additive.

XX
OS Homo sapiens.
XX
PN WO200226931-A2.
XX
PD 04-APR-2002.
XX
PF 24-SEP-2001; 2001WO-US029871.
XX
PR 25-SEP-2000; 2000US-0234925P.
PR 12-JAN-2001; 2001WO-US000911.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Komatsoulis G, Duan DR, Rosen CA, Moore PA, Shi Y;
PI Lafleur DW, Olsen H, Brewer LA, Florence KA, Young PE, Soppet DR;
PI Endress GA, Mucenski M, Ebner R;
XX
DR WPI; 2002-362489/39.
XX
PT Novel 71 isolated secreted polypeptides and polynucleotides encoding the
PT polypeptides, useful for treating Huntington's disease, sepsis,
PT meningitis, thrombocytopenia, hemolytic anemia, rheumatoid arthritis,
PT asthma.
XX
PS Example 2; Page 1434-1440; 1478pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (or its
CC fragment, homologue complement or allelic variant) encoding a human
CC secreted protein (and its fragment, domain, epitope, variant, secreted
CC form and species variant). Also included are a recombinant vector
CC comprising the nucleic acid, a recombinant host cell comprising the
CC vector, an antibody against the secreted protein, a recombinant host cell
CC that expresses the secreted protein and a method of identifying a binding
CC partner of the secreted protein. The nucleic acid and protein are used to
CC prevent, diagnose, treat or ameliorate a medical condition in e.g.
CC humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep for
CC example autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC angiogenesis, nervous system disorders e.g. Alzheimer's disease,
CC infections caused by bacteria, viruses and fungi and ocular disorders
CC e.g. corneal infection. Many other diseases and disorders are listed in
CC the specification. The polypeptides can also be used to aid wound healing
CC an epithelial cell proliferation, to prevent skin aging due to sunburn,
CC to maintain organs before transplantation, for supporting cell culture of
CC primary tissues, to regenerate tissues and in chemotaxis. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities. The present sequence
CC represents a ds DNA fragment of the gene for a novel human secreted
CC protein of the invention
XX
SQ Sequence 19965 BP; 4699 A; 5296 C; 5505 G; 4465 T; 0 U; 0 Other;

Query Match 6.3%; Score 258; DB 6; Length 19965;
Best Local Similarity 64.7%; Pred. No. 1.2e-34;
Matches 431; Conservative 0; Mismatches 230; Indels 5; Gaps 3;

```
Qy        59 TAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGA 118
             | | ||||   ||||   ||||||| ||||||||||   || ||||||||||||||||
Db      4999 TGGCCAGGAGCGGTGGCTCACGCCTATAATCCCAGCACGTTGGGAGGCCGAGGCAGGGAG 4940

Qy       119 ATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCA 178
             |  |  | |  | | |  ||          ||||  |    ||  ||| ||   |
Db      4939 GATCGCTTGAGAT-CAAGAGCTCCAGACCAGCCTAGGCAACACGGTGAAATTCCATCTCT 4881

Qy       179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATA 238
             |  |   || |  |   |||    |  |    | | | | |      | ||||| | ||
Db      4880 ACAAAAAAAATACAAAAAAAAAAAAAAAATATATATATATATATTTATATATATATGTTTA 4821

Qy       239 TATATATATATATACACACACACACATATATATATACACATATATATACACATATATA 298
             |||||||||||| || | |     | | |||| |||||    ||||||||  | ||||||
Db      4820 TATATATATATTTATATATTTTTATATATATTTATATTTTTATATATATTTATATATATT 4761

Qy       299 CATATATA-CACATATATACACATATATATACACATACATATGTACACATATATATACAC 357
             |||||||  |  |  ||| | |||| |||| | |||  |   || | |||||| | ||
Db      4760 TATATATATTATTTTCATATATATATTTATATATATATTTTCATATATATATTTTCAT 4701

Qy       358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417
             |||| |||| | ||||  | | ||||||  | | |||| | || |||||||    | ||||
Db      4700 ATATATATATATATATTCATATATATATTTATATATATTTTCATATATATATTTATATAT 4641

Qy       418 ATACACATATATACACATATATACACATATACACATATATACACATATATACATATATAC 477
             |||     ||||| | |||| | || ||||| | |||||||  |||||||| || ||||
Db      4640 ATATTTTCATATATATATATTTCATATATATATATATATTCATATATATATTTATAT 4581

Qy       478 ACATATATATA---ATATACACACATATATATACACATATATACACACATATATACACAT 534
             | || |  |||   |||||   | ||||||||   |||||||| | |  |||||| | ||
Db      4580 ATATTTTCATATATATATATTTATATATATATTTTCATATATATATATTTATATATATAT 4521

Qy       535 ATATACACATATATATACACATATATACACATATATACATATATACACATATATATACAT 594
             |  | | |||||| || | ||||||   |||||||| |||| || | | |||||| | |||
Db      4520 TTTCATATATATATTTATATATATATTTTCATATATATATATTTATATATATATTTTCAT 4461

Qy       595 ATATACACATATATACATATATACACATATATACATATATACACACATATATACACATAC 654
             ||||| | || |||| ||||||   |||||||| |||| || | | |||| | || |||
Db      4460 ATATATATATTTATATATATATTTTCATATATATATATTTATATATATATTTTCATATAT 4401

Qy       655 ATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATAT 714
             |||||    | | ||| || | ||||||||||   |||||||||  |  |||||| ||| |
Db      4400 ATATATTTATATATATATTTTCATATATATATTTATATATATATTTTCATATATATATTT 4341

Qy       715 ATATAT 720
             ||||||
Db      4340 ATATAT 4335
```

RESULT 44
ABK39930
ID ABK39930 standard; DNA; 12069 BP.
XX
AC ABK39930;
XX
DT 21-MAY-2002 (first entry)
XX
DE Human chemically pretreated gene sequence #6 strand 1.
XX
KW Human; ds; bisulphite treatment; CpG; DNA methylation; cancer; tumour;
KW cytostatic; ALDH6; CYP11A; CYP11B1; CYP3A3; DPYD; EPHX2; OCLN; TXNRD1;
KW UGT8; MRP; pharmacogenomics; SNP; single nucleotide polymorphism.
XX
OS Homo sapiens.
XX
PN WO200202806-A2.
XX
PD 10-JAN-2002.
XX
PF 29-JUN-2001; 2001WO-EP007470.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-154757/20.
XX
PT New nucleic acid, oligonucleotides and peptide nucleic acid-oligomers,
PT useful for detecting cytosine methylation state of genes associated with
PT pharmacogenomics and for therapy of diseases e.g. cancer.
XX
PS Claim 1; SEQ ID NO 11; 24pp; English.
XX
CC The invention relates to a nucleic acid comprising a sequence at least 18
CC bases in length of a segment of the chemically pretreated DNA of genes

CC associated with pharmacogenomics according to one of the sequences of the
CC genes ALDH6 (NM_000693), CYP11A (NM_000781), CYP11B1 (NM_000497), CYP3A3
CC (NM_000776 and NM_017460), DPYD (NM_000110), EPHX2 (NM_001979), OCLN
CC (NM_002538), TXNRD1 (NM_003330), UGT8 (NM_003360), MRP (NM_004996,
CC NM_019900, NM_019901, NM_019902, NM_019862, NM_019898, NM_019899) and
CC their complementary sequences, or a sequence (S1) chosen from 87
CC sequences and their complements. The chemical pretreatment is bisulphite
CC treatment to convert cytosines (but not methyl-cytosines) into uracils.
CC Also included are an oligomer (II) in particular an oligonucleotide or a
CC peptide nucleic acid (PNA)-oligomer, comprising in each case at least one
CC base sequence having a length of 9 nucleotides which hybridises to or is
CC identical to a chemically pretreated DNA of genes associated with
CC pharmacogenomics and their complements, arranged in an array for
CC analysing diseases associated with the methylation state (CpG) and/or
CC detecting SNPs (single nucleotide polymorphisms) of the 87 sequences. The
CC oligomers may also be used as PCR primers. The set of 87 nucleic acids
CC and their complements is useful for diagnosis and therapy of solid
CC tumours and cancer. The present sequence represents one the 87 DNA
CC sequences or its complement. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 12069 BP; 3210 A; 185 C; 2849 G; 5823 T; 0 U; 2 Other;

Query Match 6.3%; Score 257.6; DB 6; Length 12069;
Best Local Similarity 62.9%; Pred. No. 1.3e-34;
Matches 433; Conservative 0; Mismatches 249; Indels 6; Gaps 2;

```
Qy       226 TTTTATATATATATATATATATATATACACACACACACACATATATATATACACATATAT 285
Db      9239 TTATATATATATAGATATTTATATATATATAGATATACGTGGATATATATATTTATATAT 9298

Qy       286 ATACACATATATACATATATACACATATATACACATATATATACACATACATATGTACAC 345
Db      9299 ATATAGATATTTATATATATATAGATATACGTGGATATAGATATTTATATATATATAGAT 9358

Qy       346 ATATATATACACATATGTATACACATATATACACATATATACACACATATATACACATAT 405
Db      9359 ATTTATATATATATATAGATATTTATATATATATAGATATTTATATATATATAGATATAT 9418

Qy       406 ATACACACATATATACACATATATACACATATATACACATATACACATATATACACATAT 465
Db      9419 GTGGATATATATATTTATATATATAGAGATATTTATATATATATAGATATTTATATATAT 9478

Qy       466 ATACATATATACACATATATATA-----ATATACACACATATATATACACATATATACAC 520
Db      9479 ATAGATATACGTGGATATATATATTTATATATATATAGATATTTATATATATATAGATAT 9538

Qy       521 ACATATATACACATATATACACATATATATACACATATATACACATATATACATATATAC 580
Db      9539 ACGTGGATATATATATTTATACGTAGATATTTATATATATATAGATATACGTGGATATAT 9598

Qy       581 ACATATATATACATATATACACATATATACATATATACACATATATACATATATACACAC 640
Db      9599 ATATTTATACGTAGATATTTATATATATATAGATATACGTGGATATATATATTTATATAT 9658

Qy       641 ATATATACACATACATATACACACACATAGATATACATATATATACACATATATATACGT 700
Db      9659 ATATAGATATTTATATATATATAGATATACGTGGATATATATATTTATATATATATAGAT 9718

Qy       701 ATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGG 760
Db      9719 ATTTATATATATATAGATATACGTGGATATAGATATTTATATATATATAGATATTTATAC 9778

Qy       761 GGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATG 820
Db      9779 GTATATAGATATTTATATATATATAGATATTTATATATATATAGATATATTTATATATAT 9838

Qy       821 GCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAA 880
Db      9839 ATAGATATATATGGATATATATATTTATAT-ATATATAATATTTATATATAGATATATAT 9897
```

Qy 881 GGTAACACTATATATATATATAGTGAAA 908
Db 9898 GGATATAGTATTTATATATATATAGATA 9925

RESULT 45
ADC87552
ID ADC87552 standard; DNA; 1486 BP.
XX
AC ADC87552;
XX
DT 01-JAN-2004 (first entry)
XX
DE Human GPCR gene SEQ ID NO:2005.
XX
KW ds; gene; human; GPCR;
KW guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1270724-A2.
XX
PD 02-JAN-2003.
XX
PF 18-JUN-2002; 2002EP-00013517.
XX
PR 18-JUN-2001; 2001JP-00246789.
XX
PA (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI Suwa M, Asai K, Akiyama Y, Aburatani H;
XX
DR WPI; 2003-315783/31.
DR P-PSDB; ADC87553.
XX
PT New polynucleotide, useful for preparing a composition for treating a
PT patient in need of increased or suppressed activity or expression of the
PT guanosine triphosphate-binding protein coupled receptor.
XX
PS Claim 1; SEQ ID NO 2005; 28pp; English.
XX
CC The invention relates to a novel polynucleotide encoding a guanosine
CC triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC the invention may have a use in gene therapy. The polynucleotide and
CC polypeptide are useful for preparing a composition for treating a patient
CC in need of increased or suppressed activity or expression of the
CC guanosine triphosphate-binding protein coupled receptor. The
CC polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC invention.
XX
SQ Sequence 1486 BP; 517 A; 149 C; 242 G; 578 T; 0 U; 0 Other;

Query Match 6.2%; Score 256.4; DB 10; Length 1486;
Best Local Similarity 70.2%; Pred. No. 2e-34;
Matches 344; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

```
Qy       227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
Db       718 TGTATATATACATGTATGTGTATATATACATATGTGTATGTGTATATATACATATGTGTA 777

Qy       287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
Db       778 TGTGTATATATACATATGTGTATGTGTATATATACATATGTGTATGTGTATATATACATA 837

Qy       347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
Db       838 TGTGTATATATACATATGTGTATATATGTATATATGTATATATACATATGTGTATATATG 897

Qy       407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
Db       898 TATATATGTATATATACATATATGTGTGTATATATGTATATATGTGTATTTATATATGTG 957
```

```
Qy        467 TACATATATACACATATATATAATATACACACATATATATACACATATATACACACATAT 526
              |  | | || | |||| ||||   || |    | |||||| ||||||||  |   ||||
Db        958 CATGTGTGTATAAATATGTATATATTATATGTGTGTATATATACATATATGTATGTATAT 1017

Qy        527 ATACACATATATACACATATATATACACATATATACACATATATACATATATACACACATAT 586
              ||||| ||||| | |||||||||| | | ||| | ||| |||   |||||||| ||| ||||
Db       1018 ATACATATATACATACATATATACATATATACACACATATACGCACATATAGACATATAT 1077

Qy        587 ATATACATATATACACATATATACATATATACACATATATACATATATACACACATATAT 646
              || || ||||| ||||||||||||||||||   |||||||| | ||| | ||||||||
Db       1078 ATGTATATATACACACATATATACATATATATGTATATATACACACATATATACATATAT 1137

Qy        647 ACACATACATATATACACACACATAGATATACATATATATACACATATATATACGTATATAT 706
              |   ||| ||| ||||| | ||| |||||  ||||||||||||||||||   ||||||
Db       1138 ATGTATATATACACACATATATACATATATGTATATATACACATATATATGTATATATAC 1197

Qy        707 ATGTATATAT 716
              |  |||||||
Db       1198 ACATATATAT 1207
```

RESULT 46
AAQ39248/c
ID AAQ39248 standard; DNA; 870 BP.
XX
AC AAQ39248;
XX
DT 25-MAR-2003 (revised)
DT 10-AUG-1993 (first entry)
XX
DE Sequence of cosmid clone 4F7 isolated from a chromosome 11q-specific
DE cosmid library genomic polymorphic locus (GPL).
XX
KW Genomic polymorphic locus; restriction site; chromosome 11 q13; ss.
XX
OS Synthetic.
XX
PN WO9307166-A1.
XX
PD 15-APR-1993.
XX
PF 07-OCT-1992; 92WO-US008539.
XX
PR 09-OCT-1991; 91US-00773099.
XX
PA (SALK) SALK INST BIOLOGICAL STUDIES.
XX
PI Evans GA, Selleri L, Eubanks JH;
XX
DR WPI; 1993-134379/16.
XX
PT Oligo-nucleotide(s) useful as primers - for identifying individuals on
PT basis of presence of polymorphic locus, allows specific detection.
XX
PS Disclosure; Page 33-34; 51pp; English.
XX
CC Two cosmids, 4F7 and 7H12, were identified by screening cosmid libraries
CC with a sequence from the 2nd transmembrane domain of a rat D2 dopamine
CC recptor cDNA. The cosmids had potential structural homology to G-protein
CC receptor gene family as well as several loci on chromosome 11. A highly
CC repetitive element was also detected in the region of of overlap of the 2
CC cosmids and was subcloned into Bluescript and its DNA sequence
CC determined. It revealed a highly degenerate repetative hexamer motif
CC spanning 438 bp consisting of 73 repeats. (Updated on 25-MAR-2003 to
CC correct PN field.)
XX
SQ Sequence 870 BP; 185 A; 127 C; 247 G; 311 T; 0 U; 0 Other;

Query Match 6.2%; Score 256; DB 2; Length 870;
Best Local Similarity 73.9%; Pred. No. 2.2e-34;
Matches 340; Conservative 0; Mismatches 115; Indels 5; Gaps 1;

```
Qy        238 ATATATATATATATACACACACACACATATATATATACACATATATATACACATATAT 297
              ||| ||| | ||||||| ||||| ||||||||| ||| ||| ||| ||||||| | | ||
Db        810 ATACATACACATATACATACACATACACATATACATACACATATACATATACATACACAT 751

Qy        298 ACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACAC 357
              | | ||||||| ||| | | ||||||| ||| ||||||| || |||| ||| ||||||||
Db        750 ATACATATACATATACACATACATATACATATACATACACATATACATATACATATACAC 691

Qy        358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417
              |||  |||||| ||| |||||||||||| | | ||||||| ||| | ||||||| | | ||
Db        690 ATACATATACATATACATACACATATACATATACATATACACATACATATACATATACAT 631

Qy        418 ATACACATATATACACATATATACACATATACACATATATACACATATATACATATATAC 477
              ||||| | | ||| ||| ||| ||| | ||| ||||||| | ||| ||| ||||||| |
Db        630 ATACATACACATATACACATACACATACATATACATATACATACACATACACATATACAT 571

Qy        478 ACATATATATAATATACACACATATATATACACATATATACACACATATATACACATATA 537
              | | |      ||| | | ||||||| |||||||||| | | ||||| | | ||| |||
Db        570 ATACA-----CATACATATACATATACATACACATATACATATACATACACATACACATA 516

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATACATATA 597
              |||| | | ||| ||| | | ||||||||| | ||| | |||||| || |  ||||  ||
Db        515 TACATACACATACACATACACATACACATACACACACACATACACGTACACGTACACGTA 456

Qy        598 TACACATATATACATATATACACATATATACATATATACACACATATATACACATACATA 657
              ||  | |  ||||  |||||| ||| ||| | ||||||| | | |||||||| |||||||
Db        455 CACGTACACGTACACGTATACATATACATATACATATACATATACATATACATATACATA 396

Qy        658 TACACACACATAGATATACATATATATACACATATATATA 697
              |||| | ||||| | | | | | ||| |    | |   |||
Db        395 TACATATACATACACACATACACATACAAGAAGAAGGATA 356
```

RESULT 47
ADC87554/c
ID ADC87554 standard; DNA; 1489 BP.
XX
AC ADC87554;
XX
DT 01-JAN-2004 (first entry)
XX
DE Human GPCR gene SEQ ID NO:2007.
XX
KW ds; gene; human; GPCR;
KW guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1270724-A2.
XX
PD 02-JAN-2003.
XX
PF 18-JUN-2002; 2002EP-00013517.
XX
PR 18-JUN-2001; 2001JP-00246789.
XX
PA (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI Suwa M, Asai K, Akiyama Y, Aburatani H;
XX
DR WPI; 2003-315783/31.
DR P-PSDB; ADC87555.
XX
PT New polynucleotide, useful for preparing a composition for treating a
PT patient in need of increased or suppressed activity or expression of the
PT guanosine triphosphate-binding protein coupled receptor.
XX
PS Claim 1; SEQ ID NO 2007; 28pp; English.
XX
CC The invention relates to a novel polynucleotide encoding a guanosine
CC triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC the invention may have a use in gene therapy. The polynucleotide and
CC polypeptide are useful for preparing a composition for treating a patient

```
CC   in need of increased or suppressed activity or expression of the
CC   guanosine triphosphate-binding protein coupled receptor. The
CC   polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC   invention.
XX
SQ   Sequence 1489 BP; 553 A; 208 C; 160 G; 568 T; 0 U; 0 Other;

  Query Match             6.2%;  Score 255.4;  DB 10;  Length 1489;
  Best Local Similarity   63.8%;  Pred. No. 2.9e-34;
  Matches  438;  Conservative    0;  Mismatches  241;  Indels    8;  Gaps    3;

Qy        236 ATATATATATATATATACACACACACACATATATATATACACATATATATACACATAT 295
              | || | |||||||||||| | | ||  | |||| | |||| | | ||| |||   ||||
Db       1197 ACATGTGTATATATATACATATATACGTATATATGTGTATATATACATACATAGGTATAT 1138

Qy        296 ATACATATATACACATATATACACATATATATACACATACATATGTACACATATATATAC 355
              | | ||||| ||| | ||||| || |||||||  | |||  ||| || |||| ||||||
Db       1137 ACATATATACACATACATATATACGTATATATGTATATATGTATATATACATGTATATAT 1078

Qy        356 ACATATG-TATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              |||||  | || ||||||||||||| ||||||  ||  |||| || ||||||||||||   |
Db       1077 GTATATGTTGTATACATATATACACGTATATATACGTATATGTATACATATATACACGTA 1018

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              |||||||  |||| || |||||||||||| |||| | |   | || | | || || |||||
Db       1017 TATATACGTATATGTATACATATATACACGTATATATACGTGTATATACATGTATATATA 958

Qy        475 TACACATATATATAATATACACACATATATATACACATATATACACACATATATACACAT 534
              | || |     ||| | | ||||||||| ||   |||||||| |  |||||||| |
Db        957 CGTATATGT-----ATACATATACATATATACACGTGTATATACATATGTATATACACGT 903

Qy        535 ATATACACATATATATACACATATATACACATATATACATATATACACATATATATACAT 594
              |||| |  ||| |||| |||  | || | || ||||| |||||| ||| | |||| ||
Db        902 GTATATATGTATGTATATACACGTGTATATATGTATACGTATATATACACGTGTATATAT 843

Qy        595 ATATACACATATATACATATATACACATATATACATATATACACACATATATACACATAC 654
              |||||  |||||||||  | || | || ||||| |||||| |||| | |||| |  ||
Db        842 GTATACGTATATATACACGTGTATATATGTATACGTATATATACACGTGTATATATGTAT 783

Qy        655 ATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATAT 714
              |  || | | |||    |||| |||| |||||  ||||||| |||| ||||||||||||
Db        782 ACGTATATATACACGTGTATATATATGTATACGTATATATACACGTGTATATATGTATAC 723

Qy        715 ATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAG 774
              |||||   | | ||  ||| |       | |   |      |     || | | | |
Db        722 GTATATATACACGTGTATATATGTATGTATATATACGTGTATATATATGTATACGTATAT 663

Qy        775 AGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGG 834
              | |  | | | | |||||     |      | | |      |       | |||   |
Db        662 ATACGTATATATAAATGTAT--ATATGTATATGTGTATATATACGTATATATACATGTAT 605

Qy        835 AAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATA 894
              | |||| || || |  |   | || ||| ||  | |||  ||   |  |     |||||
Db        604 ATATACGTATATATACACATATACATATATGGTACGTATATATACATATGTGTATATATG 545

Qy        895 TATATATAGTGAAAGCAGTGTAAACCT 921
              ||||||||   | |    | || || |
Db        544 TATATATATACATACATATATATACAT 518


RESULT 48
ABD33148/c
ID   ABD33148 standard; DNA; 104399 BP.
XX
AC   ABD33148;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Murine cancer-associated (CA) gene MD07-019.
XX
KW   Mouse; cancer-associated protein; CAP; cancer-associated gene; CA; gene;
KW   ds; cancer; cytostatic.
XX
OS   Mus musculus.
XX
PN   WO2004058146-A2.
XX
PD   15-JUL-2004.
XX
PF   15-DEC-2003; 2003WO-US040081.
XX
PR   17-DEC-2002; 2002US-00322281.
XX
PA   (SAGR-) SAGRES DISCOVERY INC.
XX
PI   Morris DW,  Malandro MS;
XX
DR   WPI; 2004-499109/47.
XX
PT   Novel human cancer associated protein encoded within open reading frame
PT   of cancer associated gene, useful as targets for diagnosing cancer.
XX
PS   Disclosure; SEQ ID NO 123; 182pp; English.
XX
CC   The invention relates to cancer-associated proteins (CAP) and the cancer-
CC   associated (CA) nucleic acids encoding them. The invention also relates
CC   to a method for treating cancers involving administering to a patient an
CC   inhibitor of CAP, and a method of screening for anticancer activity in a
CC   potential drug involving providing a cell that expresses a CA gene,
CC   contacting a tissue sample derived from a cancer cell with an anticancer
CC   drug candidate and monitoring the effect of the anticancer drug candidate
CC   on expression of the CA gene. The CAP proteins are useful for detecting
CC   cancer associated with expression of a CAP protein in a test cell sample
CC   and for screening for a bioactive agent capable of modulating the
CC   activity of a CAP protein. The CA nucleic acids are useful for diagnosing
CC   cancer, involving determining the expression of a CA nucleic acid in a
CC   tissue. This sequence represents a murine CA gene of the invention. Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 104399 BP; 25540 A; 20083 C; 20918 G; 29473 T; 0 U; 8385 Other;

  Query Match             6.2%;  Score 255.4;  DB 13;  Length 104399;
  Best Local Similarity   73.3%;  Pred. No. 3.4e-34;
  Matches  367;  Conservative    0;  Mismatches  131;  Indels    3;  Gaps    3;

Qy        230 ATATATATATATATATATATACACACACACACATATATATATACACATATATATAC 289
              ||| ||| ||| ||||||| ||||  ||| | | ||||| ||| ||||  | |||||||||
Db      13721 ATACATACATACATATATACATACGTACATATATACATACATACATACGTACATATATAC 13662

Qy        290 ACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATAT 349
              | |  || | ||||| | ||||| ||||| ||||| ||||  | | ||||  | |||||
Db      13661 ATACGTACATATATACATACATACATACATATATACATACGTACATATATACATACATAC 13602

Qy        350 ATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATAC 409
              |||   |||||| || | || || ||| | | ||| | ||| | ||||| |||||  |||
Db      13601 ATACGTACATATATACATACGTACATATATACATACATACATACATATATACATACGTAC 13542

Qy        410 ACACATATATACACATATATACACATATATACACATATACACATATATACACATATATAC 469
              | | ||| ||||| | |||    |||||||||||| | ||||| | ||||| ||||| ||||
Db      13541 ATATATACATACATACATACGTACATATATACATACATACATACATATATACATACATAC 13482

Qy        470 ATATATACACATATATATAATATACACACATATATATACACATATATACACACATATATA 529
              |||||||| | ||| |||||| | | | | ||  ||| | | ||| | ||| ||||| |
Db      13481 ATATATACATACATACATAATACATATATACATTCATATATACATA-ATACATATATACA 13423

Qy        530 CACATATATACACATATATAT-ACACATATATACACATAT-ATACATATATACACATATA 587
              ||||||||||| | ||| || | |||||||||||| | || |||||| || || |||||
Db      13422 TACATATATACATACATACATAATACATATATACATACATAATACATACATGCATATATA 13363

Qy        588 TATACATATATACACATATATACATATATACACATATATACATATATACACACATATATA 647
              || |||| ||| | | ||| ||| | ||||| | ||| | | ||||| ||| |   || |
Db      13362 CATGCATACATATATACATACACACACATACATACATACATACATATATACATACGTACA 13303

Qy        648 CACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATA 707
```

```
Db      13302 TATATACATACATACATACGTACATATATACATACATACATATATACATACATACATATA 13243

Qy        708 TGTATATATATATGCTCCAGA 728
              |  ||| ||| ||  | || |
Db      13242 TACATACATACATAATACATA 13222


RESULT 49
ADC86154/c
ID   ADC86154 standard; DNA; 84510 BP.
XX
AC   ADC86154;
XX
DT   01-JAN-2004  (first entry)
XX
DE   Human GPCR gene SEQ ID NO:607.
XX
KW   ds; gene; human; GPCR;
KW   guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1270724-A2.
XX
PD   02-JAN-2003.
XX
PF   18-JUN-2002; 2002EP-00013517.
XX
PR   18-JUN-2001; 2001JP-00246789.
XX
PA   (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA   (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI   Suwa M,  Asai K,  Akiyama Y,  Aburatani H;
XX
DR   WPI; 2003-315783/31.
DR   P-PSDB; ADC86155.
XX
PT   New polynucleotide, useful for preparing a composition for treating a
PT   patient in need of increased or suppressed activity or expression of the
PT   guanosine triphosphate-binding protein coupled receptor.
XX
PS   Claim 1; SEQ ID NO 607; 28pp; English.
XX
CC   The invention relates to a novel polynucleotide encoding a guanosine
CC   triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC   the invention may have a use in gene therapy. The polynucleotide and
CC   polypeptide are useful for preparing a composition for treating a patient
CC   in need of increased or suppressed activity or expression of the
CC   guanosine triphosphate-binding protein coupled receptor. The
CC   polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC   invention.
XX
SQ   Sequence 84510 BP; 21253 A; 19989 C; 19504 G; 23764 T; 0 U; 0 Other;

  Query Match             6.2%;  Score 255;  DB 10;  Length 84510;
  Best Local Similarity   72.6%;  Pred. No. 4e-34;
  Matches  358;  Conservative    0;  Mismatches  130;  Indels    5;  Gaps    2;

Qy        232 ATATATATATATATATATATACACACACACACACATATATATATACACATATATATACAC 291
              ||||||||| ||||   |||| | | | ||||| || ||||||||||  ||  ||||| ||
Db      80292 ATATATATACATATTCTTATATATATATACACATATGTATATATACATATGAATATATAC 80233

Qy        292 ATATATACATATATACACATATATACACATATATATACACATACATATGTACACATATAT 351
              | ||||  ||||||||| ||  ||| | | | ||||||| ||  |||| ||||||||||
Db      80232 ACATATGTATATATACATATGAATATATACACATATACATATGAATATATACACATATAC 80173

Qy        352 ATACACATATGTATACACATATATACACATATATACACACATATATACACATATATACAC 411
              |||   |||| || | | ||| |||   |||||||||| |||||   |   ||||| ||
Db      80172 ATATGAATATATACATATATACATATGAATATATACACATATATACGTATGAATATATAC 80113

Qy        412 ACATATATAC----ACATATATACACATATATACACATATACACATATATACACATATAT 467
              |||||||||      ||||||| ||| ||||| || |||  | ||||||||||||||||
Db      80112 ACATATATACGTATGAATATATATACACATATATACGTATGAATATATATACACATATAT 80053

Qy        468 ACATATATACACATATATATAATATACACACATATATATACACATATATACACACATATA 527
              || |||  | | ||||| |  ||||||  |   |||||||  ||  ||||||  |   |||
Db      80052 ACGTATGAATATATATACA-CATATACGTATGAATATATATACCCATATACGTATGAATA 79994

Qy        528 TACACATATATACACATATATATACACATATATACACATATATACATATATACACATATA 587
              |||||||||||| || ||  |||| | ||| | | | |||  ||| ||||||  | | | |
Db      79993 TACACATATATATACGTACGTATATATATACACATATATACGTACGTATATATATACACA 79934

Qy        588 TATACATATATACACATATATACATATATACACATATATACATATATACACACATATATA 647
              |||| |  ||   | ||||||||| ||||| || ||   | ||||||||||| |||||
Db      79933 TATATACGTACGTATATATATACACATATATACGTACGAATATATATACACATATATACG 79874

Qy        648 CACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATA 707
              ||  | ||||| ||||| ||| |  |||  ||||||| ||| ||||||||||   |||
Db      79873 TACGAATATATATACACATATATACGTACGAATATATATACACATATATACGTACGAATA 79814

Qy        708 TGTATATATATAT 720
              | |||| | ||||
Db      79813 TATATACACATAT 79801


RESULT 50
ADL13771/c
ID   ADL13771 standard; DNA; 201766 BP.
XX
AC   ADL13771;
XX
DT   06-MAY-2004  (first entry)
XX
DE   Osteoarthritis-associated polymorphic nucleotide #303.
XX
KW   ds; gene; osteopathic; antiinflammatory; antiarthritic; gene therapy;
KW   joint space narrowing; osteophyte development; joint pain;
KW   osteoarthritis; SNP; single nucleotide polymorphism.
XX
OS   Homo sapiens.
XX
PN   WO2003054166-A2.
XX
PD   03-JUL-2003.
XX
PF   19-DEC-2002; 2002WO-US041225.
XX
PR   20-DEC-2001; 2001US-0342603P.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Jones KA,  Schafer A;
XX
DR   WPI; 2003-559141/52.
XX
PT   Determining susceptibility of an individual to joint space narrowing,
PT   osteophyte development and/or joint pain comprises identifying whether
PT   the individual has at least one polymorphism in a polynucleotide encoding
PT   a protein.
XX
PS   Disclosure; SEQ ID NO 303; 297pp; English.
XX
CC   The invention relates to a method of determining susceptibility of an
CC   individual to joint space narrowing and/or osteophyte development and/or
CC   joint pain comprising identifying whether the individual has at least one
CC   polymorphism in a polynucleotide encoding at least one of the protein
CC   listed in the specification. The methods, composition and agent are
CC   useful for modulating the susceptibility of an individual to joint space
CC   narrowing and/or osteophyte development and/or joint pain that is
CC   associated with a disease, preferably osteoarthritis. The cell line and
CC   the non-human animal are useful for screening for an agent for diagnosing
CC   an individual having susceptibility to joint space narrowing and/or
CC   osteophyte development and/or joint pain. This sequence corresponds to
CC   the polynucleotide encoding a protein listed in the specification. (Note:
```

```
CC   The sequence data for this patent did not form part of the printed
CC   specification but was obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences).
XX
SQ   Sequence 201766 BP; 57593 A; 35575 C; 36821 G; 54688 T; 0 U; 17089 Other;

  Query Match             6.2%;  Score 255;  DB 10;  Length 201766;
  Best Local Similarity   63.6%;  Pred. No. 4.1e-34;
  Matches  437;  Conservative    0;  Mismatches  242;  Indels    8;  Gaps    3;

Qy        236 ATATATATATATATATACACACACACACACATATATATATACACATATATATACACATAT 295
              | || | ||||||||||| | | ||  | |||| | |||| | | ||| |||   ||||
Db     189316 ACATGTGTATATATATACATATATACGTATATATGTGTATATATACATACATAGGTATAT 189257

Qy        296 ATACATATATACACATATATACACATATATATACACATACATATGTACACATATATATAC 355
              | | ||||| ||| | ||||| || |||||||  | |||  ||| || |||| ||||||
Db     189256 ACATATATACACATACATATATACGTATATATGTATATATGTATATATACATGTATATAT 189197

Qy        356 ACATATGTA-TACACATATATACACATATATACACACATATATACACATATATACACACA 414
              |||||   || |||||||||||| |||||| ||  |||| || ||||||||||||  |
Db     189196 GTATATGNNGTATACATATATACACGTATATATACGTATATGTATACATATATACACGTA 189137

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              |||||||  |||| || |||||||||||| |||| | |  | || | | || || |||||
Db     189136 TATATACGTATATGTATACATATATACACGTATATATACGTGTATATACATGTATATATA 189077

Qy        475 TACACATATATATAATATACACACATATATATACACATATATACACACATATATACACAT 534
              | || |     ||| | | ||||||||| ||   |||||||| |  |||||||| |
Db     189076 CGTATATGT-----ATACATATACATATATACACGTGTATATACATATGTATATACACGT 189022

Qy        535 ATATACACATATATATACACATATATACACATATATACATATATACACATATATATACAT 594
              |||| |  ||| |||| |||  | || | || ||||| |||||| |||  | |||| ||
Db     189021 GTATATATGTATGTATATACACGTGTATATATGTATACGTATATATACACGTGTATATAT 188962

Qy        595 ATATACACATATATACATATATACACATATATACATATATACACACATATATACACATAC 654
              |||||  |||||||||  | || | || |||||  |||||| |||| | |||| |  ||
Db     188961 GTATACGTATATATACACGTGTATATATGTATACGTATATATACACGTGTATATATGTAT 188902

Qy        655 ATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATAT 714
              |  ||  | | |||    |||| |||| |||||  ||||||| |||| ||||||||||||
Db     188901 ACGTATATATACACGTGTATATATATGTATACGTATATATACACGTGTATATATGTATAC 188842

Qy        715 ATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAG 774
              |||||   | | ||  ||| |         |  |      |      || | | | |
Db     188841 GTATATATACACGTGTATATATGTATGTATATATACGTGTATATATATGTATACGTATAT 188782

Qy        775 AGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGG 834
              | |  |  | |  | |||||    |     |  | |     |      | |||   |
Db     188781 ATACGTATATATAAATGTAT--ATATGTATATGTGTATATATACGTATATATACATGTAT 188724

Qy        835 AAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATA 894
              | |||| || || |   |  | || ||| ||  | |||  ||   |  |     |||||
Db     188723 ATATACGTATATATACACATATACATATATGGTACGTATATATACATATGTGTATATATG 188664

Qy        895 TATATATAGTGAAAGCAGTGTAAACCT 921
              ||||||||  | |    | || || |
Db     188663 TATATATATACATACATATATATACAT 188637


RESULT 51
ADC87564
ID   ADC87564 standard; DNA; 1114 BP.
XX
AC   ADC87564;
XX
DT   01-JAN-2004  (first entry)
XX
DE   Human GPCR gene SEQ ID NO:2017.
XX
KW   ds; gene; human; GPCR;
KW   guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1270724-A2.
XX
PD   02-JAN-2003.
XX
PF   18-JUN-2002; 2002EP-00013517.
XX
PR   18-JUN-2001; 2001JP-00246789.
XX
PA   (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA   (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI   Suwa M,  Asai K,  Akiyama Y,  Aburatani H;
XX
DR   WPI; 2003-315783/31.
DR   P-PSDB; ADC87565.
XX
PT   New polynucleotide, useful for preparing a composition for treating a
PT   patient in need of increased or suppressed activity or expression of the
PT   guanosine triphosphate-binding protein coupled receptor.
XX
PS   Claim 1; SEQ ID NO 2017; 28pp; English.
XX
CC   The invention relates to a novel polynucleotide encoding a guanosine
CC   triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC   the invention may have a use in gene therapy. The polynucleotide and
CC   polypeptide are useful for preparing a composition for treating a patient
CC   in need of increased or suppressed activity or expression of the
CC   guanosine triphosphate-binding protein coupled receptor. The
CC   polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC   invention.
XX
SQ   Sequence 1114 BP; 398 A; 107 C; 167 G; 442 T; 0 U; 0 Other;

  Query Match             6.2%;  Score 254.2;  DB 10;  Length 1114;
  Best Local Similarity   64.0%;  Pred. No. 4.6e-34;
  Matches  399;  Conservative    0;  Mismatches  223;  Indels    1;  Gaps    1;

Qy        229 TATATATATATATATATATATATACACACACACACATATATATATACACATATATATA 288
              ||||| ||||| | ||| |||||||| | | | ||| |||| |||||     || ||||||
Db        212 TATATGTATATTTGTATGTATATACAAATATATACATATATGTATATGTGTATGTATATA 271

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              || ||||||||||||||    || |||| ||| |||||||| |||  | | |  | |||
Db        272 CAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATATA 331

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              | ||| | | ||  |||    || |||| ||| |||||  ||| ||||  |  ||||||
Db        332 CAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATATA 391

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              || | |||||  | | || | ||    |||||  | | ||||||| ||||  |   ||||||
Db        392 CAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATATA 451

Qy        469 CATATATACACATATATATAATATACACACATATATATACACATATATACACACATATAT 528
              || ||||| |||||||| |  |||  | | | | ||||| ||||||||    |  |||||
Db        452 CAAATATATACATATATGT-GTATGTATATACAAATATATACATATATGTGTATGTATAT 510

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACATATAT 588
              ||| |||||||| |||| | ||   ||||| | | ||||| | |||| |  |  |||||
Db        511 ACAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATAT 570

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
              | | |||||||||  |||| |  || |||| ||| ||||| | |||||     || ||||
Db        571 ACAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATAT 630

Qy        649 ACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATAT 708
              ||| | |||||| | |  |  || || ||| | ||| | | ||||||  |||| |||||
Db        631 ACAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATAT 690

Qy        709 GTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAG 768
                | |||||||       | ||   ||      |  |  |   ||    ||    || |
Db        691 ACAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATAT 750
```

```
Qy        769 GAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAG 828
              ||| | |     | |   || | |    | ||  | ||       ||       |||
Db        751 ACAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATAT 810

Qy        829 TAGGGGAAATACATAAATGTGTA 851
                  | ||||||| |||||||
Db        811 ACAAATATATACATATATGTGTA 833


RESULT 52
ACC84012
ID   ACC84012 standard; DNA; 2200 BP.
XX
AC   ACC84012;
XX
DT   22-SEP-2003  (first entry)
XX
DE   Pacific white shrimp skeletal muscle actin gene promoter.
XX
KW   Shrimp; actin; promoter; transgenic; disease resistance;
KW   stress tolerance; cold tolerance; ds.
XX
OS   Litopenaeus vannamei.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    878. .893
FT                   /*tag=  a
FT                   /note= "CATA-rich region"
FT   misc_feature    1071. .1078
FT                   /*tag=  a
FT                   /note= "CATA-rich region"
FT   misc_feature    1538. .1549
FT                   /*tag=  a
FT                   /note= "CACA-rich region"
FT   misc_feature    1554. .1567
FT                   /*tag=  a
FT                   /note= "CACA-rich region"
FT   misc_feature    1570. .1585
FT                   /*tag=  a
FT                   /note= "CACA-rich region"
XX
PN   WO2003048325-A2.
XX
PD   12-JUN-2003.
XX
PF   04-DEC-2002; 2002WO-US038523.
XX
PR   04-DEC-2001; 2001US-0336603P.
XX
PA   (UYHA-) UNIV HAWAII.
XX
PI   Sun PS,  Arakaki KL,  Sun SSM;
XX
DR   WPI; 2003-513755/48.
XX
PT   New isolated Beta-lactin nucleic acid promoter molecule from shrimp,
PT   useful for imparting resistance against a virus and increasing stress
PT   tolerance in an animal and for regulating growth of an animal.
XX
PS   Claim 52; Page 14-15; 97pp; English.
XX
CC   The present sequence is that of the promoter region of the Pacific white
CC   shrimp skeletal muscle actin gene. This promoter is suitable for inducing
CC   expression of a protein encoded by a nucleic acid molecular operably
CC   associated with the promoter molecule in an expression vector. It
CC   contains promoter-associated TATA and CAAT boxes approximately 500 bp
CC   upstream from the translation start site. A unique characteristic of the
CC   promoter are CACA-rich and CATA-rich regions located upstream of the TATA
CC   and CAAT boxes. The invention relates to actin and beta-actin promoter
CC   molecules from shrimp, expression cassettes including the promoter,
CC   expression vectors, host cells and transgenic animals transduced with the
CC   promoter. Methods for imparting resistance to a pathogen, for regulating
CC   growth, for improving stress tolerance, and for increasing cold tolerance
CC   in an animal, especially a shrimp, involve transforming the animal with a
CC   nucleic acid construct including the isolated actin and beta-actin
CC   promoter molecules of the invention
XX
SQ   Sequence 2200 BP; 837 A; 372 C; 263 G; 728 T; 0 U; 0 Other;

  Query Match             6.2%;  Score 254;  DB 9;  Length 2200;
  Best Local Similarity   71.7%;  Pred. No. 5.1e-34;
  Matches  360;  Conservative    0;  Mismatches  140;  Indels    2;  Gaps    2;

Qy        234 ATATATATATATATATATACACACACACACATATATATATACACATATATATACACAT 293
              | || ||||||||||||||    | |    | | |||| ||| || | | ||| ||| |||
Db        647 ACATTTATATATATATATATGTATATGTATATATATGTATTTATATACATAAATATACAC 706

Qy        294 ATATACATATATACACATATATACACATATATATACACATACATATGTACACATATATAT 353
              |||| ||| |||| | |||| || | ||| ||||| |  || |||| | || ||||||||
Db        707 ATATTCATGTATATATATATGTATATATACATATATATTTATATATATGCATATATATAT 766

Qy        354 ACACATATGTATACACATATATACACATATATACACACATATATACACATATATACACAC 413
              | |||| |||| | | |||| || ||  ||| | |  |||||| | | |||||||
Db        767 GTATATATATATATATACATATGCATATTCATATATATGTATATATATACATATACATAT 826

Qy        414 ATATATACACATATATACACATATATACACATATACACATATATACACATATATACATAT 473
              |||| || | ||||  | | ||||||  | ||  | | ||||||| | ||| |||||||
Db        827 ATATGTATATATATGCATAAATATATGTATATGAATATATATATATATATACATACATAC 886

Qy        474 ATACACATATATATAATATACACACATATATATACACATATATACACACATATATACACA 533
              ||||| |||||||| ||||| | | ||||||||| | ||||||| | | ||| | |||||
Db        887 ATACATATATATAT-ATATATATATATATATATATATATATATATATATACACACACA 945

Qy        534 TATATACACATATATATACACATATATACACATATATACATATATACACATATATATACA 593
              | || |||  | |||||||| | | | ||| | | |||||||||    | ||  |||||
Db        946 CACATGCACGCACATATACACACACAAACACACACACATATATATGTAAATGCATACA 1005

Qy        594 TATATACACATATATACATATATACACATATATACATATATACACACATATATACACATA 653
              ||| |  |||   | |||||||||||||||||||||| | ||||| | | ||||| ||
Db       1006 TATGTGTACACGGACACATATATACACATATATACATACACACACA-ACAGATACATATG 1064

Qy        654 CATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATA 713
              |||||||| ||| | | |||||  | | | || | | ||||||| |||||||| ||||||
Db       1065 CATATACATACATACACATATATCTCTCTCTATATACATATATATGTATATATGTGTATA 1124

Qy        714 TATATATGCTCCAGAGTTCATA 735
              ||  |||  |   | |   || |
Db       1125 TACTTATACATAAAAACACACA 1146


RESULT 53
AAV83939/c
ID   AAV83939 standard; DNA; 80595 BP.
XX
AC   AAV83939;
XX
DT   03-MAR-1999  (first entry)
XX
DE   HC-contig derived from normal human chromosome 10q25.2 region.
XX
KW   Yeast artificial chromosome; YAC; probe; eukaryotic chromosome;
KW   neocentromere; replication; extra-chromosomal element; segregation;
KW   cell division; artificial chromosome; gene therapy; mardel(10);
KW   human artificial chromosome; transgenic; chromosome 10; 10q25.2; ss.
XX
OS   Homo sapiens.
XX
PN   WO9851790-A1.
XX
PD   19-NOV-1998.
XX
PF   13-MAY-1998;   98WO-AU000352.
XX
PR   13-MAY-1997;   97AU-00006784.
PR   26-AUG-1997;   97AU-00008791.
```

```
XX
PA   (AMRA-) AMRAD OPERATIONS PTY LTD.
XX
PI   Choo K,  Du Sart D,  Cancilla MR;
XX
DR   WPI; 1999-009773/01.
XX
PT   New isolated nucleic acid comprising neocentromere sequences from
PT   eukaryotic chromosome - used to produce replicable, segregating
PT   artificial chromosomes that can carry large amounts of DNA for gene
PT   therapy.
XX
PS   Claim 8; Fig 6; 540pp; English.
XX
CC   The present sequence represents the HC-contig derived from normal human
CC   chromosome 10, 10q25.2 region. This region can be naturally mutated to
CC   produce an unusual chromosomal marker referred to as mardel(10). The
CC   mardel(10) marker is mitotically stable and contains a functional
CC   neocentromere at a location regarded as non-centromeric. This
CC   neocentromere maps to q25.2 on chromosome 10. The specification describes
CC   nucleic acid sequences derived from a eukaryotic chromosome, including a
CC   neocentromere or its functional derivative or hybrid, that are able, in a
CC   compatible cell, of replicating, acting as extra-chromosomal element and
CC   segregating during cell division. The sequences can be used to construct
CC   artificial chromosomes for use in gene therapy comprising a replicable,
CC   segregating nucleic acid that confers a specific phenotype on cells.
CC   Human artificial chromosomes can propagate in human cells and carry large
CC   amounts of DNA (e.g. therapeutic genes), and, being extra-chromosomal,
CC   they are not mutagenic. The artificial chromosomes are also useful for
CC   generation of transgenic plants and animals, in production of proteins
CC   and to make diagnostic reagents, e.g. for expression of cytokines,
CC   receptors and growth factors, or to increase the copy number of a gene in
CC   a cell. The constructs may also be used for functional and structural
CC   analysis of chromosomes
XX
SQ   Sequence 80595 BP; 23183 A; 16613 C; 16824 G; 23975 T; 0 U; 0 Other;

  Query Match             6.2%;  Score 253.6;  DB 2;  Length 80595;
  Best Local Similarity   71.9%;  Pred. No. 6.8e-34;
  Matches  361;  Conservative    0;  Mismatches  134;  Indels    7;  Gaps    2;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | |||| |||| ||| | |||||| ||| | | ||| |  ||| | |||| ||| |||||
Db      16026 TATATACATATTTATGTCTATATACACATATATACATATTTATGTCTATATACACATATA 15967

Qy        287 TACACATATATACATATATACACATATATACACA------TATATATACACATACATATG 340
              |||| || |||   ||||||||||||||||||| |      | ||||||||||| |||
Db      15966 TACATATTTATGTCTATATACACATATATACATATTTATGTCTATATACACATATATACA 15907

Qy        341 TACACATATATATACACATATGTATACACATATATACACATATATACACACATATATACA 400
              ||   || | |||| ||| || |||||| || |||    ||||| ||| | ||| |||
Db      15906 TATTTATGTCTATATACACATATATACATATTTATGTCTATATACACATATATACATATT 15847

Qy        401 CATATATACACACATATATACACATATATACACATATATACACATATACACATATATACA 460
              || | || | ||| ||||| |||||| ||    ||||||||||||| |||||| ||
Db      15846 TATGTCTATATACACATATATACATATTTATGTCTATATACACATATATACATATTTATG 15787

Qy        461 CATATATACATATATACACATATATAT-AATATACACACATATATATACACATATATACA 519
              |||||||| ||||| |||||| |||   |||| |||||||||| | |  ||| |  |
Db      15786 TCTATATACACATATATACATATTTATGTCTATATACACATATATACATATTTATGTCTA 15727

Qy        520 CACATATATACACATATATACACATATATATACACATATATACACATATATACATATATA 579
              | | | ||| | | ||||  |  | |||||||||||||||||| || |||   ||||||
Db      15726 TATACACATATATACATATTTATGTCTATATACACATATATACATATTTATGTCTATATA 15667

Qy        580 CACATATATATACATATATACACATATATACATATATACACATATATACATATATACACA 639
              |||||||||| | || |||  | ||||| |||||||||||| || |||   ||||||||||
Db      15666 CACATATATACATATTTATGTATATATACACATATATACATATTTATGTCTATATACACA 15607

Qy        640 CATATATACACATACATATACACACACATAGATATACATATATATACACATATATATACG 699
              |||||  |  |  ||  |||| |   |  || ||||   | |||||  ||||| |||||
Db      15606 TATATACATATTTATGTATATATTTATGTATATATGTGTGTATATGTACATACATATATA 15547

Qy        700 TATATATATGTATATATATATG 721
                  ||| |||  || |||||||||
Db      15546 CATACATAAATACATATATATG 15525


RESULT 54
AAA10225/c
ID   AAA10225 standard; DNA; 106746 BP.
XX
AC   AAA10225;
XX
DT   03-JUL-2000  (first entry)
XX
DE   Human PCTA-1 genomic DNA, SEQ ID NO:1.
XX
KW   Prostate carcinoma antigen-1; PCTA-1; human; galectin-8 homologue;
KW   splice variant; prostate cancer; biallelic marker; polymorphism;
KW   detection; diagnosis; prognosis; drug screening; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   misc_signal     1. .68647
FT                   /*tag=  a
FT                   /note= "5' regulation region"
FT   promoter        66647. .68647
FT                   /*tag=  b
FT   exon            68648. .68471
FT                   /*tag=  c
FT                   /number
FT   intron          68742. .70646
FT                   /*tag=  d
FT                   /number
FT   exon            70647. .70794
FT                   /*tag=  e
FT                   /number= 1
FT   intron          70795. .82207
FT                   /*tag=  f
FT                   /number= 1
FT   exon            82208. .82296
FT                   /*tag=  g
FT                   /number= 2
FT   intron          82297. .83612
FT                   /*tag=  h
FT                   /number= 2
FT   exon            83613. .83823
FT                   /*tag=  i
FT                   /number= 3
FT   intron          83824. .85297
FT                   /*tag=  j
FT                   /number= 3
FT   exon            85298. .85417
FT                   /*tag=  k
FT                   /number= 4
FT   intron          85418. .86388
FT                   /*tag=  l
FT                   /number= 4
FT   exon            86389. .86445
FT                   /*tag=  m
FT                   /number= 5
FT   intron          86446. .87495
FT                   /*tag=  n
FT                   /number= 5
FT   exon            87496. .87522
FT                   /*tag=  o
FT                   /number= 6
FT   intron          87523. .87649
FT                   /*tag=  p
FT                   /number= 6
FT   exon            87650. .87775
FT                   /*tag=  q
FT                   /number= 6bis
FT   intron          87776. .88294
```

```
FT                   /*tag=  r
FT                   /number= 6bis
FT   exon            88295. .88383
FT                   /*tag=  s
FT                   /number= 7
FT   intron          88384. .89483
FT                   /*tag=  t
FT                   /number= 7
FT   exon            89484. .89649
FT                   /*tag=  u
FT                   /number= 8
FT   intron          89650. .92748
FT                   /*tag=  v
FT                   /number= 8
FT   exon            92749. .97155
FT                   /*tag=  w
FT                   /number= 9
FT   exon            92749. .92883
FT                   /*tag=  x
FT                   /number= 9bis
FT   intron          92884. .95820
FT                   /*tag=  y
FT                   /number= 9bis
FT   exon            95821. .97155
FT                   /*tag=  z
FT                   /number= 9ter
FT   misc_signal     97155. .106746
FT                   /*tag=  aa
FT                   /note= "3' regulation region"
XX
PN   WO9964590-A1.
XX
PD   16-DEC-1999.
XX
PF   04-JUN-1999;   99WO-IB001072.
XX
PR   05-JUN-1998;   98US-0088187P.
PR   28-SEP-1998;   98US-0102324P.
XX
PA   (GEST ) GENSET.
XX
PI   Blumenfeld M,  Bougueleret L,  Chumakov I;
XX
DR   WPI; 2000-338635/29.
DR   P-PSDB; AAY87403, AAY87404, AAY87405.
XX
PT   Novel polynucleotide used to develop products for detecting
PT   susceptibility to, diagnosis, prognosis and therapy of prostate cancer.
XX
PS   Claim 1; Page 277-306; 339pp; English.
XX
CC   This sequence represents the genomic DNA sequence encoding human prostate
CC   carcinoma tumour antigen-1 (PCTA-1). At least three PCTA-1 cDNAs
CC   (AAA10226, AAA10227, AAA10228) have been identified all of which comprise
CC   exons 1, 2, 3, 4, 5, 6, 7, 8, and additionally either contain exon 9,
CC   exons 6bis and 9, or exons 9bis and 9ter. PCTA-1 is a human homologue of
CC   rat galectin-8, which can mediate both cell-cell and cell-matrix
CC   interactions. 125 biallelic markers were identified in the genomic
CC   sequence, with 40 biallelic markers being located in exonic regions. Six
CC   of these alter the amino acid sequence of a PCTA-1 protein. 43 markers
CC   were found in the 5' regulatory region, including 16 in the promoter, 39
CC   in intronic sequences, and 3 in the 3' regulatory region. Alleles
CC   containing certain biallelic markers are associated with prostate cancer.
CC   These polymorphisms can be used as the basis of methods for determining
CC   genetic predisposition to prostate cancer, for diagnosis of this cancer
CC   and for the prognosis/detection of an eventual treatment response to
CC   therapeutic agents against prostate cancer. They can be used to design of
CC   drug screening protocols to provide accurate and efficient evaluation of
CC   therapeutic and side-effect potential of new or already existing
CC   medicaments, as well as for screening molecules which modulate or inhibit
CC   the expression of the PCTA-1 gene, preferably those active against
CC   prostate cancer
XX
SQ   Sequence 106746 BP; 29695 A; 23487 C; 23529 G; 30029 T; 0 U; 6 Other;

  Query Match             6.1%;  Score 253;  DB 3;  Length 106746;
  Best Local Similarity   69.4%;  Pred. No. 8.7e-34;
  Matches  416;  Conservative    0;  Mismatches  175;  Indels    8;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              |||||||||||||||||  |||||||||||||||| |||||||||||||| ||||||
Db      38717 AGTTCAAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAAAATT 38658

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   ||||||||| | ||||||||||||||||||||||  |||| | ||||||||||||
Db      38657 AGCTGGGCATGGTGGTGCACGCCTGTAATCCCAGCTACTCAGGAGACTGAGGCAGGAGAA 38598

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCCA 178
              ||||||||||||  || ||||||| ||| || ||| ||| |   | ||| |||||||||
Db      38597 TCACTTGAACCTG--GGAGGTGGAGTTTGTGGTGAGCCAAAACTTTCCCACTGCACTCCA 38540

Qy        179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCAT-TTTATATATAT 237
              |||      |  | |   | || |  | | | |     | |   ||    | ||||
Db      38539 GCCAGTATATATATATATATATACACACGTACACGTGTACACATATACACACATATGC 38480

Qy        238 ATATATATATATATACACACACACACATATATATATACACATATATATACACATATAT 297
              | |  | || ||| ||||||| |||||   | ||| || ||||||| || |  | || ||
Db      38479 ACACGTGTACATACACACACATACACATGCACATACATGCACATATGTACATGCGTACAT 38420

Qy        298 ACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACAC 357
              ||| | ||  |||||||||||||| || || | |  ||||    ||||||| || || | |
Db      38419 ACACACATGTACATATATACACACATGTACATATGTACACCCATACACACACATGTACATAT 38360

Qy        358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417
              ||      | ||| | ||  |||||| ||||| |||||| | |  || |||   |||
Db      38359 GTACACCCATACACACATGTACATATGTACACCCATATACATATGTACATATGTACACCC 38300

Qy        418 ATACACATATATACACATATATACACATATACACATATATACACATATATACATATATAC 477
              |||||||||| |||| || || || |||| |||  || |||   | |   | | | ||
Db      38299 ATACACATATGTACATATGTACACCCATACACATGTACATATGTACACCCATACACATGT 38240

Qy        478 ACATATATATAATATACACACATATATATACACATATATACACACATATATACACATATA 537
              |||||| |   | || |||| || || |||   ||||  | |||||||| |||| |||||
Db      38239 ACATATGT---ACATCCACATATGTACATATGTATATCCATACACATATGTACATATATA 38183

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATACATAT 596
              |||||||| || | |  || ||| | || | |  |||| ||   |||||||||   ||
Db      38182 CACACATATGTACATATGTACATATATATTTGTGTATATCTATGTATATATATAAGAAT 38124


RESULT 55
AAV83940/c
ID   AAV83940 standard; DNA; 80240 BP.
XX
AC   AAV83940;
XX
DT   03-MAR-1999  (first entry)
XX
DE   NC-contig derived from mardel(10) on chromosome 10q25.2.
XX
KW   Yeast artificial chromosome; YAC; probe; eukaryotic chromosome;
KW   neocentromere; replication; extra-chromosomal element; segregation;
KW   cell division; artificial chromosome; gene therapy; mardel(10);
KW   human artificial chromosome; transgenic; chromosome 10; 10q25.2; ss.
XX
OS   Homo sapiens.
XX
PN   WO9851790-A1.
XX
PD   19-NOV-1998.
XX
PF   13-MAY-1998;   98WO-AU000352.
XX
PR   13-MAY-1997;   97AU-00006784.
PR   26-AUG-1997;   97AU-00008791.
XX
```

```
PA   (AMRA-) AMRAD OPERATIONS PTY LTD.
XX
PI   Choo K,  Du Sart D,  Cancilla MR;
XX
DR   WPI; 1999-009773/01.
XX
PT   New isolated nucleic acid comprising neocentromere sequences from
PT   eukaryotic chromosome - used to produce replicable, segregating
PT   artificial chromosomes that can carry large amounts of DNA for gene
PT   therapy.
XX
PS   Claim 9; Fig 16A; 540pp; English.
XX
CC   The present sequence represents the NC-contig derived from a mutated
CC   human chromosome 10, 10q25.2 region. The sequence contains an unusual
CC   chromosomal marker referred to as mardel(10). The mardel(10) marker is
CC   mitotically stable and contains a functional neocentromere at a location
CC   regarded as non-centromeric. This neocentromere maps to q25.2 on
CC   chromosome 10. The specification describes nucleic acid sequences derived
CC   from a eukaryotic chromosome, including a neocentromere or its functional
CC   derivative or hybrid, that are able, in a compatible cell, of
CC   replicating, acting as extra-chromosomal element and segregating during
CC   cell division. The sequences can be used to construct artificial
CC   chromosomes for use in gene therapy comprising a replicable, segregating
CC   nucleic acid that confers a specific phenotype on cells. Human artificial
CC   chromosomes can propagate in human cells and carry large amounts of DNA
CC   (e.g. therapeutic genes), and, being extra-chromosomal, they are not
CC   mutagenic. The artificial chromosomes are also useful for generation of
CC   transgenic plants and animals, in production of proteins and to make
CC   diagnostic reagents, e.g. for expression of cytokines, receptors and
CC   growth factors, or to increase the copy number of a gene in a cell. The
CC   constructs may also be used for functional and structural analysis of
CC   chromosomes
XX
SQ   Sequence 80240 BP; 23102 A; 16537 C; 16747 G; 23846 T; 0 U; 8 Other;

  Query Match             6.1%;  Score 252.6;  DB 2;  Length 80240;
  Best Local Similarity   71.9%;  Pred. No. 1e-33;
  Matches  360;  Conservative    0;  Mismatches  134;  Indels    7;  Gaps    2;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | || |||| ||| | |||||| | ||| | | ||| |  ||| | |||| ||| | |||
Db      15787 TATACATATTTATGTCTATATACACACATATATACATATTTATGTCTATATACACACATA 15728

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              |||| || |||   ||||||||||||||||||| || ||| |  | ||||| || || |||
Db      15727 TACATATTTATGTCTATATACACATATATACATATTTATGTCTATATACACATATATACA 15668

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              ||| |||    ||||  | | | ||| |||   || | || | ||| |||||| ||||||
Db      15667 TATTTATGTCTATATACACATATATACATATTTATGTCTATATACACATATATACATATT 15608

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              |  |    |||||||||||||||||| || |||    |||||||||||||||| || ||
Db      15607 T--ATGTCTATATACACATATATACATATTTATGTCTATATACACATATATACATATTTA 15550

Qy        467 TACATATATACACATATATA-----TAATATACACACATATATATACACATATATACACA 521
              |   ||||||||||||||||     | || |  | | | | ||||| ||||||  ||
Db      15549 TGTCTATATACACATATATACATATTTATGTCTATATACACATATATACATATTTATGTC 15490

Qy        522 CATATATACACATATATACACATATATATACACATATATACACATATATACATATATACA 581
              |||||  ||| ||| |||   || | |||| ||| ||||| ||||||  ||  | |||| |
Db      15489 TATATACACATATACATATTTATGTCTATATACACATATATACATATTTATGTCTATATA 15430

Qy        582 CATATATATACATATATACACATATATACATATATACACATATATACATATATACACACA 641
              || ||||||||||||| ||    |||||||| ||||| |||||| ||  |||||| |||||
Db      15429 CACATATATACATATTTATGTCTATATACACATATATACATATTTATGTATATATACACA 15370

Qy        642 TATATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTA 701
              |||||||| ||  || |  | | ||| | |||||||||| |||  | |||| |||   ||
Db      15369 TATATACATATTTATGTCTATATACACATATATACATATTTATGTATATATTTATGTATA 15310

Qy        702 TATATATGTATATATATATGC 722
```

```
              ||| | ||||||| || || |
Db      15309 TATGTGTGTATATGTACATAC 15289


RESULT 56
ABV72623/c
ID   ABV72623 standard; DNA; 82938 BP.
XX
AC   ABV72623;
XX
DT   27-NOV-2002  (first entry)
XX
DE   Human transporter protein encoding gene.
XX
KW   Human; transporter protein; hTP; gamma-aminobutyric-acid; GABA;
KW   gene therapy; protein therapy; vaccine; gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1120. .81932
FT                   /*tag= a
FT                   /product= "Transporter protein"
FT                   /note= "Contains 8 introns"
FT   exon            1120. .1223
FT                   /*tag= b
FT                   /number= 1
FT   intron          1224. .27733
FT                   /*tag= c
FT                   /number= 1
FT   variation       replace(2419,C)
FT                   /*tag= d
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(3191,A)
FT                   /*tag= e
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(4195,T)
FT                   /*tag= f
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(5910,C)
FT                   /*tag= g
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(7119,T)
FT                   /*tag= h
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(7529,G)
FT                   /*tag= h
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(8855,C)
FT                   /*tag= i
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(8956,A)
FT                   /*tag= j
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(10220,G)
FT                   /*tag= k
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(11628,C)
FT                   /*tag= l
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(11670,A)
FT                   /*tag= m
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(11671,G)
FT                   /*tag= n
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(11931,C)
FT                   /*tag= o
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(15984,G)
FT                   /*tag= p
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(17735,A)
```

```
FT                   /*tag= q
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(17910,A)
FT                   /*tag= r
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(18130,C)
FT                   /*tag= s
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19293. .19294,TWA)
FT                   /*tag= t
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19294. .19295,ATT)
FT                   /*tag= u
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19296. .19297,CYT)
FT                   /*tag= v
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19303. .19304,TYC)
FT                   /*tag= w
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19304. .19305,CYC)
FT                   /*tag= x
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19307. .19308,TMC)
FT                   /*tag= y
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19308. .19309,CCC)
FT                   /*tag= z
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19313. .19314,TMA)
FT                   /*tag= aa
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19314. .19315,AWT)
FT                   /*tag= ab
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19315. .19316,TCC)
FT                   /*tag= ac
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19316. .19317,CYT)
FT                   /*tag= ad
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19317. .19318,TWA)
FT                   /*tag= ae
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19318. .19319,ATT)
FT                   /*tag= af
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19860,G)
FT                   /*tag= ag
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(20843,G)
FT                   /*tag= ah
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(21144,T)
FT                   /*tag= ai
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(21371,G)
FT                   /*tag= aj
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(22061. .22062,A)
FT                   /*tag= ak
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(22321,T)
FT                   /*tag= al
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(22470. .22471,A)
FT                   /*tag= am
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(22472. .22473,A)
FT                   /*tag= an
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(24990,C)
FT                   /*tag= ao
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(25284,A)
FT                   /*tag= ap
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(25934,G)
FT                   /*tag= aq
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(27509. .27510,TAT)
FT                   /*tag= ar
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(27513. .27514,TTT)
FT                   /*tag= as
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            27734. .27882
FT                   /*tag= at
FT                   /number= 2
FT   intron          27883. .41045
FT                   /*tag= au
FT                   /number= 2
FT   variation       replace(28608,G)
FT                   /*tag= av
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(28627,C)
FT                   /*tag= aw
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(29709,T)
FT                   /*tag= ax
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(29788,T)
FT                   /*tag= ay
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(29878,T)
FT                   /*tag= az
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(30294,T)
FT                   /*tag= ba
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(30896. .30897,T)
FT                   /*tag= bb
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(31815,G)
FT                   /*tag= bc
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(31999,TTT)
FT                   /*tag= bd
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(32026,T)
FT                   /*tag= be
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(32146,A)
FT                   /*tag= bf
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(32184,S)
FT                   /*tag= bg
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(32214. .32215,T)
FT                   /*tag= bh
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(32215,C)
FT                   /*tag= bi

  Query Match             6.1%;  Score 252.6;  DB 6;  Length 82938;
  Best Local Similarity   70.7%;  Pred. No. 1e-33;
  Matches  350;  Conservative    0;  Mismatches  144;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | |||| | |||| |||| ||||||||| | | |   | |||  ||||||    ||||||
Db       7454 TATATACACATATGTATACATATATACATATATATGTATATACGTATATAGCTGTATATA 7395

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || ||| || || |||  |  | ||||| | |||| |||||||| |||||||| |||| |
Db       7394 TATACACATGTATATAGCTGTATATATACATACATGTATATACATATACATATATACATA 7335
```

```
Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              || | ||| |||||  | || | | ||||||| ||||| | |||  |||||||| |||||
Db       7334 TACACATATACATACATGTATATACATATACATATATACATACATGTATATACATATATA 7275

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              | |   ||| ||    || |||| |||||||||||| || || |||| | ||   |||||
Db       7274 CATATGTATACATGTGTATCTATATACATATATACATATGTATACATGTGTATCTATATA 7215

Qy        467 TACATATATACACATATATAT-AATATACACACATATATATACACATATATACACACATA 525
              | ||||| | |  |||| ||   ||| | | | ||||||||| || ||||||    ||
Db       7214 CATATATACATATGTATACATGTGTATCTATATACATATATACATATGTATACATGTGTA 7155

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              | || | | ||||| |||||| ||||||  | |||  | ||| ||| ||| |||   |||
Db       7154 TCTATATACATATATACATATGTATACATGTGTATCTATATACATATATACATATGTATA 7095

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              || |  || |||| ||||||||||||||| || |||| | ||  |||||||| | ||| |
Db       7094 CATGTGTATCTATATACATATATACATATGTATACATGTGTATCTATATACATATATACA 7035

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              ||   |||||| |  |   | ||| ||||| | |||| || |||| | | || ||||| |
Db       7034 TATGTATACATGTGTATCTATATACATATATACATATGTATACATGTGTGTATGTATACA 6975

Qy        706 TATGTATATATATAT 720
              ||||| | | |||||
Db       6974 TATGTGTGTGTATAT 6960
```

RESULT 57
ADC87564/c
ID ADC87564 standard; DNA; 1114 BP.
XX
AC ADC87564;
XX
DT 01-JAN-2004 (first entry)
XX
DE Human GPCR gene SEQ ID NO:2017.
XX
KW ds; gene; human; GPCR;
KW guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1270724-A2.
XX
PD 02-JAN-2003.
XX
PF 18-JUN-2002; 2002EP-00013517.
XX
PR 18-JUN-2001; 2001JP-00246789.
XX
PA (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI Suwa M, Asai K, Akiyama Y, Aburatani H;
XX
DR WPI; 2003-315783/31.
DR P-PSDB; ADC87565.
XX
PT New polynucleotide, useful for preparing a composition for treating a
PT patient in need of increased or suppressed activity or expression of the
PT guanosine triphosphate-binding protein coupled receptor.
XX
PS Claim 1; SEQ ID NO 2017; 28pp; English.
XX
CC The invention relates to a novel polynucleotide encoding a guanosine
CC triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC the invention may have a use in gene therapy. The polynucleotide and
CC polypeptide are useful for preparing a composition for treating a patient
CC in need of increased or suppressed activity or expression of the
CC guanosine triphosphate-binding protein coupled receptor. The
CC polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC invention.
XX
SQ Sequence 1114 BP; 398 A; 107 C; 167 G; 442 T; 0 U; 0 Other;

```
  Query Match             6.1%;  Score 251;  DB 10;  Length 1114;
  Best Local Similarity   63.7%;  Pred. No. 1.6e-33;
  Matches  397;  Conservative    0;  Mismatches  225;  Indels    1;  Gaps    1;

Qy        230 ATATATATATATATATATATACACACACACACATATATATATACACATATATATAC 289
              ||| | |||||| ||||||| |   | | ||| |||||||||| |||| |   | |||||||
Db        864 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 805

Qy        290 ACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATAT 349
              | | | ||| || |||| |  | |||| ||||| | ||| |  || |||| |  | |||
Db        804 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 745

Qy        350 ATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATAC 409
              ||| ||| |||||||||  |  | |||| ||||| | | | |  ||||||    |||||||
Db        744 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 685

Qy        410 ACACATATATACACATATATACACATATATACACATATACACATATATACACATATATAC 469
              | ||| |||||    ||||||    ||||| | ||| ||| |  ||||||    |||||||
Db        684 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 625

Qy        470 ATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATATAT 528
              ||| | | || ||||||  | || | ||||| | ||| |  ||||||     |||||
Db        624 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 565

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACACATATAT 588
              | ||| |||||   ||||||  |  | ||| ||||||||||||  ||||||    |||||
Db        564 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 505

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
              |||||  |||||   ||||||   |||||||| | | ||| || |||| |    |||||||
Db        504 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 445

Qy        649 ACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATAT 708
              | | |||||||   | | ||   |||| | ||| | | | || ||||||  | ||||||
Db        444 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 385

Qy        709 GTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAG 768
              || | |||||||        |  |||        ||  |   ||    |  | ||| |
Db        384 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 325

Qy        769 GAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAG 828
              | | | |  | | |    |||||    | |   | || |      |  | | | |||
Db        324 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 265

Qy        829 TAGGGGAAATACATAAATGTGTA 851
                  | ||||||| |||| ||
Db        264 ATACACATATACATATATGTATA 242
```

RESULT 58
ADE95974
ID ADE95974 standard; DNA; 96594 BP.
XX
AC ADE95974;
XX
DT 12-FEB-2004 (first entry)
XX
DE Human SYK gene genomic DNA sequence.
XX
KW cancer diagnosis; cancer treatment; carcinoma; cytostatic; gene therapy;
KW lymphoma; breast cancer; prostate cancer; leukaemia; ds; human; SYK.
XX
OS Homo sapiens.
XX
PN WO2003039484-A2.
XX
PD 15-MAY-2003.
XX
PF 08-NOV-2002; 2002WO-US036071.

```
XX
PR   08-NOV-2001; 2001US-00052482.
XX
PA   (SAGR-) SAGRES DISCOVERY.
XX
PI   Morris DW,  Engelhard EK;
XX
DR   WPI; 2003-441462/41.
XX
PT   New carcinoma associated nucleic acids and proteins, useful for screening
PT   drug candidates, or for diagnosing and treating carcinomas, e.g.
PT   lymphoma, breast cancer, prostate cancer or leukemia.
XX
PS   Claim 1; SEQ ID NO 232; 793pp; English.
XX
CC   This invention relates to novel recombinant nucleic acids for use in
CC   diagnosis and treatment of cancer, especially carcinomas, as well as the
CC   use of compositions in screening methods. The compositions of the
CC   invention may have cytostatic activity whilst the disclosed sequences may
CC   be useful for gene therapy. The carcinoma associated nucleic acids and
CC   proteins are useful for diagnosing and treating carcinomas, for example
CC   lymphoma, breast cancer, prostate cancer or leukaemia, or for screening
CC   drug candidates or bioactive agents capable of binding to, or modulating
CC   the activity of, a carcinoma associated protein. The present sequence is
CC   the genomic DNA sequence of the human SYK gene which is a carcinoma
CC   associated gene of the invention.
XX
SQ   Sequence 96594 BP; 27524 A; 20558 C; 21159 G; 26914 T; 0 U; 439 Other;

  Query Match           6.1%;  Score 251;  DB 10;  Length 96594;
  Best Local Similarity   69.5%;  Pred. No. 1.9e-33;
  Matches  354;  Conservative    0;  Mismatches  152;  Indels    3;  Gaps    1;

Qy        229 TATATATATATATATATATATATACACACACACACACATATATATATACACATATATATA 288
              || ||| | |||| | |||||||||| ||| | |||||||| | | |||| ||| ||| |
Db      90981 TACATACACATATGTGTATATATACATACATATACACATATGTGTGTACATATACATACA 91040

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              || | | |||  | | |||| ||| ||||||| | ||| ||| ||   ||  |  |||||
Db      91041 CATACACATATGTGTGTACATATACATACACACACATACACATATGTGTACATGTACATA 91100

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
               ||| ||| | |  |||    | || |||| | ||| ||| ||| ||||    | || ||
Db      91101 CATACACATACACATATGTGTACATGTACATACATACACATACACATATGTGTACATGTA 91160

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              || ||||| | | ||| ||||    | || |||| |||||||||||||| ||| || | ||
Db      91161 CATACATACACATACACATATGTGTACATGTACATATATACACATATACACATATGTGTA 91220

Qy        469 CATATATACACATATATATAATATACACACATATATATACACATATATACACACATATAT 528
              ||| || | | ||| |    |||||||| || | || |    | ||||| |||||||||
Db      91221 CATGTACATATATACAC---ATATACACATATGTGTACATGTACATATATACACATATAC 91277

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACATATAT 588
              ||| || | ||||  || ||||||||||||| ||| || | ||||| || | | | | ||
Db      91278 ACATATGTGTACATGTACATATACACATATACACATATGTGTACATGTACATATACACAT 91337

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
              ||||| ||||    | ||  |||||||| |||  | || |   | ||| ||||||| |
Db      91338 ATACACATATGTGTACATGCACATATACACATGTGTACATGCACATATACACATATGTGT 91397

Qy        649 ACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATAT 708
              ||||||| ||| ||||| ||    | || ||| |||||||| |||| |  |||||| ||
Db      91398 ACATACACATATACACATATGTGTGTATATACATATACACACATATGTGTGTATATACAT 91457

Qy        709 GTATATATATATGCTCCAGAGTTCATAAG 737
               |||| |||||| |       | |||| |
Db      91458 ATATACATATATACATATATATACATATG 91486


RESULT 59
ADA02726
ID   ADA02726 standard; DNA; 96595 BP.
XX
AC   ADA02726;
XX
DT   06-NOV-2003  (first entry)
XX
DE   Human SYK carcinoma associated gene, SEQ ID NO:1244.
XX
KW   Human; carcinoma associated; oncogene; carcinoma; cancer; breast;
KW   prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO2003057146-A2.
XX
PD   17-JUL-2003.
XX
PF   26-DEC-2002; 2002WO-US041414.
XX
PR   26-DEC-2001; 2001US-00035832.
XX
PA   (SAGR-) SAGRES DISCOVERY.
XX
PI   Morris DW;
XX
DR   WPI; 2003-587068/55.
XX
PT   New recombinant nucleic acid encoding carcinoma associated protein,
PT   useful for preparing compositions for treating carcinomas.
XX
PS   Claim 1; SEQ ID NO 1244; 245pp; English.
XX
CC   The invention relates to recombinant carcinoma associated (CA) nucleic
CC   acid sequences from mouse and human (ADA01482-ADA03094), and to
CC   recombinant carcinoma associated proteins (CAP) encoded by them. The
CC   invention also encompasses expression vectors and host cells comprising a
CC   CA nucleic acid, a polypeptide (especially an antibody) that specifically
CC   binds to the protein, and a biochip comprising CA nucleic acid or
CC   fragments thereof. The sequences of the invention were identified using
CC   oncogenic retroviruses, which insert into the genome of the host organism
CC   at random. Many of these do not carry transduced host oncogenes or
CC   pathogenic trans-acting viral genes, meaning that cancer incidence is a
CC   direct consequence of the effects of proviral integration into host
CC   protooncogenes. The CA nucleic acid sequences can be used to diagnose
CC   carcinoma (especially breast cancer, prostate cancer, lymphoma or
CC   leukaemia) or a propensity to carcinoma by determination of the sequence
CC   of a CA gene, or by determination of CA gene expression in particular
CC   tissues. CA nucleic acids, proteins and antibodies are also useful as
CC   therapeutic agents and in screening and evaluating drug candidates. The
CC   present sequence represents a specifically claimed human CA nucleic acid
CC   sequence of the invention. Note: The complete sequence data for this
CC   patent did not form part of the printed specification, but was obtained
CC   in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 96595 BP; 27524 A; 20559 C; 21158 G; 26915 T; 0 U; 439 Other;

  Query Match           6.1%;  Score 251;  DB 9;  Length 96595;
  Best Local Similarity   69.5%;  Pred. No. 1.9e-33;
  Matches  354;  Conservative    0;  Mismatches  152;  Indels    3;  Gaps    1;

Qy        229 TATATATATATATATATATATATACACACACACACACATATATATATACACATATATATA 288
              || ||| | |||| | |||||||||| ||| | |||||||| | | |||| ||| ||| |
Db      90982 TACATACACATATGTGTATATATACATACATATACACATATGTGTGTACATATACATACA 91041

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              || | | |||  | | |||| ||| ||||||| | ||| ||| ||   ||  |  |||||
Db      91042 CATACACATATGTGTGTACATATACATACACACACATACACATATGTGTACATGTACATA 91101

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
               ||| ||| | |  |||    | || |||| | ||| ||| ||| ||||    | || ||
Db      91102 CATACACATACACATATGTGTACATGTACATACATACACATACACATATGTGTACATGTA 91161
```

```
Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              || ||||| | | ||| ||||    | || |||| |||||||||||||| ||| || | ||
Db      91162 CATACATACACATACACATATGTGTACATGTACATATATACACATATACACATATGTGTA 91221

Qy        469 CATATATACACATATATATAATATACACACATATATATACACATATATACACACATATAT 528
              ||| || | | ||| |    |||||||| || | || |   | ||||| |||||||||
Db      91222 CATGTACATATATACAC---ATATACACATATGTGTACATGTACATATATACACATATAC 91278

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACATATAT 588
              ||| || | |||| || ||||||||||||| ||| || | ||||| || | | | | ||
Db      91279 ACATATGTGTACATGTACATATACACATATACACATATGTGTACATGTACATATACACAT 91338

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
              ||||| ||||    | ||  ||||||| |||  | || |   | ||| |||||||| |
Db      91339 ATACACATATGTGTACATGCACATATACACATGTGTACATGCACATATACACATATGTGT 91398

Qy        649 ACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATAT 708
              ||||||| ||| ||||| ||   | || ||| |||||||| |||| |  ||||||| ||
Db      91399 ACATACACATATACACATATGTGTGTATATACATATACACACATATGTGTGTATATACAT 91458

Qy        709 GTATATATATATGCTCCAGAGTTCATAAG 737
              |||| |||||| |       | |||| |
Db      91459 ATATACATATATACATATATATACATATG 91487


RESULT 60
ADB72464
ID   ADB72464 standard; DNA; 96595 BP.
XX
AC   ADB72464;
XX
DT   04-DEC-2003  (first entry)
XX
DE   Human SYK gene.
XX
KW   human; ds; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;
KW   cancer; neoplasm; adenocarcinoma; sarcoma; gene.
XX
OS   Homo sapiens.
XX
PN   WO2003008583-A2.
XX
PD   30-JAN-2003.
XX
PF   26-DEC-2001; 2001WO-US051291.
XX
PR   02-MAR-2001; 2001US-00798586.
PR   23-OCT-2001; 2001US-00004113.
PR   08-NOV-2001; 2001US-00052482.
PR   30-NOV-2001; 2001US-00997722.
PR   20-DEC-2001; 2001US-00034650.
XX
PA   (SAGR-) SAGRES DISCOVERY.
XX
PI   Morris DW,  Engelhard EK;
XX
DR   WPI; 2003-239337/23.
XX
PT   New recombinant nucleic acid, useful for treating carcinomas, lymphomas,
PT   cancers, neoplasm, adenocarcinoma, or sarcomas.
XX
PS   Claim 1; SEQ ID NO 292; 2304pp; English.
XX
CC   The invention relates to a novel recombinant nucleic acid comprising a
CC   nucleotide sequence selected from any of the 660 sequences fully defined
CC   in the specification. A polynucleotide of the invention has cytostatic
CC   activity, and may have a use in gene therapy, or in a vaccine. The
CC   recombinant nucleic acids and polypeptides are useful for treating
CC   carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and
CC   sarcomas. The present sequence represents a human gene of the invention.
XX
SQ   Sequence 96595 BP; 27524 A; 20559 C; 21158 G; 26915 T; 0 U; 439 Other;

  Query Match             6.1%;  Score 251;  DB 10;  Length 96595;
  Best Local Similarity   69.5%;  Pred. No. 1.9e-33;
  Matches  354;  Conservative    0;  Mismatches  152;  Indels    3;  Gaps    1;

Qy        229 TATATATATATATATATATATACACACACACACATATATATATACACATATATATA 288
              || ||| | |||| | |||||||||| ||| | |||||||| | | |||| ||| ||| |
Db      90982 TACATACACATATGTGTATATATACATACATATACACATATGTGTGTACATATACATACA 91041

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              || | | ||| | | |||| ||| ||||||| | ||| ||| ||   || |  |||||
Db      91042 CATACACATATGTGTGTACATATACATACACACACATACACATATGTGTACATGTACATA 91101

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              ||| ||| | |  |||    | || |||| | ||| ||| ||| ||||    | || ||
Db      91102 CATACACATACACATATGTGTACATGTACATACATACACATACACATATGTGTACATGTA 91161

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              || ||||| | | ||| ||||    | || |||| |||||||||||||| ||| || | ||
Db      91162 CATACATACACATACACATATGTGTACATGTACATATATACACATATACACATATGTGTA 91221

Qy        469 CATATATACACATATATATAATATACACACATATATATACACATATATACACACATATAT 528
              ||| || | | ||| |    |||||||| || | || |   | ||||| |||||||||
Db      91222 CATGTACATATATACAC---ATATACACATATGTGTACATGTACATATATACACATATAC 91278

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACACATATAT 588
              ||| || | |||| || ||||||||||||| ||| || | ||||| || | | | | ||
Db      91279 ACATATGTGTACATGTACATATACACATATACACATATGTGTACATGTACATATACACAT 91338

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
              ||||| ||||    | ||  ||||||| |||  | || |   | ||| |||||||| |
Db      91339 ATACACATATGTGTACATGCACATATACACATGTGTACATGCACATATACACATATGTGT 91398

Qy        649 ACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATAT 708
              ||||||| ||| ||||| ||   | || ||| |||||||| |||| |  ||||||| ||
Db      91399 ACATACACATATACACATATGTGTGTATATACATATACACACATATGTGTGTATATACAT 91458

Qy        709 GTATATATATATGCTCCAGAGTTCATAAG 737
              |||| |||||| |       | |||| |
Db      91459 ATATACATATATACATATATATACATATG 91487


RESULT 61
ADQ97712/c
ID   ADQ97712 standard; DNA; 103661 BP.
XX
AC   ADQ97712;
XX
DT   07-OCT-2004  (first entry)
XX
DE   Mouse cancer associated sequence MD10-032, SEQ ID 689.
XX
KW   Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Mouse; ds.
XX
OS   Mus musculus.
XX
PN   WO2004060304-A2.
XX
PD   22-JUL-2004.
XX
PF   22-DEC-2003; 2003WO-US041389.
XX
PR   27-DEC-2002; 2002US-00330773.
XX
PA   (SAGR-) SAGRES DISCOVERY INC.
XX
PI   Morris DW,  Malandro MS;
XX
DR   WPI; 2004-543781/52.
XX
PT   New isolated cancer associated nucleic acids comprising at least 10
PT   contiguous nucleotides, useful for diagnosing, preventing and/or treating
PT   cancers such as leukemia and lymphoma.
XX
```

```
PS   Claim 1; SEQ ID NO 689; 199pp; English.
XX
CC   The present invention relates to cancer associated sequences (ADQ97025-
CC   ADQ98004). The sequences are useful for the diagnosis, prevention and/or
CC   treatment of cancer, such as leukemia and lymphoma. Note: The sequence
CC   data for this patent did not form part of the printed specification, but
CC   was obtained in electronic formate directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 103661 BP; 25663 A; 25358 C; 25585 G; 26717 T; 0 U; 338 Other;

  Query Match             6.1%;  Score 251;  DB 12;  Length 103661;
  Best Local Similarity   72.6%;  Pred. No. 1.9e-33;
  Matches  339;  Conservative    0;  Mismatches  125;  Indels    3;  Gaps    1;

Qy        230 ATATATATATATATATATATATACACACACACACACATATATATATACACATATATATAC 289
              | | | | | | | | | | ||| | |||||||||| | | | | ||||| | | ||||
Db      73981 ACACACACACACACACACACACACATATACACACACACACACACACACACACACATAC 73922

Qy        290 ACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATAT 349
              ||| | ||| | | | ||||| | | ||||||| | | ||||| ||| |   |||||  ||
Db      73921 ACACACATATACACACACACACACACACACATACACACACACACACACACACACACAT 73862

Qy        350 ATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATAC 409
              ||| ||||| |   | ||||| | | ||| ||| | |||||||| | | ||||| ||||||
Db      73861 ATACACACACACACACACACACACACACACATATACACACACACACACACACACACATAC 73802

Qy        410 ACACATATATACACATATATACACATATATACACATATACACATATATACACATATATAC 469
              ||||| | | ||||| | | ||||| |||||||||| | ||||| | | ||||||| | ||
Db      73801 ACACACACACACACACACACACACACACATATACACACACACACACACACACACATACAC 73742

Qy        470 ATATATACACATATATATA---ATATACACACATATATATACACATATATACACACATAT 526
              | | | ||||| | |||||    | | ||||||| | | | ||||||||| ||||||| |
Db      73741 ACACACACACACACATATACACACACACACACACACACACACACATATACACACACAC 73682

Qy        527 ATACACATATATACACATATATATACACATATATACACATATATACATATATACACATAT 586
              | ||||| ||| ||||| ||| | ||||| | | ||||||| | ||| | | ||||| |
Db      73681 ACACACACATACACACACATACACACACACACACACACATACACACACACACACACAC 73622

Qy        587 ATATACATATATACACATATATACATATATACACATATATACATATATACACACATATAT 646
              ||| ||| | | |||||||| | ||| |||||||||| | | ||| | | ||||||| ||||
Db      73621 ATACACACACACACACATACACACACATATACACACACACACACACACACACACATAT 73562

Qy        647 ACACATACATATACACACACATAGATATACATATATATACACATATA 693
              ||||| ||| | ||| ||||| | ||||||| | | | ||||| | |
Db      73561 ACACACACACACACATACACACACATATACACACACACACACACA 73515


RESULT 62
ABD32998/c
ID   ABD32998 standard; DNA; 113079 BP.
XX
AC   ABD32998;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Human cancer-associated genomic DNA HD21-021.
XX
KW   Human; ds; cancer-associated protein; gene; cytostatic; cancer;
KW   leukaemia; lymphoma; CAP.
XX
OS   Homo sapiens.
XX
PN   WO2004074320-A2.
XX
PD   02-SEP-2004.
XX
PF   17-FEB-2004; 2004WO-US004730.
XX
PR   14-FEB-2003; 2003US-00367094.
PR   14-MAR-2003; 2003US-00388838.
PR   15-APR-2003; 2003US-00417375.
PR   13-JUN-2003; 2003US-00461862.
PR   15-SEP-2003; 2003US-00663431.
PR   15-DEC-2003; 2003US-00737318.
XX
PA   (SAGR-) SAGRES DISCOVERY INC.
XX
PI   Morris DW,  Morris DW,  Malandro MS;
XX
DR   WPI; 2004-652914/63.
XX
PT   New isolated cancer-associated polynucleotides and polypeptides useful
PT   for diagnosing, preventing or treating cancers, especially lymphoma and
PT   leukemia, or in screening for agents that modulate cancer.
XX
PS   claim 16; seqid 764; 310pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising at least 10
CC   contiguous nucleotides of any of the 233 polynucleotide sequences given
CC   in the specification, or its complement. The nucleic acids encode cancer-
CC   associated proteins. Also included are an expression vector comprising
CC   the isolated nucleic acid cited above, a host cell comprising the above
CC   recombinant nucleic acid or expression vector, a microarray for detecting
CC   a cancer-associated (CA) nucleic acid comprising at least one probe
CC   comprising at least 10 contiguous nucleotides of any of the above-
CC   mentioned nucleotide sequences, an isolated polypeptide (encoded within
CC   an open reading frame of a CA sequence selected from any of the 95
CC   polynucleotide sequences as mentioned in the specification, or its
CC   complement), an isolated antibody, (or its antigen binding fragment) that
CC   binds to the above polypeptide, a hybridoma that produces the above
CC   monoclonal antibody, a pharmaceutical composition comprising the above
CC   antibody and a pharmaceutical excipient, a kit for detecting cancer
CC   cells(comprising the antibody cited above, methods for diagnosing cancer
CC   or for detecting the presence or absence of cancer cells in an
CC   individual, a method for inhibiting growth of cancer cells in an
CC   individual, a method for delivering a therapeutic agent to cancer cells
CC   in an individual, an electronic library comprising the above
CC   polynucleotide or polypeptide (or their fragments), methods of screening
CC   for anticancer activity or for a bioactive agent capable of modulating
CC   the activity of a CA protein (CAP), methods for detecting cancer
CC   associated with expression of a polypeptide in a test cell sample, a
CC   method for treating cancers and a method for inhibiting the expression of
CC   CA gene in a cell. The composition and methods are useful for detecting,
CC   diagnosing, preventing and treating cancers, especially lymphoma and
CC   leukaemia. These may also be used in screening for agents that modulate
CC   cancer. The present sequence is a human CAP genomic sequence. Note: The
CC   sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 113079 BP; 27712 A; 27684 C; 28156 G; 29527 T; 0 U; 0 Other;

  Query Match             6.1%;  Score 250.6;  DB 13;  Length 113079;
  Best Local Similarity   73.3%;  Pred. No. 2.2e-33;
  Matches  335;  Conservative    0;  Mismatches  119;  Indels    3;  Gaps    1;

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              |||||||||||||||| ||| | | | ||||| | ||| | ||||||||| |||||||||
Db      31278 ATATATATATATACAGACATATATATATATAGACATATATATATATATAGACATATATAT 31219

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACAT 359
              ||||| ||| ||||||| | | | ||||| | ||| | |   || | ||| | | ||| ||
Db      31218 ATATAGACATATATATATAGAGACATATATATATATAGACATATATATAGAGAGACATAT 31159

Qy        360 ATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATAT 419
              || |||| |||||||| | | | | | | | ||| | ||| ||||||| | ||||||||
Db      31158 ATATATAGACATATATATAGAGAGACATATATATAGAGACATATATATAGAGACATATAT 31099

Qy        420 ACACATATATACACATATATACACATATACACATATATACACATATATACATATATACAC 479
              | | | | ||| | ||| | | | ||||| |||||||||| | | ||||| ||||| |||
Db      31098 ATAGAGACATATATATAGACATATATATAGACATATATATAGACATATATATATAGACAT 31039

Qy        480 ATATATATA---ATATACACACATATATATACACATATATACACACATATATACACATAT 536
              ||||| | |   ||||| ||| ||||||| ||| |||||| || |||||||||| | | ||
Db      31038 ATATAGACATATATATAGACATATATATAGACATATATATGCAGACATATATATAGACAT 30979
```

```
Qy        537 ATACACATATATATACACATATATACACATATATACATATATACACATATATATACATAT 596
              ||| | ||| | ||| | ||  | ||| ||||||| | ||||| | | ||||||| | |
Db      30978 ATATATATAGACATATATATGCAGACATATATATAGACATATATAGACATATATATAGAC 30919

Qy        597 ATACACATATATACATATATACACATATATACATATATACACACATATATACACATACAT 656
              ||| | ||||| |||||||||  | | ||||| ||||| ||| | ||||||| | ||| ||
Db      30918 ATATATATATAGACATATATATAGACATATATATATAAACATATATATATATATATAT 30859

Qy        657 ATACACACACATAGATATACATATATATACACATATA 693
              ||| | | | | ||| | | | | | | | | | | |
Db      30858 ATATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA 30822
```

RESULT 63
ABL33999/c
ID ABL33999 standard; DNA; 21537 BP.
XX
AC ABL33999;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 1972.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 1972; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 21537 BP; 6538 A; 333 C; 4426 G; 10240 T; 0 U; 0 Other;

Query Match 6.1%; Score 250; DB 6; Length 21537;
Best Local Similarity 68.1%; Pred. No. 2.7e-33;
Matches 391; Conservative 0; Mismatches 180; Indels 3; Gaps 3;

```
Qy        181 CCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATA 240
              |||    | || |    || |  |||||     |  | |       | ||||||||||||
Db       6005 CCAAATTAAAATACAATAACGCGATCTCGATTCACTACACTATATATATATATATATATA 5946

Qy        241 TATATATATATACACACACACACATATATATATACACATATATATACACATATATACA 300
              ||||||||||||| ||||| |||| |||||||||||| | |||||||||| | |||||||
Db       5945 TATATATATATATACACATACAC-CATATATATATATATATATATATACATACATATACA 5887

Qy        301 TATATACACATATATACACATATATATACACATACATATGTACACATATATATACACATA 360
              |  |    |   || | | | | || | | |   |   | |||| | | ||| ||
Db       5886 AAACAATTACTTACTAAATACACAAATCAATAAAATAAAAAATTCATAAACAAACATATC 5827

Qy        361 TGTATACACATATATACACATATATACACACATATATACACATATATACACACATATATA 420
              | | |  |||||||| |  |||||  |      |||| | ||||||| | | |||||||
Db       5826 TCTCTTTACATATATATATCTATATCGATCACACTATATATATATATATATATATATA 5767

Qy        421 CACATATATACACATATATACACATATACACATATATACACATATATACATATATACACA 480
              |||| ||||| | ||||||| | ||| |||||||||| | ||||||| | | | | | |
Db       5766 CACACATATATATATATATATATACACACATATATATATATATATACACACATATA 5707

Qy        481 TATATATAATATACACACATATATATACACATATATACACACATATATACACATATATAC 540
              ||||||| | |||||||||||||||||| | ||||||| ||||| ||||| | |||||||
Db       5706 TATATAT-ACATACACACATATATATATATATATATACACACATATATATATATATAT 5648

Qy        541 ACATATATATACACATATATACACATATATACATATATACACATATATATACATATATAC 600
              | ||||||||| | ||||||| | ||||||| ||||||| | ||||| |  ||||||||
Db       5647 ATATATATATATATATATATATATATATATATATATATATATATATATACACGCATATATAT 5588

Qy        601 ACATATATACATATATACACATATATACATATATACACACATATATACACATACATATAC 660
              || |||||||||||||| | |   ||| ||||||| | ||||||||| | ||| |||||
Db       5587 ACGTATATACATATATATAAACTCATATATATATATATACATATATATATATATATAT 5528

Qy        661 ACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATAT 720
              | | | ||| |   ||| |||||| |  | | | ||      || | | ||||  |  ||
Db       5527 ATATATATAAACTCACA-ATATATCCTTAAACAAATTATATAATCTCTATATACCTCAAT 5469

Qy        721 GCTCCAGAGTTCATAAGAGGTAGCAGTTGATTAC 754
              ||   | |  | || |   |  | |  | |||
Db       5468 TTTCTCAAATATAAAACAAAAATAATTATAATAC 5435
```

RESULT 64
ADA02930
ID ADA02930 standard; DNA; 96596 BP.
XX
AC ADA02930;
XX
DT 06-NOV-2003 (first entry)
XX
DE Human PIK3R1 carcinoma associated gene, SEQ ID NO:1448.
XX
KW Human; carcinoma associated; oncogene; carcinoma; cancer; breast;
KW prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO2003057146-A2.
XX
PD 17-JUL-2003.
XX
PF 26-DEC-2002; 2002WO-US041414.
XX
PR 26-DEC-2001; 2001US-00035832.
XX
PA (SAGR-) SAGRES DISCOVERY.
XX
PI Morris DW;
XX
DR WPI; 2003-587068/55.
XX
PT New recombinant nucleic acid encoding carcinoma associated protein,
PT useful for preparing compositions for treating carcinomas.
XX
PS Claim 1; SEQ ID NO 1448; 245pp; English.

```
XX
CC   The invention relates to recombinant carcinoma associated (CA) nucleic
CC   acid sequences from mouse and human (ADA01482-ADA03094), and to
CC   recombinant carcinoma associated proteins (CAP) encoded by them. The
CC   invention also encompasses expression vectors and host cells comprising a
CC   CA nucleic acid, a polypeptide (especially an antibody) that specifically
CC   binds to the protein, and a biochip comprising CA nucleic acid or
CC   fragments thereof. The sequences of the invention were identified using
CC   oncogenic retroviruses, which insert into the genome of the host organism
CC   at random. Many of these do not carry transduced host oncogenes or
CC   pathogenic trans-acting viral genes, meaning that cancer incidence is a
CC   direct consequence of the effects of proviral integration into host
CC   protooncogenes. The CA nucleic acid sequences can be used to diagnose
CC   carcinoma (especially breast cancer, prostate cancer, lymphoma or
CC   leukaemia) or a propensity to carcinoma by determination of the sequence
CC   of a CA gene, or by determination of CA gene expression in particular
CC   tissues. CA nucleic acids, proteins and antibodies are also useful as
CC   therapeutic agents and in screening and evaluating drug candidates. The
CC   present sequence represents a specifically claimed human CA nucleic acid
CC   sequence of the invention. Note: The complete sequence data for this
CC   patent did not form part of the printed specification, but was obtained
CC   in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 96596 BP; 27639 A; 18627 C; 19687 G; 30643 T; 0 U; 0 Other;

  Query Match             6.1%;  Score 249.6;  DB 9;  Length 96596;
  Best Local Similarity   70.5%;  Pred. No. 3.3e-33;
  Matches  347;  Conservative    0;  Mismatches  144;  Indels    1;  Gaps    1;

Qy        229 TATATATATATATATATATATATACACACACACACACATATATATATACACATATATATA 288
              | | || ||||||| || || || ||    | | |||  | || ||||| |||| || ||
Db      57110 TGTGTACATATATACATGTAGATGCATGTAGATACATGTGTACATATATACATGTAGATG 57169

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              ||  || |||||| ||||||  || |||||  || ||| |   |||||| |  | ||||
Db      57170 CATGTAGATACATGTATACATGTAGATACATGTAGATACATGTATACATGTAGATACATG 57229

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              || |||||  |||  ||| | | || || | | || || | |||| |||||| ||||| |
Db      57230 TAGATACATGTATACATATATACATGTAGATACATGTAGATACATGTATACATATATACA 57289

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
                | ||| ||  | | ||||| |||| || | |||| || ||||||||||  || |||
Db      57290 TGTAGATACATGTATACATATATACATGTAGATACATGTATACATATATACATGTAGATA 57349

Qy        469 CATATATACACATATATATAATATACACACATATATATACACATATATACACACATATAT 528
              ||| |||||| ||||| ||  || | |||  || ||| |   ||| ||| | |||| ||
Db      57350 CATGTATACATATATACAT-GTAGATACATGTAGATACATGTATACATATATACATGTAG 57408

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACATATAT 588
              | |||| || | |||| |||| | | ||| ||  | ||| ||  ||||| | ||||| ||
Db      57409 ATACATGTAGATACATGTATACATATATACATGTAGATACATGTATATAGATACATACAT 57468

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
              |.| ||||| | ||||| |||||||||||  | |||| |||||||||  ||| ||||||
Db      57469 AGATATATAGATACATAGATACATATATTTATATATGTACATATATCTACATATATATCT 57528

Qy        649 ACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATAT 708
              | ||| | ||| |   ||||| |||| |||||||||  |  || |||||  | |||||||
Db      57529 ATATATACATATATCTACATATATATCTATATATATATATCTACATATATATCTATATAT 57588

Qy        709 GTATATATATAT 720
              ||| | ||||||
Db      57589 GTAGAAATATAT 57600


RESULT 65
ADB72668
ID   ADB72668 standard; DNA; 96596 BP.
XX
AC   ADB72668;
XX
DT   04-DEC-2003  (first entry)
XX
DE   Human PIK3R1 gene.
XX
KW   human; ds; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;
KW   cancer; neoplasm; adenocarcinoma; sarcoma; gene.
XX
OS   Homo sapiens.
XX
PN   WO2003008583-A2.
XX
PD   30-JAN-2003.
XX
PF   26-DEC-2001; 2001WO-US051291.
XX
PR   02-MAR-2001; 2001US-00798586.
PR   23-OCT-2001; 2001US-00004113.
PR   08-NOV-2001; 2001US-00052482.
PR   30-NOV-2001; 2001US-00997722.
PR   20-DEC-2001; 2001US-00034650.
XX
PA   (SAGR-) SAGRES DISCOVERY.
XX
PI   Morris DW,  Engelhard EK;
XX
DR   WPI; 2003-239337/23.
XX
PT   New recombinant nucleic acid, useful for treating carcinomas, lymphomas,
PT   cancers, neoplasm, adenocarcinoma, or sarcomas.
XX
PS   Claim 1; SEQ ID NO 496; 2304pp; English.
XX
CC   The invention relates to a novel recombinant nucleic acid comprising a
CC   nucleotide sequence selected from any of the 660 sequences fully defined
CC   in the specification. A polynucleotide of the invention has cytostatic
CC   activity, and may have a use in gene therapy, or in a vaccine. The
CC   recombinant nucleic acids and polypeptides are useful for treating
CC   carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and
CC   sarcomas. The present sequence represents a human gene of the invention.
XX
SQ   Sequence 96596 BP; 27639 A; 18627 C; 19687 G; 30643 T; 0 U; 0 Other;

  Query Match             6.1%;  Score 249.6;  DB 10;  Length 96596;
  Best Local Similarity   70.5%;  Pred. No. 3.3e-33;
  Matches  347;  Conservative    0;  Mismatches  144;  Indels    1;  Gaps    1;

Qy        229 TATATATATATATATATATATATACACACACACACACATATATATATACACATATATATA 288
              | | || ||||||| || || || ||    | | |||  | || ||||| |||| || ||
Db      57110 TGTGTACATATATACATGTAGATGCATGTAGATACATGTGTACATATATACATGTAGATG 57169

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              ||  || |||||| ||||||  || |||||  || ||| |   |||||| |  | ||||
Db      57170 CATGTAGATACATGTATACATGTAGATACATGTAGATACATGTATACATGTAGATACATG 57229

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              || |||||  |||  ||| | | || || | | || || | |||| |||||| ||||| |
Db      57230 TAGATACATGTATACATATATACATGTAGATACATGTAGATACATGTATACATATATACA 57289

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
                | ||| ||  | | ||||| |||| || | |||| || ||||||||||  || |||
Db      57290 TGTAGATACATGTATACATATATACATGTAGATACATGTATACATATATACATGTAGATA 57349

Qy        469 CATATATACACATATATATAATATACACACATATATATACACATATATACACACATATAT 528
              ||| |||||| ||||| ||  || | |||  || ||| |   ||| ||| | |||| ||
Db      57350 CATGTATACATATATACAT-GTAGATACATGTAGATACATGTATACATATATACATGTAG 57408

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACATATAT 588
              | |||| || | |||| |||| | | ||| ||  | ||| ||  ||||| | ||||| ||
Db      57409 ATACATGTAGATACATGTATACATATATACATGTAGATACATGTATATAGATACATACAT 57468

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
              | | ||||| | ||||| ||||||||||  | |||| |||||||||  ||| ||||||
Db      57469 AGATATATAGATACATAGATACATATATTTATATATGTACATATATCTACATATATATCT 57528
```

```
Qy        649 ACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATAT 708
              | ||| | ||| |   ||||| |||| |||||||| |  || |||||  | |||||||
Db      57529 ATATATACATATATCTACATATATATCTATATATATATATCTACATATATATCTATATAT 57588

Qy        709 GTATATATATAT 720
              ||| | ||||||
Db      57589 GTAGAAATATAT 57600


RESULT 66
ADC85410
ID   ADC85410 standard; DNA; 96596 BP.
XX
AC   ADC85410;
XX
DT   01-JAN-2004  (first entry)
XX
DE   Human Pik3r1 genomic sequence.
XX
KW   Cytostatic; gene therapy; vaccine; cancer; carcinoma-associated gene; CA;
KW   secreted; transmembrane; intracellular; ds.
XX
OS   Homo sapiens.
XX
PN   WO2003045230-A2.
XX
PD   05-JUN-2003.
XX
PF   02-DEC-2002; 2002WO-US038582.
XX
PR   30-NOV-2001; 2001US-00997722.
XX
PA   (SAGR-) SAGRES DISCOVERY.
XX
PI   Morris DW,  Engelhard EK;
XX
DR   WPI; 2003-513603/48.
XX
PT   New recombinant nucleic acid comprising a nucleotide sequence of any of
PT   the carcinoma-associated (CA) genes, useful for screening for drug
PT   candidates for diagnosing or treating carcinomas.
XX
PS   Claim 1; SEQ ID NO 196; 983pp; English.
XX
CC   The invention relates to a recombinant nucleic acid comprising a
CC   nucleotide sequence selected from any of the fully defined carcinoma-
CC   associated (CA) genes from the 50 tables given in the specification. The
CC   CA proteins are secreted, transmembrane or intracellular proteins. The
CC   recombinant nucleic acids are useful for screening for drug candidates
CC   for diagnosing or treating carcinomas. Sequences given in ADC85215-
CC   ADC85514 represent CA genes of the invention.
XX
SQ   Sequence 96596 BP; 27639 A; 18627 C; 19687 G; 30643 T; 0 U; 0 Other;

  Query Match             6.1%;  Score 249.6;  DB 10;  Length 96596;
  Best Local Similarity   70.5%;  Pred. No. 3.3e-33;
  Matches  347;  Conservative    0;  Mismatches  144;  Indels    1;  Gaps    1;

Qy        229 TATATATATATATATATATATATACACACACACACACATATATATATACACATATATATA 288
              | | || ||||||| || || || ||    | | |||  | || ||||| |||| || ||
Db      57110 TGTGTACATATATACATGTAGATGCATGTAGATACATGTGTACATATATACATGTAGATG 57169

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              ||  || |||||| ||||||  || |||||  || ||| |   |||||| |  | ||||
Db      57170 CATGTAGATACATGTATACATGTAGATACATGTAGATACATGTATACATGTAGATACATG 57229

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              || |||||  |||  ||| | | || || | | || || | |||| |||||| ||||| |
Db      57230 TAGATACATGTATACATATATACATGTAGATACATGTAGATACATGTATACATATATACA 57289

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
                | ||| ||  | | ||||| |||| || | |||| || ||||||||||  || |||

Db      57290 TGTAGATACATGTATACATATATACATGTAGATACATGTATACATATATACATGTAGATA 57349

Qy        469 CATATATACACATATATATAATATACACACATATATATACACATATATACACACATATAT 528
              ||| |||||| ||||| ||  || | |||  || ||| |   ||| ||| | |||| ||
Db      57350 CATGTATACATATATACAT-GTAGATACATGTAGATACATGTATACATATATACATGTAG 57408

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACATATAT 588
              | |||| || | |||| |||| | | ||| ||  | ||| ||  ||||| | ||||| ||
Db      57409 ATACATGTAGATACATGTATACATATATACATGTAGATACATGTATATAGATACATACAT 57468

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
              | | ||||| | ||||| ||||||||||  | |||| |||||||||  ||| ||||||
Db      57469 AGATATATAGATACATAGATACATATATTTATATATGTACATATATCTACATATATATCT 57528

Qy        649 ACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATAT 708
              | ||| | ||| |   ||||| |||| |||||||| |  || |||||  | |||||||
Db      57529 ATATATACATATATCTACATATATATCTATATATATATATCTACATATATATCTATATAT 57588

Qy        709 GTATATATATAT 720
              ||| | ||||||
Db      57589 GTAGAAATATAT 57600


RESULT 67
ADM74525
ID   ADM74525 standard; DNA; 96596 BP.
XX
AC   ADM74525;
XX
DT   01-JUL-2004  (first entry)
XX
DE   Human carcinoma associated (CA) nucleic acid #97.
XX
KW   Human; carcinoma associated nucleic acid; CA nucleic acid; gene; ds;
KW   carcinoma associated protein; CAP; carcinoma; leukaemia; lymphoma;
KW   cytostatic.
XX
OS   Homo sapiens.
XX
PN   US2004072154-A1.
XX
PD   15-APR-2004.
XX
PF   30-NOV-2001; 2001US-00997722.
XX
PR   22-DEC-2000; 2000US-00747377.
PR   02-MAR-2001; 2001US-00798586.
XX
PA   (MORR/) MORRIS D W.
PA   (ENGE/) ENGELHARD E K.
XX
PI   Morris DW,  Engelhard EK;
XX
DR   WPI; 2004-328562/30.
XX
PT   New carcinoma associated gene or protein, useful for preparing a
PT   composition for diagnosing or treating carcinoma e.g., leukemia or
PT   lymphoma.
XX
PS   Claim 1; SEQ ID NO 196; 29pp; English.
XX
CC   The invention relates to new recombinant nucleic acids. The invention
CC   also relates to a host cell comprising a recombinant nucleic acid or
CC   expression vector, an expression vector comprising a recombinant nucleic
CC   acid, a recombinant protein, a method of screening for drug candidates, a
CC   method of screening for a bioactive agent capable of binding to a
CC   carcinoma associated protein (CAP) encoded by a nucleotide sequence, a
CC   method of screening for a bioactive agent capable of modulating the
CC   activity of a CAP, a method of evaluating the effect of a candidate
CC   carcinoma drug, a method of diagnosing carcinoma, a method for inhibiting
CC   the activity of a CAP, a method of treating carcinomas, a method of
CC   neutralising the effect of a CAP and a method of diagnosing carcinoma or
CC   propensity to carcinoma. A method of evaluating the effect of a candidate
```

CC carcinoma drug comprises administering the drug to a patient, removing a
CC cell sample from the patient and determining alterations in the
CC expression or activation of a gene comprising the nucleotide sequence. A
CC method of diagnosing carcinoma comprises determining the expression of
CC one or more genes comprising the nucleic acid sequence in a first tissue
CC type of a first individual and comparing the expression of the gene from
CC a second normal tissue type from the first individual or a second
CC unaffected individual, where a difference in the expression indicates
CC that the first individual has carcinoma. A method of inhibiting the
CC activity of a CAP comprises binding an inhibitor to the CAP. Treating
CC carcinomas comprises administering to a patient an inhibitor of CAP.
CC Neutralising the effect of a CAP comprises contacting an agent specific
CC for the CAP. The polypeptide specifically binds to the protein encoded by
CC the nucleic acid. It comprises an antibody that specifically binds to the
CC protein encoded by the nucleic acid. The nucleic acids are useful for
CC preparing a composition for diagnosing or treating carcinoma e.g.,
CC leukaemia or lymphoma. This sequence represents a human carcinoma
CC associated (CA) nucleic acid of the invention. Note: The sequence data
CC for this patent did not form part of the printed specification but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 96596 BP; 27639 A; 18627 C; 19687 G; 30643 T; 0 U; 0 Other;

Query Match 6.1%; Score 249.6; DB 12; Length 96596;
Best Local Similarity 70.5%; Pred. No. 3.3e-33;
Matches 347; Conservative 0; Mismatches 144; Indels 1; Gaps 1;

Qy 229 TATATATATATATATATATATATACACACACACACACATATATATATACACATATATATA 288
Db 57110 TGTGTACATATATACATGTAGATGCATGTAGATACATGTGTACATATATACATGTAGATG 57169

Qy 289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
Db 57170 CATGTAGATACATGTATACATGTAGATACATGTAGATACATGTATACATGTAGATACATG 57229

Qy 349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
Db 57230 TAGATACATGTATACATATATACATGTAGATACATGTAGATACATGTATACATATATACA 57289

Qy 409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
Db 57290 TGTAGATACATGTATACATATATACATGTAGATACATGTATACATATATACATGTAGATA 57349

Qy 469 CATATATACACATATATATAATATACACACATATATATACACATATATACACACATATAT 528
Db 57350 CATGTATACATATATACAT-GTAGATACATGTAGATACATGTATACATATATACATGTAG 57408

Qy 529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACATATAT 588
Db 57409 ATACATGTAGATACATGTATACATATATACATGTAGATACATGTATATAGATACATACAT 57468

Qy 589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
Db 57469 AGATATATAGATACATAGATACATATATTTATATATGTACATATATCTACATATATATCT 57528

Qy 649 ACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATAT 708
Db 57529 ATATATACATATATCTACATATATATCTATATATATATATCTACATATATATCTATATAT 57588

Qy 709 GTATATATATAT 720
Db 57589 GTAGAAATATAT 57600

RESULT 68
ADC87556/c
ID ADC87556 standard; DNA; 1342 BP.
XX
AC ADC87556;
XX
DT 01-JAN-2004 (first entry)
XX
DE Human GPCR gene SEQ ID NO:2009.
XX
KW ds; gene; human; GPCR;
KW guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1270724-A2.
XX
PD 02-JAN-2003.
XX
PF 18-JUN-2002; 2002EP-00013517.
XX
PR 18-JUN-2001; 2001JP-00246789.
XX
PA (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI Suwa M, Asai K, Akiyama Y, Aburatani H;
XX
DR WPI; 2003-315783/31.
DR P-PSDB; ADC87557.
XX
PT New polynucleotide, useful for preparing a composition for treating a
PT patient in need of increased or suppressed activity or expression of the
PT guanosine triphosphate-binding protein coupled receptor.
XX
PS Claim 1; SEQ ID NO 2009; 28pp; English.
XX
CC The invention relates to a novel polynucleotide encoding a guanosine
CC triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC the invention may have a use in gene therapy. The polynucleotide and
CC polypeptide are useful for preparing a composition for treating a patient
CC in need of increased or suppressed activity or expression of the
CC guanosine triphosphate-binding protein coupled receptor. The
CC polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC invention.
XX
SQ Sequence 1342 BP; 397 A; 191 C; 193 G; 436 T; 0 U; 125 Other;

Query Match 6.0%; Score 248.4; DB 10; Length 1342;
Best Local Similarity 69.9%; Pred. No. 4.5e-33;
Matches 348; Conservative 0; Mismatches 143; Indels 7; Gaps 2;

Qy 229 TATATATATATATATATATATATACACACACACACACATATATATATACACATATATATA 288
Db 891 TACACACACATATGTGTATATATACATATACACACCTACACACATATATATGTGTATATA 832

Qy 289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
Db 831 TACATATACACACACATATATGTGTATATATACATATACATACATATGTGTATACACATG 772

Qy 349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
Db 771 TATATGTGCACATATGTATACGTATATGCACATATACATACATATATGTATACATATATA 712

Qy 409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
Db 711 CACATATATGTGCATGTATGTATACGTATGTGCACATATGTACATGTGTATGCATAAATA 652

Qy 469 CATATATACACATATATATA-ATATACACACATATATATACA------CATATATACACA 521
Db 651 CATATATGTACATGTGTATACATAAATACACATATATGTACATGCGTGTACATAAATACA 592

Qy 522 CATATATACACATATATACACATATATATACACATATATACACATATATACATATATACA 581
Db 591 CATATATGTACATGCGTATACATAAATACACATATATGTACATGTGTATACATATATACA 532

Qy 582 CATATATATACATATATACACATATATACATATATACACATATATACATATATACACACA 641
Db 531 CATATATGTACATGCGTATACATATATACACATATGTACATGTGTGTACATAAATACACA 472

Qy 642 TATATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTA 701

```
                ||| |||| ||   || | | | ||| | || |||||
Db        471 TATGTACATATGTGTACATAAATACACATATGTACATNNNNNNNNNNNNNNNNNNNNNNG 412

Qy        702 TATATATGTATATATATA 719
              || || | ||||||||||
Db        411 TACATGTATATATATATA 394


RESULT 69
ADC87320/c
ID   ADC87320 standard; DNA; 1363 BP.
XX
AC   ADC87320;
XX
DT   01-JAN-2004  (first entry)
XX
DE   Human GPCR gene SEQ ID NO:1773.
XX
KW   ds; gene; human; GPCR;
KW   guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1270724-A2.
XX
PD   02-JAN-2003.
XX
PF   18-JUN-2002; 2002EP-00013517.
XX
PR   18-JUN-2001; 2001JP-00246789.
XX
PA   (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA   (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI   Suwa M,  Asai K,  Akiyama Y,  Aburatani H;
XX
DR   WPI; 2003-315783/31.
DR   P-PSDB; ADC87321.
XX
PT   New polynucleotide, useful for preparing a composition for treating a
PT   patient in need of increased or suppressed activity or expression of the
PT   guanosine triphosphate-binding protein coupled receptor.
XX
PS   Claim 1; SEQ ID NO 1773; 28pp; English.
XX
CC   The invention relates to a novel polynucleotide encoding a guanosine
CC   triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC   the invention may have a use in gene therapy. The polynucleotide and
CC   polypeptide are useful for preparing a composition for treating a patient
CC   in need of increased or suppressed activity or expression of the
CC   guanosine triphosphate-binding protein coupled receptor. The
CC   polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC   invention.
XX
SQ   Sequence 1363 BP; 424 A; 175 C; 269 G; 495 T; 0 U; 0 Other;

  Query Match             6.0%;  Score 248.2;  DB 10;  Length 1363;
  Best Local Similarity   71.7%;  Pred. No. 4.8e-33;
  Matches  367;  Conservative    0;  Mismatches  139;  Indels    6;  Gaps    3;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | || | ||||||| |||||| || | | | || |||| |   |||||||   | |||||
Db       1136 TGTACACATATATACATATATGTACATATATACGCACATACGCATATATATGTACATATA 1077

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              |||||| |||  |||||||  ||||||||||||||||  ||| |  ||||||| ||||||
Db       1076 TACACACATACGCATATATGTACATATATACACATATGCATATATGTACATATATACACA 1017

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||  ||| |  ||  |||| ||| |||| || |||| |||| | ||| | || ||||||
Db       1016 TATGCATATATGTACATATATACACATATGCATATATGTACATATATACACACGCATATA 957

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
```

```
              |  ||| ||||| ||||||||    || |||||     |||||||| ||| | |  | ||
Db        956 TGTACATATATACACACATAT----ACGTATATGTGTGTATACACACATACATATGTGTA 901

Qy        467 TACATATATACACATATATATAATATACACACATATATATACACATATATACACACATAT 526
              |||| | ||||  |||||  | || |||||| ||| ||||| || |||||| ||||||||
Db        900 TACACACATACGTATATACGT-ATGTACACATATACATATATACGTATATATACACATAT 842

Qy        527 ATACACATATATACACATATATATACACATATATACACATATATACATATATACACATAT 586
              | ||  ||||| |||||||||| ||| |  || ||| ||  | ||| | |||||    ||
Db        841 ACACCTATATACACACATATACATATATGTACATATACGCACATATACATATATGTGTAC 782

Qy        587 ATATACATATATACACATATATACATATATACACATATATACA-TATATACACACATATA 645
              ||||| | | ||||| ||||  |||||||||||||| || |||| |||||||||||||| ||
Db        781 ATATATACACATACATATATGCACATATATACACACATGTACACTATATACACACATGTA 722

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              ||| || || ||| | ||||| |   | |||  || | ||||  | | ||| ||| | |
Db        721 CACATATGCACATATATACACACATGCACACACGTACACACACGCACACATATGTACACA 662

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAG 737
              ||||| | || | | | |    |  |||| |
Db        661 CATGTACACATGTGTACACATATGCCCATATG 630


RESULT 70
ABT10748
ID   ABT10748 standard; cDNA; 121600 BP.
XX
AC   ABT10748;
XX
DT   04-DEC-2002  (first entry)
XX
DE   Human breast cancer associated coding sequence SEQ ID NO: 882.
XX
KW   Human; breast specific gene; breast cancer; differential expression;
KW   cytostatic; gene therapy; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200259271-A2.
XX
PD   01-AUG-2002.
XX
PF   25-JAN-2002; 2002WO-US002176.
XX
PR   25-JAN-2001; 2001US-0263757P.
PR   25-APR-2001; 2001US-0286090P.
PR   23-MAY-2001; 2001US-0292517P.
XX
PA   (GENE-) GENE LOGIC INC.
XX
PI   Orr MS,  Nation M,  Diggans JC,  Zeng W;
XX
DR   WPI; 2002-674803/72.
XX
PT   Diagnosing breast cancer in a patient comprises detecting the level of
PT   gene expression in cell or tissue samples, where a differential gene
PT   expression is indicative of breast cancer.
XX
PS   Claim 1; SEQ ID NO 882; 260pp + Sequence Listing; English.
XX
CC   The present invention relates to methods of diagnosing breast cancer in a
CC   patient, which comprise detecting the level of expression in a tissue
CC   sample of two or more genes selected from those shown in ABT09867-
CC   ABT11112, where a differential expression of the genes indicates breast
CC   cancer. The methods are useful in diagnosing, treating, detecting the
CC   progression, and in monitoring treatment of breast cancer in patients.
CC   The methods are also useful as a screening tool for agents that modulate
CC   the onset or progression of breast cancer. The breast cancer genes may be
CC   used as diagnostic markers for the prediction or identification of the
CC   malignant state of breast tissue, for confirming the type and progression
CC   of cancer, and for drug screening and assays. The present sequence is a
CC   coding sequence of the invention. Note: The sequence data for this patent
```

```
CC   did not form part of the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub.published_pct_sequences
XX
SQ   Sequence 121600 BP; 38372 A; 24652 C; 24048 G; 34528 T; 0 U; 0 Other;

  Query Match             6.0%;  Score 247.4;  DB 6;  Length 121600;
  Best Local Similarity   68.6%;  Pred. No. 7.9e-33;
  Matches  341;  Conservative    0;  Mismatches  156;  Indels    0;  Gaps    0;

Qy        224 CATTTTATATATATATATATATATATACACACACACACACATATATATATACACATAT 283
              ||  |||||||||| |||||||  ||||||  |   | | |   ||||||  ||||  | ||  |
Db      43894 CAGCCTATATATATCTATATATCTATATATATCTATATATCTATATATCTATATATATCT 43953

Qy        284 ATATACACATATATACATATATACACATATATACACATATATATACACATACATATGTAC 343
              |||||   |||||||  ||||||   ||||||  | || | |||| |  || |||| ||
Db      43954 ATATATCTATATATATCTATATATCTATATATCTATATCTCTATATATCTATATATCTAT 44013

Qy        344 ACATATATATACACATATGTATACACATATATACACATATATACACACATATATACACAT 403
              |  | ||||||   || | ||||   |||||   ||||||  | | || |||| | ||
Db      44014 ATCTCTATATATCTATCTATATATCTCTATATCTCTATATATCTATATATCTATATATAT 44073

Qy        404 ATATACACACATATATACACATATATACACATATATACACATATACACATATATACACAT 463
               |||| |   ||||||  | |||| || | |||| || | || || | || ||||  |  |
Db      44074 CTATATATCTATATATCTATATATCTATATATATCTATATATCTATATATCTATATATCT 44133

Qy        464 ATATACATATATACACATATATATAATATACACACATATATATACACATATATACACACA 523
              |||||  |||||| |  |||||||  || || | |  |||||||  | |||| || | |
Db      44134 ATATATCTATATATATCTATATATCATCTATATATCTATATATCTATATATCTATATATC 44193

Qy        524 TATATACACATATATACACATATATATACACATATATACACATATATACATATATACACA 583
              ||||||   ||||||  | ||||||||  | ||||||  | |||| || ||||||  | |
Db      44194 TATATATCTATATATCTATATATATATCTATATATATCTATATATCTATATATATCTATA 44253

Qy        584 TATATATACATATATACACATATATACATATATACACATATATACATATATACACACATA 643
              |||  |||| || |||| | || |||| || |||| | || |||||||  |||| |   |||
Db      44254 TATCTATATATCTATATATATCTATATATCTATATATATCTATACATCTATATATCTATA 44313

Qy        644 TATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATA 703
              ||||   |||  || || | |   ||| ||||  |||||| || | |||| ||||  | ||
Db      44314 TATATCTATATATCTATATATCTATATATATCTATATATCTATATATATCTATATATCTA 44373

Qy        704 TATATGTATATATATAT 720
              ||||| | ||||||| |
Db      44374 TATATATCTATATATCT 44390


RESULT 71
ADQ18307
ID   ADQ18307 standard; DNA; 121600 BP.
XX
AC   ADQ18307;
XX
DT   26-AUG-2004  (first entry)
XX
DE   Human soft tissue sarcoma-upregulated DNA - SEQ ID 1125.
XX
KW   soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO2004048938-A2.
XX
PD   10-JUN-2004.
XX
PF   26-NOV-2003; 2003WO-US038193.
XX
PR   26-NOV-2002; 2002US-0429739P.
XX
PA   (PROT-) PROTEIN DESIGN LABS INC.
XX
PI   Aziz N,  Ginsburg WM,  Zlotnik A;
XX
DR   WPI; 2004-441208/41.
XX
PT   Early detection of soft tissue sarcoma comprises determining expression
PT   of a gene in a first soft tissue sample and a normal soft tissue sample
PT   and comparing the gene expression, also useful in treating soft tissue
PT   sarcoma.
XX
PS   Example 2; SEQ ID NO 1125; 210pp; English.
XX
CC   The invention relates to a novel method for detecting soft tissue sarcoma
CC   which comprises obtaining a first soft tissue sample from an individual
CC   and a normal soft tissue sample from the same or different individual,
CC   determining the expression of a gene in both samples and comparing the
CC   expression of the gene in both soft tissue samples, where a higher level
CC   of protein expression in the first soft tissue sample indicates the
CC   presence of soft tissue sarcoma. The method of the invention has
CC   cytostatic applications and may be useful for detecting soft tissue
CC   sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC   acid sequences may be useful in diagnostic and screening applications.
CC   The current sequence is that of a human soft tissue sarcoma-upregulated
CC   DNA of the invention. The current sequence is not shown within the
CC   specification per se but was submitted in CD format by the inventor.
XX
SQ   Sequence 121600 BP; 38372 A; 24652 C; 24048 G; 34528 T; 0 U; 0 Other;

  Query Match             6.0%;  Score 247.4;  DB 12;  Length 121600;
  Best Local Similarity   68.6%;  Pred. No. 7.9e-33;
  Matches  341;  Conservative    0;  Mismatches  156;  Indels    0;  Gaps    0;

Qy        224 CATTTTATATATATATATATATATATACACACACACACACATATATATATACACATAT 283
              ||  |||||||||| |||||||  ||||||  |   | | |   ||||||  ||||  | ||  |
Db      43894 CAGCCTATATATATCTATATATCTATATATATCTATATATCTATATATCTATATATATCT 43953

Qy        284 ATATACACATATATACATATATACACATATATACACATATATATACACATACATATGTAC 343
              |||||   |||||||  ||||||   ||||||  | || | |||| |  || |||| ||
Db      43954 ATATATCTATATATATCTATATATCTATATATCTATATCTCTATATATCTATATATCTAT 44013

Qy        344 ACATATATATACACATATGTATACACATATATACACATATATACACACATATATACACAT 403
              |  | ||||||   || | ||||   |||||   ||||||  | | || |||| | ||
Db      44014 ATCTCTATATATCTATCTATATATCTCTATATCTCTATATATCTATATATCTATATATAT 44073

Qy        404 ATATACACACATATATACACATATATACACATATATACACATATACACATATATACACAT 463
               |||| |   ||||||  | |||| || | |||| || | || || | || ||||  |  |
Db      44074 CTATATATCTATATATCTATATATCTATATATATCTATATATCTATATATCTATATATCT 44133

Qy        464 ATATACATATATACACATATATATAATATACACACATATATATACACATATATACACACA 523
              |||||  |||||| |  |||||||  || || | |  |||||||  | |||| || | |
Db      44134 ATATATCTATATATATCTATATATCATCTATATATCTATATATCTATATATCTATATATC 44193

Qy        524 TATATACACATATATACACATATATATACACATATATACACATATATACATATATACACA 583
              ||||||   ||||||  | ||||||||  | ||||||  | |||| || ||||||  | |
Db      44194 TATATATCTATATATCTATATATATATCTATATATATCTATATATCTATATATATCTATA 44253

Qy        584 TATATATACATATATACACATATATACATATATACACATATATACATATATACACACATA 643
              |||  |||| || |||| | || |||| || |||| | || |||||||  |||| |   |||
Db      44254 TATCTATATATCTATATATATCTATATATCTATATATATCTATACATCTATATATCTATA 44313

Qy        644 TATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATA 703
              ||||   |||  || || | |   ||| ||||  |||||| || | |||| ||||  | ||
Db      44314 TATATCTATATATCTATATATCTATATATATCTATATATCTATATATATCTATATATCTA 44373

Qy        704 TATATGTATATATATAT 720
              ||||| | ||||||| |
Db      44374 TATATATCTATATATCT 44390


RESULT 72
ACC79619
ID   ACC79619 standard; DNA; 71843 BP.
XX
AC   ACC79619;
XX
```

DT 27-OCT-2003 (revised)
DT 06-AUG-2003 (first entry)
XX
DE Zebrafish protein kinase C lambda related genomic DNA SEQ ID NO:3.
XX
KW Zebrafish; Danio rerio; protein kinase C lambda; PKC-lambda; enzyme;
KW cardiant; cytostatic; nephrotropic; heart disease; cancer; organ failure;
KW epithelial-epithelial cell interaction; epithelial cell polarity;
KW heart failure; kidney failure; gene; ds.
XX
OS Danio rerio.
XX
PN WO2003023048-A2.
XX
PD 20-MAR-2003.
XX
PF 06-SEP-2002; 2002WO-US028410.
XX
PR 06-SEP-2001; 2001US-0317653P.
XX
PA (GEHO) GEN HOSPITAL CORP.
XX
PI Peterson R, Fishman MC;
XX
DR WPI; 2003-313257/30.
XX
PT Diagnosing or prognosing a disease or condition related to Protein Kinase
PT C lambda, by analyzing sample nucleic acid obtained from the test subject
PT to determine mutation in gene encoding Protein kinase C lambda.
XX
PS Disclosure; Page 63-83; 88pp; English.
XX
CC The present invention describes a method (M1) for determining whether a
CC test subject has, or is at risk of developing, a disease or condition
CC related to protein kinase C lambda (PKC-lambda) (I). M1 comprises
CC analysing a nucleic acid molecule of a sample from the subject to
CC determine whether the subject has a mutation in a gene encoding (I),
CC where the presence of a mutation indicates that the subject has, or is at
CC risk of developing, a disease or condition related to (I). (I) has
CC cardiant, cytostatic and nephrotropic activities. M1 is useful for
CC determining whether a test subject has, or is at risk of developing, a
CC disease or condition related to (I), e.g. heart disease, cancer, or
CC condition associated with epithelial-epithelial cell interactions or
CC epithelial cell polarity. (I) is useful for treating organ failure such
CC as heart or kidney failure. (I) is also useful for raising antibodies
CC specific for various regions of (I). M1 is suitable for detecting an
CC increased likelihood of diseases or conditions associated with (I), so
CC that appropriate intervention can be instituted before any symptoms
CC occur. This is useful with patients in high risk groups for such diseases
CC or conditions. Diagnostic method facilitates determination of the
CC etiology of such an existing disease or condition in a patient, so that
CC an appropriate approach to treatment can be selected. The present
CC sequence represents zebrafish genomic DNA for PKC-lambda, which is used
CC in the exemplification of the present invention. (Updated on 27-OCT-2003
CC to standardise OS field)
XX
SQ Sequence 71843 BP; 24017 A; 12903 C; 12592 G; 21955 T; 0 U; 376 Other;

Query Match 6.0%; Score 246.6; DB 8; Length 71843;
Best Local Similarity 63.1%; Pred. No. 1.1e-32;
Matches 313; Conservative 0; Mismatches 181; Indels 2; Gaps 2;

```
Qy      227 TTTATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
            | |||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db    68244 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 68303

Qy      287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
            || | ||||||| |||||||||||| | | ||||| | | | ||||| ||| |
Db    68304 TATATATATATATATATATACACACACACACACACACACACACACACACACANNNNNNNN 68363

Qy      347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406

Db    68364 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 68423

Qy      407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
                                                    ||||| | ||||||| | |||||
Db    68424 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNATATATATATATATATATATA 68483

Qy      467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
            || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db    68484 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 68543

Qy      526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
            |||| | ||||||| | |||| |||| | ||||||| | ||||||| ||||||| | |||
Db    68544 TATATATATATATATATATATNTATATATATATATATATATATATATATATATATATATA 68603

Qy      586 TATATACATATATACACATATATACATATATA-CACATATATACATATATACACACATAT 644
            |||||| ||||||| | ||||||| ||||||| |||||| ||||||| | | ||||
Db    68604 TATATATATATATATATATATATATATATATAGTGTATATATGTATATATATATATATAT 68663

Qy      645 ATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATAT 704
            ||| | ||| ||||| | | ||| ||||| |||||||||| | |||||||||  |||||
Db    68664 ATATATATATATATATATNTATATATATATATATATATATATATATATATATATATATAT 68723

Qy      705 ATATGTATATATATAT 720
            |||| ||||||||||
Db    68724 ATATATATATATATAT 68739
```

RESULT 73
AAS46591/c
ID AAS46591 standard; DNA; 6286 BP.
XX
AC AAS46591;
XX
DT 18-DEC-2001 (first entry)
XX
DE Tumour suppressor gene derived chemically modified sequence #313.
XX
KW Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer;
KW tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW cytosine methylation; ds.
XX
OS Homo sapiens.
XX
PN WO200168912-A2.
XX
PD 20-SEP-2001.
XX
PF 15-MAR-2001; 2001WO-EP002955.
XX
PR 15-MAR-2000; 2000DE-01013847.
PR 06-APR-2000; 2000DE-01019058.
PR 07-APR-2000; 2000DE-01019173.
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2001-602752/68.
XX
PT Fragments of chemically modified genes associated with tumor suppressor
PT genes and oncogenes, useful in designing primers and probes for analyzing
PT diseases associated with cytosine methylation state e.g. cancer.
XX
PS Claim 1; SEQ ID NO 313; 27pp; English.
XX
CC The invention relates to a nucleic acid comprising a sequence of 18
CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC bisulphite, of genes associated with tumour suppression and oncogenes
CC having a sequence taken from 536 (actually 533 since numbers 408, 458 and
CC 500 are missing from the sequence listing) sequences (Ss) and sequences
CC complementary to (Ss). The nucleic acid may be a peptide nucleic acid-
CC oligomer (PNA) of at least 9 nucleotides and may form part of a set of

```
CC   probes for detecting the cytosine methylation state and/or single
CC   nucleotide polymorphisms and also to be used in an array for analysing
CC   diseases associated with CpG dinucleotides e.g. cancers and tumours. The
CC   probes can also be used in a method for ascertaining genetic and/or
CC   epigenetic parameters for the diagnosis and/or therapy of existing
CC   diseases or the predisposition to specific diseases, by analysing
CC   cytosine methylations. The parameters may be compared to another set of
CC   genetic and/or epigenetic parameters, the differences serving as basis
CC   for diagnosis and/or prognosis events which are disadvantageous to
CC   patients. The present sequence is one of the 533 genomic sequences
CC   derived from tumour suppressor genes and oncogenes. Note: The sequence
CC   data for this patent did not form part of the printed specification, but
CC   was obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 6286 BP; 1508 A; 124 C; 1538 G; 3116 T; 0 U; 0 Other;

  Query Match             5.9%;  Score 245;  DB 4;  Length 6286;
  Best Local Similarity   57.3%;  Pred. No. 1.8e-32;
  Matches  443;  Conservative    0;  Mismatches  330;  Indels    0;  Gaps    0;

Qy        174 CTCCAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATAT 233
              | ||  ||||  |       |    | ||  |  | | | |||  ||  |   |  | |||
Db       4589 CACCCACCCATCCCCATTCATAAACATTTACTATAAAATAAATCAATAAATAAATAATAA 4530

Qy        234 ATATATATATATATATATACACACACACACATATATATATACACATATATATACACAT 293
              ||  ||| ||| ||| ||| | | | | | ||||| ||| ||| | | ||| ||| | |
Db       4529 ATTAATAAATAAATAAATAAATAAATAAATACATAAATAAATAAATAAATAAATAAATAA 4470

Qy        294 ATATACATATATACACATATATACACATATATATACACATACATATGTACACATATATAT 353
              ||| | | ||| | | ||| |||||| | ||| ||| | | | | ||  | | ||| |||
Db       4469 ATAAATAAATAAATAAATAAATACATAAATAAATAAATAAATAAATAAATAAATAAATAA 4410

Qy        354 ACACATATGTATACACATATATACACATATATACACACATATATACACATATATACACAC 413
              | | |||  || | ||||| ||| | | ||| | | | | ||| | | ||| ||| | |
Db       4409 ATAAATAAATAAATACATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 4350

Qy        414 ATATATACACATATATACACATATATACACATATACACATATATACACATATATACATAT 473
              ||| ||||| | ||| | | ||| ||| | | ||| | | ||| | | ||| ||| |||
Db       4349 ATAAATACATAAATAAATAAATAAATAAATAAATAAATAAATACATAAATAAATAAATAA 4290

Qy        474 ATACACATATATATAATATACACACATATATATACACATATATACACACATATATACACA 533
              ||| | | ||| ||||   |  ||| ||| ||| | | ||| | | | | ||| | | |
Db       4289 ATAAATAAATAAATAAATAAACAACAAATAAATAAATAAATAAATAAATAAATAAATAAA 4230

Qy        534 TATATACACATATATATACACATATATACACATATATACATATATACACATATATATACA 593
              || ||| | | ||| ||| | | ||| | | ||| ||| ||| ||| | | ||| ||| |
Db       4229 TAAATAAATACATAAATAAATAAATAAATAAATACATAAATAAATAAATAAATAAATAAA 4170

Qy        594 TATATACACATATATACATATATACACATATATACATATATACACACATATATACACATA 653
              || ||| | | ||| | | ||| | | ||| ||| ||| ||| | | ||| ||| | | |
Db       4169 TAAATAAATAAATACATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAA 4110

Qy        654 CATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATA 713
              | ||| | ||| | | | ||| ||| ||| | | ||| ||| |   ||| |||  || |
Db       4109 TAAATAAATACATAAATAAATAAATAAATAAATAAATAAATAAATAAATACATAAATAAA 4050

Qy        714 TATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAA 773
              || |||          | |  ||||          |  | |     | | |   |||
Db       4049 TAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATACATAAATAAATAAA 3990

Qy        774 GAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGG 833
              | |       | |  | |  ||    | |   | |  | |          |||||
Db       3989 TAAATAAATAAATAAATAAATAAATAAATAAATACATAAATAAATAAATAAATACATAAA 3930

Qy        834 GAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATAT 893
              |||||  ||||||    |||  |||    ||   |||    | | || ||    | | || ||
Db       3929 TAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAACAACAAATAAAT 3870

Qy        894 ATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATG 946
              | ||| |||   |||  | |  | |  ||||  |       ||  |  |||| |
Db       3869 AAATAAATAAATAAATAAATAAATAAATAAATACATAAATAAATAAACG 3817
```

```
RESULT 74
ADC87574/c
ID   ADC87574 standard; DNA; 1309 BP.
XX
AC   ADC87574;
XX
DT   01-JAN-2004  (first entry)
XX
DE   Human GPCR gene SEQ ID NO:2027.
XX
KW   ds; gene; human; GPCR;
KW   guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1270724-A2.
XX
PD   02-JAN-2003.
XX
PF   18-JUN-2002; 2002EP-00013517.
XX
PR   18-JUN-2001; 2001JP-00246789.
XX
PA   (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA   (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI   Suwa M,  Asai K,  Akiyama Y,  Aburatani H;
XX
DR   WPI; 2003-315783/31.
DR   P-PSDB; ADC87575.
XX
PT   New polynucleotide, useful for preparing a composition for treating a
PT   patient in need of increased or suppressed activity or expression of the
PT   guanosine triphosphate-binding protein coupled receptor.
XX
PS   Claim 1; SEQ ID NO 2027; 28pp; English.
XX
CC   The invention relates to a novel polynucleotide encoding a guanosine
CC   triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC   the invention may have a use in gene therapy. The polynucleotide and
CC   polypeptide are useful for preparing a composition for treating a patient
CC   in need of increased or suppressed activity or expression of the
CC   guanosine triphosphate-binding protein coupled receptor. The
CC   polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC   invention.
XX
SQ   Sequence 1309 BP; 463 A; 205 C; 200 G; 441 T; 0 U; 0 Other;

  Query Match             5.9%;  Score 243.2;  DB 10;  Length 1309;
  Best Local Similarity   70.7%;  Pred. No. 3.4e-32;
  Matches  355;  Conservative    0;  Mismatches  138;  Indels    9;  Gaps    2;

Qy        227 TTTATATATATATATATATATATATACACACACACACATATATATATACACATATATA 286
              | || |||| || || ||| ||||||| || |    || |||| |||||||  ||||  ||
Db        553 TATACATATGTACATCTATGTATATACGCATATGTGCATATATGTATATACGCATACGTA 494

Qy        287 TACACATATATACATATATACACATATATACACATATATATACAC------ATACATATG 340
              ||    |||||||  ||||| ||||||  | || ||||  |||| ||      |||||||||
Db        493 TATGTGTATATACGCATATATACATATGTGCATATATGTATATACGCATATATACATATG 434

Qy        341 TACACATATATATACACATATGTATACACATATATACACATATATACACACATATATACA 400
              | || |||| |||| ||  || |||||| || |   | | || |||| || ||||||||||
Db        433 TGCATATATGTATATACGCATATATACATATGTGCATATATGTATATACGCATATATACA 374

Qy        401 CATATATACACACATATATACACATATATACACATATATACACATATACACATATATACA 460
              || |  | | | |||||| ||||||||| | || | || || | || ||||  | |
Db        373 TATGTGCATATATGTATATACGCATATATACATACATGTGCATATGTGCATATATTCATA 314

Qy        461 CATATATACATATATACACATATATATAATATACACACATATATATACACATATATACAC 520
              || ||  || |||| | ||||  || |||| ||| |||| |||| |||| ||||
Db        313 AGTACATCTATGTATATGTACATAT---ATGTACATACACATATGTACATATATGTACAT 257
```

```
Qy        521 ACATATATACACATATATACACATATATATACACATATATACACATATATACATATATAC 580
              ||| ||||  ||||||||  ||||| | ||| ||||||||  |||||||| | | |||||
Db        256 ACACATATGTACATATATGTACATACACATATACATATATGTACATATACATACATATAT 197

Qy        581 ACATATATATACATATATACACATATATACATATATACACATATATACATATATACACAC 640
               ||| ||| || ||||  ||| |||  || ||||| | | || |||| ||| | | | ||
Db        196 GCATGTATGTATATATGCACATATACCTATATATACATATATGTATATATACACATATAC 137

Qy        641 ATATATACACATACATATACACACACATAGATATACATATATATACACATATATATACGT 700
              ||||||  |||||  | || | ||||| | ||  | |||| ||||| |||||||  |   |
Db        136 ATATATGTACATATGTGTATATACACACATATGCATATATGTATACGCATATATGCATAT 77

Qy        701 ATATATATGTATATATATATGC 722
              || |||||  | | ||||||||
Db         76 ATGTATATACACACATATATGC 55


RESULT 75
ABX08336_08
Continuation (9 of 17) of ABX08336 from base 800001 (Human phosphodiesterase 4D (PDE4D)
WP Sequence split into 17 fragments  LOCUS ABX08336 Accession Abx08336
WP     Fragment Name      Begin     End
WP     ABX08336_00            1    110000
WP     ABX08336_01       100001    210000
WP     ABX08336_02       200001    310000
WP     ABX08336_03       300001    410000
WP     ABX08336_04       400001    510000
WP     ABX08336_05       500001    610000
WP     ABX08336_06       600001    710000
WP     ABX08336_07       700001    810000
WP     ABX08336_08       800001    910000
WP     ABX08336_09       900001   1010000
WP     ABX08336_10      1000001   1110000
WP     ABX08336_11      1100001   1210000
WP     ABX08336_12      1200001   1310000
WP     ABX08336_13      1300001   1410000
WP     ABX08336_14      1400001   1510000
WP     ABX08336_15      1500001   1610000
WP     ABX08336_16      1600001   1691080

  Query Match             5.9%;  Score 242;  DB 6;  Length 110000;
  Best Local Similarity   70.0%;  Pred. No. 6.5e-32;
  Matches  368;  Conservative    1;  Mismatches  151;  Indels    6;  Gaps    3;

Qy        196 GAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATATACAC 255
              ||  |||| |     | | | |||| | |     | |||||||| ||||||||  |||||
Db      67577 GAGTTGTGAATATATATATAAAATATGTGTGTGTATATATACATATATATGCATATATGT 67636

Qy        256 ACACACACACATATATATATACACATATATATACACATATATACATATATACACATATAT 315
              ||  | | |||||||| |||| | |  |||||  || |||||||||||||||||||||||
Db      67637 ACCTATATACATATAT-TATATATACCTATATGAACCTATATACATATATACACATATAT 67695

Qy        316 ACACATATATATACACATACATATGTACACATATAT----ATACACATATGTATACACAT 371
               || |||||| ||| ||||| || :  || |||||    ||| ||| || | |  ||
Db      67696 GAACCTATATACACATATACACATAYGTACCTATATACACATATACACATATGTGTACCC 67755

Qy        372 ATATACACATATATACACACATATATACACATATATACACACATATATACACATATATAC 431
              ||||||||||||| ||| |  ||  || | | | ||| ||||||||  |  ||  ||||||
Db      67756 ATATACACATATACACATATGTACCTATATACACATATACACATATGTGTACCCATATAC 67815

Qy        432 ACATATATACACATATACACATATATACACATATATACATATATACACATATATATAATA 491
              ||||||| ||| || |  ||  |||||||||||||| ||||||  |  ||  |||||  | |
Db      67816 ACATATACACATATGTGTACCCATATACACATATACACATATGTGTACCCATATA-CACA 67874

Qy        492 TACACACATATATATACACATATATACACACATATATACACATATATACACATATATATA 551
              || |||||||| | ||| |||||| ||| | | | |||     ||   | | | ||||||
Db      67875 TATACACATATGTGTACCCATATACACATATACACATATGTGTACCCATATACACATATA 67934

Qy        552 CACATATATACACATATATACATATATACACATATATATACATATATACACATATATACA 611
              |||||||  |  ||  |||||||  |||||||||||| | |||  ||||||||||||| |||
Db      67935 CACATATGTGTACCCATATACACATATACACATATGTGTACCCATATACACATATACACA 67994

Qy        612 TATATACACATATATACATATATACACACATATATACACATACATATACACACACATAGA 671
```

```
              | | ||| |||||| ||||||| |||    ||   | | | ||||||||||    ||
Db      67995 TGTGTACCCATATACACATATACACATGTGTACCCATATACACATATACACATGTGTACC 68054

Qy        672 TATACATATATATACACATATATATACGTATATATATGTATATATA 717
              ||| | | |||||  ||||| | |||  |||||    ||| | ||
Db      68055 CATATACACATATATGCATATGTGTACCCATATACGCATATGTGTA 68100


RESULT 76
ADJ25985_08
Continuation (9 of 17) of ADJ25985 from base 800001 (Human phosphodiesterase 4D (PDE4D) c
WP Sequence split into 17 fragments  LOCUS ADJ25985 Accession Adj25985
WP     Fragment Name      Begin     End
WP     ADJ25985_00            1    110000
WP     ADJ25985_01       100001    210000
WP     ADJ25985_02       200001    310000
WP     ADJ25985_03       300001    410000
WP     ADJ25985_04       400001    510000
WP     ADJ25985_05       500001    610000
WP     ADJ25985_06       600001    710000
WP     ADJ25985_07       700001    810000
WP     ADJ25985_08       800001    910000
WP     ADJ25985_09       900001   1010000
WP     ADJ25985_10      1000001   1110000
WP     ADJ25985_11      1100001   1210000
WP     ADJ25985_12      1200001   1310000
WP     ADJ25985_13      1300001   1410000
WP     ADJ25985_14      1400001   1510000
WP     ADJ25985_15      1500001   1610000
WP     ADJ25985_16      1600001   1691139

  Query Match             5.9%;  Score 242;  DB 12;  Length 110000;
  Best Local Similarity   70.0%;  Pred. No. 6.5e-32;
  Matches  368;  Conservative    1;  Mismatches  151;  Indels    6;  Gaps    3;

Qy        196 GAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATATACAC 255
              ||  |||| |     | | | |||| | |     | |||||||| ||||||||  |||||
Db      67636 GAGTTGTGAATATATATATAAAATATGTGTGTGTATATATACATATATATGCATATATGT 67695

Qy        256 ACACACACACATATATATATACACATATATATACACATATATACATATATACACATATAT 315
              ||  | | |||||||| |||| | |  |||||  || |||||||||||||||||||||||
Db      67696 ACCTATATACATATAT-TATATATACCTATATGAACCTATATACATATATACACATATAT 67754

Qy        316 ACACATATATATACACATACATATGTACACATATAT----ATACACATATGTATACACAT 371
               || |||||| ||| ||||| || :  || |||||    ||| ||| || | |  ||
Db      67755 GAACCTATATACACATATACACATAYGTACCTATATACACATATACACATATGTGTACCC 67814

Qy        372 ATATACACATATATACACACATATATACACATATATACACACATATATACACATATATAC 431
              ||||||||||||| ||| |  ||  || | | | ||| ||||||||  |  ||  ||||||
Db      67815 ATATACACATATACACATATGTACCTATATACACATATACACATATGTGTACCCATATAC 67874

Qy        432 ACATATATACACATATACACATATATACACATATATACATATATACACATATATATAATA 491
              ||||||| ||| || |  ||  |||||||||||||| ||||||  |  ||  |||||  | |
Db      67875 ACATATACACATATGTGTACCCATATACACATATACACATATGTGTACCCATATA-CACA 67933

Qy        492 TACACACATATATATACACATATATACACACATATATACACATATATACACATATATATA 551
              || |||||||| | ||| |||||| ||| | | | |||     ||   | | | ||||||
Db      67934 TATACACATATGTGTACCCATATACACATATACACATATGTGTACCCATATACACATATA 67993

Qy        552 CACATATATACACATATATACATATATACACATATATATACATATATACACATATATACA 611
              |||||||  |  ||  |||||||  |||||||||||| | |||  ||||||||||||| |||
Db      67994 CACATATGTGTACCCATATACACATATACACATATGTGTACCCATATACACATATACACA 68053

Qy        612 TATATACACATATATACATATATACACACATATATACACATACATATACACACACATAGA 671
              | | ||| |||||| ||||||| |||    ||   | | | ||||||||||    ||
Db      68054 TGTGTACCCATATACACATATACACATGTGTACCCATATACACATATACACATGTGTACC 68113

Qy        672 TATACATATATATACACATATATATACGTATATATATGTATATATA 717
              ||| | | |||||  ||||| | |||  |||||    ||| | ||
Db      68114 CATATACACATATATGCATATGTGTACCCATATACGCATATGTGTA 68159


RESULT 77
```

```
ADN97989_08
Continuation (9 of 17) of ADN97989 from base 800001 (Human phosphodiesterase 4D genomic
WP Sequence split into 17 fragments  LOCUS ADN97989 Accession Adn97989
WP      Fragment Name       Begin      End
WP      ADN97989_00             1    110000
WP      ADN97989_01        100001    210000
WP      ADN97989_02        200001    310000
WP      ADN97989_03        300001    410000
WP      ADN97989_04        400001    510000
WP      ADN97989_05        500001    610000
WP      ADN97989_06        600001    710000
WP      ADN97989_07        700001    810000
WP      ADN97989_08        800001    910000
WP      ADN97989_09        900001   1010000
WP      ADN97989_10       1000001   1110000
WP      ADN97989_11       1100001   1210000
WP      ADN97989_12       1200001   1310000
WP      ADN97989_13       1300001   1410000
WP      ADN97989_14       1400001   1510000
WP      ADN97989_15       1500001   1610000
WP      ADN97989_16       1600001   1691138

  Query Match             5.9%;  Score 242;  DB 12;  Length 110000;
  Best Local Similarity   70.0%;  Pred. No. 6.5e-32;
  Matches  368;  Conservative    1;  Mismatches  151;  Indels    6;  Gaps    3;

Qy        196 GAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATATACAC 255
              ||  ||| |    | | | |||| | |    | |||||||| ||||||||  |||||
Db      67635 GAGTTGTGAATATATATATAAAATATGTGTGTGTATATATACATATATATGCATATATGT 67694

Qy        256 ACACACACACATATATATATACACATATATATACACATATATACATATATACACATATAT 315
              ||  | | |||||||| |||| | |  |||||  || ||||||||||||||||||||
Db      67695 ACCTATATACATATAT-TATATATACCTATATGAACCTATATACATATATACACATATAT 67753

Qy        316 ACACATATATATACACATACATATGTACACATATAT----ATACACATATGTATACACAT 371
                || |||||| ||| ||||| || :  || |||||    ||| ||| || | |  ||
Db      67754 GAACCTATATACACATATACACATAYGTACCTATATACACATATACACATATGTGTACCC 67813

Qy        372 ATATACACATATATACACACATATATACACATATATACACACATATATACACATATATAC 431
              |||||||||||||| ||| |  ||  || | | | ||| |||||||| |  ||  ||||||
Db      67814 ATATACACATATACACATATGTACCTATATACACATATACACATATGTGTACCCATATAC 67873

Qy        432 ACATATATACACATATACACATATATACACATATATACATATATACACATATATATAATA 491
              ||||||| ||| || |  ||  |||||||||||| |||||| |  ||  |||||  | |
Db      67874 ACATATACACATATGTGTACCCATATACACATATACACATATGTGTACCCATATA-CACA 67932

Qy        492 TACACACATATATATACACATATATACACACATATATACACATATATACACATATATATA 551
              || |||||||| | ||| |||||| ||| | | | |||     ||    | | | | |||||
Db      67933 TATACACATATGTGTACCCATATACACATATACACATATGTGTACCCATATACACATATA 67992

Qy        552 CACATATATACACATATATACATATATACACATATATATACATATATACACATATATACA 611
              |||||||·|  ||  ||||||| |||||||||||| | |||  |||||||||||| |||
Db      67993 CACATATGTGTACCCATATACACATATACACATATGTGTACCCATATACACATATACACA 68052

Qy        612 TATATACACATATATACATATATACACACATATATACACATACATATACACACACATAGA 671
              | | ||| |||||| ||||||| |||    ||   | | | ||||||||||    ||
Db      68053 TGTGTACCCATATACACATATACACATGTGTACCCATATACACATATACACATGTGTACC 68112

Qy        672 TATACATATATATACACATATATATACGTATATATATGTATATATA 717
                ||| | | |||||  |||||| | |||  |||||     ||| | ||
Db      68113 CATATACACATATATGCATATGTGTACCCATATACGCATATGTGTA 68158


RESULT 78
ADO50281_08
Continuation (9 of 17) of ADO50281 from base 800001 (Human phosphodiesterase 4D (PDE4D)
WP Sequence split into 17 fragments  LOCUS ADO50281 Accession Ado50281
WP      Fragment Name       Begin      End
WP      ADO50281_00             1    110000
WP      ADO50281_01        100001    210000
WP      ADO50281_02        200001    310000
WP      ADO50281_03        300001    410000
WP      ADO50281_04        400001    510000
WP      ADO50281_05        500001    610000
WP      ADO50281_06        600001    710000
WP      ADO50281_07        700001    810000
WP      ADO50281_08        800001    910000
WP      ADO50281_09        900001   1010000
WP      ADO50281_10       1000001   1110000
WP      ADO50281_11       1100001   1210000
WP      ADO50281_12       1200001   1310000
WP      ADO50281_13       1300001   1410000
WP      ADO50281_14       1400001   1510000
WP      ADO50281_15       1500001   1610000
WP      ADO50281_16       1600001   1691134

  Query Match             5.9%;  Score 242;  DB 12;  Length 110000;
  Best Local Similarity   70.0%;  Pred. No. 6.5e-32;
  Matches  368;  Conservative    1;  Mismatches  151;  Indels    6;  Gaps    3;

Qy        196 GAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATATACAC 255
              ||  ||| |    | | | |||| | |    | |||||||| ||||||||  |||||
Db      67636 GAGTTGTGAATATATATATAAAATATGTGTGTGTATATATACATATATATGCATATATGT 67695

Qy        256 ACACACACACATATATATATACACATATATATACACATATATACATATATACACATATAT 315
              ||  | | |||||||| |||| | |  |||||  || ||||||||||||||||||||
Db      67696 ACCTATATACATATAT-TATATATACCTATATGAACCTATATACATATATACACATATAT 67754

Qy        316 ACACATATATATACACATACATATGTACACATATAT----ATACACATATGTATACACAT 371
                || |||||| ||| ||||| || :  || |||||    ||| ||| || | |  ||
Db      67755 GAACCTATATACACATATACACATAYGTACCTATATACACATATACACATATGTGTACCC 67814

Qy        372 ATATACACATATATACACACATATATACACATATATACACACATATATACACATATATAC 431
              |||||||||||||| ||| |  ||  || | | | ||| |||||||| |  ||  ||||||
Db      67815 ATATACACATATACACATATGTACCTATATACACATATACACATATGTGTACCCATATAC 67874

Qy        432 ACATATATACACATATACACATATATACACATATATACATATATACACATATATATAATA 491
              ||||||| ||| || |  ||  |||||||||||| |||||| |  ||  |||||  | |
Db      67875 ACATATACACATATGTGTACCCATATACACATATACACATATGTGTACCCATATA-CACA 67933

Qy        492 TACACACATATATATACACATATATACACACATATATACACATATATACACATATATATA 551
              || |||||||| | ||| |||||| ||| | | | |||     ||    | | | | |||||
Db      67934 TATACACATATGTGTACCCATATACACATATACACATATGTGTACCCATATACACATATA 67993

Qy        552 CACATATATACACATATATACATATATACACATATATATACATATATACACATATATACA 611
              |||||||·|  ||  ||||||| |||||||||||| | |||  |||||||||||| |||
Db      67994 CACATATGTGTACCCATATACACATATACACATATGTGTACCCATATACACATATACACA 68053

Qy        612 TATATACACATATATACATATATACACACATATATACACATACATATACACACACATAGA 671
              | | ||| |||||| ||||||| |||    ||   | | | ||||||||||    ||
Db      68054 TGTGTACCCATATACACATATACACATGTGTACCCATATACACATATACACATGTGTACC 68113

Qy        672 TATACATATATATACACATATATATACGTATATATATGTATATATA 717
                ||| | | |||||  |||||| | |||  |||||     ||| | ||
Db      68114 CATATACACATATATGCATATGTGTACCCATATACGCATATGTGTA 68159


RESULT 79
ADQ97965
ID   ADQ97965 standard; DNA; 152096 BP.
XX
AC   ADQ97965;
XX
DT   07-OCT-2004  (first entry)
XX
DE   Mouse cancer associated sequence MD11-032, SEQ ID 942.
XX
KW   Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Mouse; ds.
XX
OS   Mus musculus.
XX
PN   WO2004060304-A2.
XX
PD   22-JUL-2004.
XX
PF   22-DEC-2003; 2003WO-US041389.
```

```
XX
PR   27-DEC-2002; 2002US-00330773.
XX
PA   (SAGR-) SAGRES DISCOVERY INC.
XX
PI   Morris DW,  Malandro MS;
XX
DR   WPI; 2004-543781/52.
XX
PT   New isolated cancer associated nucleic acids comprising at least 10
PT   contiguous nucleotides, useful for diagnosing, preventing and/or treating
PT   cancers such as leukemia and lymphoma.
XX
PS   Claim 1; SEQ ID NO 942; 199pp; English.
XX
CC   The present invention relates to cancer associated sequences (ADQ97025-
CC   ADQ98004). The sequences are useful for the diagnosis, prevention and/or
CC   treatment of cancer, such as leukemia and lymphoma. Note: The sequence
CC   data for this patent did not form part of the printed specification, but
CC   was obtained in electronic formate directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 152096 BP; 42690 A; 31533 C; 31445 G; 45238 T; 0 U; 1190 Other;

  Query Match             5.9%;  Score 241.4;  DB 12;  Length 152096;
  Best Local Similarity   68.5%;  Pred. No. 8.3e-32;
  Matches  346;  Conservative    0;  Mismatches  158;  Indels    1;  Gaps    1;

Qy        230 ATATATATATATATATATATATACACACACACACACATATATATATACACATATATATAC 289
              | | | | | | | | | | ||| |||||||||| | | | ||| | ||| | | | ||
Db      78632 ACACACACACACACACACACACACACATACACACACATACACACATACATACACACACAC 78691

Qy        290 ACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATAT 349
              ||| | ||||||| | ||||| | | ||||||| | | |||    ||| |   ||| | |
Db      78692 ACACACATACATACACACACACACACACACATACACACACATGCACACACACACATACAC 78751

Qy        350 ATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATAC 409
              | ||||||| |   ||||||| | ||||||| |||||||||| | | ||| | | | ||
Db      78752 ACATACACACACACATACACACACATACACACATATACACACACACACACATACACACAC 78811

Qy        410 ACACATATATACACATATATACACATATATACACATATACACATATATACACATATATAC 469
              | ||| | ||||||| | | ||| | | | ||||| ||||||| ||| ||||| ||| ||
Db      78812 ATACACACATACACACACACACATACACACACACACATACACACATACACACACATACAC 78871

Qy        470 ATATATACACATATATATAATATACACACATATATATACACATATATACACACATATATA 529
              | | | ||||| | |     | | |||||||| | ||| ||||| | | |||||||| | | |
Db      78872 ACACACACACACACACCACACACACACACACACATACACACACACACACACACACACACA 78931

Qy        530 CACATATATACACATATATATACACATATATACACATATATACATATATACACATATATA 589
              |||||| | ||||| ||| | ||||| | | ||| | | | ||| | | | ||| | | |
Db      78932 CACATACACACACACATACACACACACACACACATACACACACACACACATACACACACA 78991

Qy        590 TACATATATACACATATATACATATATACACATATATACATATATACACACATATATACA 649
              ||||| | ||||| | | ||| ||| ||||| | | ||| | |||||||||| | | |||
Db      78992 CACATACACACACACACACACACATACACACACACACACACACATACACACACACACACA 79051

Qy        650 CATACATATACACACACATAGATATACA-TATATATACACATATATATACGTATATATAT 708
              || ||||| |||||||||| | | | |||  | | | ||||| | | | ||  | | || |
Db      79052 CACACATACACACACACACACACACACCACACACACACACACACACACACACACACATGT 79111

Qy        709 GTATATATATATGCTCCAGAGTTCA 733
              | | |  |||  |||  |  | ||
Db      79112 GCGTGTGCATACACTCGTGCATGCA 79136


RESULT 80
ADC87404/c
ID   ADC87404 standard; DNA; 1441 BP.
XX
AC   ADC87404;
XX
DT   01-JAN-2004  (first entry)
XX
DE   Human GPCR gene SEQ ID NO:1857.
XX
KW   ds; gene; human; GPCR;
KW   guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1270724-A2.
XX
PD   02-JAN-2003.
XX
PF   18-JUN-2002; 2002EP-00013517.
XX
PR   18-JUN-2001; 2001JP-00246789.
XX
PA   (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA   (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI   Suwa M,  Asai K,  Akiyama Y,  Aburatani H;
XX
DR   WPI; 2003-315783/31.
DR   P-PSDB; ADC87405.
XX
PT   New polynucleotide, useful for preparing a composition for treating a
PT   patient in need of increased or suppressed activity or expression of the
PT   guanosine triphosphate-binding protein coupled receptor.
XX
PS   Claim 1; SEQ ID NO 1857; 28pp; English.
XX
CC   The invention relates to a novel polynucleotide encoding a guanosine
CC   triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC   the invention may have a use in gene therapy. The polynucleotide and
CC   polypeptide are useful for preparing a composition for treating a patient
CC   in need of increased or suppressed activity or expression of the
CC   guanosine triphosphate-binding protein coupled receptor. The
CC   polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC   invention.
XX
SQ   Sequence 1441 BP; 518 A; 214 C; 219 G; 490 T; 0 U; 0 Other;

  Query Match             5.9%;  Score 241.2;  DB 10;  Length 1441;
  Best Local Similarity   69.5%;  Pred. No. 7.5e-32;
  Matches  357;  Conservative    0;  Mismatches  153;  Indels    4;  Gaps    2;

Qy        227 TTTATATATATATATATATATATATACACACACACACATATATATATACACATATATA 286
              | || |||||||||| ||| | |||| |   | | | |||||||||| |||||||||| |||
Db        702 TATACATATATATAGATACACATATGCGTATATATACACATATATGTATACACATACATA 643

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              | |  ||||| ||||||||  | | | ||||| || | ||||||  |||  |||  |||
Db        642 TGCGTATATACACATATATGTATACACATACATATGTGTATACATATATGTATACACATG 583

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              ||||  |  | ||||  |||  |||||||||    |||||  |||  |    ||||  |||||||
Db        582 CATATGTGTATATATTTATGAACATATATATGTATATACACATATGCGTATATACATATA 523

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              |  | |||  ||||   |  ||||  ||    | |  |||  |   ||||||||||  |||     ||||
Db        522 TGTATACACATATGTATATATTTATGTACACATATATGTATACACATACATATGTGTATA 463

Qy        467 TACATATATACACATATATATAATATACACACATATATATACACATATATACACACATAT 526
              ||  |||  ||||||||||     ||||||||  |  ||||  |   |  ||||  ||||||  ||  |
Db        462 TATTTATGTACACATATAT---GTATACACATACATATGTGTATATATTTACACATATGT 406

Qy        527 ATACACATATATACACATATATATACACATATATACACATATATACATATATACACATAT 586
              |||||||| |||  |   | ||||||    | | | |||  ||  ||||||  |||  |  |  | |
Db        405 GTACACATACATATATGTGTATATATTTACACACACATATGTATACACATACATATGTGT 346

Qy        587 ATATACATATATACACATATATACATATATACACATATATACATA-TATACACACATATA 645
              |||||  |||  ||||||||  ||  |  |  ||||  ||  |  ||  |||  |  |||||  ||  ||
Db        345 ATATATTTATGCACACATATGTATACACATACGTATGTGTATATATTGCACACATATGTA 286

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
```

```
Db        285 TACACATACATATGTGTATATATTTATGCACACATATGTATACACATACATATGTGTATA 226

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAG 739
              | | |||||| | |||         | |||| | |
Db        225 TGTTTATATACACATGTATACACATGCATATGTG 192


RESULT 81
ADC87576
ID   ADC87576 standard; DNA; 1276 BP.
XX
AC   ADC87576;
XX
DT   01-JAN-2004  (first entry)
XX
DE   Human GPCR gene SEQ ID NO:2029.
XX
KW   ds; gene; human; GPCR;
KW   guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1270724-A2.
XX
PD   02-JAN-2003.
XX
PF   18-JUN-2002; 2002EP-00013517.
XX
PR   18-JUN-2001; 2001JP-00246789.
XX
PA   (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA   (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI   Suwa M,  Asai K,  Akiyama Y,  Aburatani H;
XX
DR   WPI; 2003-315783/31.
DR   P-PSDB; ADC87577.
XX
PT   New polynucleotide, useful for preparing a composition for treating a
PT   patient in need of increased or suppressed activity or expression of the
PT   guanosine triphosphate-binding protein coupled receptor.
XX
PS   Claim 1; SEQ ID NO 2029; 28pp; English.
XX
CC   The invention relates to a novel polynucleotide encoding a guanosine
CC   triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC   the invention may have a use in gene therapy. The polynucleotide and
CC   polypeptide are useful for preparing a composition for treating a patient
CC   in need of increased or suppressed activity or expression of the
CC   guanosine triphosphate-binding protein coupled receptor. The
CC   polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC   invention.
XX
SQ   Sequence 1276 BP; 416 A; 198 C; 221 G; 441 T; 0 U; 0 Other;

  Query Match             5.8%;  Score 240.8;  DB 10;  Length 1276;
  Best Local Similarity   69.9%;  Pred. No. 8.8e-32;
  Matches  341;  Conservative    0;  Mismatches  142;  Indels    5;  Gaps    1;

Qy        230 ATATATATATATATATATATATACACACACACACATATATATATACACATATATATAC 289
              | ||||| |||| || |||||||| |   ||| | |||  |||||||| || || ||
Db        793 ACATATACATATGTACATATATACATATGTACATATATACGTATATACATATGTACATCT 852

Qy        290 ACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATAT 349
              |  |||||||  ||||   |||||||  | |||   ||||  |||  | | || ||  ||
Db        853 ATGTATATACGCATATGTGCATATATGTATATACGCATACGTATATGTGTATATACGCAT 912

Qy        350 ATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATAC 409
              |||||||  ||  ||  |||  |  |||||||  ||||||||||  |   |   |||  |  |||||||
Db        913 ATATACATATGTGCATATATGTATATACGCATATATACATATGTGCATATATGTATATAC 972

Qy        410 ACACATATATACACATATATACACATATATACACATATACACATATATACACATATATAC 469
```

```
Db        973 GCATATATACATATGTGCATATATGTATATACGCATATATACATATGTGCATATATGTAT 1032

Qy        470 ATATATACACATATATATAATATACACACATATATATACACATATATACACACATATATA 529
              |||   | | |||      ||| |    ||||| |  | | ||  |||   |||| |||
Db       1033 ATACGCATATATA-----CATACATGTGCATATGTGCATATATTCATAAGTACATCTATG 1087

Qy        530 CACATATATACACATATATATACACATATATACACATATATACATATATACACATATATA 589
              | || || | | || || |||||||||| |||| |||| |||||| | | |  || |||
Db       1088 TATATGTACATATATGTACATACACATATGTACATATATGTACATACACATATGTACATA 1147

Qy        590 TACATATATACACATATATACATATATACACATATATACATATATACACACATATATACA 649
              ||  || |||||||||| | |||| |||| ||| |||||||||| ||   || |||| |
Db       1148 TATGTACATACACATATACATATATGTACATATACATACATATATGCATGTATGTATATA 1207

Qy        650 CATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATATG 709
              ||||||||  | | ||| |||||  |||||||||||||||| |||  | || |||||
Db       1208 TGCACATATACCTATATATACATATATGTATATATACACATATACATATATGTACATATG 1267

Qy        710 TATATATA 717
              | ||||||
Db       1268 TGTATATA 1275


RESULT 82
ABL33869/c
ID   ABL33869 standard; DNA; 13573 BP.
XX
AC   ABL33869;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 1842.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP007537.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful for
PT   diagnosis and treatment of diseases associated with abnormal cytosine
PT   methylation.
XX
PS   Claim 1; SEQ ID NO 1842; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
```

```
CC   diseases. The present sequence is a gene of the invention
XX
SQ   Sequence 13573 BP; 4074 A; 204 C; 2749 G; 6546 T; 0 U; 0 Other;

  Query Match             5.8%;  Score 240.4;  DB 6;  Length 13573;
  Best Local Similarity   69.2%;  Pred. No. 1.1e-31;
  Matches  537;  Conservative    0;  Mismatches  206;  Indels   33;  Gaps   14;

Qy          4 TCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAA--TACAAAATTAG 61
              | || |||| ||| | ||| ||    ||| ||| ||||||| |||||     |||||||
Db       6406 TTAAAACCAACCTAAACAATATAACAAAATCCCGTCTCTACAAAAAAATACAAAAATTAA 6347

Qy         62 ACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAATC 121
               ||  | |  | | |||| ||| |||||||| ||||||   |  |   |  | ||||
Db       6346 CCAAACTTAATAACACACACCTATAATCCCAACTACTTAAAAAACTAAAATAAAAAAATC 6287

Qy        122 ACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAGCC 181
              |||| ||||          | |       | ||  ||| | ||| | ||||||| ||
Db       6286 ACTTAAACCCAAAAAATAAAAAC-TACAATAAACCATAATCGTACCACTACACTCCAACC 6228

Qy        182 CAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATAT---A 238
              |  ||| || |   ||   | |    | |||||| | |       | |||||||||   |
Db       6227 TAAACAATAAAACAAAACCCTAACAAAAAAAAAAAATATATATATATATATATATAA 6168

Qy        239 TATATATATATATACACACACACACATATATATAT-ACACATATATATACACATATAT 297
              || ||||||||||||  | | |||   |||||||||| | ||||||||||| || ||||||
Db       6167 TACATATATATATACGTATATACATTTATATATATATAATACATATATATATACGTATAT 6108

Qy        298 ACATATATACACATAT-ATACACATATATATACACATACATATGTACACATATATATACA 356
              |||| |||| | |||| ||||| ||||||||||  ||| | || || | ||||||| | |
Db       6107 ACATTTATATATATATAATACATATATATATACGTATATACATTTATATATATATA-ATA 6049

Qy        357 CATATGTATACACATATATACACATATAT-------ACACACATATATACACATATATAC 409
              ||||| |||| || ||||||||  |||||       | ||| ||||||| || |||||||
Db       6048 CATATATATATACGTATATACATTTATATATATATAATACATATATATATACGTATATAC 5989

Qy        410 ACACATAT-------ATACACATATATACACATATATACACATATACACATATATACACA 462
              |   ||||       ||||| ||||||| || ||||||||  |||| | ||||| | |||
Db       5988 ATTTATATATATATAATACATATATATATACGTATATACATTTATATATATATA-ATACA 5930

Qy        463 TATATACATATATACACATATATATAATATACACACATATATATACACATATATACACAC 522
              |||||| |||||||||  ||||||||     | ||| ||||||||||| ||||| |   |
Db       5929 TATATATATATATACATTTATATATATATAATACATATATATATACATATATACATTTAT 5870

Qy        523 ATATAT---ACACATATATACACATATATATACACATATATACACATATATACATATATA 579
              ||||||    | |||||||||  | |  |||||||||  ||||||  | |||  ||||||||||
Db       5869 ATATATATAATACATATATATATACATATATACATTTATATATATATA-ATACATATATA 5811

Qy        580 CACATATATATACATATATACACAT---ATATACATATATACACATATATACATATATAC 636
              | | |||||||||| |||| | ||    ||| | ||||||| ||||||||||| ||||
Db       5810 TATACATATATACATTTATATATATATAATACATATATATATACATATATACATTTATAT 5751

Qy        637 ACACATATATACACATACATATACACACACATAGATATACATATAT-ATACACATATATA 695
              | | ||| ||||| ||| ||||||| | | | |  |||| |||||| ||||| |||||||
Db       5750 ATATATA-ATACATATATATATACATATATACATTTATATATATATAATACATATATATA 5692

Qy        696 TACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGAT 751
              ||| |||||| || ||||||||||  | || |  |    | |  ||    || ||
Db       5691 TACATATATACATTTATATATATATAATACATATATATATACATATATACATTTAT 5636


RESULT 83
ABK83460/c
ID   ABK83460 standard; cDNA; 165199 BP.
XX
AC   ABK83460;
XX
DT   14-AUG-2002  (first entry)
XX
DE   Human cDNA differentially expressed in granulocytic cells #31.
XX
KW   Human; ss; granulocytic cell; DNA chip; bacterial infection;
KW   viral infection; parasitic infection; protozoal infection;
KW   fungal infection; sterile inflammatory disease; psoriasis;
KW   rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KW   cardiac reperfusion injury; renal reperfusion injury; ARDS;
KW   adult respiratory distress syndrome; inflammatory bowel disease;
KW   Crohn's disease; ulcerative colitis; periodontal disease;
KW   granulocyte activation; chronic inflammation; allergy.
XX
OS   Homo sapiens.
XX
PN   WO200228999-A2.
XX
PD   11-APR-2002.
XX
PF   03-OCT-2001; 2001WO-US030821.
XX
PR   03-OCT-2000; 2000US-0237189P.
XX
PA   (GENE-) GENE LOGIC INC.
XX
PI   Beazer-Barclay Y,  Weissman SM,  Yamaga S,  Vockley J;
XX
DR   WPI; 2002-435328/46.
XX
PT   Detecting granulocyte activation by detecting differential expression of
PT   genes associated with granulocyte activation, which serves as diagnostic
PT   markers that is useful for monitoring disease states and drug toxicity.
XX
PS   Claim 1; SEQ ID NO 31; 114pp; English.
XX
CC   The invention relates to detecting (M1) granulocyte (GC) activation
CC   (GCA), by detecting the level of expression of gene(s) (Gs) identified by
CC   DNA chip analysis as given in the specification, and comparing the
CC   expression level to an expression level in an unactivated GC, where
CC   differential expression of Gs is indicative of GCA. Also included are
CC   modulating (M2) GA by contacting GC with an agent that alters the
CC   expression of at least one gene in Gs; (2) screening (M3) for an agent
CC   capable of modulating GCA or an inflammation (especially chronic) in a
CC   tissue, an allergic response in a subject, exposure of a subject to a
CC   pathogen or sterile inflammatory disease using the gene expression
CC   profile; (3) detecting (M4) an inflammation (especially chronic) in a
CC   tissue, an allergic response in a subject, exposure of a subject to a
CC   pathogen or sterile inflammatory disease, by detecting the level of
CC   expression in a sample of the tissue of gene(s) from Gs, where the level
CC   of expression of the gene is indicative of inflammation; (4) treating
CC   (M5) an inflammation (especially chronic) or in a tissue, an allergic
CC   response in a subject, exposure of a subject to a pathogen or sterile
CC   inflammatory disease, by contacting a tissue having inflammation with an
CC   agent that modulates the expression of gene(s) from Gs in the tissue. M1
CC   is useful for detecting GCA; M2 is useful for modulating GA; M3 is useful
CC   for screening an agent capable of modulating GCA preferably in an
CC   inflammation in a tissue; M4 is useful for detecting an inflammation
CC   (especially chronic) in a tissue, an allergic response in a subject,
CC   exposure of a subject to a pathogen or sterile inflammatory disease (e.g.
CC   psoriasis, rheumatoid arthritis, glomerulonephritis, asthma, thrombosis,
CC   cardiac reperfusion injury, renal reperfusion injury, ARDS, adult
CC   respiratory distress syndrome, inflammatory bowel disease, Crohn's
CC   disease, ulcerative colitis, periodontal disease; also bacterial
CC   infection, viral infection, parasitic infection, protozoal infection,
CC   fungal infection and M5 is useful for treating one of the above
CC   conditions. The present sequence represents a gene differentially
CC   expressed in granulocytes. Note: The sequence data for this patent did
CC   not form part of the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 165199 BP; 48510 A; 33223 C; 34406 G; 49060 T; 0 U; 0 Other;

  Query Match             5.8%;  Score 240.2;  DB 6;  Length 165199;
  Best Local Similarity   66.3%;  Pred. No. 1.3e-31;
  Matches  377;  Conservative    0;  Mismatches  186;  Indels    6;  Gaps    2;

Qy        234 ATATATATATATATATATACACACACACACATATATATATACACATATATATACACAT 293
              ||  |||||||||||||| ||||||  | | | ||||||||||| | |||||||||  |||
```

```
Db     139452 ATGGATATATATATATACACACACGGATATATATATATATATATAGATATATATATCCAT 139393

Qy        294 ATATACATATATACACATATATACACATATATATACACATACATATGTACACATATATAT 353
              ||||  |||||||  |    |||||  | ||  ||||||    |||  ||||  |   |  ||||  ||
Db     139392 ATATGGATATATATATCGATATATATATCTATATATGGATATATATATCTATATATGGAT 139333

Qy        354 ACACATATGTATACACATATATACACATATATACACACATATATACACATATATACACAC 413
              | | |||||  |||  |  ||||  ||  |  ||   |||  | |    |||||  | ||   |||  | |
Db     139332 ATATATATGGATATATATATCTATATATGGATATATATGGATATATATATGGATATATAT 139273

Qy        414 ATATATACACATATATACACATATATACACATATACACATATATACACATATATACACATAT 473
              ||||||    |||||||  |    |||||  | ||  ||  | ||   |||  | ||||   | ||||
Db     139272 CTATATATGGATATATATATGGATATATATATCTATATATGGATATATATATGGATATAT 139213

Qy        474 ATACACATATATATA-ATATACACACATATATATACACATATATACACACATATATACAC 532
              |||   ||||||   | ||||  |    |||||||||    ||||||   | |  ||||||  |
Db     139212 ATATCTATATATGGATATATATATGGATATATATATCTATATATGGATATATATATGGAT 139153

Qy        533 ATATATACACATATATATACACATATATACACATATATACATATATACACATATATATAC 592
              |||||||   ||||||  | |  ||||||  | |||||||  ||||||   |||||||  |
Db     139152 ATATATATCTATATATGGATATATATATGGATATATATAACTATATATGGATATATACAT 139093

Qy        593 ATATATACACATATATACATATATACACATATATACATATATACACACATATATACACAT 652
              ||||||   |||||||  ||  ||     ||||||  | |||||  | |  |||||||   ||
Db     139092 CTATATATGGATATATATATTGAT-----TATATATAGATATATAGATATATATATGGAT 139038

Qy        653 ACATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTAT 712
              | |||||   | |  ||  ||||||  ||  |||||  | ||  ||||||  | |||||||||  ||
Db     139037 ATATATATCTATATATGGATATATATGGATATATATATCTATATATGGATATATATGGAT 138978

Qy        713 ATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAA 772
              |||||||||    | |  |    | |  |||   |  ||    | |    |||||   |
Db     138977 ATATATATGGATTATATATAGATATATATAGATATATATATGGATTATATATAGATATAT 138918

Qy        773 AGAGAGTTTGACAGCAGTGTATTGTGAGA 801
              | |||  |  | | |  | |||  | | |
Db     138917 ATAGATATATATATCTATATATGGATATA 138889


RESULT 84
ABD32673
ID   ABD32673 standard; DNA; 91359 BP.
XX
AC   ABD32673;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Human cancer-associated genomic DNA HD13-075.
XX
KW   Human; ds; cancer-associated protein; gene; cytostatic; cancer;
KW   leukaemia; lymphoma; CAP.
XX
OS   Homo sapiens.
XX
PN   WO2004074320-A2.
XX
PD   02-SEP-2004.
XX
PF   17-FEB-2004; 2004WO-US004730.
XX
PR   14-FEB-2003; 2003US-00367094.
PR   14-MAR-2003; 2003US-00388838.
PR   15-APR-2003; 2003US-00417375.
PR   13-JUN-2003; 2003US-00461862.
PR   15-SEP-2003; 2003US-00663431.
PR   15-DEC-2003; 2003US-00737318.
XX
PA   (SAGR-) SAGRES DISCOVERY INC.
XX
PI   Morris DW,  Morris DW,  Malandro MS;
XX
DR   WPI; 2004-652914/63.
XX
PT   New isolated cancer-associated polynucleotides and polypeptides useful
PT   for diagnosing, preventing or treating cancers, especially lymphoma and
PT   leukemia, or in screening for agents that modulate cancer.
XX
PS   claim 16; seqid 221; 310pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising at least 10
CC   contiguous nucleotides of any of the 233 polynucleotide sequences given
CC   in the specification, or its complement. The nucleic acids encode cancer-
CC   associated proteins. Also included are an expression vector comprising
CC   the isolated nucleic acid cited above, a host cell comprising the above
CC   recombinant nucleic acid or expression vector, a microarray for detecting
CC   a cancer-associated (CA) nucleic acid comprising at least one probe
CC   comprising at least 10 contiguous nucleotides of any of the above-
CC   mentioned nucleotide sequences, an isolated polypeptide (encoded within
CC   an open reading frame of a CA sequence selected from any of the 95
CC   polynucleotide sequences as mentioned in the specification, or its
CC   complement), an isolated antibody, (or its antigen binding fragment) that
CC   binds to the above polypeptide, a hybridoma that produces the above
CC   monoclonal antibody, a pharmaceutical composition comprising the above
CC   antibody and a pharmaceutical excipient, a kit for detecting cancer
CC   cells(comprising the antibody cited above, methods for diagnosing cancer
CC   or for detecting the presence or absence of cancer cells in an
CC   individual, a method for inhibiting growth of cancer cells in an
CC   individual, a method for delivering a therapeutic agent to cancer cells
CC   in an individual, an electronic library comprising the above
CC   polynucleotide or polypeptide (or their fragments), methods of screening
CC   for anticancer activity or for a bioactive agent capable of modulating
CC   the activity of a CA protein (CAP), methods for detecting cancer
CC   associated with expression of a polypeptide in a test cell sample, a
CC   method for treating cancers and a method for inhibiting the expression of
CC   CA gene in a cell. The composition and methods are useful for detecting,
CC   diagnosing, preventing and treating cancers, especially lymphoma and
CC   leukaemia. These may also be used in screening for agents that modulate
CC   cancer. The present sequence is a human CAP genomic sequence. Note: The
CC   sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 91359 BP; 24503 A; 18193 C; 19029 G; 29634 T; 0 U; 0 Other;

  Query Match             5.8%;  Score 238.8;  DB 13;  Length 91359;
  Best Local Similarity   68.8%;  Pred. No. 2.3e-31;
  Matches  373;  Conservative    0;  Mismatches  162;  Indels    7;  Gaps    3;

Qy        204 ATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATATACACACACACAC 263
              | || || |  |   || ||  | |     || |||||| || |||| |   | | |
Db      19526 AGCTGAGTAGCATTCAATGGTGTATATATAATCTATATACATCTATACATCTATATATCT 19585

Qy        264 ACATATATATATACACATATATATACACATATATACATATATACACATATATACACATAT 323
              | ||| ||||||||| || ||| || | || | || ||| ||| | | ||||  | ||||
Db      19586 ATATACATATATACATATGTATGTATATATTTCTATATACATATATACATATGTATATAT 19645

Qy        324 ATATACACATACATATGTACACATATATATACACATATGTATACACATATATACACATAT 383
              | ||  | |||||| | || |||||| |||||| | |||||||||  ||| || |||||
Db      19646 AGATTGATATACATCTATATACATATGTATACATAGATGTATACATCTATGTATACATAG 19705

Qy        384 ATACACACATATATACACATATATACACACATATATACACATATATACACATATATACAC 443
              ||  | |||||       || |||| ||| |||| || ||||||||||| || ||||||
Db      19706 ATGTATACATA-----CTATGTATATACATATATGTATACATATATACATATGTATACAT 19760

Qy        444 ATATACACATATATACACATATATACATATATACACATATATATA-ATATACACACATAT 502
              |||  | | ||||| |||||| || ||||||  | ||||| ||| || || | |  |||
Db      19761 CTATGTATACATATATACATATGTATATATATCTATATATACATATATGTATATATGTAT 19820

Qy        503 ATATACACATATATACACACATATATACACATATATACACATATATATACACATATATAC 562
              ||||  | |||  || | |  |||||||| ||||| | | || |||||||| || ||||
Db      19821 ATATCTATATACGTATATAGGTATATACATATATAAATATATCTATATACATATGTATAT 19880

Qy        563 ACATATATACATATATACACATATATATACATATATACACATATATACATATATACACAT 622
              |  |||||| ||||||  | |||||| |||||||||  | ||| ||| || |||| |  |
Db      19881 ATCTATATAAATATATGGATATATATCTACATATATCTATATACATATATCTATATATCT 19940
```

```
Qy        623 ATATACATATATACACACATATATACACAT-ACATATACACACACATAGATATACATATA 681
              ||||||||||| |   | || |||  ||||  |||||  | | | ||| ||||||||| ||
Db      19941 ATATACATATAGATGTATATGTATCTACATGGCATATGTATATAGATATATATACATCTA 20000

Qy        682 TATACACATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGT 741
              |||  | || |||||||  ||| |||||   | |||||||   |||  |   |    | || |
Db      20001 TATGTATATCTATATACATATCTATATAGCTGTATATATATGCCATATATATAGATAGAT 20060

Qy        742 AG 743
              ||
Db      20061 AG 20062
```

RESULT 85
ABL51954
ID ABL51954 standard; DNA; 40023 BP.
XX
AC ABL51954;
XX
DT 11-JUL-2002 (first entry)
XX
DE Human solute carrier family 18 member 2 (SLC18A2) gene SEQ ID NO:1.
XX
KW Human; solute carrier family 18 member 2; SLC18A2; vesicular monoamine;
KW vesicular monoamine transporter; VMAT2; polymorphic site; SNP;
KW single nucleotide polymorphism; antiinflammatory; neuroleptic;
KW haplotyping; genotyping; respiratory inflammatory disease;
KW neuropsychiatric disorder; monoaminergic brain system; gene; ds.
XX
OS Homo sapiens.
XX

FH	Key	Location/Qualifiers
FT	misc_feature	4024
FT		/*tag= a
FT		/note= "PS1: polymorphic base C or T"
FT	misc_feature	6142
FT		/*tag= b
FT		/note= "PS2: polymorphic base C or G"
FT	misc_feature	6211
FT		/*tag= c
FT		/note= "PS3: polymorphic base G or A"
FT	misc_feature	6213
FT		/*tag= d
FT		/note= "PS4: polymorphic base G or T"
FT	misc_feature	6397
FT		/*tag= e
FT		/note= "PS5: polymorphic base C or T"
FT	misc_feature	6398
FT		/*tag= f
FT		/note= "PS6: polymorphic base T or G"
FT	misc_feature	16326
FT		/*tag= g
FT		/note= "PS7: polymorphic base T or C"
FT	misc_feature	16464
FT		/*tag= h
FT		/note= "PS8: polymorphic base C or T"
FT	misc_feature	16855
FT		/*tag= i
FT		/note= "PS9: polymorphic base T or C"
FT	misc_feature	16959
FT		/*tag= j
FT		/note= "PS10: polymorphic base T or A"
FT	misc_feature	16997
FT		/*tag= k
FT		/note= "PS11: polymorphic base T or C"
FT	misc_feature	17761
FT		/*tag= l
FT		/note= "PS12: polymorphic base C or T"
FT	misc_feature	17763
FT		/*tag= m
FT		/note= "PS13: polymorphic base T or C"
FT	misc_feature	17780
FT		/*tag= n
FT		/note= "PS14: polymorphic base A or C"
FT	misc_feature	17897
FT		/*tag= o
FT		/note= "PS15: polymorphic base G or C"
FT	misc_feature	20047
FT		/*tag= p
FT		/note= "PS16: polymorphic base C or T"
FT	misc_feature	20070
FT		/*tag= q
FT		/note= "PS17: polymorphic base T or A"
FT	misc_feature	29050
FT		/*tag= r
FT		/note= "PS18: polymorphic base G or T"
FT	misc_feature	29230
FT		/*tag= s
FT		/note= "PS19: polymorphic base C or T"
FT	misc_feature	29395
FT		/*tag= t
FT		/note= "PS20: polymorphic base C or T"
FT	misc_feature	32343
FT		/*tag= u
FT		/note= "PS21: polymorphic base C or T"
FT	misc_feature	32572
FT		/*tag= v
FT		/note= "PS22: polymorphic base A or G"
FT	misc_feature	39464
FT		/*tag= w
FT		/note= "PS23: polymorphic base T or C"
FT	misc_feature	39766
FT		/*tag= x
FT		/note= "PS24: polymorphic base C or T"
FT	misc_feature	39848
FT		/*tag= y
FT		/note= "PS25: polymorphic base A or C"
FT	misc_feature	39910
FT		/*tag= z
FT		/note= "PS26: polymorphic base T or C"
FT	misc_feature	39984
FT		/*tag= aa
FT		/note= "PS27: polymorphic base T or C"

XX
PN WO200222652-A2.
XX
PD 21-MAR-2002.
XX
PF 17-SEP-2001; 2001WO-US042217.
XX
PR 15-SEP-2000; 2000US-0232895P.
XX
PA (GENA-) GENAISSANCE PHARM INC.
XX
PI Anastasio AE, Han J, Kliem SE, Sausker EA;
XX
DR WPI; 2002-393942/42.
XX
PT Novel genetic variants of soluble carrier family 18 (vesicular
PT monoamine), member 2 gene useful for screening drugs to treat diseases
PT e.g. neuropsychiatric disorders involving monoaminergic brain systems.
XX
PS Claim 15; Fig 1; 183pp; English.
XX
CC The present invention describes an isolated polynucleotide (I) having a
CC sequence (S1) comprising soluble carrier family 18 (vesicular monoamine),
CC member 2 (SLC18A2) isogene selected from 49 isogenes with regions of a
CC sequence (SS) of 40023 bp (see ABL51954), and defined by a corresponding
CC set of polymorphisms whose locations and identities are given in the
CC specification; or a sequence (S2) complementary to (S1). (I) has
CC antiinflammatory and neuroleptic activities, and can be used in gene
CC therapy. Methods from the present invention can be used for haplotyping
CC and genotyping the SLC18A2 gene in an individual. SLC18A2 is also known
CC as the vesicular monoamine transporter (VMAT2). (I) is useful in studying
CC the expression and function of SLC18A2, and in expressing the SLC18A2
CC protein for use in screening for candidate drugs to treat diseases

```
CC   related to SLC18A2 activity and in studying the effect of the variation
CC   on the biological activity of SLC18A2 as well as on the binding affinity
CC   of candidate drugs targeting SLC18A2 for the treatment of respiratory
CC   inflammatory diseases such as neuropsychiatric disorders involving
CC   monoaminergic brain systems. The present sequence represents the human
CC   SLC18A2 gene sequence, indicating the polymorphic sites in the sequence
CC   by the ambiguity code
XX
SQ   Sequence 40023 BP; 10020 A; 9128 C; 9557 G; 11291 T; 0 U; 27 Other;

  Query Match             5.8%;  Score 237.4;  DB 6;  Length 40023;
  Best Local Similarity   71.2%;  Pred. No. 3.8e-31;
  Matches  341;  Conservative    0;  Mismatches  136;  Indels    2;  Gaps    2;

Qy        230 ATATATATATATATATATATATACACACACACACACATATATATATACACATATATATAC 289
              | ||| | ||  ||| | | | |||||||| || ||| | ||| | ||||  | | |  |
Db      31298 ACATACACATTCATACACAAACACACACACTCATACACACATACACACACTGACACACGC 31357

Qy        290 ACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATAT 349
              | | | ||| ||| ||||||| |  | |  | ||| ||||||||| ||  ||| ||||
Db      31358 ATACACACACACATACACACATACACCCCCCCACATAAACACATACACATACACAAATAT 31417

Qy        350 ATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATAC 409
              | ||||||| ||  |  | ||||| |||||| || ||||||| | | |||||  | |   ||
Db      31418 ACATACACACATATGCATACATACATACACGTAGATACACATACACATACAAACACCCAC 31477

Qy        410 ACACATATATACACATATATACACATATATACACATATACACATATATACACATATATAC 469
              ||||||  |||  ||  | ||||||| | | | |||||||||| ||| |||||  |  |||
Db      31478 CCACATACATATGCACACATACACACACACATACATATACACACATACACACACCTAC 31537

Qy        470 ATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATATAT 528
              | | |||||||||  | |  | | | | | |||  || |||||||  | |||  |||||  | |
Db      31538 ACACATACACATACACACACACACAAATACAAGTACATACACAAACATATACACAGACAC 31597

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACATATAT 588
              | |||  ||  | |||  ||  ||  |||||||||||||||||  |||||  | |||||  |||||
Db      31598 ATACACATGCAAACACATGTACACACATATATACACATACATACACAGATACATATATAC 31657

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
              |||  | |||  | |||  |||  | |  | |  |  ||||  | |||||||  | |  |||||  | |
Db      31658 ATATACATA-ATACACATAAATACACACATGCATACACACATATACATAGACATACACAG 31716

Qy        649 ACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATA 707
              | | |  |||||||  | |||||  |||  ||||  |||  |||||||||  ||||  |  |||  |
Db      31717 ATACATATATACATATACATACATACTCATAAATACACACATATACATACAAACATACA 31775


RESULT 86
ADR07428
ID   ADR07428 standard; cDNA; 1821 BP.
XX
AC   ADR07428;
XX
DT   04-NOV-2004  (first entry)
XX
DE   Full length human cDNA useful for treating neurological disease Seq 934.
XX
KW   gene; ss; human; oligo-capping method; diagnostic marker; gene therapy;
KW   osteoporosis; neurological disease; Alzheimer's disease;
KW   Parkinson's disease; dementia; short memory; cancer;
KW   sense or motor function; emotional reaction; fear response; panic;
KW   osteopathic; neuroprotective; nootropic; antiparkinsonian; cytostatic;
KW   tranquiliser.
XX
OS   Homo sapiens.
XX
PN   EP1447413-A2.
XX
PD   18-AUG-2004.
XX
PF   12-FEB-2004; 2004EP-00003145.
XX
PR   14-FEB-2003; 2003JP-00102207.
PR   09-MAY-2003; 2003JP-00131452.
XX
PA   (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI   Isogai T,  Yamamoto J,  Nishikawa T,  Isono Y,  Sugiyama T,  Otsuki T;
PI   Wakamatsu A,  Ishii S,  Nagai K,  Irie R;
XX
DR   WPI; 2004-583265/57.
DR   P-PSDB; ADR09384.
XX
PT   New 1995 cDNA, useful for treating osteoporosis, neurological diseases,
PT   Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.
XX
PS   Claim 1; SEQ ID NO 934; 2686pp; English.
XX
CC   This invention relates to novel, isolated full length human cDNA
CC   molecules and the encoded proteins thereof. Specifically, it refers to
CC   cDNA clones obtained by an oligo-capping method, where none of these
CC   clones are identical to any known human mRNAs. The present invention
CC   describes an immunoassay to identify agonists and antagonists, as well as
CC   antibodies, antisense molecules and siRNAs that can all be used to bind
CC   to and modulate expression of the cDNA molecules. As such, these
CC   molecules are useful for diagnostic markers or therapeutic targets for
CC   the various diseases or morbid states. In particular, they are useful in
CC   gene therapy for treating osteoporosis, neurological disease, Alzheimer's
CC   disease, Parkinson's disease, dementia, short memory and various cancers,
CC   as well as for maintaining equilibrium of sense or motor function, and
CC   for treating emotional reaction, fear response and panic. Accordingly,
CC   they exhibit osteopathic, neuroprotective, nootropic, antiparkinsonian,
CC   cytostatic and tranquiliser activities. This polynucleotide is a full
CC   length human cDNA sequence of the invention. NOTE: This sequence is not
CC   given in the sequence listing of the specification but can be obtained on
CC   CD-ROM from the European Patent Office, Vienna Sub-office.
XX
SQ   Sequence 1821 BP; 752 A; 616 C; 149 G; 304 T; 0 U; 0 Other;

  Query Match             5.7%;  Score 236.6;  DB 13;  Length 1821;
  Best Local Similarity   69.0%;  Pred. No. 4.6e-31;
  Matches  339;  Conservative    0;  Mismatches  149;  Indels    3;  Gaps    1;

Qy        230 ATATATATATATATATATATATACACACACACACACATATATATATACACATATATATAC 289
              | | ||  | ||| |    | || |||||||| ||||||  | ||||||||||  | | ||
Db         44 ACACATGCACATACACTCACATGCACACACATACACATGCACATATACACATGCACACAC 103

Qy        290 ACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATAT 349
              |||  ||||||| | || || ||||||||||  | |||||||  ||||   ||||| ||
Db        104 GCATGCATACATACACACGCACATATACACATGCACATACACACTCATACACACACACAT 163

Qy        350 ATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATAC 409
              ||| |||||||   || || | | || |||  ||  |||||||  | ||||||||  | |
Db        164 ATACACACATACACATGCATACACATGCACTCATGCACACACACACATACACATGCACAT 223

Qy        410 ACACATATATACACATATATACACATATATACACATATACACATATATACACATATATAC 469
              |||||  | ||  ||  | ||||  |||||||| |||    | |  |||||  |  |||||  | ||
Db        224 ACACAAACATGCAGAGATATGCACATATACACATGCACATACATACACTCACATGCACAC 283

Qy        470 ATATATACACATATATAT---AATATACACACATATATATACACATATATACACACATAT 526
              |||  | |||||||  | ||    |||||||  ||||   | |||||||  | ||  ||||||  |
Db        284 ATAGACACACATACACATGCACATATACATACATGCACATACACAAACATGCACACAGAC 343

Qy        527 ATACACATATATACACATATATATACACATATATACACATATATACATATATACACATAT 586
              | |||  | |  |||||||  ||  |  | |||  | ||||||  ||||  ||||  |||  |||  ||
Db        344 ACACATACACATACACACATGCACATACACAGATACACTCATATGCATACATATACACAT 403

Qy        587 ATATACATATATACACATATATACATATATACACATATATACATATATACACACATATAT 646
              |   |||||||  |||    | |||||  |||  ||||||   ||  ||  ||||||||||||  |
Db        404 ACGCACATATACACATGAACATACACATACACACATGCATGCACATATACACACATGCAC 463

Qy        647 ACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATAT 706
              |||||  |    | | |  |||||  |    |||||  | |  ||  ||||  | |  |  ||  | ||
Db        464 ACACACATGCACATATACACACATGCATACACACACATGCACACAGACACACACACATGC 523

Qy        707 ATGTATATATA 717
```

```
Db        524 ACACACATACA 534


RESULT 87
ACC79619/c
ID   ACC79619 standard; DNA; 71843 BP.
XX
AC   ACC79619;
XX
DT   27-OCT-2003  (revised)
DT   06-AUG-2003  (first entry)
XX
DE   Zebrafish protein kinase C lambda related genomic DNA SEQ ID NO:3.
XX
KW   Zebrafish; Danio rerio; protein kinase C lambda; PKC-lambda; enzyme;
KW   cardiant; cytostatic; nephrotropic; heart disease; cancer; organ failure;
KW   epithelial-epithelial cell interaction; epithelial cell polarity;
KW   heart failure; kidney failure; gene; ds.
XX
OS   Danio rerio.
XX
PN   WO2003023048-A2.
XX
PD   20-MAR-2003.
XX
PF   06-SEP-2002; 2002WO-US028410.
XX
PR   06-SEP-2001; 2001US-0317653P.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Peterson R,  Fishman MC;
XX
DR   WPI; 2003-313257/30.
XX
PT   Diagnosing or prognosing a disease or condition related to Protein Kinase
PT   C lambda, by analyzing sample nucleic acid obtained from the test subject
PT   to determine mutation in gene encoding Protein kinase C lambda.
XX
PS   Disclosure; Page 63-83; 88pp; English.
XX
CC   The present invention describes a method (M1) for determining whether a
CC   test subject has, or is at risk of developing, a disease or condition
CC   related to protein kinase C lambda (PKC-lambda) (I). M1 comprises
CC   analysing a nucleic acid molecule of a sample from the subject to
CC   determine whether the subject has a mutation in a gene encoding (I),
CC   where the presence of a mutation indicates that the subject has, or is at
CC   risk of developing, a disease or condition related to (I). (I) has
CC   cardiant, cytostatic and nephrotropic activities. M1 is useful for
CC   determining whether a test subject has, or is at risk of developing, a
CC   disease or condition related to (I), e.g. heart disease, cancer, or
CC   condition associated with epithelial-epithelial cell interactions or
CC   epithelial cell polarity. (I) is useful for treating organ failure such
CC   as heart or kidney failure. (I) is also useful for raising antibodies
CC   specific for various regions of (I). M1 is suitable for detecting an
CC   increased likelihood of diseases or conditions associated with (I), so
CC   that appropriate intervention can be instituted before any symptoms
CC   occur. This is useful with patients in high risk groups for such diseases
CC   or conditions. Diagnostic method facilitates determination of the
CC   etiology of such an existing disease or condition in a patient, so that
CC   an appropriate approach to treatment can be selected. The present
CC   sequence represents zebrafish genomic DNA for PKC-lambda, which is used
CC   in the exemplification of the present invention. (Updated on 27-OCT-2003
CC   to standardise OS field)
XX
SQ   Sequence 71843 BP; 24017 A; 12903 C; 12592 G; 21955 T; 0 U; 376 Other;

  Query Match             5.7%;  Score 236;  DB 8;  Length 71843;
  Best Local Similarity   60.9%;  Pred. No. 6.7e-31;
  Matches  314;  Conservative    0;  Mismatches  198;  Indels    4;  Gaps    2;

Qy        209 AGAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACATA 268

Db      68772 AAAAAAAAAAAATNTCTATATATATATATATATATATATATATATATATATATATA 68713

Qy        269 TATATATACACATATATATACACATATATACATATATACACATATATACACATATATATA 328

Db      68712 TATATATATATATATATATATATATATANATATATATATATATATATATATATATATA 68653

Qy        329 CACATACATATGTACAC---ATATATATACACATATGTATACACATATATACACATATAT 385

Db      68652 TATATACATATATACACTATATATATATATATATATATATATATATATATATATATAT 68593

Qy        386 ACACACATATATACACATATATACACACATATATACACATATATACACATATATACACAT 445

Db      68592 ATATATATATATATATATATATATATANATATATATATATATATATATATATATATAT 68533

Qy        446 ATACACATATATACACATATATACATATATACACATATATATA-ATATACACACATATAT 504

Db      68532 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 68473

Qy        505 ATACACATATATACACACATATATACACATATATACACATATATATACACATATATACAC 564

Db      68472 ATATATATATATNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 68413

Qy        565 ATATATACATATATACACATATATATACATATATACACATATATACATATATACACATAT 624

Db      68412 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTGT 68353

Qy        625 ATACATATATACACACATATATACACATACATATACACACACATAGATATACATATATAT 684

Db      68352 GTGTGTGTGTGTGTGTGTGTGTGTGTGTATATATATATATATATATATATATATAT 68293

Qy        685 ACACATATATATACGTATATATATGTATATATATAT 720

Db      68292 ATATATATATATATATATATATATATATATATAT 68257


RESULT 88
ADC87374/c
ID   ADC87374 standard; DNA; 1192 BP.
XX
AC   ADC87374;
XX
DT   01-JAN-2004  (first entry)
XX
DE   Human GPCR gene SEQ ID NO:1827.
XX
KW   ds; gene; human; GPCR;
KW   guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1270724-A2.
XX
PD   02-JAN-2003.
XX
PF   18-JUN-2002; 2002EP-00013517.
XX
PR   18-JUN-2001; 2001JP-00246789. .
XX
PA   (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA   (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI   Suwa M,  Asai K,  Akiyama Y,  Aburatani H;
XX
DR   WPI; 2003-315783/31.
DR   P-PSDB; ADC87375.
XX
PT   New polynucleotide, useful for preparing a composition for treating a
PT   patient in need of increased or suppressed activity or expression of the
PT   guanosine triphosphate-binding protein coupled receptor.
XX
PS   Claim 1; SEQ ID NO 1827; 28pp; English.
XX
```

```
CC   The invention relates to a novel polynucleotide encoding a guanosine
CC   triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC   the invention may have a use in gene therapy. The polynucleotide and
CC   polypeptide are useful for preparing a composition for treating a patient
CC   in need of increased or suppressed activity or expression of the
CC   guanosine triphosphate-binding protein coupled receptor. The
CC   polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC   invention.
XX
SQ   Sequence 1192 BP; 384 A; 185 C; 211 G; 412 T; 0 U; 0 Other;

  Query Match             5.7%;  Score 235.8;  DB 10;  Length 1192;
  Best Local Similarity   69.2%;  Pred. No. 6.2e-31;
  Matches  355;  Conservative    0;  Mismatches  147;  Indels   11;  Gaps    2;

Qy        214 AAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACATATATAT 273
              |||||| |     | |||| |||||||| | ||||| || | ||||||| |||| |||||
Db       1000 AAAAAATATTATATATATACATATATATGTTTATATGTATATACACACATACATGTATAT 941

Qy        274 ATACACATATATATACACATATATACATATATACACATATATACACATATATATACACAT 333
              ||  | |||||  || ||| ||||||  | |  ||| ||| || ||       ||||||||||
Db        940 ATGTATATATACATGCACGTATATATACACATATACACATGTATGTGCGTATATACACAT 881

Qy        334 ACATATGTACACATATATATACACATATGTATACACATATATACACATATATACACACAT 393
              |||| | || | | | |||||||||| | | | |  ||||||||    ||||| ||||||
Db        880 ACATGTATATATACACATATACACATGTATGTGCGTATATATAC----ATATATACACAT 825

Qy        394 ATATACACATATATACACACATATATACACATATATACACATATATACACATATACACAT 453
              |||   | | |||| ||| ||||| |||  | | | |   |||||||| ||||||||||
Db        824 GTATGTGCGTGTATATACATATATACACATGTGTGTGCGTGTATATACATATATACACAT 765

Qy        454 ATATACACATATATACATATATACACATATATAT-------AATATACACACATATATAT 506
              | |    | | |||| | ||||| ||| | ||||         | || | |||||||||
Db        764 GTGTGTGCGTGTATATACATATATACACAGATATGTGTGTGCGTGTATATACATATATAC 705

Qy        507 ACACATATATACACACATATATACACATATATACACATATATATACACATATATACACAT 566
              ||| |||| |     |||||||||| ||| | ||| |||| | |    |||| |||| ||
Db        704 ACAGATATGTGTGTGTATATATACATATACACACAGATATGTGTGTGTATATGTACATAT 645

Qy        567 ATATACATATATACACATATATATACATATATACACATATATACATATATACACATATAT 626
              | | |||| | |  | ||||| ||| ||| ||| | ||   | |||| | | ||||
Db        644 ACACACATGTGTGTATATATACATATATACATATACACACATGTGTGTATATATACATAT 585

Qy        627 ACATATATATACACACATATATACACATACATATACACACACATAGATATACATATATATAC 686
              | | |||||||||| | | || | |||||||| ||| | | | |   |||||||||
Db        584 ATACATATACACACGTGTGTATATATACATATATACATATACACACGTGTGTATATATAC 525

Qy        687 ACATATATATACGTATATATATGTATATATATA 719
              | ||||| | || || || || || ||||||||
Db        524 ATATATACACACATACATGTACGTGTATATATA 492


RESULT 89
ABL33998/c
ID   ABL33998 standard; DNA; 21537 BP.
XX
AC   ABL33998;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 1971.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP007537.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful for
PT   diagnosis and treatment of diseases associated with abnormal cytosine
PT   methylation.
XX
PS   Claim 1; SEQ ID NO 1971; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention
XX
SQ   Sequence 21537 BP; 5893 A; 333 C; 4680 G; 10631 T; 0 U; 0 Other;

  Query Match             5.7%;  Score 235.2;  DB 6;  Length 21537;
  Best Local Similarity   69.6%;  Pred. No. 8.8e-31;
  Matches  348;  Conservative    0;  Mismatches  148;  Indels    4;  Gaps    2;

Qy        197 AAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACA 256
              ||||   | | || |||||  | |   | | |   | |       |||| ||  ||| | |
Db      16071 AAATTAAAACACACAAAAACTATATAACTTATCCAAAAACACACTATAAATCCATATATA 16012

Qy        257 CACACACACATATATATATACACATATATATACACATATATACATATATACACATATATA 316
              | | | | |||||||||||| | ||||| | | | ||| || |||||||| | |||||
Db      16011 TATATATATATATATATATATATATATACACATATATAAATCCATATATATATATACG 15952

Qy        317 CACATATATATACACATACATATGTACACATATATATACACATATGTATACACATATATA 376
              || |||||||   | ||| ||||  ||||| ||||||||||||||| || ||||||||||
Db      15951 CATATATATACGTATATATATATACACACACATATATACACATATATACACACATATATA 15892

Qy        377 CACATATATACACACATATATACACATATATACACACATATATACACATATATACACATA 436
              |||||||| ||| | |||||||| | |||||||||||||| | ||| | ||||||||||| |
Db      15891 CACATATACACATATATATATATATATATACACACACACATATATATATATACACACA 15832

Qy        437 TATACACATATACACATATATACACATATATACATATATACACATATATATA---ATATA 493
              | | | ||||| | | | | |||||||||||| ||||||||||| | | |||    |||||
Db      15831 CACATATATATATACACACACACATATATATATATATACACACACACATATATATATA 15772

Qy        494 CACACATATATATACACATATATACACACATATATACACATATATACACATATATATACA 553
              | | ||| | | |||||||||||| |      ||  | | | ||||| | ||||| | | |
Db      15771 TATATATACACACACACATATATATAATATAATCGATATAAATATATATATATAAAAAAA 15712

Qy        554 CATATATACACATATATACATATATACACATATATATACATATATACACATATATACATA 613
              |||||    |  |||| ||          |||  || || | || ||   |    ||   |
Db      15711 AATATACCTATTTATAAACTCTCTATTCCATTAATTTATACATCTAATAAACAATTATTT 15652

Qy        614 TATACACATATATACATATATACACACATATATACACATACATATACACACACATAGATA 673
              |||| | ||||||| |||||||||||||||||||| | ||| ||||| | | ||||| |||
Db      15651 TATATATATATATATATATATACACACATATATATA-ATATATATATATATACATACATA 15593

Qy        674 TACATATATATACACATATA 693
              | | ||| ||| |||||||
Db      15592 CATACATACATATACATATA 15573
```

```
RESULT 90
AAS46748
ID   AAS46748 standard; DNA; 26997 BP.
XX
AC   AAS46748;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Tumour suppressor gene derived chemically modified sequence #472.
XX
KW   Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer;
KW   tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW   cytosine methylation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200168912-A2.
XX
PD   20-SEP-2001.
XX
PF   15-MAR-2001; 2001WO-EP002955.
XX
PR   15-MAR-2000; 2000DE-01013847.
PR   06-APR-2000; 2000DE-01019058.
PR   07-APR-2000; 2000DE-01019173.
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2001-602752/68.
XX
PT   Fragments of chemically modified genes associated with tumor suppressor
PT   genes and oncogenes, useful in designing primers and probes for analyzing
PT   diseases associated with cytosine methylation state e.g. cancer.
XX
PS   Claim 1; SEQ ID NO 472; 27pp; English.
XX
CC   The invention relates to a nucleic acid comprising a sequence of 18
CC   bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC   bisulphite, of genes associated with tumour suppression and oncogenes
CC   having a sequence taken from 536 (actually 533 since numbers 408, 458 and
CC   500 are missing from the sequence listing) sequences (Ss) and sequences
CC   complementary to (Ss). The nucleic acid may be a peptide nucleic acid-
CC   oligomer (PNA) of at least 9 nucleotides and may form part of a set of
CC   probes for detecting the cytosine methylation state and/or single
CC   nucleotide polymorphisms and also to be used in an array for analysing
CC   diseases associated with CpG dinucleotides e.g. cancers and tumours. The
CC   probes can also be used in a method for ascertaining genetic and/or
CC   epigenetic parameters for the diagnosis and/or therapy of existing
CC   diseases or the predisposition to specific diseases, by analysing
CC   cytosine methylations. The parameters may be compared to another set of
CC   genetic and/or epigenetic parameters, the differences serving as basis
CC   for diagnosis and/or prognosis events which are disadvantageous to
CC   patients. The present sequence is one of the 533 genomic sequences
CC   derived from tumour suppressor genes and oncogenes. Sequences with even
CC   numbered Seq ID numbers are the complementary sequence of the
CC   corresponding odd numbered sequence (e.g. ID 2 and ID1, ID 536 and ID
CC   535, except for those whose partner sequence is missing). Note: The
CC   sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 26997 BP; 7650 A; 354 C; 5924 G; 13069 T; 0 U; 0 Other;

  Query Match             5.7%;  Score 234.2;  DB 4;  Length 26997;
  Best Local Similarity   69.7%;  Pred. No. 1.3e-30;
  Matches  345;  Conservative    0;  Mismatches  148;  Indels    2;  Gaps    2;

Qy        227 TTTATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||  |||||||||||||| |   | | | | ||||||  |||| |  ||||||

Db       5699 TATATATATACGTATATATATATATATATTTATATATATATATATGTATATATGTATATA 5758

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||  |||| || | || |||| | |||||||| | |||  ||| || | |

Db       5759 TATATATATATGTATATTTATATATGTATATATATATATATATA-ATATGTATATATATA 5817

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              ||||||||   |||| |||| | || |||| || |||||| | | || | |  | ||||

Db       5818 TATATATATGTATATATATATATATGTATATACGTATATATAAATATGTGTGTATATATG 5877

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | || | |||| |  |||||| | ||||||    |||||  |  ||| || |  ||||

Db       5878 TATATACGTGTATATATGTATATATATATATATGTGTATATATATGTATGTATATGTATA 5937

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || |||||||    ||||||||  |||  | |  |||||||| | |||| |  | |  ||

Db       5938 TATATATATATGTGTATATATATGTATGTATATGTATATATATATATATGTGTATATGTA 5997

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||  | ||||||  | |||||||||  |  | ||||   |||| || ||||||  | |||

Db       5998 TATGTATATATATGTATATATATATATATGTGTATATGTATATGTATATATATGTATATA 6057

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              ||||||  || | ||  | ||||||  |||||||   |||||||  |||||||      |||||

Db       6058 TATATATATGTGTATATATATATGTATATATATGTATATATATATATATATGTGTATATA 6117

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              |  | |||  | |  | | |  || ||  |  | ||||||    ||||||||  || | ||

Db       6118 TGTATATATGTGTGTATATATGTATATGTGTGTGTATATATGTTTATATATATGTGTGTA 6177

Qy        706 TATGTATATATATAT 720
              ||| ||||||||||||

Db       6178 TATATATATATATAT 6192


RESULT 91
AAV83939
ID   AAV83939 standard; DNA; 80595 BP.
XX
AC   AAV83939;
XX
DT   03-MAR-1999  (first entry)
XX
DE   HC-contig derived from normal human chromosome 10q25.2 region.
XX
KW   Yeast artificial chromosome; YAC; probe; eukaryotic chromosome;
KW   neocentromere; replication; extra-chromosomal element; segregation;
KW   cell division; artificial chromosome; gene therapy; mardel(10);
KW   human artificial chromosome; transgenic; chromosome 10; 10q25.2; ss.
XX
OS   Homo sapiens.
XX
PN   WO9851790-A1.
XX
PD   19-NOV-1998.
XX
PF   13-MAY-1998;   98WO-AU000352.
XX
PR   13-MAY-1997;   97AU-00006784.
PR   26-AUG-1997;   97AU-00008791.
XX
PA   (AMRA-) AMRAD OPERATIONS PTY LTD.
XX
PI   Choo K,  Du Sart D,  Cancilla MR;
XX
DR   WPI; 1999-009773/01.
XX
PT   New isolated nucleic acid comprising neocentromere sequences from
PT   eukaryotic chromosome - used to produce replicable, segregating
PT   artificial chromosomes that can carry large amounts of DNA for gene
PT   therapy.
XX
PS   Claim 8; Fig 6; 540pp; English.
```

```
XX
CC   The present sequence represents the HC-contig derived from normal human
CC   chromosome 10, 10q25.2 region. This region can be naturally mutated to
CC   produce an unusual chromosomal marker referred to as mardel(10). The
CC   mardel(10) marker is mitotically stable and contains a functional
CC   neocentromere at a location regarded as non-centromeric. This
CC   neocentromere maps to q25.2 on chromosome 10. The specification describes
CC   nucleic acid sequences derived from a eukaryotic chromosome, including a
CC   neocentromere or its functional derivative or hybrid, that are able, in a
CC   compatible cell, of replicating, acting as extra-chromosomal element and
CC   segregating during cell division. The sequences can be used to construct
CC   artificial chromosomes for use in gene therapy comprising a replicable,
CC   segregating nucleic acid that confers a specific phenotype on cells.
CC   Human artificial chromosomes can propagate in human cells and carry large
CC   amounts of DNA (e.g. therapeutic genes), and, being extra-chromosomal,
CC   they are not mutagenic. The artificial chromosomes are also useful for
CC   generation of transgenic plants and animals, in production of proteins
CC   and to make diagnostic reagents, e.g. for expression of cytokines,
CC   receptors and growth factors, or to increase the copy number of a gene in
CC   a cell. The constructs may also be used for functional and structural
CC   analysis of chromosomes
XX
SQ   Sequence 80595 BP; 23183 A; 16613 C; 16824 G; 23975 T; 0 U; 0 Other;

  Query Match             5.7%;  Score 234.2;  DB 2;  Length 80595;
  Best Local Similarity   66.3%;  Pred. No. 1.4e-30;
  Matches  352;  Conservative    0;  Mismatches  178;  Indels    1;  Gaps    1;

Qy        189 CAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATA 248
              ||  |   ||||| || | | |     ||||     ||   | ||| |||||||| |||
Db      15479 CATGTTTGAAATATCTTATAATTTTTAAAAATTAAAATGGGAGATACATATATATGTATT 15538

Qy        249 TATACACACACACACACATATATATATACACATATATATACACATATATACATATATACA 308
              |||  |   | | |   || || ||||||||| |||||||| | ||||| | ||| | |
Db      15539 TATGTATGTATATATGTATGTACATATACACACATATATACATAAATATATACATAAATA 15598

Qy        309 CATATATACACATATATATACACATACATATGTACACATATATATACACATATGTATACA 368
              ||||||    |||||| ||| | | || | || |  | |||||| |||||  |||  |
Db      15599 TGTATATATGTGTATATAGACATAAATATGTATATATGTGTATATATACATAAATATGTA 15658

Qy        369 CATATATACACATATATACACACATATATACACATATATACACACATATATACACATATA 428
              |||| |  | ||| | | | | || |||| |  | |||| | ||||| |||   |||||
Db      15659 TATATGTGTATATAGACATAAATATGTATATATGTGTATATAGACATAAATATGTATATA 15718

Qy        429 TACACATATATACACATATACACATATATACACATATATACATATATACACATATATATA 488
              |    ||||| ||| | |||   ||||||    ||||| ||||| |||   |||||| |
Db      15719 TGTGTATATAGACATAAATATGTATATATGTGTATATAGACATAAATATGTATATATGT- 15777

Qy        489 ATATACACACATATATATACACATATATACACACATATATACACATATATACACATATAT 548
              |||| | ||||| |||  | |||| |  | | | | ||| | || |||| |  | |||
Db      15778 GTATATAGACATAAATATGTATATATGTGTATATAGACATAAATATGTATATATGTGTAT 15837

Qy        549 ATACACATATATACACATATATACATATATACACATATATATACATATATACACATATAT 608
              ||| |||||  |||   ||||||   |||||| ||||| ||||  ||||||    |||||
Db      15838 ATAGACATAAATATGTATATATGTGTATATAGACATAAATATGTATATATGTGTATATAG 15897

Qy        609 ACATATATACACATATATACATATATACACACATATATACACATACATATACACACACAT 668
              |||||  |||   ||||||   |||||| ||| | ||||  | |||  | || | | ||||
Db      15898 ACATAAATATGTATATATGTGTATATAGACATAAATATGTATATATGTGTATATAGACAT 15957

Qy        669 AGATATACATATATATACACATATATATACGTATATATATGTATATATATA 719
              | ||||  ||||||  |  | ||| | |||  ||| |||||  | |  ||||||
Db      15958 AAATATGTATATATGTGTATATAGACATAAATATGTATATATGTGTATATA 16008


RESULT 92
ABQ88198
ID   ABQ88198 standard; cDNA; 154902 BP.
XX
AC   ABQ88198;
XX
DT   18-SEP-2002  (first entry)
XX
DE   Human osteoblast differentiation related cDNA SEQ ID NO 105.
XX
KW   Human; osteoblast; stem cell differentiation; bone tissue deposition;
KW   osteoporosis; osteopathic; ss.
XX
OS   Homo sapiens.
XX
PN   WO200250301-A2.
XX
PD   27-JUN-2002.
XX
PF   18-DEC-2001; 2001WO-US048276.
XX
PR   18-DEC-2000; 2000US-0255882P.
PR   24-APR-2001; 2001US-0285691P.
XX
PA   (GENE-) GENE LOGIC INC.
PA   (PROC ) PROCTER & GAMBLE CO.
XX
PI   Ji D,  Axelrod DW,  Cook JS,  Jaiswal N,  Einstein R,  Houghton A;
PI   Mertz L;
XX
DR   WPI; 2002-557663/59.
XX
PT   Use of genes and their expression profiles associated with osteoblast
PT   differentiation for screening modulators bone formation, for diagnosing
PT   or treating e.g. osteoporosis, or as markers for the differentiation
PT   process.
XX
PS   Claim 1; SEQ ID NO 105; 78pp + Sequence Listing; English.
XX
CC   The invention relates to genes and their expression profiles are used
CC   for: (a) screening modulators of precursor stem cell differentiation into
CC   osteoblasts, or bone tissue deposition; (b) diagnosing abnormal
CC   deposition of bone tissue, abnormal rate of osteoblast formation or
CC   osteoporosis; or (c) treating or monitoring treatment of the conditions
CC   cited in (b), or monitoring the progression of bone tissue deposition.
CC   Specific conditions include postmenopausal osteoporosis, glucocorticoid
CC   osteoporosis or male osteoporosis, osteopenia, osteodystrophy, drug-
CC   induced abnormalities in bone formation or bone loss, conditions that
CC   involve altered bone metabolism (e.g. idiopathic juvenile osteoporosis),
CC   skeletal disease linked to breast cancer, mastocytosis, Fanconi syndrome
CC   or fibrous dysplasia. The present sequence is that of an osteoblast
CC   differentiation associated cDNA marker of the invention. Note: The
CC   sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 154902 BP; 43917 A; 31458 C; 32848 G; 46679 T; 0 U; 0 Other;

  Query Match             5.7%;  Score 234.2;  DB 6;  Length 154902;
  Best Local Similarity   75.4%;  Pred. No. 1.4e-30;
  Matches  304;  Conservative    0;  Mismatches   98;  Indels    1;  Gaps    1;

Qy        286 ATACACATATATACATATATACACATATATACACATATATATACACATACATATGTACAC 345
              ||||| ||||||||| ||||| | | ||||| | | |||||||||| ||| | ||||| ||
Db      31565 ATACATATATATACACATATATATACATATATATACATATATACATATATACATGTATAC 31624

Qy        346 ATATATATACACATATGTATACACATATATACACATATATACACACATATATACACATAT 405
              | ||| ||| | | || |||||||| || ||  |||||||||| | | | | |   ||
Db      31625 ACATACATATATACATATATACACACATTTATGCATATATACATAGACACACATTTATGC 31684

Qy        406 ATACACACATATATACACATATATACACATATATACACATATACACATATATACACATAT 465
              ||||| | ||| | |||   ||||| || ||||||||||||||  |||||||||  ||||
Db      31685 ATACATATATACACACATGCATATATACCTATATACACATATATGCATATATACCTATAT 31744

Qy        466 ATACATATATACACATATATATAATATACACACATATATATACACATATATACACACATA 525
              | |||||||| || ||||| || ||||||||||||  ||||| |||||||| ||||| |||
Db      31745 ACACATATATGCATATATACAT-ATATACACACATGCATATATACATATACACACATATA 31803

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |  | | ||| ||| | | |||||||||||| |||| || ||||| | ||||| ||| | |
Db      31804 TGCATATATACATATATACATATATACACACATATGCATATATACATATATACACATACA 31863
```

```
Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||||| ||||||||| ||||| | ||||| ||| | ||| ||||| ||| | || ||
Db      31864 TATATACACATATACACACATATATACATATATACACACATATATATACACATATATGTA 31923

Qy        646 TACACATACATATACACACACATAGATATACATATATATACAC 688
              |||||| ||||||| | | | ||| | || | |    |||
Db      31924 TACACACACATATATATATATATATTTTTAGCTGGAGTCTCAC 31966


RESULT 93
ADC87318/c
ID   ADC87318 standard; DNA; 1210 BP.
XX
AC   ADC87318;
XX
DT   01-JAN-2004  (first entry)
XX
DE   Human GPCR gene SEQ ID NO:1771.
XX
KW   ds; gene; human; GPCR;
KW   guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1270724-A2.
XX
PD   02-JAN-2003.
XX
PF   18-JUN-2002; 2002EP-00013517.
XX
PR   18-JUN-2001; 2001JP-00246789.
XX
PA   (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA   (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI   Suwa M,  Asai K,  Akiyama Y,  Aburatani H;
XX
DR   WPI; 2003-315783/31.
DR   P-PSDB; ADC87319.
XX
PT   New polynucleotide, useful for preparing a composition for treating a
PT   patient in need of increased or suppressed activity or expression of the
PT   guanosine triphosphate-binding protein coupled receptor.
XX
PS   Claim 1; SEQ ID NO 1771; 28pp; English.
XX
CC   The invention relates to a novel polynucleotide encoding a guanosine
CC   triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC   the invention may have a use in gene therapy. The polynucleotide and
CC   polypeptide are useful for preparing a composition for treating a patient
CC   in need of increased or suppressed activity or expression of the
CC   guanosine triphosphate-binding protein coupled receptor. The
CC   polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC   invention.
XX
SQ   Sequence 1210 BP; 391 A; 154 C; 200 G; 465 T; 0 U; 0 Other;

  Query Match             5.7%;  Score 233.6;  DB 10;  Length 1210;
  Best Local Similarity   68.5%;  Pred. No. 1.5e-30;
  Matches  337;  Conservative    0;  Mismatches  154;  Indels    1;  Gaps    1;

Qy        230 ATATATATATATATATATATATACACACACACACATATATATATACACATATATATAC 289
              ||| || ||||   |||||||||||||| | |   | |   |||||| |||| ||| ||
Db        885 ATACATGTATACGAATATATATACACATATATGTATACGAATATATATACATGTATGTAT 826

Qy        290 ACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATAT 349
              ||  ||||| ||| | | | || |||||  ||||||||||||||  || | |||  ||||
Db        825 ACGAATATATATACACATATATGTATACGAATATATATACACATGTATGTATACGAATAT 766

Qy        350 ATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATAC 409
              ||||||||||| | | || ||  ||||| | | | ||  | |  ||||   | ||||||||
Db        765 ATATACACATGTATGTATACGAATATATATACACATGTATATGTATACGAATATATATAC 706

Qy        410 ACACATATATACACATATATACACATATATACACATATACACATATATACACATATATAC 469
              |||  |||||  | | | ||| | ||| | |||||| || |   ||| ||||| || ||
Db        705 ACATGTATATGTATACAAATATATATACACACACATGTATATGAATACACACACATGTAT 646

Qy        470 ATATATACACATATATAT-AATATACACACATATATATACACATATATACACACATATAT 528
              ||  ||||||| |||||| |||| |||||||||| ||||   ||| | |||||||||| |
Db        645 ATGAATACACACATATATGAATACACACACATATGTATATGAATACACACACACATATGT 586

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACATATAT 588
              | |   ||| ||||| |  |||  | ||  |||||| | |   || ||||   ||| |
Db        585 ATATGAATACACACACACGTATGTATATGAATACACACACACGTATGTATATGAATACAC 526

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
              | |||  |||  | ||  ||||| | | ||  || ||||   ||| |||||||  |||
Db        525 ACACACGTATGTATATGAATACACACACACGTATGTATATGAATACACACACACGTATGT 466

Qy        649 ACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATAT 708
              | ||  ||| | |||||| ||  ||||   ||| |||||||  ||| ||  |  ||| ||
Db        465 ATATGAATACATACACACGTATGTATATGAATACATACACACGTATGTATATGAATACAT 406

Qy        709 GTATATATATAT 720
              | | ||||||
Db        405 ACACACATATAT 394


RESULT 94
AAF84800/c
ID   AAF84800 standard; DNA; 110000 BP.
XX
AC   AAF84800;
XX
DT   09-JUL-2001  (first entry)
XX
DE   Nucleotide sequence of the human SPG4 gene.
XX
KW   Human; SPG4 gene; spastin; PSF-AD; gene therapy;
KW   autosomal dominant familial spastic paraplegia; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             9932. .102009
FT                   /*tag=  a
FT                   /note= "contains introns"
FT   exon            9932. .10471
FT                   /*tag=  b
FT                   /number= 1
FT   intron          10472. .33718
FT                   /*tag=  c
FT                   /number= 1
FT   exon            33719. .33805
FT                   /*tag=  d
FT                   /number= 2
FT   intron          33806. .35748
FT                   /*tag=  e
FT                   /number= 2
FT   exon            35749. .35832
FT                   /*tag=  f
FT                   /number= 3
FT   intron          35833. .45022
FT                   /*tag=  g
FT                   /number= 3
FT   exon            45023. .45118
FT                   /*tag=  h
FT                   /number= 4
FT   intron          45119. .60863
FT                   /*tag=  i
FT                   /number= 4
FT   exon            60864. .61051
FT                   /*tag=  j
FT                   /number= 5
FT   intron          61052. .61927
```

```
FT                   /*tag= k
FT                   /number= 5
FT   exon            61928. .62061
FT                   /*tag= l
FT                   /number= 6
FT   intron          62062. .62344
FT                   /*tag= m
FT                   /number= 6
FT   exon            62345. .62438
FT                   /*tag= n
FT                   /number= 7
FT   intron          62439. .73173
FT                   /*tag= o
FT                   /number= 7
FT   exon            73174. .73248
FT                   /*tag= p
FT                   /number= 8
FT   intron          73249. .74633
FT                   /*tag= q
FT                   /number= 8
FT   exon            74634. .74705
FT                   /*tag= r
FT                   /number= 9
FT   intron          74706. .82788
FT                   /*tag= s
FT                   /number= 9
FT   exon            82789. .82864
FT                   /*tag= t
FT                   /number= 10
FT   intron          82865. .83102
FT                   /*tag= u
FT                   /number= 10
FT   exon            83103. .83194
FT                   /*tag= v
FT                   /number= 11
FT   intron          83195. .83334
FT                   /*tag= w
FT                   /number= 11
FT   exon            83335. .83414
FT                   /*tag= x
FT                   /number= 12
FT   intron          83415. .88129
FT                   /*tag= y
FT                   /number= 12
FT   exon            88130. .88172
FT                   /*tag= z
FT                   /number= 13
FT   intron          88173. .89561
FT                   /*tag= aa
FT                   /number= 13
FT   exon            89562. .89641
FT                   /*tag= ab
FT                   /number= 14
FT   intron          89642. .91162
FT                   /*tag= ac
FT                   /number= 14
FT   exon            91163. .91233
FT                   /*tag= ad
FT                   /number= 15
FT   intron          91234. .93443
FT                   /*tag= ae
FT                   /number= 15
FT   exon            93444. .93484
FT                   /*tag= af
FT                   /number= 16
FT   intron          93485. .100599
FT                   /*tag= ag
FT                   /number= 16
FT   exon            100600. .102009
FT                   /*tag= ah
FT                   /number= 17
XX
PN   FR2798138-A1.
XX
PD   09-MAR-2001.
XX
PF   03-SEP-1999;   99FR-00011097.
XX
PR   03-SEP-1999;   99FR-00011097.
XX
PA   (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI   Weissenbach J,  Hazan J;
XX
DR   WPI; 2001-283966/30.
XX
PT   New human nucleic acid from the SPG4 gene, useful e.g. for diagnosis of
PT   autosomal dominant familial spastic paraplegia and in drug screening.
XX
PS   Claim 2; Page 45-106; 145pp; French.
XX
CC   The present sequence represents a human SPG4 gene. The SPG4 gene encodes
CC   a spastin polypeptide. Mutations in the SPG4 gene are responsible for
CC   autosomal dominant familial spastic paraplegia. SPG4 polynucleotides, and
CC   their fragments, are used to screen DNA banks for sequences that encode
CC   spastin (particularly sequences in other mammals, specifically mice); to
CC   identify SPG4 mutations, or other genetic anomalies, particularly for
CC   diagnosis of autosomal dominant familial spastic paraplegia (PSF-AD); to
CC   identify promoters and other regulatory elements of the SPG4 gene; for
CC   detection and amplification; for recombinant production of spastin; and
CC   for diagnostic genotyping of PSF-AD
XX
SQ   Sequence 110000 BP; 30622 A; 21640 C; 22817 G; 34921 T; 0 U; 0 Other;

  Query Match             5.6%;  Score 232.8;  DB 5;  Length 110000;
  Best Local Similarity   67.2%;  Pred. No. 2.4e-30;
  Matches  449;  Conservative    0;  Mismatches  207;  Indels   12;  Gaps    8;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC--AAAAT 58
              ||||  ||| |||||||  ||||||||||||||||||| |||||||| ||||  |||||
Db      56570 AGTTTGAGAACAGCCTGGCCAACATGGTGAAACCCCATTTCTACTAAAGATACAAAAAAT 56511

Qy         59 TAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGA 118
              ||| | ||| ||||| ||  |||||||||||||| ||||  |||| | || ||||||||
Db      56510 TAGCCTGGCGTGGTGCCACGTGCCTGTAATCCCAGGTACTCGGGAGACTGAAGCAGGAGA 56451

Qy        119 ATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCC 177
              ||  |||||||||  ||  |  ||||||||||  || ||||||    ||| |||| |||
Db      56450 ATTGCTTGAACCTG--GGAAGCAGAGGTTGCGGTAAGCCAAGATAGCACCACTGCATTCC 56393

Qy        178 AGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATAT 237
              ||||  ||  || || || | |   ||||||  |||||||| ||    |   | | | |
Db      56392 AGCCTGGGGGAC-AGGGCAAGACTCCATCTCCAAAAAAAAACAAAAAAAACAAAAAAAAA 56334

Qy        238 ATATATATATATATACACACACACACATATATATATACACATATATATACACATATAT 297
              | | | | | | || | |   | ||||| ||||| | | |||||||||     ||| ||
Db      56333 AAACAAAAAAAACTATATATGTATACACACATATACAAATACATATATAACTGTATACAT 56274

Qy        298 ACATATATACACATATATACACATATATATACACATACATATGTACACATATAT-ATACA 356
              | |  | || | ||| ||  |||  |||||| |||||  ||| |  | |||  | |||||
Db      56273 ATA-GTTTATATATACATTTACAGCTATATATACATAGTTATTTTGATATAACTAATACA 56215

Qy        357 CATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATA 416
              ||||   || | || |||| |||||| | | |||||| ||   ||||||  | |   |
Db      56214 TATATAACTATATATGTATATACATATACATATGTAACTATATATGTATATGCA 56155

Qy        417 TATACACATATATACACATATATACACATATACACATATATACACATATATACATATATA 476
              |||||| || ||   | |||| |  |   ||| |||||| ||   ||||||  ||||  |
Db      56154 TATACATATGTAACTATATATGTGTATGCATATACATATGTAACTATATATGTATATGCA 56095

Qy        477 CACATATATATAA-TATACACACATATATATACACATATATACACACATATATACACATA 535
              | | |||| ||| |||| |   ||||  ||| |||||| ||   | |||| ||    |
Db      56094 TATAGATATGTAACTATATATGTATATGCATATACATATGTAACTATATATGTATGTGTG 56035

Qy        536 TATACACATATATATACACATATATACACATATATACATATATACACATATATATACATA 595
              |||||| || ||  || | || |||   |||||| | |||||   | ||| | |||  |
```

```
Db      56034 TATACATATGTAACTATATATGTATGTGCATATACATATATAACTATATACAGATAGCT- 55976

Qy        596 TATACACATATATACATATATACACATATATACATATATACACACATATATACACATACA 655
              || |||||||||  |||||| |||||||   || | ||   | ||  | ||  |   |
Db      55975 --TATACATATATAGTTATATATACATATAGTTATTTTTAGTTATATCAAAACTAAACAA 55918

Qy        656 TATACACA 663
              |||| | |
Db      55917 TATATATA 55910


RESULT 95
AAV83940
ID   AAV83940 standard; DNA; 80240 BP.
XX
AC   AAV83940;
XX
DT   03-MAR-1999  (first entry)
XX
DE   NC-contig derived from mardel(10) on chromosome 10q25.2.
XX
KW   Yeast artificial chromosome; YAC; probe; eukaryotic chromosome;
KW   neocentromere; replication; extra-chromosomal element; segregation;
KW   cell division; artificial chromosome; gene therapy; mardel(10);
KW   human artificial chromosome; transgenic; chromosome 10; 10q25.2; ss.
XX
OS   Homo sapiens.
XX
PN   WO9851790-A1.
XX
PD   19-NOV-1998.
XX
PF   13-MAY-1998;   98WO-AU000352.
XX
PR   13-MAY-1997;   97AU-00006784.
PR   26-AUG-1997;   97AU-00008791.
XX
PA   (AMRA-) AMRAD OPERATIONS PTY LTD.
XX
PI   Choo K,  Du Sart D,  Cancilla MR;
XX
DR   WPI; 1999-009773/01.
XX
PT   New isolated nucleic acid comprising neocentromere sequences from
PT   eukaryotic chromosome - used to produce replicable, segregating
PT   artificial chromosomes that can carry large amounts of DNA for gene
PT   therapy.
XX
PS   Claim 9; Fig 16A; 540pp; English.
XX
CC   The present sequence represents the NC-contig derived from a mutated
CC   human chromosome 10, 10q25.2 region. The sequence contains an unusual
CC   chromosomal marker referred to as mardel(10). The mardel(10) marker is
CC   mitotically stable and contains a functional neocentromere at a location
CC   regarded as non-centromeric. This neocentromere maps to q25.2 on
CC   chromosome 10. The specification describes nucleic acid sequences derived
CC   from a eukaryotic chromosome, including a neocentromere or its functional
CC   derivative or hybrid, that are able, in a compatible cell, of
CC   replicating, acting as extra-chromosomal element and segregating during
CC   cell division. The sequences can be used to construct artificial
CC   chromosomes for use in gene therapy comprising a replicable, segregating
CC   nucleic acid that confers a specific phenotype on cells. Human artificial
CC   chromosomes can propagate in human cells and carry large amounts of DNA
CC   (e.g. therapeutic genes), and, being extra-chromosomal, they are not
CC   mutagenic. The artificial chromosomes are also useful for generation of
CC   transgenic plants and animals, in production of proteins and to make
CC   diagnostic reagents, e.g. for expression of cytokines, receptors and
CC   growth factors, or to increase the copy number of a gene in a cell. The
CC   constructs may also be used for functional and structural analysis of
CC   chromosomes
XX
SQ   Sequence 80240 BP; 23102 A; 16537 C; 16747 G; 23846 T; 0 U; 8 Other;

  Query Match             5.6%;  Score 232.2;  DB 2;  Length 80240;
  Best Local Similarity   64.9%;  Pred. No. 3e-30;
  Matches  359;  Conservative    0;  Mismatches  193;  Indels    1;  Gaps    1;

Qy        217 AAAAAGGCATTTTATATATATATATATATATACACACACACACATATATATATA 276
              |||||    |  |   | ||| |||||||| ||| ||   |   | | |  ||| || |||
Db      15240 AAAAATTAAAATGGGAGATACATATATATGTATTTATGTATGTATATATGTATGTACATA 15299

Qy        277 CACATATATATACACATATATACATATATACACATATATACACATATATATACACATACA 336
              ||| | ||||| ||||| ||| ||| ||| | || |||| |   | |||||| |||||
Db      15300 TACACACATATATACATAAATATATACATAAATATGTATATATGTGTATATAGACATAAA 15359

Qy        337 TATGTACACATATATATACACATATGTATACACATATATACACATATATACACACATATA 396
              |||||| | || | |||| | | ||  |||   ||||||    ||||| ||| | ||||
Db      15360 TATGTATATATGTGTATATATACATAAATATGTATATATGTGTATATAGACATAAATATG 15419

Qy        397 TACACATATATACACACATATATACACATATATACACATATATACACATATACACATATA 456
              || | || | || | | | ||| | |  ||||||    |||||| ||||| | |  ||||
Db      15420 TATATATGTGTATATAGACATAAATATGTATATATGTGTATATAGACATAAATATGTATA 15479

Qy        457 TACACATATATACATATATACACATATATATAATATACACACATATATATACACATATAT 516
              |     ||||| ||||| |||    |||||| |  |||| | ||||| ||||  | |||| |
Db      15480 TGTGTATATAGACATAAATATGTATATATGT-GTATATAGACATAAATATGTATATATGT 15538

Qy        517 ACACACATATATACACATATATACACATATATATACACATATATACACATATATACATAT 576
              | | | | ||| | || |||| |  | |||||| ||||| |||   ||||||   |||
Db      15539 GTATATAGACATAAATATGTATATATGTGTATATAGACATAAATATGTATATATGTGTAT 15598

Qy        577 ATACACATATATATACATATATACACATATATACATATATACACATATATACATATATAC 636
              ||| ||||| |||| ||||||    ||||| ||||| |||   ||||||   ||||||
Db      15599 ATAGACATAAATATGTATATATGTGTATATAGACATAAATATGTATATATGTGTATATAG 15658

Qy        637 ACACATATATACACATACATATACACACACATAGATATACATATATATACACATATATAT 696
              ||| | ||||  | |||  | || | | ||||| ||||  ||||||  |  | ||| | ||
Db      15659 ACATAAATATGTATATATGTGTATATAGACATAAATATGTATATATGTGTATATAGACAT 15718

Qy        697 ACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCA 756
              |  ||| ||||||| | ||||||  |   |   |  |||   |   | |  |   | |
Db      15719 AAATATGTATATGTGTGTATATAGACATAAATATGTATATATGTGTGTATATAGACATAA 15778

Qy        757 CTGGGGATAGAGG 769
              |  | ||| | |
Db      15779 ATATGTATATATG 15791


RESULT 96
ACN44026/c
ID   ACN44026 standard; DNA; 122923 BP.
XX
AC   ACN44026;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Human genomic sequence hCG24510.
XX
KW   Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO2003073826-A2.
XX
PD   12-SEP-2003.
XX
PF   28-FEB-2003; 2003WO-US006235.
XX
PR   01-MAR-2002; 2002US-00087192.
XX
PA   (SAGR-) SAGRES DISCOVERY.
XX
PI   Morris DW;
XX
DR   WPI; 2003-328604/31.
XX
```

```
PT   Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
PT   comprises a nucleotide sequence.
XX
PS   Claim 1; SEQ ID NO 268; 0pp; English.
XX
CC   The present invention relates to novel DNA and protein sequences which
CC   are associated with carcinomas. The sequences are useful for: (i) for
CC   screening drug candidates; (ii) for screening of bioactive agent capable
CC   of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
CC   a bioactive agent capable of modulating the activity of CAP; (iv) for
CC   evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
CC   carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
CC   carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
CC   (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
CC   determining Carcinoma Associated (CA) gene copy number. In addition, the
CC   CA genes are useful as DNA vaccines and the CAP are useful as markers of
CC   carcinoma including lymphoma. The present sequence is one such CA coding
CC   sequence. Note: This patent is an equivalent to basic patent
CC   US2002182586A1, for which no sequence data was published
XX
SQ   Sequence 122923 BP; 28628 A; 24391 C; 25218 G; 32559 T; 0 U; 12127 Other;

  Query Match             5.6%;  Score 231.8;  DB 11;  Length 122923;
  Best Local Similarity   62.3%;  Pred. No. 3.5e-30;
  Matches  447;  Conservative    0;  Mismatches  257;  Indels   13;  Gaps    6;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAATTA 60
              ||||  |||||||||||| |||||||||  |||||| |||||||| |||||||||||| |
Db      43687 AGTTTGAGACCAGCCTGGGCAACATGGCAAAACCCTATCTCTACAAAAAATACAAAAATG 43628

Qy         61 GA-CAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
               | ||||  | |||  |  | |||||| || || |||||| || |||| |||||| || |
Db      43627 CACCAGGTGTAGTGGCATGCACCTGTAGTCTCAACTACTTGGGGGGCCAAGGCAGAAGGA 43568

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCCA 178
              || ||||||||  | || |||  ||||||| | ||| ||||| ||  | | |||||||||
Db      43567 TCGCTTGAACC--CGGGAGGTCAAGGTTGCAGTGAGCCAAGACCACACTACTGCACTCCA 43510

Qy        179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATA 238
              | | |||  | || || || |     | ||||||||||||        ||||||||||||
Db      43509 GTCTAGGTGATAA-AGTGAGACCCTGTTTCAGAAAAAAAAA-------TTTATATATATA 43458

Qy        239 TATATATATATATACACACACACACATATATATATACACATATATATACACATATATA 298
              ||| | | | | |               | | | | |  |  | | | ||    ||||||
Db      43457 TATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTATGTGTGTGTATGTGTATATA 43398

Qy        299 CATATATACACATATATACACATATATATACACATACATATGTACACA-TATATATACAC 357
              | | |    || | |  |||||||||||| |   |  | || || | | ||||||   |
Db      43397 TGTGTGTGTCTATGTGTGTACATATATATAAAATTGTACATATAAAAATTATATAAGTAT 43338

Qy        358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417
              |||| |  ||  |||||||  || | |  | |||||    ||  ||     | | || | |  |||
Db      43337 ATATATGCACGTGTATATACGTATGTATGTGTGTATGTATGTGTGTGTATATATGTAT 43278

Qy        418 ATACACATATATACACATATATACACATATACACATATATACACATATATACATATATAC 477
              |  |  ||||||   || | |  | |||||    ||  ||     | |||  || |||
Db      43277 GTGTATGTATATATGTATGTGTGTATATATATGTATGCATGTGTGTGTATGTATGTATGN 43218

Qy        478 ACATATATATAATATACACACATATATATACACATATATACACACATATATACACATATA 537
                              ||| ||||||| ||| | ||| | | | | | | | |||||
Db      43217 NNNNNNNNNNNNNNNNNNACACATATATATACACACATATATATACACACATATATATA 43158

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATACATATA 597
              ||| ||||||||||||||||||| ||| ||||| ||| | ||| |||||||| |||||||
Db      43157 CACATATATATATACACATATATATACACATATATATACACACATATATATACACATATA 43098

Qy        598 TACACATATATACATATATACACATATATACATATATACACACATATATACACATACATA 657
              || ||| | ||||| | | ||||| | | ||| | | |||||||| | | | | ||| |||
Db      43097 TATACACACATACACACACACACACACACACACACACACACACACACACATATATATATA 43038

Qy        658 TACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATAT 714
              || |  |  | ||  |    ||      | | |  |  | |  | | ||| | | ||
Db      43037 TATATGCCAAGAGTAAGTATTAGGGGAATAAAGCTTAAAATAAAAAAATAAGAAAAT 42981
```

```
RESULT 97
ADL13505/c
ID   ADL13505 standard; DNA; 76146 BP.
XX
AC   ADL13505;
XX
DT   06-MAY-2004  (first entry)
XX
DE   Osteoarthritis-associated polymorphic nucleotide #37.
XX
KW   ds; gene; osteopathic; antiinflammatory; antiarthritic; gene therapy;
KW   joint space narrowing; osteophyte development; joint pain;
KW   osteoarthritis; SNP; single nucleotide polymorphism.
XX
OS   Homo sapiens.
XX
PN   WO2003054166-A2.
XX
PD   03-JUL-2003.
XX
PF   19-DEC-2002; 2002WO-US041225.
XX
PR   20-DEC-2001; 2001US-0342603P.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Jones KA,  Schafer A;
XX
DR   WPI; 2003-559141/52.
XX
PT   Determining susceptibility of an individual to joint space narrowing,
PT   osteophyte development and/or joint pain comprises identifying whether
PT   the individual has at least one polymorphism in a polynucleotide encoding
PT   a protein.
XX
PS   Disclosure; SEQ ID NO 37; 297pp; English.
XX
CC   The invention relates to a method of determining susceptibility of an
CC   individual to joint space narrowing and/or osteophyte development and/or
CC   joint pain comprising identifying whether the individual has at least one
CC   polymorphism in a polynucleotide encoding at least one of the protein
CC   listed in the specification. The methods, composition and agent are
CC   useful for modulating the susceptibility of an individual to joint space
CC   narrowing and/or osteophyte development and/or joint pain that is
CC   associated with a disease, preferably osteoarthritis. The cell line and
CC   the non-human animal are useful for screening for an agent for diagnosing
CC   an individual having susceptibility to joint space narrowing and/or
CC   osteophyte development and/or joint pain. This sequence corresponds to
CC   the polynucleotide encoding a protein listed in the specification. (Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification but was obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences).
XX
SQ   Sequence 76146 BP; 14629 A; 22827 C; 22905 G; 15785 T; 0 U; 0 Other;

  Query Match             5.6%;  Score 231;  DB 10;  Length 76146;
  Best Local Similarity   69.7%;  Pred. No. 4.8e-30;
  Matches  373;  Conservative    0;  Mismatches  150;  Indels   12;  Gaps    4;

Qy        209 AGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACATA 268
              | |||| | | | |        ||  | ||  ||  | ||| || |||||| ||||||||
Db      14496 ACAAAACACATATGTACACACATCCACATGCATCCACATACATGCACACATCCACACATA 14437

Qy        269 TATATATACACATATATATACACATATATACATATAT------ACACATATATACACATA 322
              ||  |  |||||  |||| |||||| || ||| ||||      |||||  || ||||  |
Db      14436 TACGTGCACACACCTATACACACATGTACACACATATACATGCACACACCTAGACACTCA 14377

Qy        323 TATATACACATACATATGTACACATATATATACACATATGTATACACATATATACACATA 382
              | || || |  ||| | |||||||  | ||||||||||| |||||||||| |||||||
Db      14376 CACATGCATACCCATGTATACACATTCACATACACATATG-ATACACATATACACACATA 14318
```

```
Qy        383 TATACACACATATATACACATATATACACACATATATACACATATATACACATATATACA 442
              |||||||| ||  | |||||   || ||||||| |||||||||| || ||||| || | 
Db      14317 TATACACAGATGCACACACACTCATGCACACATCTATACACATACATGCACATGTACATC 14258

Qy        443 CATATACACATATATACACATATATACATATATACACATATATATAATATACACACATAT 502
              || || | || ||  |  ||| |||||| | || |||| ||  |  | |||||||||| | 
Db      14257 CACATGCGCACATGCATGCATGTATACACACATGCACACATGCA-CACATACACACATGT 14199

Qy        503 ATATACACATATATACACACATATATACACATATATACACATATATATACACATATATAC 562
              | | ||||  | | |||| || | |||||||| |||||||| || | |||||     || 
Db      14198 ACACACACGCACACACACGCACACATACACATTTATACACACGTACACACACAC----AC 14143

Qy        563 ACATATATACATATATACACATATATATACATATATACACATATATACATATATACACAT 622
              ||||  ||||||  | ||| | ||| |||||||  | |||||||| || | ||||||| 
Db      14142 ACATGCATACATGCACACATACATACATACATACCTGCACATATACACCTGTATACACGC 14083

Qy        623 ATATACATATATACACACATATATACACATACATATACACACACATAGATATACATATAT 682
              ||| |||||| ||||||||||| || |||| |  ||||||||||   | | | ||||  | 
Db      14082 ATACACATATGTACACACATACATGCACACATCTATACACACATGCATACACACATTCAC 14023

Qy        683 ATACACATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAG 737
              ||| |||||| ||||| |||| | ||  ||| | ||||    |      |  ||||| 
Db      14022 ATATACATATGTATACATATACACATACATACACATATATCCACACATAGATAAG 13968


RESULT 98
ADJ95414
ID   ADJ95414 standard; DNA; 9800 BP.
XX
AC   ADJ95414;
XX
DT   06-MAY-2004  (first entry)
XX
DE   Human nootropic protein coding sequence, SEQ ID 3.
XX
KW   Nootropic; Gene therapy; human; mental illness; gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             2. .250
FT                   /*tag=  a
FT                   /partial     .
FT                   /note= "No start codon given"
FT                   /product= "Human nootropic protein #3"
XX
PN   JP2003310264-A.
XX
PD   05-NOV-2003.
XX
PF   19-APR-2002; 2002JP-00117143.
XX
PR   19-APR-2002; 2002JP-00117143.
XX
PA   (KAZU-) ZH KAZUSA DNA KENKYUSHO.
XX
DR   WPI; 2004-027143/03.
DR   P-PSDB; ADJ95425.
XX
PT   Novel DNA encoding recombinant protein from cDNA library derived from
PT   human adult whole brain and human fetus whole brain, useful for diagnosis
PT   of disease such as mental illness.
XX
PS   Claim 1; SEQ ID NO 3; 91pp; Japanese.
XX
CC   The present invention relates to human proteins (ADJ95423-ADJ95433) and
CC   their coding sequences (ADJ95412-ADJ95422). The sequences are useful for
CC   diagnosis and treatment of disease such as mental illness.
XX
SQ   Sequence 9800 BP; 2648 A; 2439 C; 2333 G; 2380 T; 0 U; 0 Other;

  Query Match             5.6%;  Score 229.4;  DB 12;  Length 9800;
  Best Local Similarity   68.5%;  Pred. No. 8.2e-30;

  Matches  379;  Conservative    0;  Mismatches  161;  Indels   13;  Gaps    4;

Qy        185 GCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATA 244
              ||   | ||  ||| || |   || |     |||||       || | ||| ||| | | |||
Db       5611 GCCTCCAGTTCGAGATATACCCTAAATTTAAAAATTGTATATATAATACATACACACATA 5670

Qy        245 TATATATACA-----CACACACACACATATATATATACACATATATATACACATATATAC 299
              |||||||  |     |||||| ||||| | ||| | | | | | | || || ||||| ||
Db       5671 CATATATAGACACACCACACATACACACACATACACATATACACACATGCATATATACAC 5730

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACAT 359
              ||||| ||||||  |||||||  | |||||||||| ||||   | | ||| | | ||
Db       5731 ATATACACACATGCATACACACACATATACACATGCATACACATATATACACACACCACA 5790

Qy        360 ATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATAT 419
                      ||||||| ||||||| ||| | ||| |||| ||    | |||||  ||||||
Db       5791 CATACACAACATATACACACATACATATATACACATATGCATGCACATACATGCATATAC 5850

Qy        420 ACACATATATACACATATATACACATATACACATATATACACATATATACATATATACAC 479
              |||||||||| |||| ||| |||||| || ||| | |||||||||||||||  |||||| |
Db       5851 ACACATATATGCACACATACACACATGTATACACACATACACATATATACTCATATACCC 5910

Qy        480 ATATATATA-ATATACACACATATATATACACATATATACACACATATATACACATATAT 538
                | | ||| |||| |||||| ||||| | ||||||   |||||| | ||| | | | |
Db       5911 CCACACATATATATCCACACACATATAAATACATATCCCCACACAGACATATATACACAC 5970

Qy        539 AC---ACATATATATACACATATATACACATATATACATATATACACATATATATACATA 595
              ||   ||| | | |||||    | ||| | ||||| |||| ||||| || ||| | ||| |
Db       5971 ACTATACACACACATACATGCAGATATATATATATACACATGTATACTCACATAAACACACA 6030

Qy        596 TATACACATATATACATATATACACATATATACATATATACACACATATATACACATACA 655
              |  |  ||||||| ||   ||| | ||| | |  |||||||||  |||| |||| ||||
Db       6031 TGCATGCATATATGCAGGCATATATATACACATGTATATACACAAGTATACACACGTACA 6090

Qy        656 TATACACA----CACATAGATATACATATATATACACATATATATACGTATATATATGTA 711
              ||||||||      | ||  ||||||| |  |||||||||| |||||  ||||||    |
Db       6091 TATACACATATTTAAATGTATATACACAGGTATACACATATGTATACACATATATGCACA 6150

Qy        712 TATATATATGCTC 724
              | ||| ||||| |
Db       6151 CACATGTATGCAC 6163


RESULT 99
ABD32843/c
ID   ABD32843 standard; DNA; 326002 BP.
XX
AC   ABD32843;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Human cancer-associated genomic DNA HD17-014.
XX
KW   Human; ds; cancer-associated protein; gene; cytostatic; cancer;
KW   leukaemia; lymphoma; CAP.
XX
OS   Homo sapiens.
XX
PN   WO2004074320-A2.
XX
PD   02-SEP-2004.
XX
PF   17-FEB-2004; 2004WO-US004730.
XX
PR   14-FEB-2003; 2003US-00367094.
PR   14-MAR-2003; 2003US-00388838.
PR   15-APR-2003; 2003US-00417375.
PR   13-JUN-2003; 2003US-00461862.
PR   15-SEP-2003; 2003US-00663431.
PR   15-DEC-2003; 2003US-00737318.
XX
PA   (SAGR-) SAGRES DISCOVERY INC.
XX
```

```
PI   Morris DW,  Morris DW,  Malandro MS;
XX
DR   WPI; 2004-652914/63.
XX
PT   New isolated cancer-associated polynucleotides and polypeptides useful
PT   for diagnosing, preventing or treating cancers, especially lymphoma and
PT   leukemia, or in screening for agents that modulate cancer.
XX
PS   claim 16; seqid 506; 310pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising at least 10
CC   contiguous nucleotides of any of the 233 polynucleotide sequences given
CC   in the specification, or its complement. The nucleic acids encode cancer-
CC   associated proteins. Also included are an expression vector comprising
CC   the isolated nucleic acid cited above, a host cell comprising the above
CC   recombinant nucleic acid or expression vector, a microarray for detecting
CC   a cancer-associated (CA) nucleic acid comprising at least one probe
CC   comprising at least 10 contiguous nucleotides of any of the above-
CC   mentioned nucleotide sequences, an isolated polypeptide (encoded within
CC   an open reading frame of a CA sequence selected from any of the 95
CC   polynucleotide sequences as mentioned in the specification, or its
CC   complement), an isolated antibody, (or its antigen binding fragment) that
CC   binds to the above polypeptide, a hybridoma that produces the above
CC   monoclonal antibody, a pharmaceutical composition comprising the above
CC   antibody and a pharmaceutical excipient, a kit for detecting cancer
CC   cells(comprising the antibody cited above, methods for diagnosing cancer
CC   or for detecting the presence or absence of cancer cells in an
CC   individual, a method for inhibiting growth of cancer cells in an
CC   individual, a method for delivering a therapeutic agent to cancer cells
CC   in an individual, an electronic library comprising the above
CC   polynucleotide or polypeptide (or their fragments), methods of screening
CC   for anticancer activity or for a bioactive agent capable of modulating
CC   the activity of a CA protein (CAP), methods for detecting cancer
CC   associated with expression of a polypeptide in a test cell sample, a
CC   method for treating cancers and a method for inhibiting the expression of
CC   CA gene in a cell. The composition and methods are useful for detecting,
CC   diagnosing, preventing and treating cancers, especially lymphoma and
CC   leukaemia. These may also be used in screening for agents that modulate
CC   cancer. The present sequence is a human CAP genomic sequence. Note: The
CC   sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 326002 BP; 84319 A; 72970 C; 76829 G; 91884 T; 0 U; 0 Other;

  Query Match             5.6%;  Score 229;  DB 13;  Length 326002;
  Best Local Similarity   70.1%;  Pred. No. 1.1e-29;
  Matches  323;  Conservative    0;  Mismatches  135;  Indels    3;  Gaps    1;

Qy        235 TATATATATATATATATACACACACACACATATATATATACACATATATATACACATA 294
              ||| | |||| | | | ||||||||||||||||| | | | | || | || | ||||||| |
Db      46727 TATGTGTATAGACACACACACACACACACACACACACACACGCGTACACATACACACA 46668

Qy        295 TATACATATATACACATATATACACATATATATACACATACATATGTACACATATATATA 354
              || ||  | | | ||| | ||||||| | |||  || | ||| | | | ||| ||| |
Db      46667 TACACGCACACATACACACATACACACATACGCATACACAGACGCATACACATACACG 46608

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              |||| ||  | ||||| | |||||||   | |||||||||| || |||| ||| | |||||
Db      46607 CACACATACACACACACACATACACACGCACACACACATACATGCACACATACATACACA 46548

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              |||  || | |||  || ||| | ||||||| |||| |||| ||||| ||| ||  | |
Db      46547 CATACGCATACATACGCATATACACACACATACACACGCATACACACACATACACGCACA 46488

Qy        475 TACACATATATATAATATACACACATATATATACACATATATACACACATATATACACAT 534
              | ||| | ||  ||| ||||| | ||| | ||||||||| ||  ||| | |||||||
Db      46487 CATACACACAT---ATACACACAGAGATACACACACATATACACGGACACACATACACAC 46431

Qy        535 ATATACACATATATATACACATATATACACATATATACATATATACACATATATATACAT 594
              |   | ||| ||  ||||||| ||| |  |||||| ||| ||| | |||| ||| ||||
Db      46430 ACTCATACACATGCATACACACATACATGCATATACACACATACATACATGCATACACAT 46371

Qy        595 ATATACACATATATACATATATACACATATATACATATATACACACATATATACACATAC 654
              | ||||||| ||| |||||  |  ||| |   ||| ||| ||||||||| | || |||||
Db      46370 ACATACACACATACACATACCTGGACACACGCACACATACACACACATACACACGCATAC 46311

Qy        655 ATATACACACACATAGATATACATATATATACACATATATA 695
              | |||| ||||||  |   ||||| | | | | || | | ||
Db      46310 ACATACGCACACACATGCATACACACACACAGACCTCTCTA 46270


RESULT 100
ADC87570
ID   ADC87570 standard; DNA; 1672 BP.
XX
AC   ADC87570;
XX
DT   01-JAN-2004  (first entry)
XX
DE   Human GPCR gene SEQ ID NO:2023.
XX
KW   ds; gene; human; GPCR;
KW   guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1270724-A2.
XX
PD   02-JAN-2003.
XX
PF   18-JUN-2002; 2002EP-00013517.
XX
PR   18-JUN-2001; 2001JP-00246789.
XX
PA   (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA   (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI   Suwa M,  Asai K,  Akiyama Y,  Aburatani H;
XX
DR   WPI; 2003-315783/31.
DR   P-PSDB; ADC87571.
XX
PT   New polynucleotide, useful for preparing a composition for treating a
PT   patient in need of increased or suppressed activity or expression of the
PT   guanosine triphosphate-binding protein coupled receptor.
XX
PS   Claim 1; SEQ ID NO 2023; 28pp; English.
XX
CC   The invention relates to a novel polynucleotide encoding a guanosine
CC   triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC   the invention may have a use in gene therapy. The polynucleotide and
CC   polypeptide are useful for preparing a composition for treating a patient
CC   in need of increased or suppressed activity or expression of the
CC   guanosine triphosphate-binding protein coupled receptor. The
CC   polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC   invention.
XX
SQ   Sequence 1672 BP; 499 A; 171 C; 300 G; 602 T; 0 U; 100 Other;

  Query Match             5.6%;  Score 228.8;  DB 10;  Length 1672;
  Best Local Similarity   65.6%;  Pred. No. 9.8e-30;
  Matches  349;  Conservative    0;  Mismatches  182;  Indels    1;  Gaps    1;

Qy        204 ATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACAC 263
              || |  | | | | | |       | | |||||| | |||| | | |||| ||||| | |
Db        832 ATGTGTGTATATACATATATATGTGTGTATATACACATATGTGTGTATATACACATATAT 891

Qy        264 ACATATATATATACACATATATATACACATATATACATATATACACATATATACACATAT 323
              |||||||| | || | | | ||||| ||||||||    ||||||||||||||||||  | |
Db        892 ACATATATGTGTATATACACATATACACATATATGTGTATATACACATATATACATGTGT 951

Qy        324 ATATACACATACATATGTACACATATATATACACATATGTATACACATATATACACATAT 383
              | || | | |||||| |      | ||||||||||||| | |     ||||||||||||
Db        952 GTGTATATACACATATATGTGTGTGTATATACACATATATGTGTGTGTATATACACATAT 1011
```

```
Qy        384 ATACACACATATATACACATATATACACACATATATACACATATATACACATATATACAC 443
              ||       ||||||||||||||||       ||||||||||||||||     |||||||||
Db       1012 ATGTGTGTGTATATACACATATATGTGTGTGTATATACACATATATGTGTGTATATACAC 1071

Qy        444 ATATACACATATATACACATATATACATATATACACATATATATAATATACACACATATA 503
              |||||    | | || | | | ||| ||| ||| |  ||| | |  |||| | ||| |||
Db       1072 ATATATGTGTGTGTATATACACATATATACATATATCTATGTGT-GTATATATACACATA 1130

Qy        504 TATACACATATATACACACATATATACACATATATACACATATATATACACATATATACA 563
              ||||||  | | |      ||||||| ||||||||    ||||||| ||| | |||||
Db       1131 TATACATGTCTGTGTGTGTATATATATACATATATGTGTATATATACACATACATATATG 1190

Qy        564 CATATATACATATATACACATATATATACATATATACACATATATACATATATACACATA 623
                | | || |||| ||||||||||  || ||| ||      ||||||  | || | | |||
Db.      1191 TGTGTGTATATATGTACACATATACGTATATACATGTGTGTATATATGTGTACATATATA 1250

Qy        624 TATACATATATACACACATATATACACATACATATACACACACATAGATATACATATATA 683
              | |  |||||||     |||||||    || ||||    | | |||  | |  ||||||
Db       1251 TGTGTATATATATGTGTATATATATGTGTATATATGTGTATATATATGTGTGTATATATG 1310

Qy        684 TACACATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATA 735
              |  | |||| | |  || |||||||| | | ||||||||      | ||  |||
Db       1311 TGTATATATGTGTGTGTGTATATATATGTGTATATATGTGTGTGTGTGTATA 1362
```

Search completed: April 21, 2005, 00:47:34
Job time : 1374.5 secs

This Page Blank (uspto)

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 20, 2005, 22:52:45 ; Search time 2651.72 Seconds
(without alignments)
13936.065 Million cell updates/sec

Title: US-09-939-209A-3_COPY_4970_11056
Perfect score: 6087
Sequence: 1 ccagtaggaaacagttcttc.....catcatcagaaaggcactat 6087

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5633728 seqs, 3035525691 residues

Total number of hits satisfying chosen parameters: 11267456

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq:*
19: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
20: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:*
21: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
22: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	6087	100.0	20300	10	US-09-939-209A-3	Sequence 3, Appli
	2	325.8	5.4	1980090	18	US-10-719-993-6815	Sequence 6815, Ap
	3	325.8	5.4	1980090	19	US-10-741-600-17676	Sequence 17676, A
c	4	277.6	4.6	2752	14	US-10-198-846-13843	Sequence 13843, A
	5	276	4.5	1531	14	US-10-198-846-9863	Sequence 9863, Ap
	6	274.4	4.5	3165	13	US-10-044-090-113	Sequence 113, App
	7	274.4	4.5	3167	10	US-09-918-624B-6	Sequence 6, Appli
	8	253.6	4.2	1151	13	US-10-027-632-266349	Sequence 266349,
	9	253.6	4.2	1151	17	US-10-027-632-266349	Sequence 266349,
	10	253.2	4.2	1151	13	US-10-027-632-266350	Sequence 266350,
	11	253.2	4.2	1151	17	US-10-027-632-266350	Sequence 266350,
	12	233.8	3.8	576	13	US-10-027-632-288758	Sequence 288758,
	13	233.8	3.8	576	17	US-10-027-632-288758	Sequence 288758,
	14	233.4	3.8	576	13	US-10-027-632-288755	Sequence 288755,
	15	233.4	3.8	576	13	US-10-027-632-288756	Sequence 288756,
	16	233.4	3.8	576	13	US-10-027-632-288757	Sequence 288757,
	17	233.4	3.8	576	17	US-10-027-632-288755	Sequence 288755,
	18	233.4	3.8	576	17	US-10-027-632-288756	Sequence 288756,
	19	233.4	3.8	576	17	US-10-027-632-288757	Sequence 288757,
c	20	213	3.5	547	13	US-10-027-632-233469	Sequence 233469,
c	21	213	3.5	547	13	US-10-027-632-233470	Sequence 233470,
c	22	213	3.5	547	13	US-10-027-632-233471	Sequence 233471,
c	23	213	3.5	547	17	US-10-027-632-233469	Sequence 233469,
c	24	213	3.5	547	17	US-10-027-632-233470	Sequence 233470,
c	25	213	3.5	547	17	US-10-027-632-233471	Sequence 233471,
	26	208.8	3.4	507	13	US-10-027-632-181211	Sequence 181211,
	27	208.8	3.4	507	13	US-10-027-632-181212	Sequence 181212,
	28	208.8	3.4	507	17	US-10-027-632-181211	Sequence 181211,
	29	208.8	3.4	507	17	US-10-027-632-181212	Sequence 181212,
	30	202.8	3.3	3673778	16	US-10-312-841-1	Sequence 1, Appli
	31	201.6	3.3	97952	18	US-10-741-601-5646	Sequence 5646, Ap
c	32	193	3.2	669	13	US-10-027-632-87039	Sequence 87039, A
c	33	193	3.2	669	17	US-10-027-632-87039	Sequence 87039, A
	34	192.4	3.2	25041	18	US-10-719-993-6836	Sequence 6836, Ap
c	35	192.4	3.2	141158	18	US-10-719-993-6837	Sequence 6837, Ap
c	36	188.6	3.1	39428	18	US-10-719-993-7069	Sequence 7069, Ap
	37	183.6	3.0	650	13	US-10-027-632-226077	Sequence 226077,
	38	183.6	3.0	650	17	US-10-027-632-226077	Sequence 226077,
c	39	178.4	2.9	520	13	US-10-027-632-320728	Sequence 320728,
c	40	178.4	2.9	520	17	US-10-027-632-320728	Sequence 320728,
c	41	175.6	2.9	193357	17	US-10-085-117-142	Sequence 142, App
c	42	173.8	2.9	693	10	US-09-854-867-156	Sequence 156, App
c	43	173.8	2.9	693	19	US-10-786-970A-156	Sequence 156, App
	44	170.2	2.8	1503841	9	US-09-795-668-1	Sequence 1, Appli
	45	170.2	2.8	1503841	9	US-09-795-686-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-939-209A-3
; Sequence 3, Application US/09939209A
; Publication No. US20030113721A1
; GENERAL INFORMATION:
; APPLICANT: LEVITT, PAT R.
; APPLICANT: MIRNICS, KAROLY
; APPLICANT: KODAVALI, MENKATA CHOWDARI
; APPLICANT: NIMGAONKAR, VISHWAJIT L.
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR FACILITATING THE DIAGNOSIS AND TREATMENT
; TITLE OF INVENTION: SCHIZOPHRENIA
; FILE REFERENCE: 00-539-US
; CURRENT APPLICATION NUMBER: US/09/939,209A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: 60/228,021
; PRIOR FILING DATE: 2000-08-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 20300
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: A genomic sequence containing RGS4 nucleic acid sequence and sequ
; OTHER INFORMATION: ences upstream and downstream to the RGS4 nucleic acid sequence
US-09-939-209A-3

Query Match 100.0%; Score 6087; DB 10; Length 20300;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 6087; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCAGTAGGAAACAGTTCTTCTAAAGCCTATTGTCAACAATACCTTCCAGATGTTAGCATT 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

```
Db       4970 CCAGTAGGAAACAGTTCTTCTAAAGCCTATTGTCAACAATACCTTCCAGATGTTAGCATT 5029
Qy         61 TTACAATTTAAGGAACTTAAAATAGCCTTCAAACTTTTTGCCAGTTTCTCTGATATCCAA 120
Db       5030 TTACAATTTAAGGAACTTAAAATAGCCTTCAAACTTTTTGCCAGTTTCTCTGATATCCAA 5089
Qy        121 TCTATTCTTTTACTCTGCCTCCCAAGCTTTCTTTCTAGAATGCTAACCTGATCGGCTTAA 180
Db       5090 TCTATTCTTTTACTCTGCCTCCCAAGCTTTCTTTCTAGAATGCTAACCTGATCGGCTTAA 5149
Qy        181 GTACTTGAACTACCTCTTCTCCTCCATTAACTACAGAGTAAATTCTGGTCTTCAGAGTAA 240
Db       5150 GTACTTGAACTACCTCTTCTCCTCCATTAACTACAGAGTAAATTCTGGTCTTCAGAGTAA 5209
Qy        241 CAAGAAACACCCTTTAGTTCTCAGCATATTCGTGCACCTTCATTTATCTCTCCTTCTCTC 300
Db       5210 CAAGAAACACCCTTTAGTTCTCAGCATATTCGTGCACCTTCATTTATCTCTCCTTCTCTC 5269
Qy        301 TCAAAGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGC 360
Db       5270 TCAAAGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGC 5329
Qy        361 AACCAAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTA 420
Db       5330 AACCAAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTA 5389
Qy        421 ATCAAAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAA 480
Db       5390 ATCAAAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAA 5449
Qy        481 CTGTCTGTTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTA 540
Db       5450 CTGTCTGTTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTA 5509
Qy        541 GACAAAAGGATATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTA 600
Db       5510 GACAAAAGGATATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTA 5569
Qy        601 TTCAGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGG 660
Db       5570 TTCAGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGG 5629
Qy        661 GCAGGACTCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGAT 720
Db       5630 GCAGGACTCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGAT 5689
Qy        721 TTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGG 780
Db       5690 TTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGG 5749
Qy        781 AATTCTGGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGG 840
Db       5750 AATTCTGGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGG 5809
Qy        841 AGATGGTCAGAAAGAGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAGT 900
Db       5810 AGATGGTCAGAAAGAGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAGT 5869
Qy        901 ACTTCTTAGCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAACACTGC 960
Db       5870 ACTTCTTAGCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAACACTGC 5929
Qy        961 AATCATGCTAAACTCCTCTATGACCTTCAAACATTCCACTTGCTTTTATTCTTTATGGTT 1020
Db       5930 AATCATGCTAAACTCCTCTATGACCTTCAAACATTCCACTTGCTTTTATTCTTTATGGTT 5989
Qy       1021 GTGATGGCATAGAGGTCAATAGCAAAGACCCTGGAGTCCCACTGTCTGAGCTGGCATAAC 1080
Db       5990 GTGATGGCATAGAGGTCAATAGCAAAGACCCTGGAGTCCCACTGTCTGAGCTGGCATAAC 6049
Qy       1081 ATTACTACCACTTAATCAATGTGTAAGCTCAGGTAAGTACTTAAGTCCTCTATGCTTCAT 1140
Db       6050 ATTACTACCACTTAATCAATGTGTAAGCTCAGGTAAGTACTTAAGTCCTCTATGCTTCAT 6109
```

```
Qy       1141 CTGTAAAATGAGAATCATTGAAGAACATTCTCTCAGGATGGATCATGAGGAATAAGTGAA 1200
Db       6110 CTGTAAAATGAGAATCATTGAAGAACATTCTCTCAGGATGGATCATGAGGAATAAGTGAA 6169
Qy       1201 TTAACTGGCATATAGTGCTTAAACCAGTGCCTTGCTCAGTTAGTGACAGATAAAATCATC 1260
Db       6170 TTAACTGGCATATAGTGCTTAAACCAGTGCCTTGCTCAGTTAGTGACAGATAAAATCATC 6229
Qy       1261 TGTTATTACTGTGCCCACTATTGTGATGCTCTTCTCTTCTTTGTACAACGACTACATCTC 1320
Db       6230 TGTTATTACTGTGCCCACTATTGTGATGCTCTTCTCTTCTTTGTACAACGACTACATCTC 6289
Qy       1321 TATTTATCATTTTAGGGTCTCCTTGTGAAAAACCACTCCAGATTCAAAAGATTGAGTTTA 1380
Db       6290 TATTTATCATTTTAGGGTCTCCTTGTGAAAAACCACTCCAGATTCAAAAGATTGAGTTTA 6349
Qy       1381 ATCTCTATCCTCTGTGCTTTCCTGGAGTTTTGTAAAGTAAATCTTCACTTGACATCATGG 1440
Db       6350 ATCTCTATCCTCTGTGCTTTCCTGGAGTTTTGTAAAGTAAATCTTCACTTGACATCATGG 6409
Qy       1441 ATAGGTTCTTGGAAACTACAACTTCAAGTGAAAGGACATAACTAAACCAATTTTTTTCTC 1500
Db       6410 ATAGGTTCTTGGAAACTACAACTTCAAGTGAAAGGACATAACTAAACCAATTTTTTTCTC 6469
Qy       1501 ATCAACGTTATAATGAAATGGCATTGATGAAATGATGGCATTCAAGGACCTGCTGTACCT 1560
Db       6470 ATCAACGTTATAATGAAATGGCATTGATGAAATGATGGCATTCAAGGACCTGCTGTACCT 6529
Qy       1561 TGTTTCACTTAAAGTCACTGTTTCCAATAATCTATTGATGACATTGAGGACTTACTATAT 1620
Db       6530 TGTTTCACTTAAAGTCACTGTTTCCAATAATCTATTGATGACATTGAGGACTTACTATAT 6589
Qy       1621 AATAATAAATATATATATAATCGACGAAACAGGAATCAAACTGCTAACTCTGCTAACTGG 1680
Db       6590 AATAATAAATATATATATAATCGACGAAACAGGAATCAAACTGCTAACTCTGCTAACTGG 6649
Qy       1681 TCTCCCTGCTTCCACACTCTGCCCACTCATCTCAGTCTTTCTTTCACAAGAGTCAGAATG 1740
Db       6650 TCTCCCTGCTTCCACACTCTGCCCACTCATCTCAGTCTTTCTTTCACAAGAGTCAGAATG 6709
Qy       1741 ATCAGATGAGACCCCTCCTCTGCTTCTGTTTCTTCCATGGATTTCCACTGCACTCTGATA 1800
Db       6710 ATCAGATGAGACCCCTCCTCTGCTTCTGTTTCTTCCATGGATTTCCACTGCACTCTGATA 6769
Qy       1801 AAGTCCAGCCTCTTGACCACAGCCTACAAATCCTTGCACGATCTATCGTTTACTTTTCCA 1860
Db       6770 AAGTCCAGCCTCTTGACCACAGCCTACAAATCCTTGCACGATCTATCGTTTACTTTTCCA 6829
Qy       1861 TCTCCTTTTATGCTACTTTCATCTTGTTCTCAATTCTCTAGCTATGCTGGCCCCTTCTTG 1920
Db       6830 TCTCCTTTTATGCTACTTTCATCTTGTTCTCAATTCTCTAGCTATGCTGGCCCCTTCTTG 6889
Qy       1921 TTCTTTCCCATTTTTTTTTAATTTTTAAAATTTGTATATATTTATGGGTTATAAGTGAAA 1980
Db       6890 TTCTTTCCCATTTTTTTTTAATTTTTAAAATTTGTATATATTTATGGGTTATAAGTGAAA 6949
Qy       1981 TCTTTTTAGATGCATAGGTTGTATAGTGATAAAATCAGGGCTTTTAGGGTATTCATCACC 2040
Db       6950 TCTTTTTAGATGCATAGGTTGTATAGTGATAAAATCAGGGCTTTTAGGGTATTCATCACC 7009
Qy       2041 TGAATGATGTACATTGTACCCCTTAAGTAATTTCTCACCATCCGCTGACTTCTTGCCCCC 2100
Db       7010 TGAATGATGTACATTGTACCCCTTAAGTAATTTCTCACCATCCGCTGACTTCTTGCCCCC 7069
Qy       2101 TGGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGC 2160
Db       7070 TGGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGC 7129
Qy       2161 TGACTTCTTGCCCCCTGGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAA 2220
Db       7130 TGACTTCTTGCCCCCTGGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAA 7189
```

```
Qy       2221 TTTCTCACCATCCGCTGACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCAT 2280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7190 TTTCTCACCATCCGCTGACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCAT 7249

Qy       2281 TCCACACTCTACATCTATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGC 2340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7250 TCCACACTCTACATCTATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGC 7309

Qy       2341 AATATTTGTCTTTCTGTGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCT 2400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7310 AATATTTGTCTTTCTGTGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCT 7369

Qy       2401 AGGCTGCTGCAAAAGGCATGATTTCATTCTTTTTTATGGCTATGTTCTTTCCCAATTTAG 2460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7370 AGGCTGCTGCAAAAGGCATGATTTCATTCTTTTTTATGGCTATGTTCTTTCCCAATTTAG 7429

Qy       2461 ATAAAGAACACTCGCACTTGCTCTTACTTCTATTTGGAATACTAATTCCTAGGCTTCTTG 2520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7430 ATAAAGAACACTCGCACTTGCTCTTACTTCTATTTGGAATACTAATTCCTAGGCTTCTTG 7489

Qy       2521 CATTGCTTTCTCCTTCTCACCCATCAAATCTCATTTTAGATACCACCTCTTCAAAGAGGG 2580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7490 CATTGCTTTCTCCTTCTCACCCATCAAATCTCATTTTAGATACCACCTCTTCAAAGAGGG 7549

Qy       2581 CTTTCCTGACCACCTTGGCTGAATTAGCCCTTCACCATCTGATTACTCTCTAGCACATCA 2640
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7550 CTTTCCTGACCACCTTGGCTGAATTAGCCCTTCACCATCTGATTACTCTCTAGCACATCA 7609

Qy       2641 CCTGCCCATTTTATTCATGGTACAGGTCAAAATCTGGAATCACCTGATTTGTTTATTTTC 2700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7610 CCTGCCCATTTTATTCATGGTACAGGTCAAAATCTGGAATCACCTGATTTGTTTATTTTC 7669

Qy       2701 TGACTCCTTCTACTGAGATGAAAACTCTACTAGAGCGGAGATTTTATCTGCTTGTATCAG 2760
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7670 TGACTCCTTCTACTGAGATGAAAACTCTACTAGAGCGGAGATTTTATCTGCTTGTATCAG 7729

Qy       2761 GTACTGCTTCAAACAGCACCTGATACAGAGTAGGTGGTCAAAAGATATTTCTTAAACAAA 2820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7730 GTACTGCTTCAAACAGCACCTGATACAGAGTAGGTGGTCAAAAGATATTTCTTAAACAAA 7789

Qy       2821 TGAACAAATAAAAAGTAGATCTTTTGAGAGTAAAGCTCTTCCACACTACCAGAGTCATTC 2880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7790 TGAACAAATAAAAAGTAGATCTTTTGAGAGTAAAGCTCTTCCACACTACCAGAGTCATTC 7849

Qy       2881 AGGAATGACAAATCATAGAATAACAGAATTTGATGCTTTGTGCATATCAGAGAAAGAAGG 2940
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7850 AGGAATGACAAATCATAGAATAACAGAATTTGATGCTTTGTGCATATCAGAGAAAGAAGG 7909

Qy       2941 TGGAAGGTTGTCAAGGTATCATGATGTACCAGTCCTCGCCTCCTCAAACACAATCTGCAA 3000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7910 TGGAAGGTTGTCAAGGTATCATGATGTACCAGTCCTCGCCTCCTCAAACACAATCTGCAA 7969

Qy       3001 GTCCCACAGTGAAAAAGTAAGTTAACTCATGTGAAGCGTTTTACAAACACTTTTTTAAAA 3060
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7970 GTCCCACAGTGAAAAAGTAAGTTAACTCATGTGAAGCGTTTTACAAACACTTTTTTAAAA 8029

Qy       3061 GTCTTAAAACTCCTAAGAAAGCAAGATTTAATAGTCAAAGAAGTGAGTAAACATGAAATG 3120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8030 GTCTTAAAACTCCTAAGAAAGCAAGATTTAATAGTCAAAGAAGTGAGTAAACATGAAATG 8089

Qy       3121 CCTGAACAGAGTAATGAGCTAAGCACAAAGTTAGAGACATGTTAGTTAATATGTCTTGAA 3180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8090 CCTGAACAGAGTAATGAGCTAAGCACAAAGTTAGAGACATGTTAGTTAATATGTCTTGAA 8149

Qy       3181 AGCAGCAGCTCCTGCTTTCAAGGAGCAAGAACAAATTGGGCAAGTGAACACTCCTTGAAT 3240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8150 AGCAGCAGCTCCTGCTTTCAAGGAGCAAGAACAAATTGGGCAAGTGAACACTCCTTGAAT 8209

Qy       3241 AAAATGTGTAAAATTAATTTTGGGTTATGTTCTATACTGTGTATAATAGAATGATAAAAA 3300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8210 AAAATGTGTAAAATTAATTTTGGGTTATGTTCTATACTGTGTATAATAGAATGATAAAAA 8269

Qy       3301 TTATTTGACTAGCACTTTGTAGTTTAGAAATATCTCTATTTACACAGTTTACCTTATTTG 3360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8270 TTATTTGACTAGCACTTTGTAGTTTAGAAATATCTCTATTTACACAGTTTACCTTATTTG 8329

Qy       3361 ATAAGACTGTTGAGTGATGGGATAGCATGGTGGACAATCCACATAACTGAGTATCGAGAC 3420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8330 ATAAGACTGTTGAGTGATGGGATAGCATGGTGGACAATCCACATAACTGAGTATCGAGAC 8389

Qy       3421 ACCTGTATCTGGACCCAGCTCTGTTAGTAAGAAGCTGTAACCTCAGCAAGTCACTTTCTC 3480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8390 ACCTGTATCTGGACCCAGCTCTGTTAGTAAGAAGCTGTAACCTCAGCAAGTCACTTTCTC 8449

Qy       3481 TTTCTGGGTCTCTATTTCCTTTTTGGTGAAATGAGAGTGTTAGGCTAGATTGCCTTTGAA 3540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8450 TTTCTGGGTCTCTATTTCCTTTTTGGTGAAATGAGAGTGTTAGGCTAGATTGCCTTTGAA 8509

Qy       3541 GTCCCATTTTGTCTTTAAAGTCCCATCTATTGCAGTGATTTATATTTAACTCATGACAAA 3600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8510 GTCCCATTTTGTCTTTAAAGTCCCATCTATTGCAGTGATTTATATTTAACTCATGACAAA 8569

Qy       3601 TCAGGCTTCTCTTATTCTAAGTGCAAGACATAAAACTTTTATTGTGGAATTTCAGGCATC 3660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8570 TCAGGCTTCTCTTATTCTAAGTGCAAGACATAAAACTTTTATTGTGGAATTTCAGGCATC 8629

Qy       3661 AGTAAATCTTTTTGGGTACTCACTTATGTTCCTGAAATCAATCTATTTGAGTGATCACTC 3720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8630 AGTAAATCTTTTTGGGTACTCACTTATGTTCCTGAAATCAATCTATTTGAGTGATCACTC 8689

Qy       3721 TTTTAGGTGCCCAGGTAAACAAAGAAGGCCATGGTCTTTCTTTGAGTGACCTTCTTTCCC 3780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8690 TTTTAGGTGCCCAGGTAAACAAAGAAGGCCATGGTCTTTCTTTGAGTGACCTTCTTTCCC 8749

Qy       3781 TTTTAATTAGTCTGACCTCTTTAATGTCAGTTCTGACTGATTCATTTCCCTGGTCCATCT 3840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8750 TTTTAATTAGTCTGACCTCTTTAATGTCAGTTCTGACTGATTCATTTCCCTGGTCCATCT 8809

Qy       3841 TCCTTGGTCTGAGGGCCTTCCTAGTTTCATATTGCACTTCAGTTCCTTCCACACCACCAT 3900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8810 TCCTTGGTCTGAGGGCCTTCCTAGTTTCATATTGCACTTCAGTTCCTTCCACACCACCAT 8869

Qy       3901 CAAGGATGGCTGTCAACATTCATTTGTTCTATGTTATAATTCAAGGAAAAGTTGCCCAGT 3960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8870 CAAGGATGGCTGTCAACATTCATTTGTTCTATGTTATAATTCAAGGAAAAGTTGCCCAGT 8929

Qy       3961 AGCTAATCCAATAAATGCCCTCTTATGGGCGGCTAGAGACTTTTTCCTATAATTTAAATG 4020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8930 AGCTAATCCAATAAATGCCCTCTTATGGGCGGCTAGAGACTTTTTCCTATAATTTAAATG 8989

Qy       4021 CATCTTCTGTAGATTATGGTCCCTCCACCACTTTACATTTGTCTGCTGTCTCCTTGCTCT 4080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8990 CATCTTCTGTAGATTATGGTCCCTCCACCACTTTACATTTGTCTGCTGTCTCCTTGCTCT 9049

Qy       4081 GCTAGTCATGGAACGTGTTGGTAGTGGGGGCAGTGTGGGATGTTCAAGGGCACGTATTGG 4140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9050 GCTAGTCATGGAACGTGTTGGTAGTGGGGGCAGTGTGGGATGTTCAAGGGCACGTATTGG 9109

Qy       4141 GTAGGGCCACATATGGGCATTGCTTTGTGCCATTCTTTCTATATTTTTGGTATTTTGCAT 4200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9110 GTAGGGCCACATATGGGCATTGCTTTGTGCCATTCTTTCTATATTTTTGGTATTTTGCAT 9169

Qy       4201 CTCACTGGAACCCAACTATTTTTCATCTCTTCCACCTAAACTATTTGATGCCTCTGTTTC 4260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9170 CTCACTGGAACCCAACTATTTTTCATCTCTTCCACCTAAACTATTTGATGCCTCTGTTTC 9229

Qy       4261 TTATATATAAAGTATAGCTCACTGTAGCCTATGATCAGGAACCTATCTGCTTTCTAAATG 4320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9230 TTATATATAAAGTATAGCTCACTGTAGCCTATGATCAGGAACCTATCTGCTTTCTAAATG 9289

Qy       4321 AAAGCTGTTTTGGTCAGATCTAGCAATTAATTCCCTTCTTCCACTTATAGCTTTCCTCTG 4380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9290 AAAGCTGTTTTGGTCAGATCTAGCAATTAATTCCCTTCTTCCACTTATAGCTTTCCTCTG 9349

Qy       4381 TAACTCTGGTGTAGGTATTTGGTTTATGGCTATAAGATGTGAAACACCTGAATGATTCTG 4440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       9350 TAACTCTGGTGTAGGTATTTGGTTTATGGCTATAAGATGTGAAACACCTGAATGATTCTG 9409

Qy       4441 TCCATGCAGGCATTTCAGTTCATGATATTGTATGTAAAAGATACTGATTGTCTAGGTGTT 4500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9410 TCCATGCAGGCATTTCAGTTCATGATATTGTATGTAAAAGATACTGATTGTCTAGGTGTT 9469

Qy       4501 CAGAAACACCTATAGGGCTTAATATTCTTACAATCAGTTTGAAGGCTGGTGATACGCAAA 4560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9470 CAGAAACACCTATAGGGCTTAATATTCTTACAATCAGTTTGAAGGCTGGTGATACGCAAA 9529

Qy       4561 GCAAACTACATATTTTTCTGCCTGCTCTCTCTCTTTCTCTCTACATCTCTCTTTCTTTAT 4620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9530 GCAAACTACATATTTTTCTGCCTGCTCTCTCTCTTTCTCTCTACATCTCTCTTTCTTTAT 9589

Qy       4621 CTTTTGAAATATCAGTTTGGAGACTTAGAATTACATAAGACATAAACCCATTTGATATAA 4680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9590 CTTTTGAAATATCAGTTTGGAGACTTAGAATTACATAAGACATAAACCCATTTGATATAA 9649

Qy       4681 GAATTGCTGTGTATATTTGCTCATCTACTCCCTCCTTTGGTCCTCGAGCTGCCGGTTTAG 4740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9650 GAATTGCTGTGTATATTTGCTCATCTACTCCCTCCTTTGGTCCTCGAGCTGCCGGTTTAG 9709

Qy       4741 ACTTTTTACAGGACGCAGGCATGTGAAGGAGAAACTGTCAGTGCTAGGCTGAATTCTGTT 4800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9710 ACTTTTTACAGGACGCAGGCATGTGAAGGAGAAACTGTCAGTGCTAGGCTGAATTCTGTT 9769

Qy       4801 GTTACCAAGATTTCTAGAAAAGTATTCCTCAGTCAGGTTGATTACAGATATAGCAAATCT 4860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9770 GTTACCAAGATTTCTAGAAAAGTATTCCTCAGTCAGGTTGATTACAGATATAGCAAATCT 9829

Qy       4861 ATTTTTCCTAGGGTAGTTTCTGTATGCTGCCGGGCTTATAACTGTCTGTCATCCAGCTAT 4920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9830 ATTTTTCCTAGGGTAGTTTCTGTATGCTGCCGGGCTTATAACTGTCTGTCATCCAGCTAT 9889

Qy       4921 TTCTCTCCACCTTCTTGTTTGCATAACAACCAAGGCAACTTCCGCAAATCACTGCGTGGA 4980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9890 TTCTCTCCACCTTCTTGTTTGCATAACAACCAAGGCAACTTCCGCAAATCACTGCGTGGA 9949

Qy       4981 GACGATGATCCTGCCAGCTCCCTTTTGGAAATCGTGAGGATCAGATCTTGGACCATGTAT 5040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9950 GACGATGATCCTGCCAGCTCCCTTTTGGAAATCGTGAGGATCAGATCTTGGACCATGTAT 10009

Qy       5041 AATATGATGCTTCTAATCCAAAAGAGGAAAGGCATTGGGAGTCAGCTCCTAAGTAAGCTC 5100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10010 AATATGATGCTTCTAATCCAAAAGAGGAAAGGCATTGGGAGTCAGCTCCTAAGTAAGCTC 10069

Qy       5101 CAGAATTCCTGCTGGTACTTTTCCTTCCAGGAAGCAACTTCCTTGATATTTTTTTTTTAC 5160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10070 CAGAATTCCTGCTGGTACTTTTCCTTCCAGGAAGCAACTTCCTTGATATTTTTTTTTTAC 10129

Qy       5161 AGGCATATGAATAAAAACTATATTTTGCAGCATTGTACACTTTTTTTCCTTTTCTAGAAA 5220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10130 AGGCATATGAATAAAAACTATATTTTGCAGCATTGTACACTTTTTTTCCTTTTCTAGAAA 10189

Qy       5221 TTCTAAACCTCTGACATTGGTGGAGACATTGAGTACATTTTTTCCCATATCCCTACTTTT 5280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10190 TTCTAAACCTCTGACATTGGTGGAGACATTGAGTACATTTTTTCCCATATCCCTACTTTT 10249

Qy       5281 CAGAAGGATTTTCTCTGCTCGTTCACTTAACATTGCTGATGCGTCAGTCTTTTCTTCCTC 5340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10250 CAGAAGGATTTTCTCTGCTCGTTCACTTAACATTGCTGATGCGTCAGTCTTTTCTTCCTC 10309

Qy       5341 ATCTCTTTCAGGGGCTGGAGAGGCAGAGGGAGACAGAGGAGCTGGTACTGCAGAGCGGTC 5400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10310 ATCTCTTTCAGGGGCTGGAGAGGCAGAGGGAGACAGAGGAGCTGGTACTGCAGAGCGGTC 10369

Qy       5401 GTCTGATTGGCTGGACGGTCGTAGCTGGGCTATAAAAGAGACCCCTACAGGCTTAGCAGG 5460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10370 GTCTGATTGGCTGGACGGTCGTAGCTGGGCTATAAAAGAGACCCCTACAGGCTTAGCAGG 10429

Qy       5461 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 5520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10430 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 10489
```

```
Qy       5521 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 5580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10490 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 10549

Qy       5581 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAGGTAAGATTGCTT 5640
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10550 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAGGTAAGATTGCTT 10609

Qy       5641 TCAGCCATTAACCATATTAAACTTTTGGCTAGACTTTCTCAGTTATTTACATGTTGTACT 5700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10610 TCAGCCATTAACCATATTAAACTTTTGGCTAGACTTTCTCAGTTATTTACATGTTGTACT 10669

Qy       5701 TACTAACCTAGTTCTGTGCAATTAGAAACAGTGTGGTCAGGAGAGCACGACTTTCTAACT 5760
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10670 TACTAACCTAGTTCTGTGCAATTAGAAACAGTGTGGTCAGGAGAGCACGACTTTCTAACT 10729

Qy       5761 TTCCTCCAAGACTAGCTAGATATTGTGACTTAAGACATGTGCTCCCCAAATTTCAGCCCT 5820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10730 TTCCTCCAAGACTAGCTAGATATTGTGACTTAAGACATGTGCTCCCCAAATTTCAGCCCT 10789

Qy       5821 TATGTGTTGTTTTGTGTGACCTCAGTTTTGAGAACTGTTCTATTCTTTAAGCCAGGTCTA 5880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10790 TATGTGTTGTTTTGTGTGACCTCAGTTTTGAGAACTGTTCTATTCTTTAAGCCAGGTCTA 10849

Qy       5881 AGAAAGCTAGTTTTAATTAAGAAGCGAGATGAGGTTTGAGGCTATGTACAGTGATCTGTA 5940
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10850 AGAAAGCTAGTTTTAATTAAGAAGCGAGATGAGGTTTGAGGCTATGTACAGTGATCTGTA 10909

Qy       5941 ATATCTCCATCTGTGATTACTACTGCTATTTGAGCATCCCTGGAGTACATAGAAGCCTGG 6000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10910 ATATCTCCATCTGTGATTACTACTGCTATTTGAGCATCCCTGGAGTACATAGAAGCCTGG 10969

Qy       6001 CTCTGGGCTTTCTGATTGTATGCTACAACTTGTTTCAGGAAAGGTACCCCAGAATGAGGT 6060
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10970 CTCTGGGCTTTCTGATTGTATGCTACAACTTGTTTCAGGAAAGGTACCCCAGAATGAGGT 11029

Qy       6061 TTGGCTCCATCATCAGAAAGGCACTAT 6087
             |||||||||||||||||||||||||||
Db      11030 TTGGCTCCATCATCAGAAAGGCACTAT 11056
```

RESULT 2
US-10-719-993-6815
; Sequence 6815, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6815
; LENGTH: 1980090
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1980090)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-2
US-10-719-993-6815

Query Match 5.4%; Score 325.8; DB 18; Length 1980090;
Best Local Similarity 75.4%; Pred. No. 4.4e-67;
Matches 494; Conservative 1; Mismatches 128; Indels 32; Gaps 6;

```
Qy        309 GCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAA 368
              || ||| |||| |||| |||||||| |||| |||  |||||||||||||| |   |    |
Db     116699 GCTGTAAGGGTAAAAAAGTGTGATATATTTCCTCACCCATCATAAGGGTTGTGGCTGGCA 116758
```

```
Qy        369 CTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGT 428
              |||||||| |||||||||||||||||||||| ||| | |||||||: || || ||||||||
Db     116759 CTCCTATAATAAAAGACAGGTTAATAAGAGAAAAGCATAACAAAYGTACTTCATCAAAGT 116818

Qy        429 TTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGT 488
              ||||  |||||   | |||||||||||||||||||| |||||||||| | | |||||
Db     116819 TTTAGGTGACACCAGTGTCTTCAGAAATGAAGACCAAAAGACCCAGAGAACACTGT---G 116875

Qy        489 TTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAG 548
              | | | |||| |||| |||||||||||| ||| |     |    |||| ||||||||
Db     116876 TCTCTGTGCTTAGGTCTGATGAAGAATGGGCAGCGGTGTAGGAATGTGATTGGACAAAAG 116935

Qy        549 GATATGATCTAGTGGTAAAGGACTCAG-GGGGAAACACAGCAAGGCCTGTCTATTCAGAT 607
              | ||||||||| |||||| ||||| || |||||||| |||||||| || ||| |||||||
Db     116936 GGTATGATCTAATGGTAATGGACTAAGCGGGGAAACCCAGCAAGGTCTCTCTGTTCAGAT 116995

Qy        608 TCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGAC 667
              |||||||  |||||        |||  |||||||||||||||| || ||||||||||
Db     116996 CCTTCTTGGCCTCTC--------TGTGCAGCATTCTTTCCTCCTGGGTTTGGGGCAGGAC 117047

Qy        668 TCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATGG 727
              |||||  | |||||||||||||| ||| ||  |||||||| ||||| ||| |||||||||
Db     117048 CCTTCTGGAGTGAGGGTCTTCAATGGACAACAGAGAAAGTGACCTTTCTAGGTTTTATGG 117107

Qy        728 CTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCTG 787
              |||| ||  ||| |||| |||||||||||||| ||||| | |||||||||||||||| ||
Db     117108 CTTGTTTGTGGGGAGAGAAGTTCTAGTTTCTCTGACCAACCTTGGGGAAGAGGAATTATG 117167

Qy        788 GTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGGT 847
              ||||||||||||  | |    | ||||||||  || |||| ||||||||||||||| |||
Db     117168 GTTTCTGTGACT--CCTCAGGGGAGAAAGAGAGGTGAGAGACAGGAGGGCAGGAGAAGGT 117225

Qy        848 CAGAAAGAGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAGTACTTCTT 907
              ||||  |           |||||||||||           | |||   ||     | | ||
Db     117226 CAGAGA-------CTTTGCTTCTGAGG-----------CTTTTTAGTCTCTAAAGTACTC 117267

Qy        908 AGCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAACACTGCAA 962
              ||||| ||||||| | |||||||||  |||  |||||||||||| ||||  || |
Db     117268 AGCATGCCAAAGCGCCATACTTTGGGGTATCATTTTCTGAGCTCCAACAAAGCTA 117322


RESULT 3
US-10-741-600-17676
; Sequence 17676, Application US/10741600
; Publication No. US20050026169A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001499
;  CURRENT APPLICATION NUMBER: US/10/741,600
;  CURRENT FILING DATE:  2003-12-22
;  NUMBER OF SEQ ID NOS: 73997
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17676
;   LENGTH: 1980090
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1980090)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-741-600-17676

  Query Match             5.4%;  Score 325.8;  DB 19;  Length 1980090;
  Best Local Similarity   75.4%;  Pred. No. 4.4e-67;
  Matches  494;  Conservative    1;  Mismatches  128;  Indels   32;  Gaps    6;

Qy        309 GCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAA 368
              || ||| |||| |||| |||||||| |||| |||  |||||||||||| |   |   |
Db     116699 GCTGTAAGGGTAAAAAAGTGTGATATATTTCCTCACCCATCATAAGGGTTGTGGCTGGCA 116758

Qy        369 CTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGT 428
              |||||||| |||||||||||||||||||||| ||| | |||||||: || || ||||||||
Db     116759 CTCCTATAATAAAAGACAGGTTAATAAGAGAAAAGCATAACAAAYGTACTTCATCAAAGT 116818

Qy        429 TTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGT 488
              ||||  |||||   | |||||||||||||||||||| |||||||||| | | |||||
Db     116819 TTTAGGTGACACCAGTGTCTTCAGAAATGAAGACCAAAAGACCCAGAGAACACTGT---G 116875

Qy        489 TTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAG 548
              | | | |||| |||| |||||||||||| ||| |     |    |||| ||||||||
Db     116876 TCTCTGTGCTTAGGTCTGATGAAGAATGGGCAGCGGTGTAGGAATGTGATTGGACAAAAG 116935

Qy        549 GATATGATCTAGTGGTAAAGGACTCAG-GGGGAAACACAGCAAGGCCTGTCTATTCAGAT 607
              | ||||||||| |||||| ||||| || |||||||| |||||||| || ||| |||||||
Db     116936 GGTATGATCTAATGGTAATGGACTAAGCGGGGAAACCCAGCAAGGTCTCTCTGTTCAGAT 116995

Qy        608 TCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGAC 667
              |||||||  |||||        |||  |||||||||||||||| || ||||||||||
Db     116996 CCTTCTTGGCCTCTC--------TGTGCAGCATTCTTTCCTCCTGGGTTTGGGGCAGGAC 117047

Qy        668 TCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATGG 727
              |||||  | |||||||||||||| ||| ||  |||||||| ||||| ||| |||||||||
Db     117048 CCTTCTGGAGTGAGGGTCTTCAATGGACAACAGAGAAAGTGACCTTTCTAGGTTTTATGG 117107

Qy        728 CTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCTG 787
              |||| ||  ||| |||| |||||||||||||| ||||| | |||||||||||||||| ||
Db     117108 CTTGTTTGTGGGGAGAGAAGTTCTAGTTTCTCTGACCAACCTTGGGGAAGAGGAATTATG 117167

Qy        788 GTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGGT 847
              ||||||||||||  | |    | ||||||||  || |||| ||||||||||||||| |||
Db     117168 GTTTCTGTGACT--CCTCAGGGGAGAAAGAGAGGTGAGAGACAGGAGGGCAGGAGAAGGT 117225

Qy        848 CAGAAAGAGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAGTACTTCTT 907
              ||||  |           |||||||||||           | |||   ||     | | ||
Db     117226 CAGAGA-------CTTTGCTTCTGAGG-----------CTTTTTAGTCTCTAAAGTACTC 117267

Qy        908 AGCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAACACTGCAA 962
              ||||| ||||||| | |||||||||  |||  |||||||||||| ||||  || |
Db     117268 AGCATGCCAAAGCGCCATACTTTGGGGTATCATTTTCTGAGCTCCAACAAAGCTA 117322


RESULT 4
US-10-198-846-13843/c
; Sequence 13843, Application US/10198846
; Publication No. US20030099974A1
; GENERAL INFORMATION:
;  APPLICANT: Lillie, James
;  APPLICANT:  Xu, Yongyao
;  APPLICANT:  Wang, Youzhen
;  APPLICANT:  Steinmann, Kathleen
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
;  TITLE OF INVENTION:  FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
;  TITLE OF INVENTION:  THERAPY OF BREAST CANCER
;  FILE REFERENCE: MRI-049
;  CURRENT APPLICATION NUMBER: US/10/198,846
;  CURRENT FILING DATE:  2002-07-18
;  PRIOR APPLICATION NUMBER: 60/306,220
;  PRIOR FILING DATE: 2001-07-18
;  NUMBER OF SEQ ID NOS: 14084
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13843
;   LENGTH: 2752
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-198-846-13843

  Query Match             4.6%;  Score 277.6;  DB 14;  Length 2752;
  Best Local Similarity   98.6%;  Pred. No. 6e-57;
  Matches  280;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       5345 CTTTCAGGGGCTGGAGAGGCAGAGGGAGACAGAGGAGCTGGTACTGCAGAGCGGTCGTCT 5404
              ||   | ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2696 CTCCTAAGGGCTGGAGAGGCAGAGGGAGACAGAGGAGCTGGTACTGCAGAGCGGTCGTCT 2637
```

```
Qy       5405 GATTGGCTGGACGGTCGTAGCTGGGCTATAAAAGAGACCCCTACAGGCTTAGCAGGAAGA 5464
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2636 GATTGGCTGGACGGTCGTAGCTGGGCTATAAAAGAGACCCCTACAGGCTTAGCAGGAAGA 2577

Qy       5465 CGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGCCGA 5524
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2576 CGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGCCGA 2517

Qy       5525 AGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAATAAG 5584
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2516 AGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAATAAG 2457

Qy       5585 ATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             |||||||||||||||||||||||||||||||||||||||||||||
Db       2456 ATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 2413


RESULT 5
US-10-198-846-9863
; Sequence 9863, Application US/10198846
; Publication No. US20030099974A1
; GENERAL INFORMATION:
;  APPLICANT: Lillie, James
;  APPLICANT:  Xu, Yongyao
;  APPLICANT:  Wang, Youzhen
;  APPLICANT:  Steinmann, Kathleen
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
;  TITLE OF INVENTION:  FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
;  TITLE OF INVENTION:  THERAPY OF BREAST CANCER
;  FILE REFERENCE: MRI-049
;  CURRENT APPLICATION NUMBER: US/10/198,846
;  CURRENT FILING DATE:  2002-07-18
;  PRIOR APPLICATION NUMBER: 60/306,220
;  PRIOR FILING DATE: 2001-07-18
;  NUMBER OF SEQ ID NOS: 14084
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9863
;   LENGTH: 1531
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 1, 2, 3, 4, 1531
;   OTHER INFORMATION: n = A,T,C or G
US-10-198-846-9863

  Query Match             4.5%;  Score 276;  DB 14;  Length 1531;
  Best Local Similarity   98.2%;  Pred. No. 1e-56;
  Matches  279;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       5345 CTTTCAGGGGCTGGAGAGGCAGAGGGAGACAGAGGAGCTGGTACTGCAGAGCGGTCGTCT 5404
             ||   | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         58 CTCCTAAGGGCTGGAGAGGCAGAGGGAGACAGAGGAGCTGGTACTGCAGAGCGGTCGTCT 117

Qy       5405 GATTGGCTGGACGGTCGTAGCTGGGCTATAAAAGAGACCCCTACAGGCTTAGCAGGAAGA 5464
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        118 GATTGGCTGGACGGTCGTAGCTGGGCTATAAAAGAGACCCCTACAGGCTTAGCAGGAACA 177

Qy       5465 CGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGCCGA 5524
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        178 CGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGCCGA 237

Qy       5525 AGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAATAAG 5584
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        238 AGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAATAAG 297

Qy       5585 ATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             |||||||||||||||||||||||||||||||||||||||||||||
Db        298 ATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 341


RESULT 6
US-10-044-090-113
; Sequence 113, Application US/10044090
; Publication No. US20020137081A1
; GENERAL INFORMATION:
;  APPLICANT: Olga Bandman
;  TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
;  FILE REFERENCE: PA-0028 US
;  CURRENT APPLICATION NUMBER: US/10/044,090
;  CURRENT FILING DATE:  2002-01-09
;  NUMBER OF SEQ ID NOS: 850
;  SOFTWARE: PERL Program
; SEQ ID NO 113
;   LENGTH: 3165
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20020137081A1 1290877CB1
;   NAME/KEY: unsure
;   LOCATION: 3157
;   OTHER INFORMATION: a, t, c, g, or other
US-10-044-090-113

  Query Match             4.5%;  Score 274.4;  DB 13;  Length 3165;
  Best Local Similarity   97.9%;  Pred. No. 4e-56;
  Matches  278;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       5345 CTTTCAGGGGCTGGAGAGGCAGAGGGAGACAGAGGAGCTGGTACTGCAGAGCGGTCGTCT 5404
             | |  ||||  ||||||||||||||||||||||||||||||||||||||||||||||||||
Db         63 CGTAGAGGGCTCGGAGAGGCAGAGGGAGACAGAGGAGCTGGTACTGCAGAGCGGTCGTCT 122

Qy       5405 GATTGGCTGGACGGTCGTAGCTGGGCTATAAAAGAGACCCCTACAGGCTTAGCAGGAAGA 5464
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        123 GATTGGCTGGACGGTCGTAGCTGGGCTATAAAAGAGACCCCTACAGGCTTAGCAGGAAGA 182

Qy       5465 CGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGCCGA 5524
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        183 CGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGCCGA 242

Qy       5525 AGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAATAAG 5584
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        243 AGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAATAAG 302

Qy       5585 ATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             |||||||||||||||||||||||||||||||||||||||||||||
Db        303 ATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 346


RESULT 7
US-09-918-624B-6
; Sequence 6, Application US/09918624B
; Publication No. US20030113720A1
; GENERAL INFORMATION:
;  APPLICANT: Schebye, Xiao Min
;  APPLICANT:  Sornasse, Thierry
;  TITLE OF INVENTION: CDNAS EXPRESSED IN ADIPOCYTE DIFFERENTIATION
;  FILE REFERENCE: PA-0033 US
;  CURRENT APPLICATION NUMBER: US/09/918,624B
;  CURRENT FILING DATE:  2002-12-03
;  PRIOR APPLICATION NUMBER: 60/222,470
;  PRIOR FILING DATE: 2000-07-28
;  NUMBER OF SEQ ID NOS: 71
;  SOFTWARE: PERL Program
; SEQ ID NO 6
;   LENGTH: 3167
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20030113720A1 232915.1
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: 1465-1487, 3159
;   OTHER INFORMATION: a, t, c, g, or other
```

```
US-09-918-624B-6

  Query Match             4.5%;  Score 274.4;  DB 10;  Length 3167;
  Best Local Similarity   97.9%;  Pred. No. 4e-56;
  Matches  278;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       5345 CTTTCAGGGGCTGGAGAGGCAGAGGGAGACAGAGGAGCTGGTACTGCAGAGCGGTCGTCT 5404
              | |  ||||   ||||||||||||||||||||||||||||||||||||||||||||||||||
Db         63 CGTAGAGGGCTCGGAGAGGCAGAGGGAGACAGAGGAGCTGGTACTGCAGAGCGGTCGTCT 122

Qy       5405 GATTGGCTGGACGGTCGTAGCTGGGCTATAAAAGAGACCCCTACAGGCTTAGCAGGAAGA 5464
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        123 GATTGGCTGGACGGTCGTAGCTGGGCTATAAAAGAGACCCCTACAGGCTTAGCAGGAAGA 182

Qy       5465 CGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGCCGA 5524
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        183 CGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGCCGA 242

Qy       5525 AGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAATAAG 5584
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        243 AGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAATAAG 302

Qy       5585 ATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
              |||||||||||||||||||||||||||||||||||||||||||||
Db        303 ATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 346


RESULT 8
US-10-027-632-266349
; Sequence 266349, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 266349
;   LENGTH: 1151
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-266349

  Query Match             4.2%;  Score 253.6;  DB 13;  Length 1151;
  Best Local Similarity   73.7%;  Pred. No. 2.8e-51;
  Matches  407;  Conservative    0;  Mismatches  129;  Indels   16;  Gaps    6;

Qy        313 TAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCA-ACCAAAACTC 371
              | ||||   ||| ||||||| |  || |||  ||||||||   |   || | ||| |||
Db        593 TGGGGGCAGAAAGGTGTGATGCCCTTGCTCACCCATCATATAAGCATCACAACAACACTG 652

Qy        372 CTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTT 431
              ||||   |||||||||||||| ||||| ||| | ||||||||||||||||||||||||
Db        653 CTATCACAAAAGACAGGTTAACAAGAG-AAAGCGTAACAAATTTATTTAATCAAAGTTTC 711

Qy        432 ACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTT 491
              ||||||||  |||| ||||||||||||||||||||||||  |||||||  |||||  |||
Db        712 ACATGACAAAGGAGCCTTCAGAAATGAAGACCCAAAGACCTGGGGGAAATTGTCTG-TTT 770

Qy        492 TTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGAT 551
              ||  ||| ||||| || ||| |||||| |||         |      ||| |||||| || |
Db        771 TTACGCTTAGGTTTGAGGAATAATGGACAGCCATGTAGAAACGTAATTGGACAAATGGCT 830

Qy        552 ATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTT 611
              |||||||| ||||||  | || ||||||| || ||||||||||||||   |||||||||
Db        831 ATGATCTAATGGTAATAGTCTAAGGGGGAGACCCAGCAAGGCCTGTCTGCCCAGATTCTT 890

Qy        612 CTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTT 671
              ||||  ||||||  |        |     |       |||||     | |||| || |||
Db        891 CTTGGCCTCTCTGTGAAGCATTTCTTCTCCACCTCCTC-----TGTGGGACAAGACCTCA 945

Qy        672 CTTCAATGAGGGTCTTCAAGGGAGAAGGGAG-------AAAGTGGCCTTTTTAGATTTTA 724
              ||  |||||||||| ||||||||||||||        |||||  |||||  ||| |||||
Db        946 CTGGAATGAGGGTCCTCAAGGGAGAAGGGAGGAGGGAAAAAGTTACCTTTCTAGGTTTTA 1005

Qy        725 TGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATT 784
              |||||||||| | |||| |||||||||||||||||| |||    | || ||||||||||||
Db       1006 TGGCTTGCTTTGAGGAAAAGGAGTTCTAGTTTCTATTACCTGCCCTGAGGAAGAGGAATT 1065

Qy        785 CTGGTTTCTGTGACTTGCTT-TCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGA 843
              ||| |||||||||  |||    | | ||||  |||    | |||   |||||  | | |||
Db       1066 CTGATTTCTGTGAATCACTTCAAAGGGAGAAGAAGGGTCAGGAGATGGGAGGACTGCAGA 1125

Qy        844 TGGTCAGAAAGA 855
              ||||||||||||
Db       1126 TGGTCAGAAAGA 1137


RESULT 9
US-10-027-632-266349
; Sequence 266349, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 266349
;   LENGTH: 1151
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-266349

  Query Match             4.2%;  Score 253.6;  DB 17;  Length 1151;
  Best Local Similarity   73.7%;  Pred. No. 2.8e-51;
  Matches  407;  Conservative    0;  Mismatches  129;  Indels   16;  Gaps    6;

Qy        313 TAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCA-ACCAAAACTC 371
              | ||||   ||| ||||||| |  || |||  ||||||||   |   || | ||| |||
Db        593 TGGGGGCAGAAAGGTGTGATGCCCTTGCTCACCCATCATATAAGCATCACAACAACACTG 652
```

```
Qy        372 CTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTT 431
              ||||   |||||||||||||| ||||| ||| | ||||||||||||||||||||||||
Db        653 CTATCACAAAAGACAGGTTAACAAGAG-AAAGCGTAACAAATTTATTTAATCAAAGTTTC 711

Qy        432 ACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTT 491
              ||||||||  |||| ||||||||||||||||||||||||  ||||||| |||||| |||
Db        712 ACATGACAAAGGAGCCTTCAGAAATGAAGACCCAAAGACCTGGGGGAAATTGTCTG-TTT 770

Qy        492 TTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGAT 551
              ||  ||| ||||| || ||| |||||| |||         |     ||| |||||| || |
Db        771 TTACGCTTAGGTTTGAGGAATAATGGACAGCCATGTAGAAACGTAATTGGACAAATGGCT 830

Qy        552 ATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTT 611
              |||||||| ||||||  | || ||||||| || ||||||||||||||   |||||||||
Db        831 ATGATCTAATGGTAATAGTCTAAGGGGGAGACCCAGCAAGGCCTGTCTGCCCAGATTCTT 890

Qy        612 CTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTT 671
              ||||  |||||| |       |     |      |||||     | |||| || |||
Db        891 CTTGGCCTCTCTGTGAAGCATTTCTTCTCCACCTCCTC-----TGTGGGACAAGACCTCA 945

Qy        672 CTTCAATGAGGGTCTTCAAGGGAGAAGGGAG-------AAAGTGGCCTTTTTAGATTTTA 724
              ||  |||||||||| ||||||||||||||       |||||  ||||| ||| |||||
Db        946 CTGGAATGAGGGTCCTCAAGGGAGAAGGGAGGAGGGAAAAAGTTACCTTTCTAGGTTTTA 1005

Qy        725 TGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATT 784
              |||||||||| | |||| |||||||||||||||||| |||    | || ||||||||||||
Db       1006 TGGCTTGCTTTGAGGAAAAGGAGTTCTAGTTTCTATTACCTGCCCTGAGGAAGAGGAATT 1065

Qy        785 CTGGTTTCTGTGACTTGCTT-TCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGA 843
              |||  |||||||||  |  |||   | | ||||  |||    | |||    |||||  | | |||
Db       1066 CTGATTTCTGTGAATCACTTCAAAGGGAGAAGAAGGGTCAGGAGATGGGAGGACTGCAGA 1125

Qy        844 TGGTCAGAAAGA 855
              ||||||||||||
Db       1126 TGGTCAGAAAGA 1137
```

RESULT 10
US-10-027-632-266350
; Sequence 266350, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 266350
; LENGTH: 1151
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-266350

Query Match 4.2%; Score 253.2; DB 13; Length 1151;
Best Local Similarity 73.6%; Pred. No. 3.6e-51;
Matches 406; Conservative 1; Mismatches 129; Indels 16; Gaps 6;

```
Qy        313 TAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCA-ACCAAAACTC 371
              | ||||   ||| ||||||| |  || |||  ||||||||    |   || | ||| |||
Db        593 TGGGGGCAGAAAGGTGTGATGCCCTTGCTCACCCATCATATAAGCATCACAACAACACTG 652

Qy        372 CTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTT 431
              ||||   |||||||||||||| ||||| ||| | ||||||||||||||||||||||||
Db        653 CTATCACAAAAGACAGGTTAACAAGAG-AAAGCGTAACAAATTTATTTAATCAAAGTTTC 711

Qy        432 ACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTT 491
              ||||||||  |||| ||||||||||||||||||||||||  ||||||| |||||| |||
Db        712 ACATGACAAAGGAGCCTTCAGAAATGAAGACCCAAAGACCTGGGGGAAATTGTCTG-TTT 770

Qy        492 TTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGAT 551
              ||  ||| ||||| || ||| |||||| |||         |     ||| |||||| || |
Db        771 TTACGCTTAGGTTTGAGGAATAATGGACAGCCATGTAGAAACGTAATTGGACAAATGGCT 830

Qy        552 ATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTT 611
              |||||||| ||||||  | || ||||||| || ||||||||||||||   |||||||||
Db        831 ATGATCTAATGGTAATAGTCTAAGGGGGAGACCCAGCAAGGCCTGTCTGCCCAGATTCTT 890

Qy        612 CTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTT 671
              ||||  |||||| |       |     |      |||||     | |||| || |||
Db        891 CTTGGCCTCTCTGTGAAGCATTTCTTCTCCACCTCCTC-----TGTGGGACAAGACCTCA 945

Qy        672 CTTCAATGAGGGTCTTCAAGGGAGAAGGGAG-------AAAGTGGCCTTTTTAGATTTTA 724
              ||  |||||||||| ||||||||||||||       |||||  ||||| ||| |||||
Db        946 CTGGAATGAGGGTCCTCAAGGGAGAAGGGAGGAGGGAAAAAGTTACCTTTCTAGGTTTTA 1005

Qy        725 TGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATT 784
              |||||:|||| | |||| |||||||||||||||||| |||    | || ||||||||||||
Db       1006 TGGCTKGCTTTGAGGAAAAGGAGTTCTAGTTTCTATTACCTGCCCTGAGGAAGAGGAATT 1065

Qy        785 CTGGTTTCTGTGACTTGCTT-TCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGA 843
              |||  |||||||||  |  |||   | | ||||  |||    | |||    |||||  | | |||
Db       1066 CTGATTTCTGTGAATCACTTCAAAGGGAGAAGAAGGGTCAGGAGATGGGAGGACTGCAGA 1125

Qy        844 TGGTCAGAAAGA 855
              ||||||||||||
Db       1126 TGGTCAGAAAGA 1137
```

RESULT 11
US-10-027-632-266350
; Sequence 266350, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 266350
; LENGTH: 1151
; TYPE: DNA
; ORGANISM: Human

US-10-027-632-266350

Query Match 4.2%; Score 253.2; DB 17; Length 1151;
Best Local Similarity 73.6%; Pred. No. 3.6e-51;
Matches 406; Conservative 1; Mismatches 129; Indels 16; Gaps 6;

```
Qy        313 TAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCA-ACCAAAACTC 371
              | ||||   ||| |||||||  |  || |||  ||||||||    |    || | ||| |||
Db        593 TGGGGGCAGAAAGGTGTGATGCCCTTGCTCACCCATCATATAAGCATCACAACAACACTG 652

Qy        372 CTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTT 431
              ||||   |||||||||||||| ||||| ||| | ||||||||||||||||||||||||
Db        653 CTATCACAAAAGACAGGTTAACAAGAG-AAAGCGTAACAAATTTATTTAATCAAAGTTTC 711

Qy        432 ACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTT 491
              ||||||||  |||| |||||||||||||||||||||||||  |||||||  ||||||  |||
Db        712 ACATGACAAAGGAGCCTTCAGAAATGAAGACCCAAAGACCTGGGGGAAATTGTCTG-TTT 770

Qy        492 TTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGAT 551
              ||  ||| ||||| || ||| |||||| |||         |     ||| ||||||  || |
Db        771 TTACGCTTAGGTTTGAGGAATAATGGACAGCCATGTAGAAACGTAATTGGACAAATGGCT 830

Qy        552 ATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTT 611
              |||||||| ||||||  | || ||||||| || ||||||||||||||   |||||||||
Db        831 ATGATCTAATGGTAATAGTCTAAGGGGGAGACCCAGCAAGGCCTGTCTGCCCAGATTCTT 890

Qy        612 CTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTT 671
              ||||  ||||||  |        |     |||||     | |||| || |||
Db        891 CTTGGCCTCTCTGTGAAGCATTTCTTCTCCACCTCCTC-----TGTGGGACAAGACCTCA 945

Qy        672 CTTCAATGAGGGTCTTCAAGGGAGAAGGGAG-------AAAGTGGCCTTTTTAGATTTTA 724
              ||  |||||||||| ||||||||||||||       |||||  |||||| ||| |||||
Db        946 CTGGAATGAGGGTCCTCAAGGGAGAAGGGAGGAGGGAAAAAGTTACCTTTCTAGGTTTTA 1005

Qy        725 TGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATT 784
              |||||:|||| | |||| ||||||||||||||||| |||   | || ||||||||||||
Db       1006 TGGCTKGCTTTGAGGAAAAGGAGTTCTAGTTTCTATTACCTGCCCTGAGGAAGAGGAATT 1065

Qy        785 CTGGTTTCTGTGACTTGCTT-TCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGA 843
              ||| |||||||||| |  |||   | | ||||  |||   | |||   ||||| | | |||
Db       1066 CTGATTTCTGTGAATCACTTCAAAGGGAGAAGAAGGGTCAGGAGATGGGAGGACTGCAGA 1125

Qy        844 TGGTCAGAAAGA 855
              ||||||||||||
Db       1126 TGGTCAGAAAGA 1137
```

RESULT 12
US-10-027-632-288758
; Sequence 288758, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720

; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 288758
; LENGTH: 576
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-288758

Query Match 3.8%; Score 233.8; DB 13; Length 576;
Best Local Similarity 76.1%; Pred. No. 1.4e-46;
Matches 386; Conservative 0; Mismatches 97; Indels 24; Gaps 7;

```
Qy        313 TAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTCC 372
              || ||||| ||| |||||||||| ||| ||| |||||||  | ||       | || |||||
Db         89 TACGGGTGGAAAGGTGTGATACCTTTCCTCAGCCATCATTAAGG-----GCTAACACTCC 143

Qy        373 TATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTA 432
              ||||  |||| ||| ||||| ||||| ||||| ||||| |||||||||||||| |||||||
Db        144 TATAACAAAA-ACAAGTTAACAAGAGAAAAACATAACACATTTATTTAATCAGAGTTTTA 202

Qy        433 CATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTTT 492
              |||| ||  |||| ||||||||||||||||||||||||||||||| |||||      ||||
Db        203 CATGGCACAGGAGGCTTCAGAAATGAAGACCCAAAGACCCAGGGAAAACT-----ATTTT 257

Qy        493 TTTGCTGAGGTTCGATGAAGAATGGATAGC-ATGTAGCCATGTAGATTAGACAAAAGGAT 551
              | |||| || ||  ||| |||| |  ||| ||| ||  ||||    |      ||| || | |
Db        258 TATGCTCAGATTTAATGCAGAATAGTCAGCTATGAAGGAATGTGATTGGACAAAATGGGT 317

Qy        552 ATGATCTAGTGGTAAAGGACTCAGGGGGAAAC---ACAGCAAGGCCTGTCTATTCAGATT 608
              |||||||| |||| |  ||||  ||||  | |    ||||||| ||||||| ||||||||
Db        318 ATGATCTAATGGTTATAGACTGGGGGGTGAGCGGACCAGCAAGCCCTGTCTGTTCAGATT 377

Qy        609 CTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACT 668
              |||||||  | |||         || |||| |||  ||||||  || ||||||||||||||
Db        378 CTTCTTGGACGCTC---------TGCATAGTATTTATTCCTCTTGGGTATGGGGCAGGACT 429

Qy        669 CTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATGGC 728
              | |||  |  |||||||||||||||||||||| ||| |||| ||||| ||| |||||| |
Db        430 CCTCTAAATGGAGGGTCTTCAAGGGAGAAGGAAGAGAGTGACCTTTSTAGGTTTTAT-GT 488

Qy        729 TTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCTGG 788
              |||||| | || |||||||| |||||||||  |||     ||| ||||||||||||||||
Db        489 TTGCTTTGAGGGAGAGGAGTCCTAGTTTCTAAGACTGGCTTTGAGGAAGAGGAATTCTGG 548

Qy        789 TTTCTGTGACTTGCTTTCATGAAGAAA 815
              ||||| || ||||| | || | |||||
Db        549 TTTCTATGGCTTGCATCCAGGGAGAAA 575
```

RESULT 13
US-10-027-632-288758
; Sequence 288758, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09

```
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 288758
;   LENGTH: 576
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-288758

  Query Match             3.8%;  Score 233.8;  DB 17;  Length 576;
  Best Local Similarity   76.1%;  Pred. No. 1.4e-46;
  Matches  386;  Conservative    0;  Mismatches   97;  Indels   24;  Gaps    7;

Qy        313 TAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTCC 372
              || ||||| ||| |||||||||| ||| |||  ||||||| | ||       | || |||||
Db         89 TACGGGTGGAAAGGTGTGATACCTTTCCTCAGCCATCATTAAGG-----GCTAACACTCC 143

Qy        373 TATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTA 432
              ||||  |||| ||| ||||| ||||| ||||| ||||| |||||||||||| |||||||
Db        144 TATAACAAAA-ACAAGTTAACAAGAGAAAAACATAACACATTTATTTAATCAGAGTTTTA 202

Qy        433 CATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTTT 492
              |||| ||  |||| |||||||||||||||||||||||||||| |||||      ||||
Db        203 CATGGCACAGGAGGCTTCAGAAATGAAGACCCAAAGACCCAGGGAAAACT-----ATTTT 257

Qy        493 TTTGCTGAGGTTCGATGAAGAATGGATAGC-ATGTAGCCATGTAGATTAGACAAAAGGAT 551
              | |||| || ||  ||| ||||| |  ||| ||| ||  ||||   |      ||| || |
Db        258 TATGCTCAGATTTAATGCAGAATAGTCAGCTATGAAGGAATGTGATTGGACAAAATGGGT 317

Qy        552 ATGATCTAGTGGTAAAGGACTCAGGGGGAAAC---ACAGCAAGGCCTGTCTATTCAGATT 608
              |||||||| |||| |  ||||  ||||  | |    ||||||| ||||||| ||||||||
Db        318 ATGATCTAATGGTTATAGACTGGGGGGTGAGCGGACCAGCAAGCCCTGTCTGTTCAGATT 377

Qy        609 CTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACT 668
              |||||||  | |||         || |||| |||  |||||| || ||||||||||||||
Db        378 CTTCTTGGACGCTC--------TGCATAGTATTTATTCCTCTTGGGTATGGGGCAGGACT 429

Qy        669 CTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATGGC 728
              | |||  |  |||||||||||||||||||| ||| |||| ||||| ||| |||||| |
Db        430 CCTCTAAATGGAGGGTCTTCAAGGGAGAAGGAAGAGAGTGACCTTTSTAGGTTTTAT-GT 488

Qy        729 TTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCTGG 788
              |||||| | || |||||||| |||||||||| |||     ||| ||||||||||||||||
Db        489 TTGCTTTGAGGGAGAGGAGTCCTAGTTTCTAAGACTGGCTTTGAGGAAGAGGAATTCTGG 548

Qy        789 TTTCTGTGACTTGCTTTCATGAAGAAA 815
              ||||| || ||||| | || | |||||
Db        549 TTTCTATGGCTTGCATCCAGGGAGAAA 575


RESULT 14
US-10-027-632-288755
; Sequence 288755, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 288755
;   LENGTH: 576
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-288755

  Query Match             3.8%;  Score 233.4;  DB 13;  Length 576;
  Best Local Similarity   75.9%;  Pred. No. 1.7e-46;
  Matches  385;  Conservative    1;  Mismatches   97;  Indels   24;  Gaps    7;

Qy        313 TAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTCC 372
              || ||||| ||| ||||||||| ||| |||  ||||||| | ||       | || |||||
Db         89 TACGGGTGGAAAGGTGTGATACCTTTCCTCAGCCATCATTAAGG-----GCTAACACTCC 143

Qy        373 TATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTA 432
              ||||  |||| ||| ||||| ||||| ||||| ||||| |||||||||||| |||||||
Db        144 TATAACAAAA-ACAAGTTAACAAGAGAAAAACATAACACATTTATTTAATCAGAGTTTTA 202

Qy        433 CATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTTT 492
              |||| ||  |||| |||||||||||||||||||||||||||||: |||||      ||||
Db        203 CATGGCACAGGAGGCTTCAGAAATGAAGACCCAAAGACCCAGGRAAAACT-----ATTTT 257

Qy        493 TTTGCTGAGGTTCGATGAAGAATGGATAGC-ATGTAGCCATGTAGATTAGACAAAAGGAT 551
              | |||| || ||  ||| ||||| |  ||| ||| ||  ||||   |      ||| || |
Db        258 TATGCTCAGATTTAATGCAGAATAGTCAGCTATGAAGGAATGTGATTGGACAAAATGGGT 317

Qy        552 ATGATCTAGTGGTAAAGGACTCAGGGGGAAAC---ACAGCAAGGCCTGTCTATTCAGATT 608
              |||||||| |||| |  ||||  ||||  | |    ||||||| ||||||| ||||||||
Db        318 ATGATCTAATGGTTATAGACTGGGGGGTGAGCGGACCAGCAAGCCCTGTCTGTTCAGATT 377

Qy        609 CTTCTTGATCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACT 668
              |||||||  | |||         || |||| |||  |||||| || ||||||||||||||
Db        378 CTTCTTGGACGCTC--------TGCATAGTATTTATTCCTCTTGGGTATGGGGCAGGACT 429

Qy        669 CTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATGGC 728
              | |||  |  |||||||||||||||||||| ||| |||| ||||| ||| |||||| |
Db        430 CCTCTAAATGGAGGGTCTTCAAGGGAGAAGGAAGAGAGTGACCTTTGTAGGTTTTAT-GT 488

Qy        729 TTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCTGG 788
              |||||| | || |||||||| |||||||||| |||     ||| ||||||||||||||||
Db        489 TTGCTTTGAGGGAGAGGAGTCCTAGTTTCTAAGACTGGCTTTGAGGAAGAGGAATTCTGG 548

Qy        789 TTTCTGTGACTTGCTTTCATGAAGAAA 815
              ||||| || ||||| | || | |||||
Db        549 TTTCTATGGCTTGCATCCAGGGAGAAA 575


RESULT 15
US-10-027-632-288756
; Sequence 288756, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
```

```
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 288756
;   LENGTH: 576
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-288756

  Query Match             3.8%;  Score 233.4;  DB 13;  Length 576;
  Best Local Similarity   75.9%;  Pred. No. 1.7e-46;
  Matches  385;  Conservative    1;  Mismatches   97;  Indels   24;  Gaps    7;

Qy        313 TAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTCC 372
              || ||||| ||| ||||||||| ||| |||  ||||||| | ||      | || |||||
Db         89 TACGGGTGGAAAGGTGTGATACCTTTCCTCAGCCATCATTAAGG-----GCTAACACTCC 143

Qy        373 TATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTA 432
              ||||  |||| ||| ||||| ||||| ||||| ||||| |||||||||||| |||||||
Db        144 TATAACAAAA-ACAAGTTAACAAGAGAAAAACATAACACATTTATTTAATCAGAGTTTTA 202

Qy        433 CATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTTT 492
              |||| ||  |||| |||||||||||||||||||||||||||||| |||||      ||||
Db        203 CATGGCACAGGAGGCTTCAGAAATGAAGACCCAAAGACCCAGGGAAAACT-----ATTTT 257

Qy        493 TTTGCTGAGGTTCGATGAAGAATGGATAGC-ATGTAGCCATGTAGATTAGACAAAAGGAT 551
              | |||| || ||  ||| ||||| |  ||| ||| ||  ||||  |     ||| || |
Db        258 TATGCTCAGATTTAATGCAGAATAGTCAGCTATGAAGGAATGTGATTGGACAAAATGGGT 317

Qy        552 ATGATCTAGTGGTAAAGGACTCAGGGGGAAAC---ACAGCAAGGCCTGTCTATTCAGATT 608
              |||||||| |||| |  ||||  |||| | |    ||||||| ||||||| ||||||||
Db        318 ATGATCTAATGGTTATAGACTGGGGGGTGAGCGGACCAGCAAGCCCTGTCTGTTCAGATT 377

Qy        609 CTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACT 668
              |||||||  | |||         || |||| |||  |||||| || |||||||||:|||||
Db        378 CTTCTTGGACGCTC--------TGCATAGTATTTATTCCTCTTGGGTATGGGGCMGGACT 429

Qy        669 CTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATGGC 728
              | |||  |  |||||||||||||||||||||| ||| |||| ||||| ||| |||||| |
Db        430 CCTCTAAATGGAGGGTCTTCAAGGGAGAAGGAAGAGAGTGACCTTTGTAGGTTTTAT-GT 488

Qy        729 TTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCTGG 788
              |||||| | || |||||||| |||||||||| |||    ||| ||||||||||||||||
Db        489 TTGCTTTGAGGGAGAGGAGTCCTAGTTTCTAAGACTGGCTTTGAGGAAGAGGAATTCTGG 548

Qy        789 TTTCTGTGACTTGCTTTCATGAAGAAA 815
              ||||| || ||||| | || | |||||
Db        549 TTTCTATGGCTTGCATCCAGGGAGAAA 575


RESULT 16
US-10-027-632-288757
; Sequence 288757, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 288757
;   LENGTH: 576
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-288757

  Query Match             3.8%;  Score 233.4;  DB 13;  Length 576;
  Best Local Similarity   75.9%;  Pred. No. 1.7e-46;
  Matches  385;  Conservative    1;  Mismatches   97;  Indels   24;  Gaps    7;

Qy        313 TAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTCC 372
              || ||||| ||| ||||||||| ||| |||  ||||||| | ||      | || |||||
Db         89 TACGGGTGGAAAGGTGTGATACCTTTCCTCAGCCATCATTAAGG-----GCTAACACTCC 143

Qy        373 TATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTA 432
              ||||  |||| ||| ||||| ||||| ||||| ||||| |||||||||||| |||||||
Db        144 TATAACAAAA-ACAAGTTAACAAGAGAAAAACATAACACATTTATTTAATCAGAGTTTTA 202

Qy        433 CATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTTT 492
              |||| ||  |||| |||||||||||||||||||||||||||||| |||||      ||||
Db        203 CATGGCACAGGAGGCTTCAGAAATGAAGACCCAAAGACCCAGGGAAAACT-----ATTTT 257

Qy        493 TTTGCTGAGGTTCGATGAAGAATGGATAGC-ATGTAGCCATGTAGATTAGACAAAAGGAT 551
              | |||| || ||  ||| ||||| |  ||| ||| ||  ||||  |     ||| || |
Db        258 TATGCTCAGATTTAATGCAGAATAGTCAGCTATGAAGGAATGTGATTGGACAAAATGGGT 317

Qy        552 ATGATCTAGTGGTAAAGGACTCAGGGGGAAAC---ACAGCAAGGCCTGTCTATTCAGATT 608
              |||||||| |||| |  ||||  |||| | |    ||||||| ||||||| ||||||||
Db        318 ATGATCTAATGGTTATAGACTGGGGGGTGAGCGGACCAGCAAGCCCTGTCTGTTCAGATT 377

Qy        609 CTTCTTGATCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACT 668
              |||||||  | |||         || |||| |||  |||||| || |||||||||||||||
Db        378 CTTCTTGGACGCTC--------TGCATAGTATTTATTCCTCTTGGGTATGGGGCAGGACT 429

Qy        669 CTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATGGC 728
              | |||  |  |||||:||||||||||||||| ||| |||| ||||| ||| |||||| |
Db        430 CCTCTAAATGGAGGGYCTTCAAGGGAGAAGGAAGAGAGTGACCTTTGTAGGTTTTAT-GT 488

Qy        729 TTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCTGG 788
              |||||| | || |||||||| |||||||||| |||    ||| ||||||||||||||||
Db        489 TTGCTTTGAGGGAGAGGAGTCCTAGTTTCTAAGACTGGCTTTGAGGAAGAGGAATTCTGG 548

Qy        789 TTTCTGTGACTTGCTTTCATGAAGAAA 815
              ||||| || ||||| | || | |||||
Db        549 TTTCTATGGCTTGCATCCAGGGAGAAA 575


RESULT 17
US-10-027-632-288755
; Sequence 288755, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
```

```
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 288755
;   LENGTH: 576
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-288755

  Query Match             3.8%;  Score 233.4;  DB 17;  Length 576;
  Best Local Similarity   75.9%;  Pred. No. 1.7e-46;
  Matches  385;  Conservative    1;  Mismatches   97;  Indels   24;  Gaps    7;

Qy        313 TAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTCC 372
              || ||||| ||| ||||||||| ||| |||  ||||||| | ||      | || |||||
Db         89 TACGGGTGGAAAGGTGTGATACCTTTCCTCAGCCATCATTAAGG-----GCTAACACTCC 143

Qy        373 TATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTA 432
              ||||  |||| ||| ||||| ||||| ||||| ||||| |||||||||||| |||||||
Db        144 TATAACAAAA-ACAAGTTAACAAGAGAAAAACATAACACATTTATTTAATCAGAGTTTTA 202

Qy        433 CATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTTT 492
              |||| ||  |||| |||||||||||||||||||||||||||||: |||||      ||||
Db        203 CATGGCACAGGAGGCTTCAGAAATGAAGACCCAAAGACCCAGGRAAAACT-----ATTTT 257

Qy        493 TTTGCTGAGGTTCGATGAAGAATGGATAGC-ATGTAGCCATGTAGATTAGACAAAAGGAT 551
              | |||| || ||  ||| ||||| |  ||| ||| ||  |||| |      ||| || |
Db        258 TATGCTCAGATTTAATGCAGAATAGTCAGCTATGAAGGAATGTGATTGGACAAAATGGGT 317

Qy        552 ATGATCTAGTGGTAAAGGACTCAGGGGGAAAC---ACAGCAAGGCCTGTCTATTCAGATT 608
              |||||||| |||| |  ||||  ||||  | |    ||||||| ||||||| ||||||||
Db        318 ATGATCTAATGGTTATAGACTGGGGGGTGAGCGGACCAGCAAGCCCTGTCTGTTCAGATT 377

Qy        609 CTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACT 668
              |||||||  | |||        || |||| |||  |||||| || ||||||||||||||||
Db        378 CTTCTTGGACGCTC--------TGCATAGTATTTATTCCTCTTGGGTATGGGGCAGGACT 429

Qy        669 CTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATGGC 728
              | |||  |  |||||||||||||||||||| ||| |||| ||||| ||| |||||| |
Db        430 CCTCTAAATGGAGGGTCTTCAAGGGAGAAGGAAGAGAGTGACCTTTGTAGGTTTTAT-GT 488

Qy        729 TTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCTGG 788
              |||||| | || |||||||| |||||||||  |||     ||| ||||||||||||||||
Db        489 TTGCTTTGAGGGAGAGGAGTCCTAGTTTCTAAGACTGGCTTTGAGGAAGAGGAATTCTGG 548

Qy        789 TTTCTGTGACTTGCTTTCATGAAGAAA 815
              ||||| || ||||| | || | |||||
Db        549 TTTCTATGGCTTGCATCCAGGGAGAAA 575


RESULT 18
US-10-027-632-288756
; Sequence 288756, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 288756
;   LENGTH: 576
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-288756

  Query Match             3.8%;  Score 233.4;  DB 17;  Length 576;
  Best Local Similarity   75.9%;  Pred. No. 1.7e-46;
  Matches  385;  Conservative    1;  Mismatches   97;  Indels   24;  Gaps    7;

Qy        313 TAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTCC 372
              || ||||| ||| ||||||||| ||| |||  ||||||| | ||      | || |||||
Db         89 TACGGGTGGAAAGGTGTGATACCTTTCCTCAGCCATCATTAAGG-----GCTAACACTCC 143

Qy        373 TATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTA 432
              ||||  |||| ||| ||||| ||||| ||||| ||||| |||||||||||| |||||||
Db        144 TATAACAAAA-ACAAGTTAACAAGAGAAAAACATAACACATTTATTTAATCAGAGTTTTA 202

Qy        433 CATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTTT 492
              |||| ||  |||| |||||||||||||||||||||||||||||| |||||      ||||
Db        203 CATGGCACAGGAGGCTTCAGAAATGAAGACCCAAAGACCCAGGGAAAACT-----ATTTT 257

Qy        493 TTTGCTGAGGTTCGATGAAGAATGGATAGC-ATGTAGCCATGTAGATTAGACAAAAGGAT 551
              | |||| || ||  ||| ||||| |  ||| ||| ||  |||| |      ||| || |
Db        258 TATGCTCAGATTTAATGCAGAATAGTCAGCTATGAAGGAATGTGATTGGACAAAATGGGT 317

Qy        552 ATGATCTAGTGGTAAAGGACTCAGGGGGAAAC---ACAGCAAGGCCTGTCTATTCAGATT 608
              |||||||| |||| |  ||||  ||||  | |    ||||||| ||||||| ||||||||
Db        318 ATGATCTAATGGTTATAGACTGGGGGGTGAGCGGACCAGCAAGCCCTGTCTGTTCAGATT 377

Qy        609 CTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACT 668
              |||||||  | |||        || |||| |||  |||||| || |||||||||:|||||
Db        378 CTTCTTGGACGCTC--------TGCATAGTATTTATTCCTCTTGGGTATGGGGCMGGACT 429

Qy        669 CTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATGGC 728
              | |||  |  |||||||||||||||||||| ||| |||| ||||| ||| |||||| |
Db        430 CCTCTAAATGGAGGGTCTTCAAGGGAGAAGGAAGAGAGTGACCTTTGTAGGTTTTAT-GT 488

Qy        729 TTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCTGG 788
              |||||| | || |||||||| |||||||||  |||     ||| ||||||||||||||||
Db        489 TTGCTTTGAGGGAGAGGAGTCCTAGTTTCTAAGACTGGCTTTGAGGAAGAGGAATTCTGG 548

Qy        789 TTTCTGTGACTTGCTTTCATGAAGAAA 815
              ||||| || ||||| | || | |||||
Db        549 TTTCTATGGCTTGCATCCAGGGAGAAA 575


RESULT 19
US-10-027-632-288757
; Sequence 288757, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
```

```
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 288757
;   LENGTH: 576
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-288757

  Query Match             3.8%;  Score 233.4;  DB 17;  Length 576;
  Best Local Similarity   75.9%;  Pred. No. 1.7e-46;
  Matches  385;  Conservative    1;  Mismatches   97;  Indels   24;  Gaps    7;

Qy        313 TAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTCC 372
              || ||||| ||| ||||||||| ||| |||  ||||||| | ||      | || |||||
Db         89 TACGGGTGGAAAGGTGTGATACCTTTCCTCAGCCATCATTAAGG-----GCTAACACTCC 143

Qy        373 TATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTA 432
              ||||  |||| ||| ||||| ||||| ||||| ||||| |||||||||||||| |||||||
Db        144 TATAACAAAA-ACAAGTTAACAAGAGAAAAACATAACACATTTATTTAATCAGAGTTTTA 202

Qy        433 CATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTTT 492
              |||| ||  |||| |||||||||||||||||||||||||||||| |||||      ||||
Db        203 CATGGCACAGGAGGCTTCAGAAATGAAGACCCAAAGACCCAGGGAAAACT-----ATTTT 257

Qy        493 TTTGCTGAGGTTCGATGAAGAATGGATAGC-ATGTAGCCATGTAGATTAGACAAAAGGAT 551
              | |||| || ||  ||| ||||| |  ||| ||| ||  |||| |     ||| || |
Db        258 TATGCTCAGATTTAATGCAGAATAGTCAGCTATGAAGGAATGTGATTGGACAAAATGGGT 317

Qy        552 ATGATCTAGTGGTAAAGGACTCAGGGGGAAAC---ACAGCAAGGCCTGTCTATTCAGATT 608
              |||||||| |||| |  ||||  ||||  | |    ||||||| ||||||| ||||||||
Db        318 ATGATCTAATGGTTATAGACTGGGGGGTGAGCGGACCAGCAAGCCCTGTCTGTTCAGATT 377

Qy        609 CTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACT 668
              |||||||  | |||          || |||| |||  |||||| || ||||||||||||||||
Db        378 CTTCTTGGACGCTC--------TGCATAGTATTTATTCCTCTTGGGTATGGGGCAGGACT 429

Qy        669 CTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATGGC 728
              | |||  |  |||||:||||||||||||||| ||| |||| ||||| ||| |||||| |
Db        430 CCTCTAAATGGAGGGYCTTCAAGGGAGAAGGAAGAGAGTGACCTTTGTAGGTTTTAT-GT 488

Qy        729 TTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCTGG 788
              |||||| | || |||||||| |||||||||| |||     ||| ||||||||||||||||
Db        489 TTGCTTTGAGGGAGAGGAGTCCTAGTTTCTAAGACTGGCTTTGAGGAAGAGGAATTCTGG 548

Qy        789 TTTCTGTGACTTGCTTTCATGAAGAAA 815
              ||||| || ||||| | || | |||||
Db        549 TTTCTATGGCTTGCATCCAGGGAGAAA 575


RESULT 20
US-10-027-632-233469/c
; Sequence 233469, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 233469
;   LENGTH: 547
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-233469

  Query Match             3.5%;  Score 213;  DB 13;  Length 547;
  Best Local Similarity   73.2%;  Pred. No. 1.8e-41;
  Matches  358;  Conservative    1;  Mismatches  111;  Indels   19;  Gaps    6;

Qy        374 ATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTAC 433
              |||  |||||||  |||| |||||  |||||  | ||||||||||||||| ||||||||
Db        547 ATAACAAAAGACAYATTAACAAGAGAAAAACAGAGCAAATTTATTTAATCCAAGTTTTAT 488

Qy        434 ATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTTTT 493
              |||||||||| ||||||||||||  |||||||| ||||||| |||||||  |  |||||
Db        487 GTGACATGGGAGGCTTCAGAAATGACAACCCAAAGGCCCAGGGAAAACTGT-TAATTTTT 429

Qy        494 TTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGATAT 553
              |||| || || |||||||||||| |||  |    |    |||| |||||| || |||
Db        428 ATGCTTAGATTAGATGAAGAATGGACAGCCAGATAGAAATGTGATTGGACAAACGGGTAT 369

Qy        554 GATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTTCT 613
              || ||| ||| ||  |||| |  ||| | |      ||   ||||||  |||||||||
Db        368 GACCTACTGGGAACAGACTGAAAGGGGACCCAGCAAA--ACTGTCTGCTCAGATTCT--- 314

Qy        614 TGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTC-TTC 672
                 ||||   |||| ||| |||| | | |||||||||  |||||||||:|||| | |||
Db        313 -----TCTCAGCCTCTCTGTGTAGCCTCCATTCCTCCTGGGTATGGGGCRGGACCCTTTC 259

Qy        673 TTCAATGAGGGTCTTCAAGGGAGAAGG------GAGAAAGTGGCCTTTTTAGATTTTATG 726
              |  |||||| || ||||| ||||||||       ||| |||| ||||| ||| |||||||
Db        258 TGAAATGAGAGTTTTCAAAGGAGAAGGAGAAGATAGAGAGTGACCTTTCTAGGTTTTATG 199

Qy        727 GCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCT 786
              | |||||   | | ||||  |||||| |||||||||||||||||| ||||||  ||||||
Db        198 GTTTGCTT-TAGCAGGAGGGTTTCTAGGTTCTATGACCCATCTTGAGGAAGATAAATTCT 140

Qy        787 GGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGG 846
              |||||||||||||  ||||   | |  |||   || |  ||  |||     |||||| ||
Db        139 GGTTTCTGTGACTCACTTTGGAGGATGAAGGCAAGCAGAAGAGAGGGTMTGAGGAGAAGG 80

Qy        847 TCAGAAAGA 855
              |  ||||||
Db         79 TTGGAAAGA 71


RESULT 21
US-10-027-632-233470/c
; Sequence 233470, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
```

```
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 233470
;   LENGTH: 547
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-233470

  Query Match             3.5%;  Score 213;  DB 13;  Length 547;
  Best Local Similarity   73.2%;  Pred. No. 1.8e-41;
  Matches  358;  Conservative    1;  Mismatches  111;  Indels   19;  Gaps    6;

Qy        374 ATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTAC 433
              |||  ||||||||  ||||  |||||  |||||  | ||||||||||||||| ||||||||
Db        547 ATAACAAAAGACAYATTAACAAGAGAAAAACAGAGCAAATTTATTTAATCCAAGTTTTAT 488

Qy        434 ATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTTTT 493
              |||||||||||  ||||||||||||  ||||||||  |||||||  |||||||  |  |||||
Db        487 GTGACATGGGAGGCTTCAGAAATGACAACCCAAAGGCCCAGGGAAAACTGT-TAATTTTT 429

Qy        494 TTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGATAT 553
              ||||  ||  ||  ||||||||||||||  |||  |      |    ||||  ||||||  ||  |||
Db        428 ATGCTTAGATTAGATGAAGAATGGACAGCCAGATAGAAATGTGATTGGACAAACGGGTAT 369

Qy        554 GATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTTCT 613
              ||  |||  |||  ||  ||||  |  |||  | |      ||   ||||||  |||||||||
Db        368 GACCTACTGGGAACAGACTGAAAGGGGACCCAGCAAA--ACTGTCTGCTCAGATTCT--- 314

Qy        614 TGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTC-TTC 672
                  ||||   |||| |||  |||| | | |||||||||  |||||||||:|||| | |||
Db        313 -----TCTCAGCCTCTCTGTGTAGCCTCCATTCCTCCTGGGTATGGGGCRGGACCCTTTC 259

Qy        673 TTCAATGAGGGTCTTCAAGGGAGAAGG------GAGAAAGTGGCCTTTTTAGATTTTATG 726
              |  |||||| ||  |||||  ||||||||       |||  ||||  |||||  |||  |||||||
Db        258 TGAAATGAGAGTTTTCAAAGGAGAAGGAGAAGATAGAGAGTGACCTTTCTAGGTTTTATG 199

Qy        727 GCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCT 786
              | ||||||   | | ||||  ||||||  ||||||||||||||||||  ||||||  ||||||
Db        198 GTTTGCTT-TAGCAGGAGGGTTTCTAGGTTCTATGACCCATCTTGAGGAAGATAAATTCT 140

Qy        787 GGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGG 846
              ||||||||||||  ||||   | |  |||   || |  ||  |||     ||||||  ||
Db        139 GGTTTCTGTGACTCACTTTGGAGGATGAAGGCAAGCAGAAGAGAGGGTMTGAGGAGAAGG 80

Qy        847 TCAGAAAGA 855
              |  ||||||
Db         79 TTGGAAAGA 71


RESULT 22
US-10-027-632-233471/c
; Sequence 233471, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 233471
;   LENGTH: 547
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-233471

  Query Match             3.5%;  Score 213;  DB 13;  Length 547;
  Best Local Similarity   73.2%;  Pred. No. 1.8e-41;
  Matches  358;  Conservative    1;  Mismatches  111;  Indels   19;  Gaps    6;

Qy        374 ATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTAC 433
              |||  ||||||||  ||||  |||||  |||||  | ||||||||||||||| ||||||||
Db        547 ATAACAAAAGACAYATTAACAAGAGAAAAACAGAGCAAATTTATTTAATCCAAGTTTTAT 488

Qy        434 ATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTTTT 493
              |||||||||||  ||||||||||||  ||||||||  |||||||  |||||||  |  |||||
Db        487 GTGACATGGGAGGCTTCAGAAATGACAACCCAAAGGCCCAGGGAAAACTGT-TAATTTTT 429

Qy        494 TTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGATAT 553
              ||||  ||  ||  ||||||||||||||  |||  |      |    ||||  ||||||  ||  |||
Db        428 ATGCTTAGATTAGATGAAGAATGGACAGCCAGATAGAAATGTGATTGGACAAACGGGTAT 369

Qy        554 GATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTTCT 613
              ||  |||  |||  ||  ||||  |  |||  | |      ||   ||||||  |||||||||
Db        368 GACCTACTGGGAACAGACTGAAAGGGGACCCAGCAAA--ACTGTCTGCTCAGATTCT--- 314

Qy        614 TGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTC-TTC 672
                  ||||   |||| |||  |||| | | |||||||||  |||||||||:|||| | |||
Db        313 -----TCTCAGCCTCTCTGTGTAGCCTCCATTCCTCCTGGGTATGGGGCRGGACCCTTTC 259

Qy        673 TTCAATGAGGGTCTTCAAGGGAGAAGG------GAGAAAGTGGCCTTTTTAGATTTTATG 726
              |  |||||| ||  |||||  ||||||||       |||  ||||  |||||  |||  |||||||
Db        258 TGAAATGAGAGTTTTCAAAGGAGAAGGAGAAGATAGAGAGTGACCTTTCTAGGTTTTATG 199

Qy        727 GCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCT 786
              | ||||||   | | ||||  ||||||  ||||||||||||||||||  ||||||  ||||||
Db        198 GTTTGCTT-TAGCAGGAGGGTTTCTAGGTTCTATGACCCATCTTGAGGAAGATAAATTCT 140

Qy        787 GGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGG 846
              ||||||||||||  ||||   | |  |||   || |  ||  |||     ||||||  ||
Db        139 GGTTTCTGTGACTCACTTTGGAGGATGAAGGCAAGCAGAAGAGAGGGTMTGAGGAGAAGG 80

Qy        847 TCAGAAAGA 855
              |  ||||||
Db         79 TTGGAAAGA 71


RESULT 23
US-10-027-632-233469/c
; Sequence 233469, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
```

```
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 233469
;   LENGTH: 547
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-233469

  Query Match             3.5%;  Score 213;  DB 17;  Length 547;
  Best Local Similarity   73.2%;  Pred. No. 1.8e-41;
  Matches  358;  Conservative    1;  Mismatches  111;  Indels   19;  Gaps    6;

Qy        374 ATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTAC 433
              |||  ||||||||  ||||  |||||  |||||  | ||||||||||||||| ||||||||
Db        547 ATAACAAAAGACAYATTAACAAGAGAAAAACAGAGCAAATTTATTTAATCCAAGTTTTAT 488

Qy        434 ATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTTTT 493
              ||||||||||| ||||||||||||  |||||||| |||||||  |||||||  |  |||||
Db        487 GTGACATGGGAGGCTTCAGAAATGACAACCCAAAGGCCCAGGGAAAACTGT-TAATTTTT 429

Qy        494 TTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGATAT 553
              ||||  ||  || ||||||||||||| |||  |     |    ||||  ||||||  || |||
Db        428 ATGCTTAGATTAGATGAAGAATGGACAGCCAGATAGAAATGTGATTGGACAAACGGGTAT 369

Qy        554 GATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTTCT 613
              ||  |||  |||  ||  ||||  |  |||  | |      ||   ||||||  |||||||||
Db        368 GACCTACTGGGAACAGACTGAAAGGGGACCCAGCAAA--ACTGTCTGCTCAGATTCT--- 314

Qy        614 TGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTC-TTC 672
                  ||||   |||| |||  |||| | |  |||||||| |||||||||:|||| | |||
Db        313 -----TCTCAGCCTCTCTGTGTAGCCTCCATTCCTCCTGGGTATGGGGCRGGACCCTTTC 259

Qy        673 TTCAATGAGGGTCTTCAAGGGAGAAGG------GAGAAAGTGGCCTTTTTAGATTTTATG 726
              |  |||||| || ||||| ||||||||       ||| |||| ||||| ||| |||||||
Db        258 TGAAATGAGAGTTTTCAAAGGAGAAGGAGAAGATAGAGAGTGACCTTTCTAGGTTTTATG 199

Qy        727 GCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCT 786
              | ||||||   | | ||||  |||||| |||||||||||||||||| ||||||  ||||||
Db        198 GTTTGCTT-TAGCAGGAGGGTTTCTAGGTTCTATGACCCATCTTGAGGAAGATAAATTCT 140

Qy        787 GGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGG 846
              |||||||||||||  ||||   | |  |||   || |  ||  |||     |||||| ||
Db        139 GGTTTCTGTGACTCACTTTGGAGGATGAAGGCAAGCAGAAGAGAGGGTMTGAGGAGAAGG 80

Qy        847 TCAGAAAGA 855
              |  ||||||
Db         79 TTGGAAAGA 71


RESULT 24
US-10-027-632-233470/c
; Sequence 233470, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 233470
;   LENGTH: 547
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-233470

  Query Match             3.5%;  Score 213;  DB 17;  Length 547;
  Best Local Similarity   73.2%;  Pred. No. 1.8e-41;
  Matches  358;  Conservative    1;  Mismatches  111;  Indels   19;  Gaps    6;

Qy        374 ATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTAC 433
              |||  ||||||||  ||||  |||||  |||||  | ||||||||||||||| ||||||||
Db        547 ATAACAAAAGACAYATTAACAAGAGAAAAACAGAGCAAATTTATTTAATCCAAGTTTTAT 488

Qy        434 ATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTTTT 493
              ||||||||||| ||||||||||||  |||||||| |||||||  |||||||  |  |||||
Db        487 GTGACATGGGAGGCTTCAGAAATGACAACCCAAAGGCCCAGGGAAAACTGT-TAATTTTT 429

Qy        494 TTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGATAT 553
              ||||  ||  || ||||||||||||| |||  |     |    ||||  ||||||  || |||
Db        428 ATGCTTAGATTAGATGAAGAATGGACAGCCAGATAGAAATGTGATTGGACAAACGGGTAT 369

Qy        554 GATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTTCT 613
              ||  |||  |||  ||  ||||  |  |||  | |      ||   ||||||  |||||||||
Db        368 GACCTACTGGGAACAGACTGAAAGGGGACCCAGCAAA--ACTGTCTGCTCAGATTCT--- 314

Qy        614 TGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTC-TTC 672
                  ||||   |||| |||  |||| | |  |||||||| |||||||||:|||| | |||
Db        313 -----TCTCAGCCTCTCTGTGTAGCCTCCATTCCTCCTGGGTATGGGGCRGGACCCTTTC 259

Qy        673 TTCAATGAGGGTCTTCAAGGGAGAAGG------GAGAAAGTGGCCTTTTTAGATTTTATG 726
              |  |||||| || ||||| ||||||||       ||| |||| ||||| ||| |||||||
Db        258 TGAAATGAGAGTTTTCAAAGGAGAAGGAGAAGATAGAGAGTGACCTTTCTAGGTTTTATG 199

Qy        727 GCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCT 786
              | ||||||   | | ||||  |||||| |||||||||||||||||| ||||||  ||||||
Db        198 GTTTGCTT-TAGCAGGAGGGTTTCTAGGTTCTATGACCCATCTTGAGGAAGATAAATTCT 140

Qy        787 GGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGG 846
              |||||||||||||  ||||   | |  |||   || |  ||  |||     |||||| ||
Db        139 GGTTTCTGTGACTCACTTTGGAGGATGAAGGCAAGCAGAAGAGAGGGTMTGAGGAGAAGG 80

Qy        847 TCAGAAAGA 855
              |  ||||||
Db         79 TTGGAAAGA 71


RESULT 25
US-10-027-632-233471/c
; Sequence 233471, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
```

```
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 233471
;   LENGTH: 547
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-233471

  Query Match             3.5%;  Score 213;  DB 17;  Length 547;
  Best Local Similarity   73.2%;  Pred. No. 1.8e-41;
  Matches  358;  Conservative    1;  Mismatches  111;  Indels   19;  Gaps    6;

Qy        374 ATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTAC 433
              |||  ||||||||  ||||  |||||  |||||   | ||||||||||||||| ||||||||
Db        547 ATAACAAAAGACAYATTAACAAGAGAAAAACAGAGCAAATTTATTTAATCCAAGTTTTAT 488

Qy        434 ATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTTTT 493
              ||||||||||| ||||||||||||  |||||||| |||||| ||||||| |   |||||
Db        487 GTGACATGGGAGGCTTCAGAAATGACAACCCAAAGGCCCAGGGAAAACTGT-TAATTTTT 429

Qy        494 TTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGATAT 553
              ||||  || || |||||||||||||  |||  |     |    |||| ||||||  || |||
Db        428 ATGCTTAGATTAGATGAAGAATGGACAGCCAGATAGAAATGTGATTGGACAAACGGGTAT 369

Qy        554 GATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTTCT 613
              || ||| ||| ||  |||| |  ||| | |      ||   ||||||  |||||||||
Db        368 GACCTACTGGGAACAGACTGAAAGGGGACCCAGCAAA--ACTGTCTGCTCAGATTCT--- 314

Qy        614 TGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTC-TTC 672
              ||||    |||| ||| |||| | | |||||||||  |||||||||:|||| | |||
Db        313 -----TCTCAGCCTCTCTGTGTAGCCTCCATTCCTCCTGGGTATGGGGCRGGACCCTTTC 259

Qy        673 TTCAATGAGGGTCTTCAAGGGAGAAGG------GAGAAAGTGGCCTTTTTAGATTTATG 726
              |  |||||| || ||||| ||||||||       ||| |||| ||||| ||| |||||||
Db        258 TGAAATGAGAGTTTTCAAAGGAGAAGGAGAAGATAGAGAGTGACCTTTCTAGGTTTTATG 199

Qy        727 GCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCT 786
              | ||||||   | | ||||  |||||| ||||||||||||||||||| ||||||  ||||||
Db        198 GTTTGCTT-TAGCAGGAGGGTTTCTAGGTTCTATGACCCATCTTGAGGAAGATAAATTCT 140

Qy        787 GGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGG 846
              ||||||||||||  ||||   | |  |||   || |  ||  |||     ||||| ||
Db        139 GGTTTCTGTGACTCACTTTGGAGGATGAAGGCAAGCAGAAGAGAGGGTMTGAGGAGAAGG 80

Qy        847 TCAGAAAGA 855
              |  ||||||
Db         79 TTGGAAAGA 71


RESULT 26
US-10-027-632-181211
; Sequence 181211, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 181211
;   LENGTH: 507
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-181211

  Query Match             3.4%;  Score 208.8;  DB 13;  Length 507;
  Best Local Similarity   70.6%;  Pred. No. 1.8e-40;
  Matches  339;  Conservative    2;  Mismatches  119;  Indels   20;  Gaps    4;

Qy        305 AGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACC 364
              || |   |||  ||  ||| ||||||||| ||| |||  | ||||||||||  || ||
Db         40 AGATAGTGTAAAAGTACAAAAGTGTGATACCTTTCCTCACCAATCATAAGGGTAACAGCC 99

Qy        365 AAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCA 424
              || |||||||||  |||||| |||||||  |||||  || | | ||||||||||||||||
Db        100 AAGACTCCTATAACAAAAGATAGGTTAACAAGAGACAAGCATGACAAATTTATTTAATCA 159

Qy        425 AAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGT 484
              || ||| |||||||||  | || |||| |||||||  |||||  | |  ||  : | ||
Db        160 AAATTTCACATGACACAGCAGCCTTCGGAAATGAGGACCCCCAAAACCTAKAGGAA---- 215

Qy        485 CTGTTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACA 544
                | | | | || |||||| || |||||||||||||          |    |||||||||
Db        216 ACCATCTATATACTTAGGTTCAATAAAGAATGGATAGCCACGTAGAAATGTGATTAGACA 275

Qy        545 AAAGGATATGATCTAGTGGTAAAGGACT-CAGGGGGAAACACAGCAAGGCCTGTCTATTC 603
              ||||  ||||| ||| ||||||  || | |  ||||  |  ||||||||||| ||| |||
Db        276 AAAGAGTATGACCTAATGGTAATAGATTGCGTGGGGGGATCCAGCAAGGCCTATCTGTTC 335

Qy        604 AGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCA 663
              |||  ||  :|||  ||||         ||| |||||||| ||| |||| ||||||   ||
Db        336 AGAGACTGGKTGACTTCTC--------TGTGTAGCATTCCTTCTTCCTCAGTATGATACA 387

Qy        664 GGACTCTTCTTCAATGAGGGTCTTCAAGGGAGA-------AGGGAGAAAGTGGCCTTTTT 716
              |||   |||  ||||| |||||||||||||| |       |||||| |||  ||||| |
Db        388 CGACCTCTCTGGAATGAAGGTCTTCAAGGGAAAGGGGAACAGGGAGAGAGTTACCTTTCT 447

Qy        717 AGATTTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAA 776
              || ||| ||||||||||  ||| ||||| ||||||||||  |||||||   |||| |||
Db        448 AGGTTTAATGGCTTGCTTTTGGGGAGAGGGGTTCTAGTTTTCATGACCCTCTTTGGAGAA 507


RESULT 27
US-10-027-632-181212
; Sequence 181212, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
```

```
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 181212
;   LENGTH: 507
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-181212

  Query Match             3.4%;  Score 208.8;  DB 13;  Length 507;
  Best Local Similarity   70.6%;  Pred. No. 1.8e-40;
  Matches  339;  Conservative    2;  Mismatches  119;  Indels   20;  Gaps    4;

Qy        305 AGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACC 364
              || |   |||   ||  ||| ||||||||| ||| |||  | ||||||||||  || ||
Db         40 AGATAGTGTAAAAGTACAAAAGTGTGATACCTTTCCTCACCAATCATAAGGGTAACAGCC 99

Qy        365 AAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCA 424
              || |||||||||  |||||| ||||||| |||||  || | | |||||||||||||||||
Db        100 AAGACTCCTATAACAAAAGATAGGTTAACAAGAGACAAGCATGACAAATTTATTTAATCA 159

Qy        425 AAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGT 484
              || ||| ||||||||  | || |||| ||||||| |||||  | | ||  : | ||
Db        160 AAATTTCACATGACACAGCAGCCTTCGGAAATGAGGACCCCCAAAACCTAKAGGAA---- 215

Qy        485 CTGTTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACA 544
                | | | | || |||||| || ||||||||||||||        |    |||||||||
Db        216 ACCATCTATATACTTAGGTTCAATAAAGAATGGATAGCCACGTAGAAATGTGATTAGACA 275

Qy        545 AAAGGATATGATCTAGTGGTAAAGGACT-CAGGGGGAAACACAGCAAGGCCTGTCTATTC 603
              ||||  ||||| ||| |||||||  || | |  ||||  |  ||||||||||| ||| |||
Db        276 AAAGAGTATGACCTAATGGTAATAGATTGCGTGGGGGGATCCAGCAAGGCCTATCTGTTC 335

Qy        604 AGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCA 663
              |||  ||  :|||  ||||        ||| |||||||| ||| |||| ||||||   ||
Db        336 AGAGACTGGKTGACTTCTC--------TGTGTAGCATTCCTTCTTCCTCAGTATGATACA 387

Qy        664 GGACTCTTCTTCAATGAGGGTCTTCAAGGGAGA-------AGGGAGAAAGTGGCCTTTTT 716
               |||   |||  ||||| |||||||||||||  |       |||||| |||  ||||| |
Db        388 CGACCTCTCTGGAATGAAGGTCTTCAAGGGAAAGGGGAACAGGGAGAGAGTTACCTTTCT 447

Qy        717 AGATTTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAA 776
              || ||| ||||||||||  ||| ||||| ||||||||||  ||||||   |||| |||
Db        448 AGGTTTAATGGCTTGCTTTTGGGGAGAGGGGTTCTAGTTTTCATGACCCTCTTTGGAGAA 507


RESULT 28
US-10-027-632-181211
; Sequence 181211, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 181211
;   LENGTH: 507
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-181211

  Query Match             3.4%;  Score 208.8;  DB 17;  Length 507;
  Best Local Similarity   70.6%;  Pred. No. 1.8e-40;
  Matches  339;  Conservative    2;  Mismatches  119;  Indels   20;  Gaps    4;

Qy        305 AGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACC 364
              || |   |||   ||  ||| ||||||||| ||| |||  | ||||||||||  || ||
Db         40 AGATAGTGTAAAAGTACAAAAGTGTGATACCTTTCCTCACCAATCATAAGGGTAACAGCC 99

Qy        365 AAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCA 424
              || |||||||||  |||||| ||||||| |||||  || | | |||||||||||||||||
Db        100 AAGACTCCTATAACAAAAGATAGGTTAACAAGAGACAAGCATGACAAATTTATTTAATCA 159

Qy        425 AAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGT 484
              || ||| ||||||||  | || |||| ||||||| |||||  | | ||  : | ||
Db        160 AAATTTCACATGACACAGCAGCCTTCGGAAATGAGGACCCCCAAAACCTAKAGGAA---- 215

Qy        485 CTGTTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACA 544
                | | | | || |||||| || ||||||||||||||        |    |||||||||
Db        216 ACCATCTATATACTTAGGTTCAATAAAGAATGGATAGCCACGTAGAAATGTGATTAGACA 275

Qy        545 AAAGGATATGATCTAGTGGTAAAGGACT-CAGGGGGAAACACAGCAAGGCCTGTCTATTC 603
              ||||  ||||| ||| |||||||  || | |  ||||  |  ||||||||||| ||| |||
Db        276 AAAGAGTATGACCTAATGGTAATAGATTGCGTGGGGGGATCCAGCAAGGCCTATCTGTTC 335

Qy        604 AGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCA 663
              |||  ||  :|||  ||||        ||| |||||||| ||| |||| ||||||   ||
Db        336 AGAGACTGGKTGACTTCTC--------TGTGTAGCATTCCTTCTTCCTCAGTATGATACA 387

Qy        664 GGACTCTTCTTCAATGAGGGTCTTCAAGGGAGA-------AGGGAGAAAGTGGCCTTTTT 716
               |||   |||  ||||| |||||||||||||  |       |||||| |||  ||||| |
Db        388 CGACCTCTCTGGAATGAAGGTCTTCAAGGGAAAGGGGAACAGGGAGAGAGTTACCTTTCT 447

Qy        717 AGATTTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAA 776
              || ||| ||||||||||  ||| ||||| ||||||||||  ||||||   |||| |||
Db        448 AGGTTTAATGGCTTGCTTTTGGGGAGAGGGGTTCTAGTTTTCATGACCCTCTTTGGAGAA 507


RESULT 29
US-10-027-632-181212
; Sequence 181212, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
```

```
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 181212
;   LENGTH: 507
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-181212

  Query Match             3.4%;  Score 208.8;  DB 17;  Length 507;
  Best Local Similarity   70.6%;  Pred. No. 1.8e-40;
  Matches  339;  Conservative    2;  Mismatches  119;  Indels   20;  Gaps    4;

Qy        305 AGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACC 364
              ||  |   |||   ||  ||| ||||||||| ||| |||  | ||||||||||  || ||
Db         40 AGATAGTGTAAAAGTACAAAAGTGTGATACCTTTCCTCACCAATCATAAGGGTAACAGCC 99

Qy        365 AAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCA 424
              || |||||||||  ||||||| ||||||| |||||  || | | ||||||||||||||||
Db        100 AAGACTCCTATAACAAAAGATAGGTTAACAAGAGACAAGCATGACAAATTTATTTAATCA 159

Qy        425 AAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGT 484
              || ||| ||||||||  | || |||| ||||||| |||||  | | ||  : | ||
Db        160 AAATTTCACATGACACAGCAGCCTTCGGAAATGAGGACCCCCAAAACCTAKAGGAA---- 215

Qy        485 CTGTTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACA 544
                | | | | || |||||| || ||||||||||||          |    |||||||||
Db        216 ACCATCTATATACTTAGGTTCAATAAAGAATGGATAGCCACGTAGAAATGTGATTAGACA 275

Qy        545 AAAGGATATGATCTAGTGGTAAAGGACT-CAGGGGGAAACACAGCAAGGCCTGTCTATTC 603
              ||||  ||||| ||| ||||||  || | |  ||||  |  |||||||||| ||| |||
Db        276 AAAGAGTATGACCTAATGGTAATAGATTGCGTGGGGGGATCCAGCAAGGCCTATCTGTTC 335

Qy        604 AGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCA 663
              |||  ||  :|||  ||||         ||| |||||||| ||| |||| ||||||   ||
Db        336 AGAGACTGGKTGACTTCTC--------TGTGTAGCATTCCTTCTTCCTCAGTATGATACA 387

Qy        664 GGACTCTTCTTCAATGAGGGTCTTCAAGGGAGA-------AGGGAGAAAGTGGCCTTTTT 716
                |||   |||  ||||| |||||||||||||  |       ||||||| |||  ||||| |
Db        388 CGACCTCTCTGGAATGAAGGTCTTCAAGGGAAAGGGGAACAGGGAGAGAGTTACCTTTCT 447

Qy        717 AGATTTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAA 776
              || ||| ||||||||||  ||| ||||| ||||||||||  |||||||   |||| |||
Db        448 AGGTTTAATGGCTTGCTTTTGGGGAGAGGGGTTCTAGTTTTCATGACCCTCTTTGGAGAA 507


RESULT 30
US-10-312-841-1
; Sequence 1, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
;  APPLICANT: Epigenomics AG
;  TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
;  FILE REFERENCE: E01/1208/WO
;  CURRENT APPLICATION NUMBER: US/10/312,841
;  CURRENT FILING DATE:  2002-12-30
;  NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 1
;   LENGTH: 3673778
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (3294164)
US-10-312-841-1

  Query Match             3.3%;  Score 202.8;  DB 16;  Length 3673778;
  Best Local Similarity   61.9%;  Pred. No. 1.2e-36;
  Matches  400;  Conservative    0;  Mismatches  222;  Indels   24;  Gaps    4;

Qy        215 AGAGTAAATTCTGGTCTTCAGAGTAACAAGAAACACCCTTTAGTTCTCAGCATATTCGTG 274
              ||||| || | | || |  | | ||| || ||  |   ||   ||  |     | || ||
Db    2555152 AGAGTGAAATTTTGTTTAAAAAATAATAATAATAATTTTTGTTTTTTTTTTTTTTTAGTT 2555211

Qy        275 CACCTTCATTTATCTCTCCTTCTCTCTCAAAGCTGCAGTAGGGGTGAAAACGTGTGATAC 334
                  | |||  | | |  ||   | | | |  |  | ||||||  ||| |||||  ||
Db    2555212 TTTGGTAATTATTATTTTATT---TTTTATATTTATTGAAGGGGTAGAAAAGTGTGGTAT 2555268

Qy        335 ATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTCCTATAGTAAAAGACAGGTTAATA 394
              ||||  |  |  || ||||| ||        || | |  | || ||||||| |||||||||
Db    2555269 TTTTTTTTATTTATTATAAGAGTTATCGTTAATATTTTTGTAATAAAAGATAGGTTAATA 2555328

Qy        395 AGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTACATGACATGGGAGTCTTCAGAA 454
              |||| |||   ||| ||||||||||||| |||||||||  ||| ||| ||    | ||||
Db    2555329 AGAGAAAAGTATAATAAATTTATTTAATTAAAGTTTTATGTGATATGAGAAATATTAGAA 2555388

Qy        455 ATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTTTTTTGCTGAGGTTCGATGAAGAA 514
              |||||||   |||||   ||| ||| ||| |  ||||| || | ||| |  |||||||||
Db    2555389 ATGAAGATTTAAAGATTTAGGAAAAATTGT-TTATTTTTATGTTTAGGATTAATGAAGAA 2555447

Qy        515 TGGATAGCATGTAGCCATGTAGATTAGACAAAAGGATATGATCTAGTGGTAAAGGACTCA 574
              || ||||          |     ||| ||  ||||| |||||| || |||| |  |  | |
Db    2555448 TGAATAGTTATGTAGAAATGTTATTGGATTAAAGGGTATGATTTAATGGTGATAGGTTGA 2555507

Qy        575 GGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTTCTTGATCTCTCTCTCTCTATGTA 634
              ||||||||   || ||||  ||| | ||  ||||| || ||| | | | | | |    |
Db    2555508 GGGGGAAATTTAGTAAGGTTTGTTTGTTTAGATTTTTTTTGGTTTTTTTGTGTAGTTTGT 2555567

Qy        635 TAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTTCTTCAATGAGGGTCTTCAAGGGA 694
              |   ||| |           |  | ||| |||| | || |  ||||| ||| || ||| ||
Db    2555568 TTTTATTTTG---------GGTTTGGGTAGGATTTTTGTGGAATGAAGGTTTTTAAGAGA 2555618

Qy        695 GAAGGGAGAAAGTGGCCTTTTTAGATTTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGT 754
              ||  |              |||||  |  |||||| ||| || |||||| ||  ||| ||||
Db    2555619 GAGTGA-----------TTTTTTTAGGTTATGGTTTGTTTGGGGGAAAAGAGGTTTTAGT 2555667

Qy        755 TTCTATGACCCATCTTGGGGAAGAGGAATTCTGGTTTCTGTGACTTGCTTTCATGAAGAA 814
              || |||||   || ||||| |||||||||| |||||| | ||| ||  |||   | ||||
Db    2555668 TTTTATGATTTATTTTGGGAAAGAGGAATTTTGGTTTTTATGATTTATTTTATGGGAGAA 2555727

Qy        815 AGAGGAGTAAGAGGCAGGAGGGCAGGAGATGGTCAGAAAGAGACTT 860
              ||||  |  |||   ||||||  ||||| ||| ||| |||   ||
Db    2555728 AGAGAGGCGAGAAATAGGAGGATGGGAGAGGGTTAGAGAGATTGTT 2555773


RESULT 31
US-10-741-601-5646
; Sequence 5646, Application US/10741601
; Publication No. US20040166519A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  STENOSIS, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001500
;  CURRENT APPLICATION NUMBER: US/10/741,601
;  CURRENT FILING DATE:  2003-12-22
;  NUMBER OF SEQ ID NOS: 26415
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5646
;   LENGTH: 97952
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-741-601-5646

  Query Match             3.3%;  Score 201.6;  DB 18;  Length 97952;
  Best Local Similarity   62.4%;  Pred. No. 2.6e-37;
  Matches  448;  Conservative    0;  Mismatches  184;  Indels   86;  Gaps    5;

Qy        312 GTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTC 371
              ||||| ||||| | ||||||||| ||| |||  | ||||||| ||| |  |||| |||
Db      92323 GTAGGAGTGAAGAGGTGTGATACCTTTCCTCACCTATCATAAGAGTCACGGCCAATACTA 92382
```

```
Qy        372 CTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTT 431
Db      92383 TGACAAAAGAT----GGGTAACAAGAGAAAAGGGTAACATATTTCTTTCATCAAAGTTTT 92438

Qy        432 ACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTT 491
Db      92439 ACATGACATGGGAGCTTTTAGAAATGAAAACCCAAAGATCCAAGGAAAACTGT-TCATTT 92497

Qy        492 TTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGAT 551
Db      92498 TTATGCTTAGATTTAATGCAGAATGAACGGCCAGGCAGAAATGTGATTGGGCAAAAGGCT 92557

Qy        552 ATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTT 611
Db      92558 GTGATCTAAGGCTAATAAATGGAAGGGGAAGCCCAGCAAGGCCTGTGTGTTCCGATTCTT 92617

Qy        612 CTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTT 671
Db      92618 CTTGGCTGCAGTC--------TGTAGCATTTTTTCCTCCAGGGATCAGGGCAGGACCCCC 92669

Qy        672 CTTCAATGAGGGTCTT---------------------------------------------- 687
Db      92670 CTGGAATGAGGGTCTTATTACCTACTATCAGGCAAGGTAGGCCAGAGAATTTCTTTACCG 92729

Qy        688 ------------------CAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATGGCT 729
Db      92730 TCACTGTCTACACAGAAAGGCAGGGGAAGGTTAGAGTAATTTTTCTAGGGTTTTATGGCT 92789

Qy        730 TGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCTGGT 789
Db      92790 TGCTTTGGGGGAAAGGGGTTCTAGTTTCTATGACCCTCCTTGAGAAAGAGGAATTCTGCT 92849

Qy        790 TTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGGTCA 849
Db      92850 TTCTGCGAGTTGCTTCAGGGGAGAAAGAAGAGGAGGAGACAGGAGGGCAGGAGAAGATCG 92909

Qy        850 GAAAGAGACTT-----------GGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAA 898
Db      92910 GAGAGGGAGACCTTGTTTCTGAGGTTTCCTTGGAGGCTTTCCAGTCTCCTTTAGTTCAAA 92969

Qy        899 GTACTTCTTAGCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAACA 956
Db      92970 GTACTCTGCACACCAAAGTGCCACACTTTGGGGTTTCATTTTTTTCTGAGCCCCAACA 93027
```

RESULT 32
US-10-027-632-87039/c
; Sequence 87039, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 87039
; LENGTH: 669
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-87039

Query Match 3.2%; Score 193; DB 13; Length 669;
Best Local Similarity 71.2%; Pred. No. 1.7e-36;
Matches 364; Conservative 1; Mismatches 111; Indels 35; Gaps 7;

```
Qy        286 ATCTCTCCTTCTCTCTCAAAGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTC 345
Db        487 ACCTGTATTTTCTTCTACGAAGGAGTGTAGAGGCAGAAAGATGTGATACATTCCTTCACC 428

Qy        346 CATCATAAGGGTCGCAACCAAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACC 405
Db        427 CATCATAAGGGTCACGGTTGACATTCCTAAAATGAAAGACTGTTTAACAA-AAAAAGGTG 369

Qy        406 TAACAAATTTATTTAATCAAAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCA 465
Db        368 TAACAAACTTATTTAATCAAAGTTTTACATGACACAGGGACCTTCAGAAATGAAAACCCA 309

Qy        466 AAGACCCAGGGGAAACTGTCTGTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATG 525
Db        308 AAGACCCAGGGAAAACT-----ATTTTTATGCTTAGGCTCGATAAAGTGTTATAGAAATG 254

Qy        526 TAGCCATGTAGATTAGACAAAAGGATATGATCTAGTGGTAAAGGACTCAGG-GGGAAACA 584
Db        253 T---------GATKGGACAAAAGGGTGTAATTTGATTGTAATTGACTGAGGAGGGAAACG 203

Qy        585 CAGCAAGGCCTGTCTATTCAGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTT 644
Db        202 CAGCAAGGTCTGTCCATTCAGATTCTTCTTGGACTGTC--------AGTGTAGCATTCCA 151

Qy        645 TCCTCCTGAGTATGGGGCAGGACTCTTCTTCAATGAGGGTCTTC----------AAGGGA 694
Db        150 TCCTCCTGAGTATGGGGAAGGACGCCTCTGGAATGAAGGTCTTCAAAGAAGAAGAAGGCA 91

Qy        695 GAAGGGAGAAAGTGGCCTTTTTAGATTTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGT 754
Db         90 GAAGATAGAGAATGACTTTTATAGGTTTTATGGCTTGCTTTGGAGAAGAGGAGTTCTA-T 32

Qy        755 TTCTATGACCCATCTTGGGGAAGAGGAATTC 785
Db         31 TTCTATGACCCTCCTTGGGAAACAGGAATTC 1
```

RESULT 33
US-10-027-632-87039/c
; Sequence 87039, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720

```
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 87039
;   LENGTH: 669
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-87039

  Query Match             3.2%;  Score 193;  DB 17;  Length 669;
  Best Local Similarity   71.2%;  Pred. No. 1.7e-36;
  Matches  364;  Conservative    1;  Mismatches  111;  Indels   35;  Gaps    7;

Qy        286 ATCTCTCCTTCTCTCTCAAAGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTC 345
              | || |  ||   |||   |      |||| ||   |||  ||||||||||   ||  |
Db        487 ACCTGTATTTTCTTCTACGAAGGAGTGTAGAGGCAGAAAGATGTGATACATTCCTTCACC 428

Qy        346 CATCATAAGGGTCGCAACCAAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACC 405
              ||||||||||||| |      | | ||||| | | |||||| | |||| || |  ||
Db        427 CATCATAAGGGTCACGGTTGACATTCCTAAAATGAAAGACTGTTTAACAA-AAAAAGGTG 369

Qy        406 TAACAAATTTATTTAATCAAAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCA 465
              ||||||| ||||||||||||||||||||||||||  ||    |||||||||||||| |||||
Db        368 TAACAAACTTATTTAATCAAAGTTTTACATGACACAGGGACCTTCAGAAATGAAAACCCA 309

Qy        466 AAGACCCAGGGGAAACTGTCTGTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATG 525
              |||||||||| |||||      ||||| |||| ||| ||||| |||  |        |||
Db        308 AAGACCCAGGGAAAACT-----ATTTTTATGCTTAGGCTCGATAAAGTGTTATAGAAATG 254

Qy        526 TAGCCATGTAGATTAGACAAAAGGATATGATCTAGTGGTAAAGGACTCAGG-GGGAAACA 584
              |          |||: |||||||| | | || |  | ||||  |||| ||| |||||||
Db        253 T---------GATKGGACAAAAGGGTGTAATTTGATTGTAATTGACTGAGGAGGGAAACG 203

Qy        585 CAGCAAGGCCTGTCTATTCAGATTCTTCTTGATCTCTCTCTCTATGTATAGCATTCTT 644
              |||||||| ||||| |||||||||||||||  || ||          || ||||||||
Db        202 CAGCAAGGTCTGTCCATTCAGATTCTTCTTGGACTGTC--------AGTGTAGCATTCCA 151

Qy        645 TCCTCCTGAGTATGGGGCAGGACTCTTCTTCAATGAGGGTCTTC----------AAGGGA 694
              |||||||||||||||| ||||| | |||  ||||| |||||||          |||| |
Db        150 TCCTCCTGAGTATGGGGAAGGACGCCTCTGGAATGAAGGTCTTCAAAGAAGAAGAAGGCA 91

Qy        695 GAAGGGAGAAAGTGGCCTTTTTAGATTTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGT 754
              ||||  ||| | || | ||| ||| ||||||||||||||| || ||||||||||||| |
Db         90 GAAGATAGAGAATGACTTTTATAGGTTTTATGGCTTGCTTTGGAGAAGAGGAGTTCTA-T 32

Qy        755 TTCTATGACCCATCTTGGGGAAGAGGAATTC 785
              ||||||||||  |||||| || ||||||||
Db         31 TTCTATGACCCTCCTTGGGAAACAGGAATTC 1


RESULT 34
US-10-719-993-6836
; Sequence 6836, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001496
;  CURRENT APPLICATION NUMBER: US/10/719,993
;  CURRENT FILING DATE:  2003-11-24
;  NUMBER OF SEQ ID NOS: 55342
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6836
;   LENGTH: 25041
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-719-993-6836

  Query Match             3.2%;  Score 192.4;  DB 18;  Length 25041;
  Best Local Similarity   68.6%;  Pred. No. 2.1e-35;
  Matches  420;  Conservative    0;  Mismatches  136;  Indels   56;  Gaps    9;

Qy        328 GTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTCCTATAGTAAAAGACAG 387
              || |||| | | ||    |||  |||| |||    |  | |||||||||  ||||||||
Db      21142 GTAATACCTCTCCTTACTCATTGTAAGAGTCATGGCTGACACTCCTATAACAAAAGACAA 21201

Qy        388 GTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTACATGACATGGGAGTC 447
              ||||| ||||| ||| | |||||||||||| ||||||||||||||  ||| ||| ||| |
Db      21202 GTTAACAAGAGAAAAGCATAACAAATTTACTTAATCAAAGTTTTATGTGATATGAGAGCC 21261

Qy        448 TTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTTTTTTTGCTGAGGTTCGA 507
              |||||||||||||||||||||||| |||||| ||||||     ||||| |||| |||||| |
Db      21262 TTCAGAAATGAAGACCCAAAGATCCAGGGAAAACTG-----TTTTTGTGCTTAGGTTCCA 21316

Qy        508 TGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGATATGATCTAGTGGTAAA 567
              |||||||||  |||               | || ||||| ||  |||||||| ||||||
Db      21317 TGAAGAATGGGCAGCTCTGTAGAGATGTGTTTGGACAAGAG--TATGATCTAATGGTAAT 21374

Qy        568 GGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTTCTTGATCTCTCTCTCT 627
              ||||  |                    |||| ||||||||||  |||||
Db      21375 AGACTGGG----------------------TTCAAATTCTTCTTGGACTCTC----- 21404

Qy        628 CTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTTCTTCAATGAGGGTCTT 687
              |||  |||||||  |||  |    |||||||||||||| || |||  | | ||||||||
Db      21405 ---TGTGTAGCATTTCTTCTTTGCAAGTATGGGGCAGGATTCCTCTGGATTAAGGGTCTT 21461

Qy        688 C-------AAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATGGCTTGCTTCGGGGA 740
              |       |||||||||||||||| |||| ||||   ||  |||||||||||||| |||||
Db      21462 CAAATGAGAAGGGAGAAGGGAGAGAGTGACCTTCCTAAGTTTTATGGCTTGCTTTGGGGA 21521

Qy        741 AGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCTGGTTTCTGTGACTT 800
              | |||  |||||||||||||| |    ||||| |||||||||||||| |||||  || | |
Db      21522 A-AGGGCTTCTAGTTTCTATCA-TTGTCTTGCGGAAGAGGAATTCTAGTTTCCATGGTTC 21579

Qy        801 GCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGGTCAGAAAGAGACTT 860
              ||      | ||||||| |  | | || || |||||||||||| ||||| | |||    |
Db      21580 ACTCCAGGGGAGAAAGAAGCATGAAAGACAAGAGGGCAGGAGAAGGTCAAAGAGA----T 21635

Qy        861 GGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAGTACTTCTTAGCATACCAAAGC 920
              |  |      |  ||||  | || || |||| ||||   || |||| ||| | ||
Db      21636 CTTTCTTCGGAGGCCCTTCCAATGTCTTTCAGTTCAAAGTA---CTCAGCACACCGAGGC 21692

Qy        921 ACTATACTTTGG 932
              ||||||||||||
Db      21693 GCTATACTTTGG 21704


RESULT 35
US-10-719-993-6837/c
; Sequence 6837, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001496
;  CURRENT APPLICATION NUMBER: US/10/719,993
;  CURRENT FILING DATE:  2003-11-24
;  NUMBER OF SEQ ID NOS: 55342
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6837
;   LENGTH: 141158
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-719-993-6837

  Query Match             3.2%;  Score 192.4;  DB 18;  Length 141158;
  Best Local Similarity   68.6%;  Pred. No. 6e-35;
  Matches  420;  Conservative    0;  Mismatches  136;  Indels   56;  Gaps    9;

Qy        328 GTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTCCTATAGTAAAAGACAG 387
              || |||| | | ||    |||  |||| |||    |  | |||||||||  ||||||||
Db       5639 GTAATACCTCTCCTTACTCATTGTAAGAGTCATGGCTGACACTCCTATAACAAAAGACAA 5580

Qy        388 GTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTACATGACATGGGAGTC 447
              ||||| ||||| ||| | |||||||||||| ||||||||||||||  ||| ||| ||| |
Db       5579 GTTAACAAGAGAAAAGCATAACAAATTTACTTAATCAAAGTTTTATGTGATATGAGAGCC 5520
```

```
Qy       448 TTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTTTTTTGCTGAGGTTCGA 507
Db      5519 TTCAGAAATGAAGACCCAAAGATCCAGGGAAAACTG-----TTTTTGTGCTTAGGTTCCA 5465

Qy       508 TGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGATATGATCTAGTGGTAAA 567
Db      5464 TGAAGAATGGGCAGCTCTGTAGAGATGTGTTTGGACAAGAG--TATGATCTAATGGTAAT 5407

Qy       568 GGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTTCTTGATCTCTCTCTCT 627
Db      5406 AGACTGGG-----------------------TTCAAATTCTTCTTGGACTCTC----- 5377

Qy       628 CTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTTCTTCAATGAGGGTCTT 687
Db      5376 ---TGTGTAGCATTTCTTCTTTGCAAGTATGGGGCAGGATTCCTCTGGATTAAGGGTCTT 5320

Qy       688 C-------AAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATGGCTTGCTTCGGGGA 740
Db      5319 CAAATGAGAAGGGAGAAGGGAGAGAGTGACCTTCCTAAGTTTTATGGCTTGCTTTGGGGA 5260

Qy       741 AGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCTGGTTTCTGTGACTT 800
Db      5259 A-AGGGCTTCTAGTTTCTATCA-TTGTCTTGCGGAAGAGGAATTCTAGTTTCCATGGTTC 5202

Qy       801 GCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGGTCAGAAAGAGACTT 860
Db      5201 ACTCCAGGGGAGAAAGAAGCATGAAAGACAAGAGGGCAGGAGAAGGTCAAAGAGA----T 5146

Qy       861 GGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAGTACTTCTTAGCATACCAAAGC 920
Db      5145 CTTTCTTCGGAGGCCCTTCCAATGTCTTTCAGTTCAAAGTA---CTCAGCACACCGAGGC 5089

Qy       921 ACTATACTTTGG 932
Db      5088 GCTATACTTTGG 5077
```

RESULT 36
US-10-719-993-7069/c
; Sequence 7069, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7069
; LENGTH: 39428
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(39428)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-719-993-7069

Query Match 3.1%; Score 188.6; DB 18; Length 39428;
Best Local Similarity 79.7%; Pred. No. 2.4e-34;
Matches 235; Conservative 0; Mismatches 59; Indels 1; Gaps 1;

```
Qy      2177 GGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCT 2236
Db      1251 GGGTATCCATTACCCAAACAATATGCATTGTACCCATTAAGTAATTTCTCATCATCCAGC 1192

Qy      2237 GACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTCTACATCT 2296
Db      1191 CCCCTCAAACTCCATCATCCTTCTGAGTCTCCATTGTCTATTATTCCACACTCTATGTCC 1132

Qy      2297 ATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGTCTTTCTG 2356
Db      1131 ATGTGTACACATTGTTTAGCTCCCACTTATAAGTTAGAACATGCAACATTTGTCTTTCTG 1072

Qy      2357 TGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAAGG 2416
Db      1071 TTTCTGACTTGTTTCACTTAAGATAATGGTCTCCCATTCCATCCATGTTTATGCAAAAGA 1012

Qy      2417 CATGATTTCATTCTTTTTTATGGCTATGTTCTTTCCCAATTTAGATAAAGAACAC 2471
Db      1011 CA-GATTTCATTCTTTTTTATGGCTGAATAGTATTTCATTGTATATATAGACCAC 958
```

RESULT 37
US-10-027-632-226077
; Sequence 226077, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 226077
; LENGTH: 650
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-226077

Query Match 3.0%; Score 183.6; DB 13; Length 650;
Best Local Similarity 72.6%; Pred. No. 3.4e-34;
Matches 308; Conservative 0; Mismatches 104; Indels 12; Gaps 5;

```
Qy       284 TTATCTCTCCTTCTCTCTCAAAGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCT 343
Db       238 TTAATTCATTTATTCATTCAATACATGCACACTGTAGGAAAGTTGTGGTATCCTTCTTCA 297

Qy       344 TCCATCATAAGGGTCGCAACCAAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAA 403
Db       298 TCCATCATAAGGGTCATGGCCAACACTCCTATAAC-AAAGACAGATTAATAAGAGAAAAG 356

Qy       404 CCTAACAAATTTATTTAATCAAAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACC 463
Db       357 CATAGCAAATGTATTTAATCAAAGTTTTA-GTGACATGGGGAACTTCAGAAATGAAGATT 415

Qy       464 CAAAGACCCAGGGGAAACTGTCTGTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCA 523
Db       416 CAAAGACCCAGGGAAAACTGAC-CATTTTAATGCTTAGATTTGATGAAGAATGAACAGCC 474

Qy       524 TGTAGCCATGTAGATTAGACAAAAGGATATGATCTAGTGGTAAAGGACTCAGGGGG-AAA 582
Db       475 ATGTAGAAATGAAATTGGACAAAGGGGTATAATCTAATGGTAATTGACTGAGGGGGAAAA 534

Qy       583 CACAGCAAGGCCTGTCTATTCAGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTC 642
Db       535 CTCAGCAAGGCCTGTCTGTTCATAGTCTTCTTGGCCTGTC--------TGTGTAGCACCC 586
```

```
Qy        643 TTTCCTCCTGAGTATGGGGCAGGACTCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAG 702
              |||||||   |||| || |||||| | |||  ||||||||||| |||||||| ||||
Db        587 CTTCCTCCCAGGTATAGGACAGGACCCCTCTGGAATGAGGGTCTTTAAGGGAGAAAGGAG 646

Qy        703 AAAG 706
              | ||
Db        647 AGAG 650
```

RESULT 38
US-10-027-632-226077
; Sequence 226077, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 226077
; LENGTH: 650
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-226077

Query Match 3.0%; Score 183.6; DB 17; Length 650;
Best Local Similarity 72.6%; Pred. No. 3.4e-34;
Matches 308; Conservative 0; Mismatches 104; Indels 12; Gaps 5;

```
Qy        284 TTATCTCTCCTTCTCTCTCAAAGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCT 343
              |||  ||   |  ||  ||||  |     |  |  | |||  |||| ||   ||  ||
Db        238 TTAATTCATTTATTCATTCAATACATGCACACTGTAGGAAAGTTGTGGTATCCTTCTTCA 297

Qy        344 TCCATCATAAGGGTCGCAACCAAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAA 403
              ||||||||||||||    |||| |||||||||   |||||||| ||||||||| |||
Db        298 TCCATCATAAGGGTCATGGCCAACACTCCTATAAC-AAAGACAGATTAATAAGAGAAAAG 356

Qy        404 CCTAACAAATTTATTTAATCAAAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACC 463
              | || ||||| ||||||||||||||||||  |||||||||   |||||||||||||||
Db        357 CATAGCAAATGTATTTAATCAAAGTTTTA-GTGACATGGGGAACTTCAGAAATGAAGATT 415

Qy        464 CAAAGACCCAGGGGAAACTGTCTGTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCA 523
              |||||||||||| |||||| |   ||||  |||| || || |||||||||| | |||
Db        416 CAAAGACCCAGGGAAAACTGAC-CATTTTAATGCTTAGATTTGATGAAGAATGAACAGCC 474

Qy        524 TGTAGCCATGTAGATTAGACAAAAGGATATGATCTAGTGGTAAAGGACTCAGGGGG-AAA 582
              |    | ||| |||||| || ||| ||||| ||||||  |||| ||||||| |||
Db        475 ATGTAGAAATGAAATTGGACAAAGGGGTATAATCTAATGGTAATTGACTGAGGGGGAAAA 534

Qy        583 CACAGCAAGGCCTGTCTATTCAGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTC 642
              | |||||||||||||| |||| | ||||||||  || ||         ||| |||||  |
Db        535 CTCAGCAAGGCCTGTCTGTTCATAGTCTTCTTGGCCTGTC--------TGTGTAGCACCC 586

Qy        643 TTTCCTCCTGAGTATGGGGCAGGACTCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAG 702
              |||||||   |||| || |||||| | |||  ||||||||||| |||||||| ||||
```

```
Db        587 CTTCCTCCCAGGTATAGGACAGGACCCCTCTGGAATGAGGGTCTTTAAGGGAGAAAGGAG 646

Qy        703 AAAG 706
              | ||
Db        647 AGAG 650
```

RESULT 39
US-10-027-632-320728/c
; Sequence 320728, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 320728
; LENGTH: 520
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-320728

Query Match 2.9%; Score 178.4; DB 13; Length 520;
Best Local Similarity 70.3%; Pred. No. 5.6e-33;
Matches 336; Conservative 1; Mismatches 107; Indels 34; Gaps 6;

```
Qy        286 ATCTCTCCTTCTCTCTCAAAGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTC 345
              | || |  ||  |||   |     |||| ||   |||  ||||||||||    ||  |
Db        455 ACCTGTATTTTCTTCTACGAAGGAGTGTAGAGGCAGAAAGATGTGATACATTCCTTCACC 396

Qy        346 CATCATAAGGGTCGCAACCAAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACC 405
              ||||||||||||  |     | | ||||| | | |||||| | |||| || |  ||
Db        395 CATCATAAGGGTCACGGTTGACATTCCTAAAATGAAAGACTGTTTAACAA-AAAAAGGTG 337

Qy        406 TAACAAATTTATTTAATCAAAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCA 465
              ||||||| ||||||||||||||||||||||||||   ||   |||||||||||||| |||||
Db        336 TAACAAACTTATTTAATCAAAGTTTTACATGACACAGGGACCTTCAGAAATGAAAACCCA 277

Qy        466 AAGACCCAGGGGAAACTGTCTGTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATG 525
              |||||||||| |||||      |||||| |||| ||| ||||| |||  |  .     |||
Db        276 AAGACCCAGGGAAAACT-----ATTTTTATGCTTAGGCTCGATAAAGTGTTATAGAAATG 222

Qy        526 TAGCCATGTAGATTAGACAAAAGGATATGATCTAGTGGTAAAGGACTCAGG-GGGAAACA 584
              |         |||: |||||||| | | || |   | ||||  |||| ||| |||||||
Db        221 T---------GATKGGACAAAAGGGTGTAATTTGATTGTAATTGACTGAGGAGGGAAACG 171

Qy        585 CAGCAAGGCCTGTCTATTCAGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTT 644
              |||||||| ||||| |||||||||||||||  || ||          || ||||||||
Db        170 CAGCAAGGTCTGTCCATTCAGATTCTTCTTGGACTGTC--------AGTGTAGCATTCCA 119

Qy        645 TCCTCCTGAGTATGGGGCAGGACTCTTCTTCAATGAGGGTCTTC----------AAGGGA 694
              |||||||||||||||| ||||| | |||  ||||| |||||||          |||| |
Db        118 TCCTCCTGAGTATGGGGAAGGACGCCTCTGGAATGAAGGTCTTCAAAGAAGAAGAAGGCA 59

Qy        695 GAAGGGAGAAAGTGGCCTTTTTAGATTTTATGGCTTGCTTCGGGGAAGAGGAGTTCTA 752
```

```
Db         58 GAAGATAGAGAATGACTTTTATAGGTTTTATGGCTTGCTTTGGAGAAGAGGAGTTCTA 1
```

RESULT 40
US-10-027-632-320728/c
; Sequence 320728, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 320728
; LENGTH: 520
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-320728

Query Match 2.9%; Score 178.4; DB 17; Length 520;
Best Local Similarity 70.3%; Pred. No. 5.6e-33;
Matches 336; Conservative 1; Mismatches 107; Indels 34; Gaps 6;

```
Qy        286 ATCTCTCCTTCTCTCTCAAAGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTC 345
Db        455 ACCTGTATTTTCTTCTACGAAGGAGTGTAGAGGCAGAAAGATGTGATACATTCCTTCACC 396

Qy        346 CATCATAAGGGTCGCAACCAAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACC 405
Db        395 CATCATAAGGGTCACGGTTGACATTCCTAAAATGAAAGACTGTTTAACAA-AAAAAGGTG 337

Qy        406 TAACAAATTTATTTAATCAAAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCA 465
Db        336 TAACAAACTTATTTAATCAAAGTTTTACATGACACAGGGACCTTCAGAAATGAAAACCCA 277

Qy        466 AAGACCCAGGGGAAACTGTCTGTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATG 525
Db        276 AAGACCCAGGGAAAACT-----ATTTTTATGCTTAGGCTCGATAAAGTGTTATAGAAATG 222

Qy        526 TAGCCATGTAGATTAGACAAAAGGATATGATCTAGTGGTAAAGGACTCAGG-GGGAAACA 584
Db        221 T---------GATKGGACAAAAGGGTGTAATTTGATTGTAATTGACTGAGGAGGGAAACG 171

Qy        585 CAGCAAGGCCTGTCTATTCAGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTT 644
Db        170 CAGCAAGGTCTGTCCATTCAGATTCTTCTTGGACTGTC--------AGTGTAGCATTCCA 119

Qy        645 TCCTCCTGAGTATGGGGCAGGACTCTTCTTCAATGAGGGTCTTC----------AAGGGA 694
Db        118 TCCTCCTGAGTATGGGGAAGGACGCCTCTGGAATGAAGGTCTTCAAAGAAGAAGAAGGCA 59

Qy        695 GAAGGGAGAAAGTGGCCTTTTTAGATTTTATGGCTTGCTTCGGGGAAGAGGAGTTCTA 752
Db         58 GAAGATAGAGAATGACTTTTATAGGTTTTATGGCTTGCTTTGGAGAAGAGGAGTTCTA 1
```

RESULT 41
US-10-085-117-142/c
; Sequence 142, Application US/10085117
; Publication No. US20030232334A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: 529452000121
; CURRENT APPLICATION NUMBER: US/10/085,117
; CURRENT FILING DATE: 2002-02-27
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 361
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 142
; LENGTH: 193357
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-085-117-142

Query Match 2.9%; Score 175.6; DB 17; Length 193357;
Best Local Similarity 68.4%; Pred. No. 9.9e-31;
Matches 350; Conservative 0; Mismatches 119; Indels 43; Gaps 6;

```
Qy        389 TTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTACATGACATGGGAGTCT 448
Db     191709 TTTATTGCTTCAAACAATGAACAGCCTATTTGATCATAGTTTTACATGACAGGTGAGCCT 191650

Qy        449 TCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTTTTTTTGCTGAGGTTCGAT 508
Db     191649 TCAGAAC-GAAGACCCAGAGACTCAGGAGAAGCTCTCTG-ATTTTATGCTTAGGTTTGAT 191592

Qy        509 GAAGAATGGATAGCATGTAGCCATGTAGATTAGAC--AAAAGGATATGATCTAGTGGTAA 566
Db     191591 GAAGCAGGGACAGTCATGTGGAAATGTGATTGGACAAAAAAGGCTATGATCTAATGGAAA 191532

Qy        567 AGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTTCTTGATCTCTCTCTC 626
Db     191531 TAGACTGCGTGGGGAAACCTACAAGGCCTGGCTGTTCAGATCCTTCTTGGCCTCTC---- 191476

Qy        627 TCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTTCTTCAATGAGGGTCT 686
Db     191475 ----CATGCAGCATTCCTTCCTCCCAGGTATGGGGTAGGACCCCTCTGGGATGAGGGTCT 191420

Qy        687 T------------------------------CAAGGGAGAAGGGAGAAAGTGGCCTTTTT 716
Db     191419 TAGTTTCTTAATGGCCAAGTGTTACACAGAAGAATGGTGGAAGGTTAGAATAATATTTCT 191360

Qy        717 AGATTTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAA 776
Db     191359 AGGTTTTATGGCTGGCTTTGGGGAAAAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAA 191300

Qy        777 GAGGAATTCTGGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAG-GAGG 835
Db     191299 GAGGGATTACAGTTTCTATGGCTTATCTCCGGGGAGAATGAGGGGTGAGAGACAGAGCGG 191240

Qy        836 GCAGGAGATGGTCAGAAAGAGACTTGGCTGCT 867
Db     191239 GCAGCAGAAGGTCAGAGAGAGACTTTGCTTCT 191208
```

RESULT 42
US-09-854-867-156/c
; Sequence 156, Application US/09854867
; Publication No. US20030224356A1
; GENERAL INFORMATION:
; APPLICANT: JOAN, KNOLL H
; APPLICANT: ROGAN, PETER K
; TITLE OF INVENTION: SINGLE COPY GENOMIC HYBRIDIZATION PROBES AND METHOD OF GENERATING
; FILE REFERENCE: 30307
; CURRENT APPLICATION NUMBER: US/09/854,867

```
;  CURRENT FILING DATE:  2003-05-08
;  NUMBER OF SEQ ID NOS: 613
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 156
;   LENGTH: 693
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: repeat_region
;   LOCATION: (1)..(693)
;   OTHER INFORMATION: mer39
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (57)..(57)
;   OTHER INFORMATION: n is a, c, g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (63)..(63)
;   OTHER INFORMATION: n is a, c, g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (79)..(79)
;   OTHER INFORMATION: n is a, c, g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (83)..(83)
;   OTHER INFORMATION: n is a, c, g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (144)..(144)
;   OTHER INFORMATION: n is a, c, g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (151)..(151)
;   OTHER INFORMATION: n is a, c, g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (161)..(161)
;   OTHER INFORMATION: n is a, c, g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (193)..(193)
;   OTHER INFORMATION: n is a, c, g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (209)..(210)
;   OTHER INFORMATION: n is a, c, g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (230)..(230)
;   OTHER INFORMATION: n is a, c, g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (477)..(477)
;   OTHER INFORMATION: n is a, c, g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (610)..(610)
;   OTHER INFORMATION: n is a, c, g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (628)..(629)
;   OTHER INFORMATION: n is a, c, g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (659)..(659)
;   OTHER INFORMATION: n is a, c, g or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (683)..(685)
;   OTHER INFORMATION: n is a, c, g or t
US-09-854-867-156
```

```
  Query Match             2.9%;  Score 173.8;  DB 10;  Length 693;
  Best Local Similarity   72.8%;  Pred. No. 9e-32;
  Matches  283;  Conservative   13;  Mismatches   78;  Indels   15;  Gaps    6;

Qy        302 CAAAGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCA 361
              ||       | |||||||   ||| ||||||| | ||  ||   |||||||||||||| |
Db        690 CATTANNNCTGTAGGGGCAGAAAGGTGTGATNCCTTCCCT-ACCCATCATAAGGGTCACG 632

Qy        362 ACCAAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAA 421
              |  | ||||||||| ||||| |:|||||  ||||| ||| | |||:| ||||||||||
Db        631 GCNNACACTCCTATAATAAAANAYAGGTTAGCAAGAGAAAAGCATAAYAGATTTATTTAA 572

Qy        422 TCAAAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAAC 481
              |||||||||||  |||||||:||| |||||||||||||||:||||:|| ||||| ||||
Db        571 TCAAAGTTTTATGTGACATGRGAGCCTTCAGAAATGAAGACMCAAAKACTCAGGGAAAAC 512

Qy        482 TGTCTGTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAG 541
              | |    ||||| ||:| || || ||||||||||||| |||         |    ||||:|
Db        511 TCT---ATTTTTATGYTTAGATTTGATGAAGAATGGANAGCTGCGCAGAAATGTGATTRG 455

Qy        542 ACAAAA-GGATATGATCTAGTGGTAAAGGACTCAG-GGGGAAACACAGCAAGGCCTGTCT 599
              |||||| || :|||||||| ||||:|  :||| :| | ||:||| ||||||||| |||||
Db        454 ACAAAAGGGGWATGATCTAATGGTMATARACTGRGTGTGGRAACCCAGCAAGGCTTGTCT 395

Qy        600 ATTCAGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGG 659
              |||  |||||||||| |||:||         |||   |||||| ||||||| | |||| |
Db        394 GTTCGTATTCTTCTTGGTCTMTC--------TGTGCGGCATTCCTTCCTCCCGGGTATAG 343

Qy        660 GGCAGGA-CTCTTCTTCAATGAGGGTCTT 687
              ||| ||| | |||||  ||||||||||||
Db        342 GGCGGGACCCCTTCTGGAATGAGGGTCTT 314
```

```
RESULT 43
US-10-786-970A-156/c
; Sequence 156, Application US/10786970A
; Publication No. US20050064449A1
; GENERAL INFORMATION:
;  APPLICANT: JOAN, KNOLL
;  APPLICANT:  ROGAN, PETER
;  TITLE OF INVENTION: SINGLE COPY GENOMIC HYBRIDIZATION PROBES AND METHOD OF GENERATING
;  FILE REFERENCE: 30307
;  CURRENT APPLICATION NUMBER: US/10/786,970A
;  CURRENT FILING DATE:  2004-02-24
;  PRIOR APPLICATION NUMBER: US/09/573,080
;  PRIOR FILING DATE: 2000-05-16
;  NUMBER OF SEQ ID NOS: 479
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 156
;   LENGTH: 693
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: repeat_region
;   LOCATION: (1)..(693)
;   OTHER INFORMATION: mer39
;  PUBLICATION INFORMATION:
;  PUBLICATION INFORMATION:
;   AUTHORS: Jurka, J; Walichiewicz, J; Milosavljevic, A
;   TITLE: Prototypic sequences for human repetitive DNA
;   JOURNAL: Journal of Molecular Evolution
;   VOLUME: 35
;   ISSUE: 4
;   PAGES: 286-291
;   DATE: 1992-10-__
;   DATABASE ACCESSION NUMBER: Database of repetitive elements (repbase)
;   DATABASE ENTRY DATE: ____-__-__
;   DATABASE ENTRY DATE: 1996-01-26
US-10-786-970A-156
```

```
  Query Match             2.9%;  Score 173.8;  DB 19;  Length 693;
  Best Local Similarity   72.8%;  Pred. No. 9e-32;
  Matches  283;  Conservative   13;  Mismatches   78;  Indels   15;  Gaps    6;
```

```
Qy        302 CAAAGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCA 361
              ||      | ||||||||.  ||| |||||||| | ||  ||   ||||||||||||| |
Db        690 CATTANNNCTGTAGGGGCAGAAAGGTGTGATNCCTTCCCT-ACCCATCATAAGGGTCACG 632

Qy        362 ACCAAAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAA 421
              |  | |||||||||| ||||| |:||||||  ||||| ||| | |||:| ||||||||||
Db        631 GCNNACACTCCTATAATAAAANAYAGGTTAGCAAGAGAAAAGCATAAYAGATTTATTTAA 572

Qy        422 TCAAAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAAC 481
              |||||||||||  |||||||:||| ||||||||||||||||:||||:|| ||||| ||||
Db        571 TCAAAGTTTTATGTGACATGRGAGCCTTCAGAAATGAAGACMCAAAKACTCAGGGAAAAC 512

Qy        482 TGTCTGTTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAG 541
              | |    ||||| ||:| || || |||||||||||||| |||        |    ||||:|
Db        511 TCT---ATTTTTATGYTTAGATTTGATGAAGAATGGANAGCTGCGCAGAAATGTGATTRG 455

Qy        542 ACAAAA-GGATATGATCTAGTGGTAAAGGACTCAG-GGGGAAACACAGCAAGGCCTGTCT 599
              |||||| || :|||||||| ||||:|  :||| :| | ||:||| |||||||| |||||
Db        454 ACAAAAGGGGWATGATCTAATGGTMATARACTGRGTGTGGRAACCCAGCAAGGCTTGTCT 395

Qy        600 ATTCAGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGG 659
              |||  |||||||||| |||:||       |||   ||||| ||||||| | |||| |
Db        394 GTTCGTATTCTTCTTGGTCTMTC--------TGTGCGGCATTCCTTCCTCCCGGGTATAG 343

Qy        660 GGCAGGA-CTCTTCTTCAATGAGGGTCTT 687
              ||| ||| | |||||  ||||||||||||
Db        342 GGCGGGACCCCTTCTGGAATGAGGGTCTT 314


RESULT 44
US-09-795-668-1
; Sequence 1, Application US/09795668
; Patent No. US20020045577A1
; GENERAL INFORMATION:
;  APPLICANT: Stefansson, Hreinn
;  APPLICANT:  Steinthorsdottir, Valgerdur
;  APPLICANT:  Gulcher, Jeffrey R.
;  TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
;  FILE REFERENCE: 2345.2004-001
;  CURRENT APPLICATION NUMBER: US/09/795,668
;  CURRENT FILING DATE:  2001-02-28
;  PRIOR APPLICATION NUMBER: US 09/515,716
;  PRIOR FILING DATE: 2000-02-28
;  NUMBER OF SEQ ID NOS: 1531
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 1503841
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: r=g or a
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: y=t/u or c
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: m=a or c
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: k=g or t/u
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: s=g or c
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: w=a or t/u
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: b=g or c or t/u
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: d=a or g or t/u
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: h=a or c or t/u
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: v=a or g or c
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: n=a or g or c or t/u
US-09-795-668-1

  Query Match             2.8%;  Score 170.2;  DB 9;  Length 1503841;
  Best Local Similarity   68.8%;  Pred. No. 7.3e-29;
  Matches  232;  Conservative    1;  Mismatches  104;  Indels    0;  Gaps    0;

Qy       2146 TTTCTCACCATCCGCTGACTTCTTGCCCCCTGGGTATTCATCACCTGAATGATGTGCATT 2205
              ||| | ||         | |||      |   | | |||  ||||||   |||  ||| | ||
Db    1013790 TTTGTTACATGGATATAACTCTGGTCTTTCAGTGTAACCATCACTCAAATAGTGTACCTT 1013849

Qy       2206 GTACCCCTTAAGTAATTTCTCACCATCCGCTGACTTCTTGCCCCCTCATCCTTCTGAGGC 2265
              |||||  ||| |||| || |||      | |   | ||   ||| | |||| ||:|||| |
Db    1013850 GTACCTATTACGTAACTTGTCATTCCTCACCCTCCTCCCACCCTCCCATCTTTYTGAGTC 1013909

Qy       2266 TCCATTGTCCATCATTCCACACTCTACATCTATGTGTACACATTATTTAGCTCCTACTTA 2325
              |||| |||| || ||||||| |||||  || ||||||||| |||||||||| || ||||||
Db    1013910 TCCAATGTCTATTATTCCACCCTCTATGTCCATGTGTACTCATTATTTAGTTCTTACTTA 1013969

Qy       2326 TAAGTGATAACATGCAATATTTGTCTTTCTGTGTCTGTCTTGTTTTACTTATGATAATGG 2385
              ||||| | |||||||||||||||| |||||||| ||||  || ||| ||||| |||||| |
Db    1013970 TAAGTAAGAACATGCAATATTTGACTTTCTGTTTCTGAGTTATTTCACTTAAGATAATAG 1014029

Qy       2386 CCCCCAGTTCTATCTAGGCTGCTGCAAAAGGCATGATTTCATTCTTTTTTATGGCTATGT 2445
              ||  ||| || ||| | | |||||||||| ||||| || |||||||| |||||||  ||
Db    1014030 CCTTCAGGTCCATCCATGTTGCTGCAAAAGACATGAATTGATTCTTTTGTATGGCTGAGT 1014089

Qy       2446 TCTTTCCCAATTTAGATAAAGAACACTCGCACTTGCT 2482
                | | |||   |  ||| |    ||     |||| |
Db    1014090 AGTATTCCATGGTGTATATACGTTACATTTTCTTGAT 1014126


RESULT 45
US-09-795-686-1
; Sequence 1, Application US/09795686
; Patent No. US20020094954A1
; GENERAL INFORMATION:
;  APPLICANT: Stefansson, Hreinn
;  APPLICANT:  Steinthorsdottir, Valgerdur
;  APPLICANT:  Gulcher, Jeffrey R.
;  TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
;  FILE REFERENCE: 2345.2005-001
;  CURRENT APPLICATION NUMBER: US/09/795,686
;  CURRENT FILING DATE:  2001-02-28
;  PRIOR APPLICATION NUMBER: US 09/515,715
;  PRIOR FILING DATE: 2000-02-28
;  NUMBER OF SEQ ID NOS: 1531
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 1503841
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: r=g or a
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: y=t/u or c
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: m=a or c
;   NAME/KEY: misc_feature
```

```
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: k=g or t/u
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: s=g or c
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: w=a or t/u
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: b=g or c or t/u
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: d=a or g or t/u
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: h=a or c or t/u
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: v=a or g or c
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: n=a or g or c or t/u
US-09-795-686-1

  Query Match             2.8%;  Score 170.2;  DB 9;  Length 1503841;
  Best Local Similarity   68.8%;  Pred. No. 7.3e-29;
  Matches  232;  Conservative    1;  Mismatches  104;  Indels    0;  Gaps    0;

Qy       2146 TTTCTCACCATCCGCTGACTTCTTGCCCCCTGGGTATTCATCACCTGAATGATGTGCATT 2205
             ||| | ||       | |||     |   | | |||  ||||||   |||  ||| | ||
Db    1013790 TTTGTTACATGGATATAACTCTGGTCTTTCAGTGTAACCATCACTCAAATAGTGTACCTT 1013849

Qy       2206 GTACCCCTTAAGTAATTTCTCACCATCCGCTGACTTCTTGCCCCCTCATCCTTCTGAGGC 2265
             |||||  ||| |||| || |||     | |   | ||   ||| | |||| ||:|||| |
Db    1013850 GTACCTATTACGTAACTTGTCATTCCTCACCCTCCTCCCACCCTCCCATCTTTYTGAGTC 1013909

Qy       2266 TCCATTGTCCATCATTCCACACTCTACATCTATGTGTACACATTATTTAGCTCCTACTTA 2325
             |||| |||| || ||||||| |||||  || |||||||| |||||||||| || ||||||
Db    1013910 TCCAATGTCTATTATTCCACCCTCTATGTCCATGTGTACTCATTATTTAGTTCTTACTTA 1013969

Qy       2326 TAAGTGATAACATGCAATATTTGTCTTTCTGTGTCTGTCTTGTTTTACTTATGATAATGG 2385
             ||||| | |||||||||||||||| |||||||| ||||  || ||| |||| |||||| |
Db    1013970 TAAGTAAGAACATGCAATATTTGACTTTCTGTTTCTGAGTTATTTCACTTAAGATAATAG 1014029

Qy       2386 CCCCCAGTTCTATCTAGGCTGCTGCAAAAGGCATGATTTCATTCTTTTTTATGGCTATGT 2445
             ||  ||| || ||| | | |||||||||| ||||| || |||||||| |||||||  ||
Db    1014030 CCTTCAGGTCCATCCATGTTGCTGCAAAAGACATGAATTGATTCTTTTGTATGGCTGAGT 1014089

Qy       2446 TCTTTCCCAATTTAGATAAAGAACACTCGCACTTGCT 2482
                | | |||   |  ||| |    ||     |||| |
Db    1014090 AGTATTCCATGGTGTATATACGTTACATTTTCTTGAT 1014126


RESULT 46
US-09-946-807-1
; Sequence 1, Application US/09946807
; Patent No. US20020165144A1
; GENERAL INFORMATION:
;  APPLICANT: Stefansson, Hreinn
;  APPLICANT:  Steinthorsdottir, Valgerdur
;  APPLICANT:  Gulcher, Jeffrey R.
;  TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
;  FILE REFERENCE: 2345.2004-001
;  CURRENT APPLICATION NUMBER: US/09/946,807
;  CURRENT FILING DATE:  2001-09-05
;  PRIOR APPLICATION NUMBER: US/09/795,668
;  PRIOR FILING DATE: 2001-02-28
;  PRIOR APPLICATION NUMBER: US 09/515,716
;  PRIOR FILING DATE: 2000-02-28
;  NUMBER OF SEQ ID NOS: 1531
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 1503841
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: r=g or a
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: y=t/u or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: m=a or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: k=g or t/u
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: s=g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: w=a or t/u
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: b=g or c or t/u
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: d=a or g or t/u
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: h=a or c or t/u
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: v=a or g or c
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1531)
;   OTHER INFORMATION: n=a or g or c or t/u
US-09-946-807-1

  Query Match             2.8%;  Score 170.2;  DB 9;  Length 1503841;
  Best Local Similarity   68.8%;  Pred. No. 7.3e-29;
  Matches  232;  Conservative    1;  Mismatches  104;  Indels    0;  Gaps    0;

Qy       2146 TTTCTCACCATCCGCTGACTTCTTGCCCCCTGGGTATTCATCACCTGAATGATGTGCATT 2205
             ||| | ||       | |||     |   | | |||  ||||||   |||  ||| | ||
Db    1013790 TTTGTTACATGGATATAACTCTGGTCTTTCAGTGTAACCATCACTCAAATAGTGTACCTT 1013849

Qy       2206 GTACCCCTTAAGTAATTTCTCACCATCCGCTGACTTCTTGCCCCCTCATCCTTCTGAGGC 2265
             |||||  ||| |||| || |||     | |   | ||   ||| | |||| ||:|||| |
Db    1013850 GTACCTATTACGTAACTTGTCATTCCTCACCCTCCTCCCACCCTCCCATCTTTYTGAGTC 1013909

Qy       2266 TCCATTGTCCATCATTCCACACTCTACATCTATGTGTACACATTATTTAGCTCCTACTTA 2325
             |||| |||| || ||||||| |||||  || |||||||| |||||||||| || ||||||
Db    1013910 TCCAATGTCTATTATTCCACCCTCTATGTCCATGTGTACTCATTATTTAGTTCTTACTTA 1013969

Qy       2326 TAAGTGATAACATGCAATATTTGTCTTTCTGTGTCTGTCTTGTTTTACTTATGATAATGG 2385
             ||||| | |||||||||||||||| |||||||| ||||  || ||| |||| |||||| |
Db    1013970 TAAGTAAGAACATGCAATATTTGACTTTCTGTTTCTGAGTTATTTCACTTAAGATAATAG 1014029

Qy       2386 CCCCCAGTTCTATCTAGGCTGCTGCAAAAGGCATGATTTCATTCTTTTTTATGGCTATGT 2445
             ||  ||| || ||| | | |||||||||| ||||| || |||||||| |||||||  ||
Db    1014030 CCTTCAGGTCCATCCATGTTGCTGCAAAAGACATGAATTGATTCTTTTGTATGGCTGAGT 1014089
```

```
Qy       2446 TCTTTCCCAATTTAGATAAAGAACACTCGCACTTGCT 2482
Db    1014090 AGTATTCCATGGTGTATATACGTTACATTTTCTTGAT 1014126
```

RESULT 47
US-10-027-632-266134
; Sequence 266134, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 266134
; LENGTH: 667
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-266134

Query Match 2.8%; Score 169.8; DB 13; Length 667;
Best Local Similarity 68.5%; Pred. No. 8.5e-31;
Matches 342; Conservative 1; Mismatches 133; Indels 23; Gaps 7;

```
Qy        308 TGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAA 367
Db        182 TGGTGAAGGGGTAGAAAGGTGTGAGATATCTTTCTTCACACATCATAGGTCATGGCTGAC 241

Qy        368 ACTCCTATAGTAAAAGACAG-GTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAA 426
Db        242 ATTCCTGTAACAAAAGACAGAGATTAACAAGAGAAAAGGATAGCAATTTATTTAATCAAT 301

Qy        427 GTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCT 486
Db        302 TTTTTACATGACACAGGAGCCTTCAGAAATGAAGACCCAATGACCCAGGGGGAACTGTCT 361

Qy        487 GTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAA 546
Db        362 ATTTTGATGCTTAGGCT--CATGAAGAATGGGTATCCACGTAGAAAGTTGACTGGACAAA 419

Qy        547 AGGATATGATCTAGTGGTAAAGGACT-CAGGGGGAAACACAGCAAGGCCTGTCTATTCAG 605
Db        420 AGTGTATAATCCATTGGTAACAGACTGAAGGGGGAAACACAATAATGCCTTTCTATTCAA 479

Qy        606 ATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGG 665
Db        480 AT--------ATTTATTAGCCTCTTTATGTAGGGTTTCTTCCTCCTAGGTAT-GGGCAGG 530

Qy        666 ACTCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAA-------AGTGGCCTTTTTAG 718
Db        531 ACCTTCCTAGAATAAGTGTCTTCAAGGGAGAAGGGAAARAAGTTAGAATGAGCCTTCTAG 590

Qy        719 ATTTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGA 778
Db        591 GTTTTATGGCTTGCTTTGG---AGAAGAATTTTTGTTTCCATGACGTGCCTTGGGGAAGA 647
```

```
Qy        779 GGAATTCTGGTTTCTGTGA 797
Db        648 AGAATTTTACTTTCTATGA 666
```

RESULT 48
US-10-027-632-266134
; Sequence 266134, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 266134
; LENGTH: 667
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-266134

Query Match 2.8%; Score 169.8; DB 17; Length 667;
Best Local Similarity 68.5%; Pred. No. 8.5e-31;
Matches 342; Conservative 1; Mismatches 133; Indels 23; Gaps 7;

```
Qy        308 TGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAA 367
Db        182 TGGTGAAGGGGTAGAAAGGTGTGAGATATCTTTCTTCACACATCATAGGTCATGGCTGAC 241

Qy        368 ACTCCTATAGTAAAAGACAG-GTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAA 426
Db        242 ATTCCTGTAACAAAAGACAGAGATTAACAAGAGAAAAGGATAGCAATTTATTTAATCAAT 301

Qy        427 GTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCT 486
Db        302 TTTTTACATGACACAGGAGCCTTCAGAAATGAAGACCCAATGACCCAGGGGGAACTGTCT 361

Qy        487 GTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAA 546
Db        362 ATTTTGATGCTTAGGCT--CATGAAGAATGGGTATCCACGTAGAAAGTTGACTGGACAAA 419

Qy        547 AGGATATGATCTAGTGGTAAAGGACT-CAGGGGGAAACACAGCAAGGCCTGTCTATTCAG 605
Db        420 AGTGTATAATCCATTGGTAACAGACTGAAGGGGGAAACACAATAATGCCTTTCTATTCAA 479

Qy        606 ATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGG 665
Db        480 AT--------ATTTATTAGCCTCTTTATGTAGGGTTTCTTCCTCCTAGGTAT-GGGCAGG 530

Qy        666 ACTCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAA-------AGTGGCCTTTTTAG 718
Db        531 ACCTTCCTAGAATAAGTGTCTTCAAGGGAGAAGGGAAARAAGTTAGAATGAGCCTTCTAG 590

Qy        719 ATTTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGA 778
```

```
Db        591 GTTTTATGGCTTGCTTTGG---AGAAGAATTTTTGTTTCCATGACGTGCCTTGGGGAAGA 647

Qy        779 GGAATTCTGGTTTCTGTGA 797
              ||||| |  ||||| |||
Db        648 AGAATTTTACTTTCTATGA 666


RESULT 49
US-10-027-632-107002/c
; Sequence 107002, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 107002
;   LENGTH: 634
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-107002

  Query Match             2.8%;  Score 168.4;  DB 13;  Length 634;
  Best Local Similarity   67.2%;  Pred. No. 1.8e-30;
  Matches  238;  Conservative    0;  Mismatches  116;  Indels    0;  Gaps    0;

Qy       2159 GCTGACTTCTTGCCCCCTGGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGT 2218
              | |||  ||  | |  ||||    |       |  ||| || | ||||||||||| |||| |
Db        395 GGTGAAGTCCRGGCTTCTGGTGTATTCATCACCAAATAATATACATTGTACCCATTAAAT 336

Qy       2219 AATTTCTCACCATCCGCTGACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATC 2278
              |||||||||    ||      | |||  | | |||| ||||| ||| || || |||| ||
Db        335 AATTTCTCATTCCCCATCCCCCTCTGACTCTCTCACCCTTCCGAGTCTTCAGTGTCTATT 276

Qy       2279 ATTCCACACTCTACATCTATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACAT 2338
              | |||||||| ||  || ||||||| |||| ||||||||  ||| |||| ||| |||||
Db        275 ACTCCACACTGTATGTCCATGTGTACCCATTTTTTAGCTCTCACTGATAAATGAGAACAT 216

Qy       2339 GCAATATTTGTCTTTCTGTGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTAT 2398
              ||| |||||| |||||||| ||||| |||||| ||||||||||| | | |||||||  ||
Db        215 GCAGTATTTGACTTTCTGTTTCTGTGTTGTTTCACTTATGATAATAGACTCCAGTTCCAT 156

Qy       2399 CTAGGCTGCTGCAAAAGGCATGATTTCATTCTTTTTTATGGCTATGTTCTTTCCCAATTT 2458
              | | | |||||||||||| |||| |||||||| |||||||||    |  | ||   | ||
Db        155 CCATGTTGCTGCAAAAGACATGGTTTCATTCTCTTTTATGGCTGAATAATATCATGAGTT 96

Qy       2459 AGATAAAGAACACTCGCACTTGCTCTTACTTCTATTTGGAATACTAATTCCTAG 2512
                |  | |    ||    ||  | |   ||  |  ||  | | |     ||||
Db         95 CATTTTAAATGCATCAATTTTAATATGTATTTCAAATGTTAAAATGCACACTAG 42


RESULT 50
US-10-027-632-107002/c
; Sequence 107002, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 107002
;   LENGTH: 634
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-107002

  Query Match             2.8%;  Score 168.4;  DB 17;  Length 634;
  Best Local Similarity   67.2%;  Pred. No. 1.8e-30;
  Matches  238;  Conservative    0;  Mismatches  116;  Indels    0;  Gaps    0;

Qy       2159 GCTGACTTCTTGCCCCCTGGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGT 2218
              | |||  ||  | |  ||||    |       |  ||| || | ||||||||||| |||| |
Db        395 GGTGAAGTCCRGGCTTCTGGTGTATTCATCACCAAATAATATACATTGTACCCATTAAAT 336

Qy       2219 AATTTCTCACCATCCGCTGACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATC 2278
              |||||||||    ||      | |||  | | |||| ||||| ||| || || |||| ||
Db        335 AATTTCTCATTCCCCATCCCCCTCTGACTCTCTCACCCTTCCGAGTCTTCAGTGTCTATT 276

Qy       2279 ATTCCACACTCTACATCTATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACAT 2338
              | |||||||| ||  || ||||||| |||| ||||||||  ||| |||| ||| |||||
Db        275 ACTCCACACTGTATGTCCATGTGTACCCATTTTTTAGCTCTCACTGATAAATGAGAACAT 216

Qy       2339 GCAATATTTGTCTTTCTGTGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTAT 2398
              ||| |||||| |||||||| ||||| |||||| ||||||||||| | | |||||||  ||
Db        215 GCAGTATTTGACTTTCTGTTTCTGTGTTGTTTCACTTATGATAATAGACTCCAGTTCCAT 156

Qy       2399 CTAGGCTGCTGCAAAAGGCATGATTTCATTCTTTTTTATGGCTATGTTCTTTCCCAATTT 2458
              | | | |||||||||||| |||| |||||||| |||||||||    |  | ||   | ||
Db        155 CCATGTTGCTGCAAAAGACATGGTTTCATTCTCTTTTATGGCTGAATAATATCATGAGTT 96

Qy       2459 AGATAAAGAACACTCGCACTTGCTCTTACTTCTATTTGGAATACTAATTCCTAG 2512
                |  | |    ||    ||  | |   ||  |  ||  | | |     ||||
Db         95 CATTTTAAATGCATCAATTTTAATATGTATTTCAAATGTTAAAATGCACACTAG 42


RESULT 51
US-10-027-632-25806
; Sequence 25806, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
```

```
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 25806
;   LENGTH: 703
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-25806

  Query Match             2.8%;  Score 168.4;  DB 13;  Length 703;
  Best Local Similarity   67.2%;  Pred. No. 1.9e-30;
  Matches  238;  Conservative    0;  Mismatches  116;  Indels    0;  Gaps    0;

Qy       2159 GCTGACTTCTTGCCCCCTGGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGT 2218
             | |||  ||  | |  ||||     |       |  ||| || | ||||||||||| |||| |
Db        240 GGTGAAGTCCRGGCTTCTGGTGTATTCATCACCAAATAATATACATTGTACCCATTAAAT 299

Qy       2219 AATTTCTCACCATCCGCTGACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATC 2278
             |||||||||     ||      | |||  | | |||| ||||| ||| || || |||| ||
Db        300 AATTTCTCATTCCCCATCCCCCTCTGACTCTCTCACCCTTCCGAGTCTTCAGTGTCTATT 359

Qy       2279 ATTCCACACTCTACATCTATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACAT 2338
             | |||||||| ||  || |||||||| |||| ||||||||  ||| |||| ||| |||||
Db        360 ACTCCACACTGTATGTCCATGTGTACCCATTTTTTAGCTCTCACTGATAAATGAGAACAT 419

Qy       2339 GCAATATTTGTCTTTCTGTGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTAT 2398
             |||·||||||| |||||||| ||||| |||||| |||||||||||| | | |||||||| ||
Db        420 GCAGTATTTGACTTTCTGTTTCTGTGTTGTTTCACTTATGATAATAGACTCCAGTTCCAT 479

Qy       2399 CTAGGCTGCTGCAAAAGGCATGATTTCATTCTTTTTTATGGCTATGTTCTTTCCCAATTT 2458
             | | | |||||||||||| |||| |||||||| |||||||||    |  | ||   | ||
Db        480 CCATGTTGCTGCAAAAGACATGGTTTCATTCTCTTTTATGGCTGAATAATATCATGAGTT 539

Qy       2459 AGATAAAGAACACTCGCACTTGCTCTTACTTCTATTTGGAATACTAATTCCTAG 2512
                | | |    ||    ||  | |   ||  |  ||  | | |     ||||
Db        540 CATTTTAAATGCATCAATTTTAATATGTATTTCAAATGTTAAAATGCACACTAG 593


RESULT 52
US-10-027-632-25806
; Sequence 25806, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 25806
;   LENGTH: 703
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-25806

  Query Match             2.8%;  Score 168.4;  DB 17;  Length 703;
  Best Local Similarity   67.2%;  Pred. No. 1.9e-30;
  Matches  238;  Conservative    0;  Mismatches  116;  Indels    0;  Gaps    0;

Qy       2159 GCTGACTTCTTGCCCCCTGGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGT 2218
             | |||  ||  | |  ||||     |       |  ||| || | ||||||||||| |||| |
Db        240 GGTGAAGTCCRGGCTTCTGGTGTATTCATCACCAAATAATATACATTGTACCCATTAAAT 299

Qy       2219 AATTTCTCACCATCCGCTGACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATC 2278
             |||||||||     ||      | |||  | | |||| ||||| ||| || || |||| ||
Db        300 AATTTCTCATTCCCCATCCCCCTCTGACTCTCTCACCCTTCCGAGTCTTCAGTGTCTATT 359

Qy       2279 ATTCCACACTCTACATCTATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACAT 2338
             | |||||||| ||  || |||||||| |||| ||||||||  ||| |||| ||| |||||
Db        360 ACTCCACACTGTATGTCCATGTGTACCCATTTTTTAGCTCTCACTGATAAATGAGAACAT 419

Qy       2339 GCAATATTTGTCTTTCTGTGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTAT 2398
             |||·||||||| |||||||| ||||| |||||| |||||||||||| | | |||||||| ||
Db        420 GCAGTATTTGACTTTCTGTTTCTGTGTTGTTTCACTTATGATAATAGACTCCAGTTCCAT 479

Qy       2399 CTAGGCTGCTGCAAAAGGCATGATTTCATTCTTTTTTATGGCTATGTTCTTTCCCAATTT 2458
             | | | |||||||||||| |||| |||||||| |||||||||    |  | ||   | ||
Db        480 CCATGTTGCTGCAAAAGACATGGTTTCATTCTCTTTTATGGCTGAATAATATCATGAGTT 539

Qy       2459 AGATAAAGAACACTCGCACTTGCTCTTACTTCTATTTGGAATACTAATTCCTAG 2512
                | | |    ||    ||  | |   ||  |  ||  | | |     ||||
Db        540 CATTTTAAATGCATCAATTTTAATATGTATTTCAAATGTTAAAATGCACACTAG 593


RESULT 53
US-10-719-993-6822
; Sequence 6822, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001496
;  CURRENT APPLICATION NUMBER: US/10/719,993
;  CURRENT FILING DATE:  2003-11-24
;  NUMBER OF SEQ ID NOS: 55342
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6822
;   LENGTH: 786452
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(786452)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-719-993-6822

  Query Match             2.8%;  Score 168.4;  DB 18;  Length 786452;
  Best Local Similarity   65.9%;  Pred. No. 1.4e-28;
  Matches  432;  Conservative    0;  Mismatches  176;  Indels   48;  Gaps   11;

Qy        312 GTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTC 371
              | |||||   ||  || ||||| ||| ||| || || ||||||||| ||         |
Db     553254 GCAGGGGCAGAATGATGCGATACTTTTCCTCATCTATTATAAGGGTCACA--------GC 553305

Qy        372 CTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTT 431
              | |||  |||||||||||||| ||||| | |   ||| ||||||| ||||||||||||||
Db     553306 CCATAACAAAAGACAGGTTAACAAGAGAAGAGGATAAGAAATTTACTTAATCAAAGTTTT 553365

Qy        432 ACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTT 491
                 ||   ||| | ||| |||||||||     |||                     | |
```

```
Db     553366 CTGAGATGGGGGGGCCTTTAGAAATGAAAGGGAAAA------------------CTAGT 553406

Qy        492 TTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGAT 551

Db     553407 TTTATGCTTATGTTGATGAAGAATGGACAGCTGTGTAGAAACATGGTTAGAC-AAAGGGT 553465

Qy        552 ATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTT 611

Db     553466 ATGATCTAATGGGAATAGATTGA-GGGGAAACCTAGCAAGGCCTGTCTGTTCAGATTCTT 553524

Qy        612 CTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTT 671

Db     553525 CTTGGCCTCTC--------TATGTAGCATTCCTTCC-CCTGGTTATAGGACAGAAC-CCT 553574

Qy        672 CTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATGGCTTG 731

Db     553575 CTGAAATGAGAGTCTTCAAAGGAGAAGAGAGAGTGTGAATTTTCTAGGTTTTATGGCTGG 553634

Qy        732 CTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCTGGTTT 791

Db     553635 CTTTAAGGGATAGGA-TTCTAGTTTCGATGACCCCACTTAGGGAAGAGGAATTCTGGTTT 553693

Qy        792 CTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGGTCAGA 851

Db     553694 CTATAACTCACTGCAGGGCAGAGACAGAGGGATGAGGGAGCA-GGTAAGAGAAGGCCAGA 553752

Qy        852 AAGAGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAGTACTTCTTAGCA 911

Db     553753 GAGA-----CCTTGCTTCTGAGGCCCTTCCAGTGTCTTTTGTTTCAA--AGCACTCAGCA 553805

Qy        912 TACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAACACTGCAATCATG 967

Db     553806 CGTCAAGGTGCCACAATTTCGGGTATTGTGTTTTGAGCCCTGTCAGAGGCAGTATG 553861
```

RESULT 54
US-09-795-668-941
; Sequence 941, Application US/09795668
; Patent No. US20020045577A1
; GENERAL INFORMATION:
; APPLICANT: Stefansson, Hreinn
; APPLICANT: Steinthorsdottir, Valgerdur
; APPLICANT: Gulcher, Jeffrey R.
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
; FILE REFERENCE: 2345.2004-001
; CURRENT APPLICATION NUMBER: US/09/795,668
; CURRENT FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 09/515,716
; PRIOR FILING DATE: 2000-02-28
; NUMBER OF SEQ ID NOS: 1531
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 941
; LENGTH: 401
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-795-668-941

Query Match 2.8%; Score 167.8; DB 9; Length 401;
Best Local Similarity 71.2%; Pred. No. 1.9e-30;
Matches 220; Conservative 1; Mismatches 88; Indels 0; Gaps 0;

```
Qy       2146 TTTCTCACCATCCGCTGACTTCTTGCCCCCTGGGTATTCATCACCTGAATGATGTGCATT 2205

Db         88 TTTGTTACATGGATATAACTCTGGTCTTTCAGTGTAACCATCACTCAAATAGTGTACCTT 147

Qy       2206 GTACCCCTTAAGTAATTTCTCACCATCCGCTGACTTCTTGCCCCCTCATCCTTCTGAGGC 2265

Db        148 GTACCTATTACGTAACTTGTCATTCCTCACCCTCCTCCCACCCTCCCATCTTTYTGAGTC 207

Qy       2266 TCCATTGTCCATCATTCCACACTCTACATCTATGTGTACACATTATTTAGCTCCTACTTA 2325

Db        208 TCCAATGTCTATTATTCCACCCTCTATGTCCATGTGTACTCATTATTTAGTTCTTACTTA 267

Qy       2326 TAAGTGATAACATGCAATATTTGTCTTTCTGTGTCTGTCTTGTTTTACTTATGATAATGG 2385

Db        268 TAAGTAAGAACATGCAATATTTGACTTTCTGTTTCTGAGTTATTTCACTTAAGATAATAG 327

Qy       2386 CCCCCAGTTCTATCTAGGCTGCTGCAAAAGGCATGATTTCATTCTTTTTTATGGCTATGT 2445

Db        328 CCTTCAGGTCCATCCATGTTGCTGCAAAAGACATGAATTGATTCTTTTGTATGGCTGAGT 387

Qy       2446 TCTTTCCCA 2454

Db        388 AGTATTCCA 396
```

RESULT 55
US-09-795-686-941
; Sequence 941, Application US/09795686
; Patent No. US20020094954A1
; GENERAL INFORMATION:
; APPLICANT: Stefansson, Hreinn
; APPLICANT: Steinthorsdottir, Valgerdur
; APPLICANT: Gulcher, Jeffrey R.
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
; FILE REFERENCE: 2345.2005-001
; CURRENT APPLICATION NUMBER: US/09/795,686
; CURRENT FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 09/515,715
; PRIOR FILING DATE: 2000-02-28
; NUMBER OF SEQ ID NOS: 1531
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 941
; LENGTH: 401
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-795-686-941

Query Match 2.8%; Score 167.8; DB 9; Length 401;
Best Local Similarity 71.2%; Pred. No. 1.9e-30;
Matches 220; Conservative 1; Mismatches 88; Indels 0; Gaps 0;

```
Qy       2146 TTTCTCACCATCCGCTGACTTCTTGCCCCCTGGGTATTCATCACCTGAATGATGTGCATT 2205

Db         88 TTTGTTACATGGATATAACTCTGGTCTTTCAGTGTAACCATCACTCAAATAGTGTACCTT 147

Qy       2206 GTACCCCTTAAGTAATTTCTCACCATCCGCTGACTTCTTGCCCCCTCATCCTTCTGAGGC 2265

Db        148 GTACCTATTACGTAACTTGTCATTCCTCACCCTCCTCCCACCCTCCCATCTTTYTGAGTC 207

Qy       2266 TCCATTGTCCATCATTCCACACTCTACATCTATGTGTACACATTATTTAGCTCCTACTTA 2325

Db        208 TCCAATGTCTATTATTCCACCCTCTATGTCCATGTGTACTCATTATTTAGTTCTTACTTA 267

Qy       2326 TAAGTGATAACATGCAATATTTGTCTTTCTGTGTCTGTCTTGTTTTACTTATGATAATGG 2385

Db        268 TAAGTAAGAACATGCAATATTTGACTTTCTGTTTCTGAGTTATTTCACTTAAGATAATAG 327

Qy       2386 CCCCCAGTTCTATCTAGGCTGCTGCAAAAGGCATGATTTCATTCTTTTTTATGGCTATGT 2445

Db        328 CCTTCAGGTCCATCCATGTTGCTGCAAAAGACATGAATTGATTCTTTTGTATGGCTGAGT 387

Qy       2446 TCTTTCCCA 2454

Db        388 AGTATTCCA 396
```

RESULT 56
US-09-946-807-941
; Sequence 941, Application US/09946807
; Patent No. US20020165144A1
; GENERAL INFORMATION:
; APPLICANT: Stefansson, Hreinn
; APPLICANT: Steinthorsdottir, Valgerdur
; APPLICANT: Gulcher, Jeffrey R.
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE

```
;  FILE REFERENCE: 2345.2004-001
;  CURRENT APPLICATION NUMBER: US/09/946,807
;  CURRENT FILING DATE:  2001-09-05
;  PRIOR APPLICATION NUMBER: US/09/795,668
;  PRIOR FILING DATE: 2001-02-28
;  PRIOR APPLICATION NUMBER: US 09/515,716
;  PRIOR FILING DATE: 2000-02-28
;  NUMBER OF SEQ ID NOS: 1531
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 941
;   LENGTH: 401
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-946-807-941

  Query Match             2.8%;  Score 167.8;  DB 9;  Length 401;
  Best Local Similarity   71.2%;  Pred. No. 1.9e-30;
  Matches  220;  Conservative    1;  Mismatches   88;  Indels    0;  Gaps    0;

Qy       2146 TTTCTCACCATCCGCTGACTTCTTGCCCCCTGGGTATTCATCACCTGAATGATGTGCATT 2205
             ||| | ||       | |||     |   | | |||  ||||||   |||  ||| | ||
Db         88 TTTGTTACATGGATATAACTCTGGTCTTTCAGTGTAACCATCACTCAAATAGTGTACCTT 147

Qy       2206 GTACCCCTTAAGTAATTTCTCACCATCCGCTGACTTCTTGCCCCCTCATCCTTCTGAGGC 2265
             |||||  ||| |||| || |||      | |    | ||   ||| | |||| ||:|||| |
Db        148 GTACCTATTACGTAACTTGTCATTCCTCACCCTCCTCCCACCCTCCCATCTTTYTGAGTC 207

Qy       2266 TCCATTGTCCATCATTCCACACTCTACATCTATGTGTACACATTATTTAGCTCCTACTTA 2325
             |||| |||| || ||||||| |||||  || ||||||| |||||||| || ||||||
Db        208 TCCAATGTCTATTATTCCACCCTCTATGTCCATGTGTACTCATTATTTAGTTCTTACTTA 267

Qy       2326 TAAGTGATAACATGCAATATTTGTCTTTCTGTGTCTGTCTTGTTTTACTTATGATAATGG 2385
             ||||| | |||||||||||||||| |||||||| ||||  || ||| |||| |||||| |
Db        268 TAAGTAAGAACATGCAATATTTGACTTTCTGTTTCTGAGTTATTTCACTTAAGATAATAG 327

Qy       2386 CCCCCAGTTCTATCTAGGCTGCTGCAAAAGGCATGATTTCATTCTTTTTTATGGCTATGT 2445
             ||  ||| || ||| | | |||||||||||| ||||| || |||||||| |||||||  ||
Db        328 CCTTCAGGTCCATCCATGTTGCTGCAAAAGACATGAATTGATTCTTTTGTATGGCTGAGT 387

Qy       2446 TCTTTCCCA 2454
                | | |||
Db        388 AGTATTCCA 396


RESULT 57
US-10-235-192A-47/c
; Sequence 47, Application US/10235192A
; Publication No. US20040043389A1
; GENERAL INFORMATION:
;  APPLICANT: McCarthy, Jeanette
;  TITLE OF INVENTION: Methods and Compositions for Identifying
;  TITLE OF INVENTION:  Risk Factors for Abnormal Lipid Levels and the Diseases
;  TITLE OF INVENTION:  and Disorders Associated Therewith
;  FILE REFERENCE: MMI-011
;  CURRENT APPLICATION NUMBER: US/10/235,192A
;  CURRENT FILING DATE:  2002-09-04
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 47
;   LENGTH: 182328
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-235-192A-47

  Query Match             2.8%;  Score 167.8;  DB 17;  Length 182328;
  Best Local Similarity   62.9%;  Pred. No. 7.9e-29;
  Matches  397;  Conservative    0;  Mismatches  187;  Indels   47;  Gaps    7;

Qy        312 GTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTC 371
              ||| |||   ||| ||| ||||| |||||||  ||||||||||||||     ||| | |
Db     127069 GTATGGGCAGAAAGGTGCGATACTTTTTCTCACCCATCATAAGGGTCATGGTCAATGCCC 127010

Qy        372 CTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTT 431
              ||||| ||||| |||| |||| ||||| ||| | |||||| | ||||||| || ||||||
Db     127009 CTATAATAAAATACAGTTTAACAAGAGAAAAGCATAACAAGTGTATTTAAGCACAGTTTT 126950

Qy        432 ACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTT 491
              |  ||| |||||||||||||||||||||||||||||| | |   ||| ||      |  |||
Db     126949 ATGTGATATGGGAGTCTTCAGAAATGAAGACCCAAAAAACTCAGGGGAAACTCTTCATTT 126890

Qy        492 TTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAA--GG 549
              ||  ||| | ||| |||||||||||||| ||        |    |||   ||||||  ||
Db     126889 TTAGGCTTACGTTTGATGAAGAATGGACAGTCCTGTAGAAAAGTGATGGAACAAAAAGGG 126830

Qy        550 ATATGATCTAGTGGTAAAGGACT-CAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATT 608
              | |    |||||||||||  | ||  | ||||||| ||||||| || ||| ||| ||||
Db     126829 AAAGATCCTAGTGGTAATAGTCTGAATGGGGAAACTCAGCAAGGTCTATCTGTTCTGATT 126770

Qy        609 CTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTAT-GGGGCAGGAC 667
              |||||| ||||          ||| | ||||||||||||||  | | | | ||| |
Db     126769 CTTCTTGGTCTC----------TGTGTGGCATTCTTTCCTCCTGGATGTAGTGTCAGACC 126720

Qy        668 TCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATGG 727
              ||| |  ||||| |||||| |    |      | || |   |    | | ||| |
Db     126719 CCTTTTGGAATGACGGTCTTAAGATCAACTATGGGACAAGGTAGGTCAGATAATTTCTTT 126660

Qy        728 CTTGC--------------------------TTCGGGGAAGAGGAGTTCTAGTTTCTAT 760
              | |                          |    || | || ||| |  ||
Db     126659 GTAACCAGCTCCTATACAGAAAGGAAGGGGAAGGTTAGATTTAGTATTTTTAGGCTTCAT 126600

Qy        761 GACCCATCTTGGGGAAGAGGAATTCTGGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGA 820
              | |    ||  |||||| |  | || |||||| |||| ||| ||   | ||||||||||
Db     126599 GGCTGGCTTTCAGGAAGAAGGGTCCTAGTTTCTATGACCTGCCTTGGGGGAGAAAGAGGA 126540

Qy        821 GTAAGAGGCAGGAGGGCAGGAGATGGTCAGAAAGAGACTTGGCTGCTTCTGAGGGCTTCC 880
              ||| |||  || | | ||| ||| |    |||||||||| ||| ||           |||||
Db     126539 GTAGGAGAGAGAAAGACAGAAGAGG----GAAAGAGACTTTGCTTCT--GAGACTCTTCC 126486

Qy        881 GCTCTCCTTTAGTTCCAAGTACTTCTTAGCA 911
                | | |||  |||||  |||||||  |  || |
Db     126485 AGTGTTCTTCAGTTCAAAGTACTCCCAAGGA 126455


RESULT 58
US-10-027-632-200319/c
; Sequence 200319, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 200319
;   LENGTH: 637
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-200319
```

```
  Query Match             2.7%;  Score 167.2;  DB 13;  Length 637;
  Best Local Similarity   72.3%;  Pred. No. 3.6e-30;
  Matches  217;  Conservative    0;  Mismatches   83;  Indels    0;  Gaps    0;

Qy       2159 GCTGACTTCTTGCCCCCTGGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGT 2218
             | |||  ||  | |  ||||    |       |  ||| || | |||||||||| |||| |
Db        302 GGTGAAGTCCRGGCTTCTGGTGTATTCATCACCAAATAATATACATTGTACCCATTAAAT 243

Qy       2219 AATTTCTCACCATCCGCTGACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATC 2278
             |||||||||    ||     | |||  | | |||| ||||| ||| || || |||| ||
Db        242 AATTTCTCATTCCCCATCCCCCTCTGACTCTCTCACCCTTCCGAGTCTTCAGTGTCTATT 183

Qy       2279 ATTCCACACTCTACATCTATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACAT 2338
             | |||||||| ||  || |||||||| |||| ||||||||  ||| |||| ||| |||||
Db        182 ACTCCACACTGTATGTCCATGTGTACCCATTTTTTAGCTCTCACTGATAAATGAGAACAT 123

Qy       2339 GCAATATTTGTCTTTCTGTGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTAT 2398
             |||  |||||| |||||||| ||||| |||||| |||||||||||| | | ||||||| ||
Db        122 GCAGTATTTGACTTTCTGTTTCTGTGTTGTTTCACTTATGATAATAGACTCCAGTTCCAT 63

Qy       2399 CTAGGCTGCTGCAAAAGGCATGATTTCATTCTTTTTTATGGCTATGTTCTTTCCCAATTT 2458
             | | | |||||||||||  |||| |||||||| ||||||||||   |  | ||   | ||
Db         62 CCATGTTGCTGCAAAAGACATGGTTTCATTCTCTTTTATGGCTGAATAATATCATGAGTT 3


RESULT 59
US-10-027-632-200319/c
; Sequence 200319, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 200319
;   LENGTH: 637
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-200319

  Query Match             2.7%;  Score 167.2;  DB 17;  Length 637;
  Best Local Similarity   72.3%;  Pred. No. 3.6e-30;
  Matches  217;  Conservative    0;  Mismatches   83;  Indels    0;  Gaps    0;

Qy       2159 GCTGACTTCTTGCCCCCTGGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGT 2218
             | |||  ||  | |  ||||    |       |  ||| || | |||||||||| |||| |
Db        302 GGTGAAGTCCRGGCTTCTGGTGTATTCATCACCAAATAATATACATTGTACCCATTAAAT 243

Qy       2219 AATTTCTCACCATCCGCTGACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATC 2278
             |||||||||    ||     | |||  | | |||| ||||| ||| || || |||| ||
Db        242 AATTTCTCATTCCCCATCCCCCTCTGACTCTCTCACCCTTCCGAGTCTTCAGTGTCTATT 183

Qy       2279 ATTCCACACTCTACATCTATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACAT 2338
             | |||||||| ||  || |||||||| |||| ||||||||  ||| |||| ||| |||||
Db        182 ACTCCACACTGTATGTCCATGTGTACCCATTTTTTAGCTCTCACTGATAAATGAGAACAT 123

Qy       2339 GCAATATTTGTCTTTCTGTGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTAT 2398
             |||  |||||| |||||||| ||||| |||||| |||||||||||| | | ||||||| ||
Db        122 GCAGTATTTGACTTTCTGTTTCTGTGTTGTTTCACTTATGATAATAGACTCCAGTTCCAT 63

Qy       2399 CTAGGCTGCTGCAAAAGGCATGATTTCATTCTTTTTTATGGCTATGTTCTTTCCCAATTT 2458
             | | | |||||||||||  |||| |||||||| ||||||||||   |  | ||   | ||
Db         62 CCATGTTGCTGCAAAAGACATGGTTTCATTCTCTTTTATGGCTGAATAATATCATGAGTT 3


RESULT 60
US-10-108-260A-1134
; Sequence 1134, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
;  APPLICANT: HELIX RESEARCH INSTITUTE
;  TITLE OF INVENTION: No. US20040005560A1el full length cDNA
;  FILE REFERENCE: H1-A0106
;  CURRENT APPLICATION NUMBER: US/10/108,260A
;  CURRENT FILING DATE:  2002-03-27
;  NUMBER OF SEQ ID NOS: 5458
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1134
;   LENGTH: 3769
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-108-260A-1134

  Query Match             2.7%;  Score 166;  DB 17;  Length 3769;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-29;
  Matches  166;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5463 GACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGCC 5522
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGCC 60

Qy       5523 GAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAATA 5582
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAATA 120

Qy       5583 AGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             |||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 166


RESULT 61
US-10-027-632-264720
; Sequence 264720, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 264720
```

```
;   LENGTH: 542
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-264720

  Query Match             2.7%;  Score 164.2;  DB 13;  Length 542;
  Best Local Similarity   73.5%;  Pred. No. 1.8e-29;
  Matches  208;  Conservative    1;  Mismatches   74;  Indels    0;  Gaps    0;

Qy       2177 GGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCT 2236
              | ||| |   ||||  ||| |||| |||||||||| |||||||||| |||       |
Db        244 GTGTAATGGCCACCCAAATAATGTACATTGTACCCATTAAGTAATTTTTCATTCCTTACC 303

Qy       2237 GACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTCTACATCT 2296
                | ||:    | |  | ||||| ||| |  || |||| || |||||||||||| ||
Db        304 CCCCTCYCAGTCTCCTACCCTTCAGAGTCCTCAGTGTCTATTATTCCACACTCTATATTG 363

Qy       2297 ATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGTCTTTCTG 2356
              |||||||||||||||||||||  ||||| ||||||  ||||  | ||||| |
Db        364 ATGTGTACACATTATTTAGCTCTCACTTACAAGTGAGAACATGTGGTATTCGACTTTCCG 423

Qy       2357 TGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAAGG 2416
              | ||||  || || |||||| || ||||||| ||||||| | | | |  ||||||||||
Db        424 TTTCTGGATTTTTATACTTAAGACAATGGCCTCCAGTTCCAACCATGTGGCTGCAAAAGA 483

Qy       2417 CATGATTTCATTCTTTTTTATGGCTATGTTCTTTCCCAATTTA 2459
              |||||||||||||||||||||||    | || | ||| | ||
Db        484 CATGATTTCATTCTTTTTTATGGCAGAATGCTATTCCATTATA 526


RESULT 62
US-10-027-632-264720
; Sequence 264720, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 264720
;   LENGTH: 542
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-264720

  Query Match             2.7%;  Score 164.2;  DB 17;  Length 542;
  Best Local Similarity   73.5%;  Pred. No. 1.8e-29;
  Matches  208;  Conservative    1;  Mismatches   74;  Indels    0;  Gaps    0;

Qy       2177 GGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCT 2236
              | ||| |   ||||  ||| |||| |||||||||| |||||||||| |||       |
Db        244 GTGTAATGGCCACCCAAATAATGTACATTGTACCCATTAAGTAATTTTTCATTCCTTACC 303

Qy       2237 GACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTCTACATCT 2296
                | ||:    | |  | ||||| ||| |  || |||| || |||||||||||| ||
```

```
Db        304 CCCCTCYCAGTCTCCTACCCTTCAGAGTCCTCAGTGTCTATTATTCCACACTCTATATTG 363

Qy       2297 ATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGTCTTTCTG 2356
              |||||||||||||||||||||  ||||| ||||||  ||||  | ||||| |
Db        364 ATGTGTACACATTATTTAGCTCTCACTTACAAGTGAGAACATGTGGTATTCGACTTTCCG 423

Qy       2357 TGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAAGG 2416
              | ||||  || || |||||| || ||||||| ||||||| | | | |  ||||||||||
Db        424 TTTCTGGATTTTTATACTTAAGACAATGGCCTCCAGTTCCAACCATGTGGCTGCAAAAGA 483

Qy       2417 CATGATTTCATTCTTTTTTATGGCTATGTTCTTTCCCAATTTA 2459
              |||||||||||||||||||||||    | || | ||| | ||
Db        484 CATGATTTCATTCTTTTTTATGGCAGAATGCTATTCCATTATA 526


RESULT 63
US-10-719-993-6969
; Sequence 6969, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001496
;  CURRENT APPLICATION NUMBER: US/10/719,993
;  CURRENT FILING DATE:  2003-11-24
;  NUMBER OF SEQ ID NOS: 55342
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6969
;   LENGTH: 26920
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-719-993-6969

  Query Match             2.7%;  Score 163.8;  DB 18;  Length 26920;
  Best Local Similarity   74.9%;  Pred. No. 2.4e-28;
  Matches  218;  Conservative    0;  Mismatches   72;  Indels    1;  Gaps    1;

Qy       2177 GGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCT 2236
              | |||  |||||||  ||| || | | |||||||| |||||| |||||||| |   | |
Db      19925 GTGTAACCATCACCCTAATAATATACGTTGTACCCATTAAGTTATTTCTCATCCCTCACC 19984

Qy       2237 GACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTCTACATCT 2296
                || |   ||   ||| ||||||||| ||||| |||| || |||||||||| |  ||
Db      19985 CCCTCCCA-CCTTGTCACCCTTCTGAGTCTCCARTGTCTATTATTCCACACTCCATGTCC 20043

Qy       2297 ATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGTCTTTCTG 2356
              | |||||||||||||||||||| ||||| |||||| ||||||   |||||| |||||||
Db      20044 ACGTGTACACATTATTTAGCTCCCACTTACAAGTGAGAACATGTGGTATTTGACTTTCTG 20103

Qy       2357 TGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAAGG 2416
              | | ||  || ||| |||||  |||||| || ||||||  ||| | | ||||||||||
Db      20104 TTTTTGAGTTATTTCACTTAAAATAATGACCTCCAGTTTCATCCATGTTGCTGCAAAAGA 20163

Qy       2417 CATGATTTCATTCTTTTTTATGGCTATGTTCTTTCCCAATTTAGATAAAGA 2467
              |||||| || |||||||||||||  |||  | | |||  |  | ||| | |
Db      20164 CATGATCTCGTTCTTTTTTATGGCAGTGTAGTATTCCATTCCATATACATA 20214


RESULT 64
US-09-741-149-3
; Sequence 3, Application US/09741149
; Patent No. US20020031800A1
; GENERAL INFORMATION:
;  APPLICANT: LI, Zhenya et al
;  TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
;  TITLE OF INVENTION:  NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
;  TITLE OF INVENTION:  AND USES THEREOF
;  FILE REFERENCE: CL000780
;  CURRENT APPLICATION NUMBER: US/09/741,149
;  CURRENT FILING DATE:  2000-12-21
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
```

```
; SEQ ID NO 3
;   LENGTH: 34337
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(34337)
;   OTHER INFORMATION: n = A,T,C or G
US-09-741-149-3

  Query Match             2.7%;  Score 163.2;  DB 9;  Length 34337;
  Best Local Similarity   72.8%;  Pred. No. 3.9e-28;
  Matches  268;  Conservative    0;  Mismatches   88;  Indels   12;  Gaps    4;

Qy        489 TTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAG 548
              ||||| |||| || ||   ||||| || |  |||        |    |||| ||||||||
Db       4905 TTTTTATGCTTAGATTACCTGAAGTATAGGCAGCCATATAGAAATGTGATTGGACAAAAG 4964

Qy        549 GATATGATCTAGTGGTAAAGGACTCAGGG--GGAAACACAGCAAGGCCTGTCTATTCAGA 606
              | || |||||||||  ||  |||| | ||  | |||| ||||||| ||||||| ||||||
Db       4965 GGTA-GATCTAGTGAGAATAGACTAAAGGAAGCAAACCCAGCAAGTCCTGTCTGTTCAGA 5023

Qy        607 TTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGA 666
              || ||||||  |||||         ||  | |||||| |||||||| ||||   |||||
Db       5024 TTTTTCTTGGCCTCTC--------TGAGTGGCATTCCTTCCTCCTGGATATGAAACAGGA 5075

Qy        667 CTCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATG 726
              |   ||   ||||||| ||||||||||||||||||||  || ||||| ||| |||| ||
Db       5076 CGTCTCCGGAATGAGGATCTTCAAGGGAGAAGGGAGAAGATGACCTTTCTAGGTTTTGTG 5135

Qy        727 GCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCT 786
              |||||||| |||||||||| |||||||||||||||||| |||| ||  || ||| |||
Db       5136 GCTTGCTTTGGGGAAGAGGCATTCTAGTTTCTATGACCCACCTTGAGG-GGAAGAAATCT 5194

Qy        787 GGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGG 846
               || |||||||||||||     ||  ||| | ||||  ||| ||||||   || | | ||
Db       5195 TGTGTCTGTGACTTGCTTCAGGGACAAAAAAAGAGTGGGAGACAGGAGTAAAGAAAAAGG 5254

Qy        847 TCAGAAAG 854
               ||| |||
Db       5255 CCAGGAAG 5262


RESULT 65
US-10-385-614-3
; Sequence 3, Application US/10385614
; Publication No. US20030157649A1
; GENERAL INFORMATION:
;  APPLICANT: LI, Zhenya et al
;  TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
;  TITLE OF INVENTION:  NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
;  TITLE OF INVENTION:  AND USES THEREOF
;  FILE REFERENCE: CL000780
;  CURRENT APPLICATION NUMBER: US/10/385,614
;  CURRENT FILING DATE:  2003-03-12
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 34337
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(34337)
;   OTHER INFORMATION: n = A,T,C or G
US-10-385-614-3

  Query Match             2.7%;  Score 163.2;  DB 16;  Length 34337;
  Best Local Similarity   72.8%;  Pred. No. 3.9e-28;
  Matches  268;  Conservative    0;  Mismatches   88;  Indels   12;  Gaps    4;

Qy        489 TTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAG 548
              ||||| |||| || ||   ||||| || |  |||        |    |||| ||||||||

Db       4905 TTTTTATGCTTAGATTACCTGAAGTATAGGCAGCCATATAGAAATGTGATTGGACAAAAG 4964

Qy        549 GATATGATCTAGTGGTAAAGGACTCAGGG--GGAAACACAGCAAGGCCTGTCTATTCAGA 606
              | || |||||||||  ||  |||| | ||  | |||| ||||||| ||||||| ||||||

Db       4965 GGTA-GATCTAGTGAGAATAGACTAAAGGAAGCAAACCCAGCAAGTCCTGTCTGTTCAGA 5023

Qy        607 TTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGA 666
              || ||||||  |||||         ||  | |||||| |||||||| ||||   |||||

Db       5024 TTTTTCTTGGCCTCTC--------TGAGTGGCATTCCTTCCTCCTGGATATGAAACAGGA 5075

Qy        667 CTCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATG 726
              |   ||   ||||||| ||||||||||||||||||||  || ||||| ||| |||| ||

Db       5076 CGTCTCCGGAATGAGGATCTTCAAGGGAGAAGGGAGAAGATGACCTTTCTAGGTTTTGTG 5135

Qy        727 GCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCT 786
              |||||||| |||||||||| |||||||||||||||||| |||| ||  || ||| |||

Db       5136 GCTTGCTTTGGGGAAGAGGCATTCTAGTTTCTATGACCCACCTTGAGG-GGAAGAAATCT 5194

Qy        787 GGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGG 846
               || |||||||||||||     ||  ||| | ||||  ||| ||||||   || | | ||

Db       5195 TGTGTCTGTGACTTGCTTCAGGGACAAAAAAAGAGTGGGAGACAGGAGTAAAGAAAAAGG 5254

Qy        847 TCAGAAAG 854
               ||| |||

Db       5255 CCAGGAAG 5262


RESULT 66
US-10-027-632-264719
; Sequence 264719, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 264719
;   LENGTH: 542
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-264719

  Query Match             2.7%;  Score 163;  DB 13;  Length 542;
  Best Local Similarity   73.5%;  Pred. No. 3.5e-29;
  Matches  208;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy       2177 GGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCT 2236
              | ||| |   ||||  ||| |||| |||||||||| |||||||||| |||       |
Db        244 GTGTAATGGCCACCCAAATAATGTACATTGTACCCATTAAGTAATTTTTCATTCCTTACC 303

Qy       2237 GACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTCTACATCT 2296
                | ||    | |  | ||||| ||| |  || |||| || ||||||||||||| ||
Db        304 CCCCTCCCAGTCTCCTACCCTTCAGAGTCCTCAGTGTCTATTATTCCACACTCTATATTG 363

Qy       2297 ATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGTCTTTCTG 2356

Db       4905 TTTTTATGCTTAGATTACCTGAAGTATAGGCAGCCATATAGAAATGTGATTGGACAAAAG 4964

Qy        549 GATATGATCTAGTGGTAAAGGACTCAGGG--GGAAACACAGCAAGGCCTGTCTATTCAGA 606
              | || |||||||||  ||  |||| | ||  | |||| ||||||| ||||||| ||||||

Db       4965 GGTA-GATCTAGTGAGAATAGACTAAAGGAAGCAAACCCAGCAAGTCCTGTCTGTTCAGA 5023

Qy        607 TTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGA 666
              || ||||||  |||||         ||  | |||||| |||||||| ||||   |||||

Db       5024 TTTTTCTTGGCCTCTC--------TGAGTGGCATTCCTTCCTCCTGGATATGAAACAGGA 5075

Qy        667 CTCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATG 726
              |   ||   ||||||| ||||||||||||||||||||  || ||||| ||| |||| ||

Db       5076 CGTCTCCGGAATGAGGATCTTCAAGGGAGAAGGGAGAAGATGACCTTTCTAGGTTTTGTG 5135

Qy        727 GCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCT 786
              |||||||| |||||||||| |||||||||||||||||| |||| ||  || ||| |||

Db       5136 GCTTGCTTTGGGGAAGAGGCATTCTAGTTTCTATGACCCACCTTGAGG-GGAAGAAATCT 5194

Qy        787 GGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGG 846
               || |||||||||||||     ||  ||| | ||||  ||| ||||||   || | | ||

Db       5195 TGTGTCTGTGACTTGCTTCAGGGACAAAAAAAGAGTGGGAGACAGGAGTAAAGAAAAAGG 5254

Qy        847 TCAGAAAG 854
               ||| |||

Db       5255 CCAGGAAG 5262
```

```
Db        364 ATGTGTACACATTATTTAGCTCTCACTTACAAGTGAGAACATGTGGTATTCGACTTTCCG 423

Qy       2357 TGTCTGTCTTGTTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAAGG 2416

Db        424 TTTCTGGATTTTTATACTTAAGACAATGGCCTCCAGTTCCAACCATGTGGCTGCAAAAGA 483

Qy       2417 CATGATTTCATTCTTTTTTATGGCTATGTTCTTTCCCAATTTA 2459

Db        484 CATGATTTCATTCTTTTTTATGGCAGAATGCTATTCCATTATA 526


RESULT 67
US-10-027-632-264719
; Sequence 264719, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 264719
;   LENGTH: 542
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-264719

  Query Match             2.7%;  Score 163;  DB 17;  Length 542;
  Best Local Similarity   73.5%;  Pred. No. 3.5e-29;
  Matches  208;  Conservative    0;  Mismatches   75;  Indels    0;  Gaps    0;

Qy       2177 GGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCT 2236

Db        244 GTGTAATGGCCACCCAAATAATGTACATTGTACCCATTAAGTAATTTTTCATTCCTTACC 303

Qy       2237 GACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTCTACATCT 2296

Db        304 CCCCTCCCAGTCTCCTACCCTTCAGAGTCCTCAGTGTCTATTATTCCACACTCTATATTG 363

Qy       2297 ATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGTCTTTCTG 2356

Db        364 ATGTGTACACATTATTTAGCTCTCACTTACAAGTGAGAACATGTGGTATTCGACTTTCCG 423

Qy       2357 TGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAAGG 2416

Db        424 TTTCTGGATTTTTATACTTAAGACAATGGCCTCCAGTTCCAACCATGTGGCTGCAAAAGA 483

Qy       2417 CATGATTTCATTCTTTTTTATGGCTATGTTCTTTCCCAATTTA 2459

Db        484 CATGATTTCATTCTTTTTTATGGCAGAATGCTATTCCATTATA 526


RESULT 68
US-10-027-632-49405
; Sequence 49405, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 49405
;   LENGTH: 670
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(670)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-49405

  Query Match             2.7%;  Score 162.4;  DB 13;  Length 670;
  Best Local Similarity   64.0%;  Pred. No. 5.7e-29;
  Matches  310;  Conservative    2;  Mismatches  158;  Indels   14;  Gaps    4;

Qy        314 AGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTCCT 373

Db         90 AGGGTGGAAAAGGTGTGATACTTTTCTTTACCCATCTTAAGGGTCACAGCTGACACTACC 149

Qy        374 ATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTAC 433

Db        150 ATAACAAAAAACAGGTTAGCAAGATAAATGGATAAC----TTATTGAATCACAGTTTTAT 205

Qy        434 ATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTTTT 493

Db        206 GGGATGTAAGTGCCTTCASAAATGAAGACCCAAAGACCCAGGGAAAGCTGTC-CATTTTT 264

Qy        494 TTGCTGAGGTTCGATGAAGAATGGATAGC-ATGTAGCCATGTAGATTAGACAAAAGGATA 552

Db        265 ATGCTTAGATGAGATGAAGAATGGACAGCTGTGTAGAAATGTGATTGTACAAAAGGAGTA 324

Qy        553 TGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTTC 612

Db        325 TGATCTAATAGTAACAGACTGAAGGAGGAATCCAGCAAGTCCTGTCTGTTCAGATTCTTC 384

Qy        613 TTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTTC 672

Db        385 TTGGCCTMTC--------TGTGCAGCTTTCCTTCTTCCTGAGTGTAGGGCAAGACCCTTT 436

Qy        673 TTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATGGCTTGC 732

Db        437 TGGAACGAGGGTCTTATGACAAGGTACATCAAAGAATGTCTTTATGGCCAGCTCTTACAC 496

Qy        733 TTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCTGGTTTC 792

Db        497 AAAAAGGTGGGGGAAGGCTAGATTAATATTTCTAGGTTTTTCTAGAGGATCCCCGGGTAC 556

Qy        793 TGTG 796

Db        557 CGAG 560
```

```
RESULT 69
US-10-027-632-49406
; Sequence 49406, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 49406
;   LENGTH: 670
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(670)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-49406

  Query Match             2.7%;  Score 162.4;  DB 13;  Length 670;
  Best Local Similarity   64.0%;  Pred. No. 5.7e-29;
  Matches  310;  Conservative    2;  Mismatches  158;  Indels   14;  Gaps    4;

Qy        314 AGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTCCT 373
              ||||  ||||| |||||||||| |||   |    ||||| |||||||| || |  | ||| |
Db         90 AGGGTGGAAAAGGTGTGATACTTTTCTTTACCCATCTTAAGGGTCACAGCTGACACTACC 149

Qy        374 ATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTAC 433
              |||  |||| ||||||||  ||||  ||     ||||     ||||| ||||| |||||||
Db        150 ATAACAAAAAACAGGTTAGCAAGATAAATGGATAAC----TTATTGAATCACAGTTTTAT 205

Qy        434 ATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTTTT 493
              ||  |  | | |||||:||||||||||||||||||||||||| || |||||   |||||
Db        206 GGGATGTAAGTGCCTTCASAAATGAAGACCCAAAGACCCAGGGAAAGCTGTC-CATTTTT 264

Qy        494 TTGCTGAGGTTCGATGAAGAATGGATAGC-ATGTAGCCATGTAGATTAGACAAAAGGATA 552
               |||| || |  |||||||||||||| |||  |||||  ||||    |     ||| |  ||
Db        265 ATGCTTAGATGAGATGAAGAATGGACAGCTGTGTAGAAATGTGATTGTACAAAAGGAGTA 324

Qy        553 TGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTTC 612
              ||||||| | ||||  |||| | || | ||  |||||| ||||||| ||||||||||||
Db        325 TGATCTAATAGTAACAGACTGAAGGAGGAATCCAGCAAGTCCTGTCTGTTCAGATTCTTC 384

Qy        613 TTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTTC 672
              |||  ||:||        |||  ||| ||| ||| |||||||| | ||||| ||| |||
Db        385 TTGGCCTMTC--------TGTGCAGCTTTCCTTCTTCCTGAGTGTAGGGCAAGACCCTTT 436

Qy        673 TTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATGGCTTGC 732
              |  || |||||||||       ||       ||||      |||  |      |      |
Db        437 TGGAACGAGGGTCTTATGACAAGGTACATCAAAGAATGTCTTTATGGCCAGCTCTTACAC 496

Qy        733 TTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCTGGTTTC 792
                  ||  | |||   |||| ||       | |  ||     ||||||   | || | |
```

```
Db        497 AAAAAGGTGGGGGAAGGCTAGATTAATATTTCTAGGTTTTTCTAGAGGATCCCCGGGTAC 556

Qy        793 TGTG 796
                | |
Db        557 CGAG 560


RESULT 70
US-10-027-632-49405
; Sequence 49405, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 49405
;   LENGTH: 670
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(670)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-49405

  Query Match             2.7%;  Score 162.4;  DB 17;  Length 670;
  Best Local Similarity   64.0%;  Pred. No. 5.7e-29;
  Matches  310;  Conservative    2;  Mismatches  158;  Indels   14;  Gaps    4;

Qy        314 AGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTCCT 373
              ||||  ||||| |||||||||| |||   |    ||||| |||||||| || |  | ||| |
Db         90 AGGGTGGAAAAGGTGTGATACTTTTCTTTACCCATCTTAAGGGTCACAGCTGACACTACC 149

Qy        374 ATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTAC 433
              |||  |||| ||||||||  ||||  ||     ||||     ||||| ||||| |||||||
Db        150 ATAACAAAAAACAGGTTAGCAAGATAAATGGATAAC----TTATTGAATCACAGTTTTAT 205

Qy        434 ATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTTTT 493
              ||  |  | | |||||:||||||||||||||||||||||||| || |||||   |||||
Db        206 GGGATGTAAGTGCCTTCASAAATGAAGACCCAAAGACCCAGGGAAAGCTGTC-CATTTTT 264

Qy        494 TTGCTGAGGTTCGATGAAGAATGGATAGC-ATGTAGCCATGTAGATTAGACAAAAGGATA 552
               |||| || |  |||||||||||||| |||  |||||  ||||    |     ||| |  ||
Db        265 ATGCTTAGATGAGATGAAGAATGGACAGCTGTGTAGAAATGTGATTGTACAAAAGGAGTA 324

Qy        553 TGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTTC 612
              ||||||| | ||||  |||| | || | ||  |||||| ||||||| ||||||||||||
Db        325 TGATCTAATAGTAACAGACTGAAGGAGGAATCCAGCAAGTCCTGTCTGTTCAGATTCTTC 384

Qy        613 TTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTTC 672
              |||  ||:||        |||  ||| ||| ||| |||||||| | ||||| ||| |||
Db        385 TTGGCCTMTC--------TGTGCAGCTTTCCTTCTTCCTGAGTGTAGGGCAAGACCCTTT 436

Qy        673 TTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATGGCTTGC 732
```

```
Db        437 TGGAACGAGGGTCTTATGACAAGGTACATCAAAGAATGTCTTTATGGCCAGCTCTTACAC 496

Qy        733 TTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCTGGTTTC 792

Db        497 AAAAAGGTGGGGGAAGGCTAGATTAATATTTCTAGGTTTTTCTAGAGGATCCCCGGGTAC 556

Qy        793 TGTG 796

Db        557 CGAG 560
```

RESULT 71
US-10-027-632-49406
; Sequence 49406, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 49406
; LENGTH: 670
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(670)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-49406

Query Match 2.7%; Score 162.4; DB 17; Length 670;
Best Local Similarity 64.0%; Pred. No. 5.7e-29;
Matches 310; Conservative 2; Mismatches 158; Indels 14; Gaps 4;

```
Qy        314 AGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTCCT 373

Db         90 AGGGTGGAAAAGGTGTGATACTTTTCTTTACCCATCTTAAGGGTCACAGCTGACACTACC 149

Qy        374 ATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTAC 433

Db        150 ATAACAAAAAACAGGTTAGCAAGATAAATGGATAAC----TTATTGAATCACAGTTTTAT 205

Qy        434 ATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTTTT 493

Db        206 GGGATGTAAGTGCCTTCASAAATGAAGACCCAAAGACCCAGGGAAAGCTGTC-CATTTTT 264

Qy        494 TTGCTGAGGTTCGATGAAGAATGGATAGC-ATGTAGCCATGTAGATTAGACAAAAGGATA 552

Db        265 ATGCTTAGATGAGATGAAGAATGGACAGCTGTGTAGAAATGTGATTGTACAAAAGGAGTA 324

Qy        553 TGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTTC 612

Db        325 TGATCTAATAGTAACAGACTGAAGGAGGAATCCAGCAAGTCCTGTCTGTTCAGATTCTTC 384

Qy        613 TTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTTC 672

Db        385 TTGGCCTMTC--------TGTGCAGCTTTCCTTCTTCCTGAGTGTAGGGCAAGACCCTTT 436

Qy        673 TTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATGGCTTGC 732

Db        437 TGGAACGAGGGTCTTATGACAAGGTACATCAAAGAATGTCTTTATGGCCAGCTCTTACAC 496

Qy        733 TTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCTGGTTTC 792

Db        497 AAAAAGGTGGGGGAAGGCTAGATTAATATTTCTAGGTTTTTCTAGAGGATCCCCGGGTAC 556

Qy        793 TGTG 796

Db        557 CGAG 560
```

RESULT 72
US-09-918-995-17569
; Sequence 17569, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; TITLE OF INVENTION: FROM VARIOUS cDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 17569
; LENGTH: 414
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(414)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-17569

Query Match 2.7%; Score 162.2; DB 10; Length 414;
Best Local Similarity 77.9%; Pred. No. 4.7e-29;
Matches 208; Conservative 0; Mismatches 58; Indels 1; Gaps 1;

```
Qy       2180 TATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCTGAC 2239

Db         28 TAACCATTACCCAAATAATGTGCATTGTACCCTTTAGGTAATTTCTCATCCCTGACTCCC 87

Qy       2240 TTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTCTACATCTATG 2299

Db         88 TTCTTCCCCTCTCACCCTTCCAAGTTTCCATAGTCTATTATTCCGCACTCTGTGTCCGTG 147

Qy       2300 TGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGTCTTT-CTGTG 2358

Db        148 TGTACACATTGTTTAGCTCCTACTTATACGTGAGAACATGCAGTATTTGACTTTGCTGCT 207

Qy       2359 TCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAAGGCA 2418

Db        208 TCTGAGTTGTTTCACTTAAGATAATAGCTACCAGCTCCATCCATGTTGCTTCAAAAGATA 267

Qy       2419 TGATTTCATTCTTTTTTATGGCTATGT 2445

Db        268 AGATTTTATTCTTTTTTATGGCTGCGT 294
```

RESULT 73
US-10-085-117-316/c
; Sequence 316, Application US/10085117
; Publication No. US20030232334A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.

```
;  APPLICANT:  Engelhard, Eric K.
;  TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
;  FILE REFERENCE: 529452000121
;  CURRENT APPLICATION NUMBER: US/10/085,117
;  CURRENT FILING DATE:  2002-02-27
;  PRIOR APPLICATION NUMBER: US 09/798,586
;  PRIOR FILING DATE: 2001-03-02
;  NUMBER OF SEQ ID NOS: 361
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 316
;   LENGTH: 45855
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: variation
;   LOCATION: (1)...(45855)
;   OTHER INFORMATION: n = any nucleotide
US-10-085-117-316

  Query Match             2.7%;  Score 162;  DB 17;  Length 45855;
  Best Local Similarity   75.7%;  Pred. No. 9.2e-28;
  Matches  227;  Conservative    0;  Mismatches   70;  Indels    3;  Gaps    2;

Qy       2170 GCCCCCTGGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACC 2229
              ||    | |  ||| ||||||| ||||  ||| |||||| ||| ||||||||||||||| |
Db       7338 GCTTTCAGTTTATCCATCACCAGAATAGTGTACATTGTGCCCTTTAAGTAATTTCTCATC 7279

Qy       2230 ATCCGC--TGACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACAC 2287
                 | |  || || |   | | | || ||||   || ||||| ||||  ||||||||||
Db       7278 CCTCACCATGCCTCCCCACTCTCCCACCCTTGCAAGTCTCCAGTGTCTGTCATTCCACAT 7219

Qy       2288 TCTACATCTATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTT 2347
              ||||  || ||||||||||||||||||||||||||||||  ||| |||| ||| |||||
Db       7218 TCTATGTCCATGTGTACACATTATTTAGCTCCTACTTATACATGAGAACAGGCAGTATTT 7159

Qy       2348 GTCTTTCTGTGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGC 2407
              | |||||||| ||||  ||||||  |||||||||||||| ||| ||| ||| | | |||
Db       7158 GACTTTCTGTTTCTGCATTGTTTCGCTTATGATAATGGCCTCCAATTCCATCCATGTTGC 7099

Qy       2408 TGCAAAAGGCATGATTTCATTCTTTTTTATGGCTATGTTCTTTCCCAATTTAGATAAAGA 2467
              ||| |||| ||| |||||||||||||||||||||   |  | |  || | || ||| | |
Db       7098 TGC-AAAGACATAATTTCATTCTTTTTTATGGCTGAATAGTATTTCATTATATATATATA 7040


RESULT 74
US-10-027-632-124644/c
; Sequence 124644, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 124644
;   LENGTH: 1224
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-124644

  Query Match             2.6%;  Score 160.6;  DB 13;  Length 1224;
  Best Local Similarity   76.2%;  Pred. No. 2.3e-28;
  Matches  211;  Conservative    0;  Mismatches   64;  Indels    2;  Gaps    1;

Qy       2177 GGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCT 2236
              |||||| |||||||| ||| |||| || ||||||| ||||||||||||||| |    | ||
Db       1193 GGGTATCCATCACCTTAATAATGTACACTGTACCCATTAAGTAATTTCTCATCCCTCACT 1134

Qy       2237 GACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTCTACATCT 2296
                || |    | || ||| ||||| || ||||| |||  || ||||||||||| |    |
Db       1133 CCCTGCCCAGCTCCCCATGCTTCTCAGTCTCCAGTGTTTATTATTCCACACTTTGTGACC 1074

Qy       2297 ATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGTCTTTCTG 2356
              |||| ||||||||||||||||  ||||||  |||| ||||||   |||||| |||||||
Db       1073 ATGTTTACACATTATTTAGCTCTCACTTAT--GTGAGAACATGTGGTATTTGACTTTCTG 1016

Qy       2357 TGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAAGG 2416
              | | ||  |||||| ||||| |||||||||| ||||||| ||| | | |||||| |||
Db       1015 TTTTTGAGTTGTTTCACTTAAGATAATGGCCTCCAGTTCCATCCATGTTGCTGCGAAACA 956

Qy       2417 CATGATTTCATTCTTTTTTATGGCTATGTTCTTTCCC 2453
              |||||||||||||||||| ||||||    |  | | ||
Db        955 TATGATTTCATTCTTTTTAATGGCTGAATAGTATTCC 919


RESULT 75
US-10-027-632-124644/c
; Sequence 124644, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 124644
;   LENGTH: 1224
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-124644

  Query Match             2.6%;  Score 160.6;  DB 17;  Length 1224;
  Best Local Similarity   76.2%;  Pred. No. 2.3e-28;
  Matches  211;  Conservative    0;  Mismatches   64;  Indels    2;  Gaps    1;

Qy       2177 GGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCT 2236
              |||||| |||||||| ||| |||| || ||||||| ||||||||||||||| |    | ||
Db       1193 GGGTATCCATCACCTTAATAATGTACACTGTACCCATTAAGTAATTTCTCATCCCTCACT 1134

Qy       2237 GACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTCTACATCT 2296
                || |    | || ||| ||||| || ||||| |||  || ||||||||||| |    |
Db       1133 CCCTGCCCAGCTCCCCATGCTTCTCAGTCTCCAGTGTTTATTATTCCACACTTTGTGACC 1074
```

```
Qy       2297 ATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGTCTTTCTG 2356
             |||| |||||||||||||||||  ||||||  |||| ||||||   |||||| |||||||
Db       1073 ATGTTTACACATTATTTAGCTCTCACTTAT--GTGAGAACATGTGGTATTTGACTTTCTG 1016

Qy       2357 TGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAAGG 2416
             | | ||  |||||| ||||| |||||||||| ||||||| ||| | | |||||| |||
Db       1015 TTTTTGAGTTGTTTCACTTAAGATAATGGCCTCCAGTTCCATCCATGTTGCTGCGAAACA 956

Qy       2417 CATGATTTCATTCTTTTTTATGGCTATGTTCTTTCCC 2453
              |||||||||||||||| ||||||   |  | | ||
Db        955 TATGATTTCATTCTTTTTAATGGCTGAATAGTATTCC 919


RESULT 76
US-10-322-696-34
; Sequence 34, Application US/10322696
; Publication No. US20040166490A1
; GENERAL INFORMATION:
;  APPLICANT: Morris, David W.
;  APPLICANT:  Malandro, Marc
;  TITLE OF INVENTION: NOVEL THERAPEUTIC TARGETS IN CANCER
;  FILE REFERENCE: 529452001200
;  CURRENT APPLICATION NUMBER: US/10/322,696
;  CURRENT FILING DATE:  2003-10-17
;  NUMBER OF SEQ ID NOS: 186
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 34
;   LENGTH: 357652
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(357652)
;   OTHER INFORMATION: n = A,T,C or G
US-10-322-696-34

  Query Match             2.6%;  Score 158.2;  DB 18;  Length 357652;
  Best Local Similarity   71.5%;  Pred. No. 2.8e-26;
  Matches  208;  Conservative    0;  Mismatches   83;  Indels    0;  Gaps    0;

Qy       2177 GGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCT 2236
             | |||  ||| |||||| | || | ||||||   | ||||||||||||||| |   | |
Db     287029 GTGTAACCATGACCTGAGTAATATACATTGTGATCGTTAAGTAATTTCTCATCCCTCACC 287088

Qy       2237 GACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTCTACATCT 2296
                ||   |||   || ||| |  || ||| | |||| ||||| |||||||||  |
Db     287089 CCTCTCCCACCCTGGCACCCTCCCAAGTCTCTAATGTCTATCATGCCACACTCTAGGTTC 287148

Qy       2297 ATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGTCTTTCTG 2356
             ||||  |||||||| |||||| | | ||||||||||| ||| ||   |||||| |||||||
Db     287149 ATGTACACACATTACTTAGCTACGAGTTATAAGTGAGAACTTGTGGTATTTGACTTTCTG 287208

Qy       2357 TGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAAGG 2416
             | ||||  |||||| ||||| |||||||||| ||||||| ||||| | ||| |||||||
Db     287209 TTTCTGCGTTGTTTCACTTAAGATAATGGCCTCCAGTTCCATCTATGTTGCAGCAAAAGA 287268

Qy       2417 CATGATTTCATTCTTTTTTATGGCTATGTTCTTTCCCAATTTAGATAAAGA 2467
             ||| |||| ||| ||| | ||||||   |  | | ||| | |   || | |
Db     287269 CATTATTTTATTGTTTATAATGGCTGAATGGTATTCCATTGTGTGTATACA 287319


RESULT 77
US-10-027-632-29984
; Sequence 29984, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 29984
;   LENGTH: 865
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-29984

  Query Match             2.5%;  Score 155;  DB 13;  Length 865;
  Best Local Similarity   77.4%;  Pred. No. 4.4e-27;
  Matches  199;  Conservative    1;  Mismatches   56;  Indels    1;  Gaps    1;

Qy       2185 ATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCTGACTTCTT 2244
             |||||  || | |  | ||||| |||| |||||| |||||||| ||||| |   | ||
Db          1 ATCACTGGAGTAACATACATTGAACCCATTAAGTGATTTCTCATCATCCACCCCCCTC-C 59

Qy       2245 GCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTCTACATCTATGTGTAC 2304
              ||| |||  ||||| ||   |||||||||  |||||||||||||||||| || | :|||||
Db         60 ACCCTCTCGCCCTTCCGAATTTCCATTGTCTGTCATTCCACACTCTACTTCCACRTGTAC 119

Qy       2305 ACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGTCTTTCTGTGTCTGTC 2364
             |||||||||||||| |||||||||  | |||| |   ||||||| ||||||||||||
Db        120 ACATTATTTAGCTCCCACTTATAAGACAGAACACGTGGTATTTGTTTTTCTGTGTCTGAG 179

Qy       2365 TTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAAGGCATGATTT 2424
             ||||| ||||| |||||||||| |||||||   | | ||| | ||||| |||||||||
Db        180 GTGTTTCACTTAGGATAATGGCCTCCAGTTCCTCCCATGTTGCAGTAAAAGACATGATTT 239

Qy       2425 CATTCTTTTTTATGGCT 2441
             | || ||||||||||||
Db        240 CGTTATTTTTTATGGCT 256


RESULT 78
US-10-027-632-29985
; Sequence 29985, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
```

```
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 29985
;   LENGTH: 865
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-29985

  Query Match             2.5%;  Score 155;  DB 13;  Length 865;
  Best Local Similarity   77.4%;  Pred. No. 4.4e-27;
  Matches  199;  Conservative    1;  Mismatches   56;  Indels    1;  Gaps    1;

Qy       2185 ATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCTGACTTCTT 2244
              |||||  || | |  | ||||| |||| |||||| |||||||| ||||| |   | ||
Db          1 ATCACTGGAGTAACATACATTGAACCCATTAAGTGATTTCTCATCATCCACCCCCCTC-C 59

Qy       2245 GCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTCTACATCTATGTGTAC 2304
              ||| |||  ||||| ||   ||||||||| |||||||||||||||| || | :|||||
Db         60 ACCCTCTCGCCCTTCCGAATTTCCATTGTCTGTCATTCCACACTCTACTTCCACRTGTAC 119

Qy       2305 ACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGTCTTTCTGTGTCTGTC 2364
              |||||||||||||| |||||||||  | |||| |   ||||||| ||||||||||||
Db        120 ACATTATTTAGCTCCCACTTATAAGACAGAACACGTGGTATTTGTTTTTCTGTGTCTGAG 179

Qy       2365 TTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAAGGCATGATTT 2424
              ||||| ||||| |||||||||| |||||||   | | | ||| | ||||| ||||||||
Db        180 GTGTTTCACTTAGGATAATGGCCTCCAGTTCCTCCCATGTTGCAGTAAAAGACATGATTT 239

Qy       2425 CATTCTTTTTTATGGCT 2441
              | || ||||||||||||
Db        240 CGTTATTTTTTATGGCT 256


RESULT 79
US-10-027-632-29986
; Sequence 29986, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 29986
;   LENGTH: 865
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-29986

  Query Match             2.5%;  Score 155;  DB 13;  Length 865;
  Best Local Similarity   77.4%;  Pred. No. 4.4e-27;
  Matches  199;  Conservative    1;  Mismatches   56;  Indels    1;  Gaps    1;

Qy       2185 ATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCTGACTTCTT 2244
              |||||  || | |  | ||||| |||| |||||| |||||||| ||||| |   | ||
Db          1 ATCACTGGAGTAACATACATTGAACCCATTAAGTGATTTCTCATCATCCACCCCCCTC-C 59

Qy       2245 GCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTCTACATCTATGTGTAC 2304
              ||| |||  ||||| ||   ||||||||| |||||||||||||||| || | :|||||
Db         60 ACCCTCTCGCCCTTCCGAATTTCCATTGTCTGTCATTCCACACTCTACTTCCACRTGTAC 119

Qy       2305 ACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGTCTTTCTGTGTCTGTC 2364
              |||||||||||||| |||||||||  | |||| |   ||||||| ||||||||||||
Db        120 ACATTATTTAGCTCCCACTTATAAGACAGAACACGTGGTATTTGTTTTTCTGTGTCTGAG 179

Qy       2365 TTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAAGGCATGATTT 2424
              ||||| ||||| |||||||||| |||||||   | | | ||| | ||||| ||||||||
Db        180 GTGTTTCACTTAGGATAATGGCCTCCAGTTCCTCCCATGTTGCAGTAAAAGACATGATTT 239

Qy       2425 CATTCTTTTTTATGGCT 2441
              | || ||||||||||||
Db        240 CGTTATTTTTTATGGCT 256


RESULT 80
US-10-027-632-29984
; Sequence 29984, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 29984
;   LENGTH: 865
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-29984

  Query Match             2.5%;  Score 155;  DB 17;  Length 865;
  Best Local Similarity   77.4%;  Pred. No. 4.4e-27;
  Matches  199;  Conservative    1;  Mismatches   56;  Indels    1;  Gaps    1;

Qy       2185 ATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCTGACTTCTT 2244
              |||||  || | |  | ||||| |||| |||||| |||||||| ||||| |   | ||
Db          1 ATCACTGGAGTAACATACATTGAACCCATTAAGTGATTTCTCATCATCCACCCCCCTC-C 59

Qy       2245 GCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTCTACATCTATGTGTAC 2304
              ||| |||  ||||| ||   ||||||||| |||||||||||||||| || | :|||||
Db         60 ACCCTCTCGCCCTTCCGAATTTCCATTGTCTGTCATTCCACACTCTACTTCCACRTGTAC 119

Qy       2305 ACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGTCTTTCTGTGTCTGTC 2364
              |||||||||||||| |||||||||  | |||| |   ||||||| ||||||||||||
Db        120 ACATTATTTAGCTCCCACTTATAAGACAGAACACGTGGTATTTGTTTTTCTGTGTCTGAG 179

Qy       2365 TTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAAGGCATGATTT 2424
              ||||| ||||| |||||||||| |||||||   | | | ||| | ||||| ||||||||
Db        180 GTGTTTCACTTAGGATAATGGCCTCCAGTTCCTCCCATGTTGCAGTAAAAGACATGATTT 239

Qy       2425 CATTCTTTTTTATGGCT 2441
              | || ||||||||||||
Db        240 CGTTATTTTTTATGGCT 256
```

```
RESULT 81
US-10-027-632-29985
; Sequence 29985, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 29985
;   LENGTH: 865
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-29985

  Query Match             2.5%;  Score 155;  DB 17;  Length 865;
  Best Local Similarity   77.4%;  Pred. No. 4.4e-27;
  Matches  199;  Conservative    1;  Mismatches   56;  Indels    1;  Gaps    1;

Qy       2185 ATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCTGACTTCTT 2244
              |||||  || | |  | ||||| |||| |||||| |||||||| ||||| |   | ||
Db          1 ATCACTGGAGTAACATACATTGAACCCATTAAGTGATTTCTCATCATCCACCCCCCTC-C 59

Qy       2245 GCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTCTACATCTATGTGTAC 2304
               ||| |||  ||||| ||   |||||||||  |||||||||||||||| || | :|||||
Db         60 ACCCTCTCGCCCTTCCGAATTTCCATTGTCTGTCATTCCACACTCTACTTCCACRTGTAC 119

Qy       2305 ACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGTCTTTCTGTGTCTGTC 2364
              |||||||||||||| |||||||||  | |||| |   ||||||| ||||||||||||
Db        120 ACATTATTTAGCTCCCACTTATAAGACAGAACACGTGGTATTTGTTTTTCTGTGTCTGAG 179

Qy       2365 TTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAAGGCATGATTT 2424
               ||||| ||||| |||||||||| |||||||   | | | ||| | ||||| ||||||||
Db        180 GTGTTTCACTTAGGATAATGGCCTCCAGTTCCTCCCATGTTGCAGTAAAAGACATGATTT 239

Qy       2425 CATTCTTTTTTATGGCT 2441
              | || ||||||||||||
Db        240 CGTTATTTTTTATGGCT 256


RESULT 82
US-10-027-632-29986
; Sequence 29986, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 29986
;   LENGTH: 865
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-29986

  Query Match             2.5%;  Score 155;  DB 17;  Length 865;
  Best Local Similarity   77.4%;  Pred. No. 4.4e-27;
  Matches  199;  Conservative    1;  Mismatches   56;  Indels    1;  Gaps    1;

Qy       2185 ATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCTGACTTCTT 2244
              |||||  || | |  | ||||| |||| |||||| |||||||| ||||| |   | ||
Db          1 ATCACTGGAGTAACATACATTGAACCCATTAAGTGATTTCTCATCATCCACCCCCCTC-C 59

Qy       2245 GCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTCTACATCTATGTGTAC 2304
               ||| |||  ||||| ||   |||||||||  |||||||||||||||| || | :|||||
Db         60 ACCCTCTCGCCCTTCCGAATTTCCATTGTCTGTCATTCCACACTCTACTTCCACRTGTAC 119

Qy       2305 ACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGTCTTTCTGTGTCTGTC 2364
              |||||||||||||| |||||||||  | |||| |   ||||||| ||||||||||||
Db        120 ACATTATTTAGCTCCCACTTATAAGACAGAACACGTGGTATTTGTTTTTCTGTGTCTGAG 179

Qy       2365 TTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAAGGCATGATTT 2424
               ||||| ||||| |||||||||| |||||||   | | | ||| | ||||| ||||||||
Db        180 GTGTTTCACTTAGGATAATGGCCTCCAGTTCCTCCCATGTTGCAGTAAAAGACATGATTT 239

Qy       2425 CATTCTTTTTTATGGCT 2441
              | || ||||||||||||
Db        240 CGTTATTTTTTATGGCT 256


RESULT 83
US-10-712-795-334
; Sequence 334, Application US/10712795
; Publication No. US20040214325A1
; GENERAL INFORMATION:
;  APPLICANT: Crooke et al.
;  TITLE OF INVENTION: ANTISENSE MODULATION OF APOLIPOPROTEIN B EXPRESSION
;  FILE REFERENCE: 30566/39662
;  CURRENT APPLICATION NUMBER: US/10/712,795
;  CURRENT FILING DATE:  2003-11-13
;  PRIOR APPLICATION NUMBER: US 60/426,234
;  PRIOR FILING DATE: 2002-11-13
;  PRIOR APPLICATION NUMBER: PCT/US03/15493
;  PRIOR FILING DATE: 2003-05-13
;  NUMBER OF SEQ ID NOS: 892
; SEQ ID NO 334
;   LENGTH: 43445
;   TYPE: DNA
;   ORGANISM: H. sapiens
US-10-712-795-334

  Query Match             2.5%;  Score 155;  DB 18;  Length 43445;
  Best Local Similarity   76.6%;  Pred. No. 4.7e-26;
  Matches  203;  Conservative    0;  Mismatches   60;  Indels    2;  Gaps    1;

Qy       2170 GCCCCCTGGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACC 2229
              ||   | |||||| |||||||   |  |  | | |||||||  ||||||||||||||||
Db      12940 GCTTTCAGGGTATCCATCACCCAGACAACATATAGTGTACCCACTAAGTAATTTCTCACC 12999
```

```
Qy       2230 ATCCGCTGACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTC 2289
             ||||     | ||     ||     ||||||||| ||| |||||| |||||||||||| |
Db      13000 ATCCATCTCCCTCCACTTCCACA--CCTTCTGAGTCTCAATTGTCTATCATTCCACACAC 13057

Qy       2290 TACATCTATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGT 2349
             ||  ||  ||||| ||||||||||  |||| |||||||| ||| |||| ||||||||||
Db      13058 TATGTCCTTGTGTGCACATTATTTCACTCCCACTTATAAATGACAACACGCAATATTTGT 13117

Qy       2350 CTTTCTGTGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTG 2409
             ||||||||| ||||| ||||| ||||| || |||| || ||||||| ||| | | |||||
Db      13118 CTTTCTGTGACTGTCCTGTTTCACTTAAGACAATGACCTCCAGTTCCATCCATGTTGCTG 13177

Qy       2410 CAAAAGGCATGATTTCATTCTTTTT 2434
             |||| | |||||||| |||||||||
Db      13178 CAAATGACATGATTTTATTCTTTTT 13202


RESULT 84
US-10-920-612-334
; Sequence 334, Application US/10920612
; Publication No. US20050009088A1
; GENERAL INFORMATION:
;  APPLICANT: Crooke et al.
;  TITLE OF INVENTION: ANTISENSE MODULATION OF APOLIPOPROTEIN B EXPRESSION
;  FILE REFERENCE: 30566/39634A
;  CURRENT APPLICATION NUMBER: US/10/920,612
;  CURRENT FILING DATE:  2004-08-17
;  PRIOR APPLICATION NUMBER: PCT/US03/15493
;  PRIOR FILING DATE: 2003-11-15
;  PRIOR APPLICATION NUMBER: US 10/712,795
;  PRIOR FILING DATE: 2003-11-13
;  PRIOR APPLICATION NUMBER: US 60/426,234
;  PRIOR FILING DATE: 2002-11-13
;  NUMBER OF SEQ ID NOS: 892
; SEQ ID NO 334
;   LENGTH: 43445
;   TYPE: DNA
;   ORGANISM: H. sapiens
US-10-920-612-334

  Query Match             2.5%;  Score 155;  DB 19;  Length 43445;
  Best Local Similarity   76.6%;  Pred. No. 4.7e-26;
  Matches  203;  Conservative    0;  Mismatches   60;  Indels    2;  Gaps    1;

Qy       2170 GCCCCCTGGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACC 2229
             ||   | |||||| |||||||   |  |   |  | |||||||  ||||||||||||||||
Db      12940 GCTTTCAGGGTATCCATCACCCAGACAACATATAGTGTACCCACTAAGTAATTTCTCACC 12999

Qy       2230 ATCCGCTGACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTC 2289
             ||||     | ||     ||     ||||||||| ||| |||||| |||||||||||| |
Db      13000 ATCCATCTCCCTCCACTTCCACA--CCTTCTGAGTCTCAATTGTCTATCATTCCACACAC 13057

Qy       2290 TACATCTATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGT 2349
             ||  ||  ||||| ||||||||||  |||| |||||||| ||| |||| ||||||||||
Db      13058 TATGTCCTTGTGTGCACATTATTTCACTCCCACTTATAAATGACAACACGCAATATTTGT 13117

Qy       2350 CTTTCTGTGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTG 2409
             ||||||||| ||||| ||||| ||||| || |||| || ||||||| ||| | | |||||
Db      13118 CTTTCTGTGACTGTCCTGTTTCACTTAAGACAATGACCTCCAGTTCCATCCATGTTGCTG 13177

Qy       2410 CAAAAGGCATGATTTCATTCTTTTT 2434
             |||| | |||||||| |||||||||
Db      13178 CAAATGACATGATTTTATTCTTTTT 13202


RESULT 85
US-10-741-601-5659
; Sequence 5659, Application US/10741601
; Publication No. US20040166519A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION: STENOSIS, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001500
;  CURRENT APPLICATION NUMBER: US/10/741,601
;  CURRENT FILING DATE:  2003-12-22
;  NUMBER OF SEQ ID NOS: 26415
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5659
;   LENGTH: 54775
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-741-601-5659

  Query Match             2.5%;  Score 155;  DB 18;  Length 54775;
  Best Local Similarity   76.6%;  Pred. No. 5.4e-26;
  Matches  203;  Conservative    0;  Mismatches   60;  Indels    2;  Gaps    1;

Qy       2170 GCCCCCTGGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACC 2229
             ||   | |||||| |||||||   |  |   |  | |||||||  ||||||||||||||||
Db      18545 GCTTTCAGGGTATCCATCACCCAGACAACATATAGTGTACCCACTAAGTAATTTCTCACC 18604

Qy       2230 ATCCGCTGACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTC 2289
             ||||     | ||     ||     ||||||||| ||| |||||| |||||||||||| |
Db      18605 ATCCATCTCCCTCCACTTCCACA--CCTTCTGAGTCTCAATTGTCTATCATTCCACACAC 18662

Qy       2290 TACATCTATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGT 2349
             ||  ||  ||||| ||||||||||  |||| |||||||| ||| |||| ||||||||||
Db      18663 TATGTCCTTGTGTGCACATTATTTCACTCCCACTTATAAATGACAACACGCAATATTTGT 18722

Qy       2350 CTTTCTGTGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTG 2409
             ||||||||| ||||| ||||| ||||| || |||| || ||||||| ||| | | |||||
Db      18723 CTTTCTGTGACTGTCCTGTTTCACTTAAGACAATGACCTCCAGTTCCATCCATGTTGCTG 18782

Qy       2410 CAAAAGGCATGATTTCATTCTTTTT 2434
             |||| | |||||||| |||||||||
Db      18783 CAAATGACATGATTTTATTCTTTTT 18807


RESULT 86
US-10-741-600-17684
; Sequence 17684, Application US/10741600
; Publication No. US20050026169A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001499
;  CURRENT APPLICATION NUMBER: US/10/741,600
;  CURRENT FILING DATE:  2003-12-22
;  NUMBER OF SEQ ID NOS: 73997
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17684
;   LENGTH: 54775
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-741-600-17684

  Query Match             2.5%;  Score 155;  DB 19;  Length 54775;
  Best Local Similarity   76.6%;  Pred. No. 5.4e-26;
  Matches  203;  Conservative    0;  Mismatches   60;  Indels    2;  Gaps    1;

Qy       2170 GCCCCCTGGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACC 2229
             ||   | |||||| |||||||   |  |   |  | |||||||  ||||||||||||||||
Db      18545 GCTTTCAGGGTATCCATCACCCAGACAACATATAGTGTACCCACTAAGTAATTTCTCACC 18604

Qy       2230 ATCCGCTGACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTC 2289
             ||||     | ||     ||     ||||||||| ||| |||||| |||||||||||| |
Db      18605 ATCCATCTCCCTCCACTTCCACA--CCTTCTGAGTCTCAATTGTCTATCATTCCACACAC 18662

Qy       2290 TACATCTATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGT 2349
             ||  ||  ||||| ||||||||||  |||| |||||||| ||| |||| ||||||||||
Db      18663 TATGTCCTTGTGTGCACATTATTTCACTCCCACTTATAAATGACAACACGCAATATTTGT 18722

Qy       2350 CTTTCTGTGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTG 2409
```

```
              ||||||||| ||||| ||||| ||||| || |||| || ||||||| ||| | | |||||
Db      18723 CTTTCTGTGACTGTCCTGTTTCACTTAAGACAATGACCTCCAGTTCCATCCATGTTGCTG 18782

Qy       2410 CAAAAGGCATGATTTCATTCTTTTT 2434
              |||| | |||||||| |||||||||
Db      18783 CAAATGACATGATTTTATTCTTTTT 18807


RESULT 87
US-10-741-600-17806/c
; Sequence 17806, Application US/10741600
; Publication No. US20050026169A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001499
;  CURRENT APPLICATION NUMBER: US/10/741,600
;  CURRENT FILING DATE:  2003-12-22
;  NUMBER OF SEQ ID NOS: 73997
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17806
;   LENGTH: 14036
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(14036)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-741-600-17806

  Query Match             2.5%;  Score 154.8;  DB 19;  Length 14036;
  Best Local Similarity   67.6%;  Pred. No. 2.7e-26;
  Matches  296;  Conservative    0;  Mismatches  127;  Indels   15;  Gaps    5;

Qy        252 CTTTAGTTCTCAGCATATTCGTGCACCTTCATTTATCTCTCCTTCTCTCTCAAAGCTGCA 311
              || || ||     | |   | |   ||||   |  | |     |||     |    | |
Db       3988 CTCTACTTGAGGACCTGCCCCTTTGCCTTTTATGCTTTACTAGTCTTAGCTATTATTTCT 3929

Qy        312 GTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTC 371
              |||||||   ||| | ||||| |  || ||   |||||||||||||     |||| ||||
Db       3928 GTAGGGGCAGAAAGGGGTGAT-CCCTTCCTGACCCATCATAAGGGTTATGGCCAATACTC 3870

Qy        372 CTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTT 431
              ||||| |||||||||| |||| ||||| ||| | ||| | ||||||||||||||||||||
Db       3869 CTATAATAAAAGACAGATTAACAAGAGAAAAGCATAATACATTTATTTAATCAAAGTTTT 3810

Qy        432 ACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTT 491
              |  |||||||||| |||||||||||||||||||| ||||||||||||     |||
Db       3809 AGGTGACATGGGAGCCTTCAGAAATGAAGACCCAAGGACCCAGGGGAAA----ACTATTT 3754

Qy        492 TTTTGCTGAGGTTCGATGAAGAATGGATAGC-ATGTAGCCATGTAGATTAGACAAAAGGA 550
              || |||| || || |||||||||| | |||  || ||  |||||  | |   ||| |
Db       3753 TTATGCTTAGATTTGATGAAGAATGAACAGCTGTGCAGAAATGTAATTGAACAAAAGGAG 3694

Qy        551 TATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCT 610
              ||| ||||| |||| |  ||||  |   |         | ||    | ||   || |||||
Db       3693 TATCATCTAATGGTCACAGACTGGGACTGGGGGGATCCCAGCAAGGCCTGGCCATATTCT 3634

Qy        611 TCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGA-CTC 669
              ||||| ||||||         |||| |||||||  |||||||   |||| ||||||  | |||
Db       3633 TCTTGGTCTCTC--------TGTACAGCATTCCTTCCTCCCAGGTATAGGGCAGAACCTC 3582

Qy        670 TTCTTCAATGAGGGTCTT 687
              ||||  ||||||||||||
Db       3581 TTCTGGAATGAGGGTCTT 3564


RESULT 88
US-10-741-600-18012/c
; Sequence 18012, Application US/10741600
; Publication No. US20050026169A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001499
;  CURRENT APPLICATION NUMBER: US/10/741,600
;  CURRENT FILING DATE:  2003-12-22
;  NUMBER OF SEQ ID NOS: 73997
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 18012
;   LENGTH: 14769
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-741-600-18012

  Query Match             2.5%;  Score 154.8;  DB 19;  Length 14769;
  Best Local Similarity   67.6%;  Pred. No. 2.7e-26;
  Matches  296;  Conservative    0;  Mismatches  127;  Indels   15;  Gaps    5;

Qy        252 CTTTAGTTCTCAGCATATTCGTGCACCTTCATTTATCTCTCCTTCTCTCTCAAAGCTGCA 311
              || || ||     | |   | |   ||||   |  | |     |||     |    | |
Db        706 CTCTACTTGAGGACCTGCCCCTTTGCCTTTTATGCTTTACTAGTCTTAGCTATTATTTCT 647

Qy        312 GTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTC 371
              |||||||   ||| | ||||| |  || ||   |||||||||||||     |||| ||||
Db        646 GTAGGGGCAGAAAGGGGTGAT-CCCTTCCTGACCCATCATAAGGGTTATGGCCAATACTC 588

Qy        372 CTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTT 431
              ||||| |||||||||| |||| ||||| ||| | ||| | ||||||||||||||||||||
Db        587 CTATAATAAAAGACAGATTAACAAGAGAAAAGCATAATACATTTATTTAATCAAAGTTTT 528

Qy        432 ACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTT 491
              |  |||||||||| |||||||||||||||||||| ||||||||||||     |||
Db        527 AGGTGACATGGGAGCCTTCAGAAATGAAGACCCAAGGACCCAGGGGAAA----ACTATTT 472

Qy        492 TTTTGCTGAGGTTCGATGAAGAATGGATAGC-ATGTAGCCATGTAGATTAGACAAAAGGA 550
              || |||| || || |||||||||| | |||  || ||  |||||  | |   ||| |
Db        471 TTATGCTTAGATTTGATGAAGAATGAACAGCTGTGCAGAAATGTAATTGAACAAAAGGAG 412

Qy        551 TATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCT 610
              ||| ||||| |||| |  ||||  |   |         | ||    | ||   || |||||
Db        411 TATCATCTAATGGTCACAGACTGGGACTGGGGGGATCCCAGCAAGGCCTGGCCATATTCT 352

Qy        611 TCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGA-CTC 669
              ||||| ||||||         |||| |||||||  |||||||   |||| ||||||  | |||
Db        351 TCTTGGTCTCTC--------TGTACAGCATTCCTTCCTCCCAGGTATAGGGCAGAACCTC 300

Qy        670 TTCTTCAATGAGGGTCTT 687
              ||||  ||||||||||||
Db        299 TTCTGGAATGAGGGTCTT 282


RESULT 89
US-09-854-867-74/c
; Sequence 74, Application US/09854867
; Publication No. US20030224356A1
; GENERAL INFORMATION:
;  APPLICANT: JOAN, KNOLL H
;  APPLICANT:  ROGAN, PETER K
;  TITLE OF INVENTION: SINGLE COPY GENOMIC HYBRIDIZATION PROBES AND METHOD OF GENERATING
;  FILE REFERENCE: 30307
;  CURRENT APPLICATION NUMBER: US/09/854,867
;  CURRENT FILING DATE:  2003-05-08
;  NUMBER OF SEQ ID NOS: 613
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 74
;   LENGTH: 1042
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: repeat_region
;   LOCATION: (1)..(1042)
;   OTHER INFORMATION: 11ma5
US-09-854-867-74
```

```
  Query Match             2.5%;  Score 154.4;  DB 10;  Length 1042;
  Best Local Similarity   58.5%;  Pred. No. 6.9e-27;
  Matches  324;  Conservative    7;  Mismatches  198;  Indels   25;  Gaps    3;

Qy       1931 TTTTTTTTTAATTTTTAAAATTTGTATATATTTATGGGTTATA-AGTGAAATCTTTTTAG 1989
             |||||  || |    ||| | ||||| |||||||||| || |  |||| || |   ||
Db       1042 TTTTTAATTGACACATAATAATTGTACATATTTATGGGGTACAGTGTGATATTTCGATAC 983

Qy       1990 ATGCATAGGTTGTATAGTGATAAAATCAGGGCTTTTAGGGTATTCATCACCTGAATGATG 2049
             ||| |||  :||| || |||| |||||||||   |||| ||| |||||||| ||  ||
Db        982 ATGTATACAWTGTGTAATGATCAAATCAGGGTAATTAGCATATCCATCACCTCAAACATT 923

Qy       2050 TACATTGTACCCCTTAAGTAATTTCTCACCATCCGCTGACTTCTTGCCCCCTGGGTATTC 2109
             ||   | |      |   | |    ||| |||| ||     |              ||| |
Db        922 TATCATTTCTTTGTGTTGGGAACATTCAAAATCCTCTCTTCTAGCTATTTGAAAATATAC 863

Qy       2110 ATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCTGACTTCTT 2169
             |  | | : || |    || ||  ||||   ||  |      ||  |            ||| |
Db        862 AATAAATTRTTGTTAACTATAGTCACCCTACGGTGCTACAGAACACTAGAACTTATTCCT 803

Qy       2170 GCCCCCTGGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACC 2229
               |   ||   | |   |                  |||||:|| ||||  ||  |||| |
Db        802 CCTATCTAACTGTAATT----------------TTGTAYCCWTTAACCAACCTCTCCCT 760

Qy       2230 ATCCGCTGACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTC 2289
             ||:|        |    |:||  | |||||  || ||| | |  |||  |||| || |||
Db        759 ATYC-------CCCCCTCYCCCTACCCTTCCCAGCCTCTAGTAACCACTATTCTACTCTC 707

Qy       2290 TACATCTATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGT 2349
             ||| |||||| |  ||  || ||||||| | || ||| ||||| |||||||  |||||||
Db        706 TACTTCTATGAGATCAACTTTTTTAGCTTCCACATATGAGTGAGAACATGCGGTATTTGT 647

Qy       2350 CTTTCTGTGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTG 2409
             |||||||| ||| ||| ||| |||||  |||||| || |||| |  ||| | | ||:|:
Db        646 CTTTCTGTGCCTGGCTTATTTCACTTAGCATAATGTCCTCCAGGTTCATCCATGTTGYTR 587

Qy       2410 CAAAAGGCATGATTTCATTCTTTTTTATGGCTATGTTCTTTCCCAATTTAGATAAAGAAC 2469
             |||| | || ||||||||||||||||||||||||    | | | ||| | |  ||| | | |
Db        586 CAAATGACAGGATTTCATTCTTTTTTATGGCTGAATAGTATTCCATTGTGTATATATACC 527

Qy       2470 ACTCGCACTTGCTC 2483
             ||     |||  ||
Db        526 ACATTTTCTTTATC 513


RESULT 90
US-10-786-970A-74/c
; Sequence 74, Application US/10786970A
; Publication No. US20050064449A1
; GENERAL INFORMATION:
;  APPLICANT: JOAN, KNOLL
;  APPLICANT:  ROGAN, PETER
;  TITLE OF INVENTION: SINGLE COPY GENOMIC HYBRIDIZATION PROBES AND METHOD OF GENERATING
;  FILE REFERENCE: 30307
;  CURRENT APPLICATION NUMBER: US/10/786,970A
;  CURRENT FILING DATE:  2004-02-24
;  PRIOR APPLICATION NUMBER: US/09/573,080
;  PRIOR FILING DATE: 2000-05-16
;  NUMBER OF SEQ ID NOS: 479
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 74
;   LENGTH: 1042
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: repeat_region
;   LOCATION: (1)..(1042)
;   OTHER INFORMATION: l1ma5
;  PUBLICATION INFORMATION:
;  PUBLICATION INFORMATION:
;   AUTHORS: Jurka, J; Walichiewicz, J; Milosavljevic, A
;   TITLE: Prototypic sequences for human repetitive DNA
;   JOURNAL: Journal of Molecular Evolution
;   VOLUME: 35
;   ISSUE: 4
;   PAGES: 286-291
;   DATE: 1992-10-__
;   DATABASE ACCESSION NUMBER: Database of repetitive elements (repbase)
;   DATABASE ENTRY DATE: ____-__-__
;   DATABASE ENTRY DATE: 1996-01-26
US-10-786-970A-74

  Query Match             2.5%;  Score 154.4;  DB 19;  Length 1042;
  Best Local Similarity   58.5%;  Pred. No. 6.9e-27;
  Matches  324;  Conservative    7;  Mismatches  198;  Indels   25;  Gaps    3;

Qy       1931 TTTTTTTTTAATTTTTAAAATTTGTATATATTTATGGGTTATA-AGTGAAATCTTTTTAG 1989
             |||||  || |    ||| | ||||| |||||||||| || |  |||| || |   ||
Db       1042 TTTTTAATTGACACATAATAATTGTACATATTTATGGGGTACAGTGTGATATTTCGATAC 983

Qy       1990 ATGCATAGGTTGTATAGTGATAAAATCAGGGCTTTTAGGGTATTCATCACCTGAATGATG 2049
             ||| |||  :||| || |||| |||||||||   |||| ||| |||||||| ||  ||
Db        982 ATGTATACAWTGTGTAATGATCAAATCAGGGTAATTAGCATATCCATCACCTCAAACATT 923

Qy       2050 TACATTGTACCCCTTAAGTAATTTCTCACCATCCGCTGACTTCTTGCCCCCTGGGTATTC 2109
             ||   | |      |   | |    ||| |||| ||     |              ||| |
Db        922 TATCATTTCTTTGTGTTGGGAACATTCAAAATCCTCTCTTCTAGCTATTTGAAAATATAC 863

Qy       2110 ATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCTGACTTCTT 2169
             |  | | : || |    || ||  ||||   ||  |      ||  |            ||| |
Db        862 AATAAATTRTTGTTAACTATAGTCACCCTACGGTGCTACAGAACACTAGAACTTATTCCT 803

Qy       2170 GCCCCCTGGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACC 2229
               |   ||   | |   |                  |||||:|| ||||  ||  |||| |
Db        802 CCTATCTAACTGTAATT----------------TTGTAYCCWTTAACCAACCTCTCCCT 760

Qy       2230 ATCCGCTGACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTC 2289
             ||:|        |    |:||  | |||||  || ||| | |  |||  |||| || |||
Db        759 ATYC-------CCCCCTCYCCCTACCCTTCCCAGCCTCTAGTAACCACTATTCTACTCTC 707

Qy       2290 TACATCTATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGT 2349
             ||| |||||| |  ||  || ||||||| | || ||| ||||| |||||||  |||||||
Db        706 TACTTCTATGAGATCAACTTTTTTAGCTTCCACATATGAGTGAGAACATGCGGTATTTGT 647

Qy       2350 CTTTCTGTGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTG 2409
             |||||||| ||| ||| ||| |||||  |||||| || |||| |  ||| | | ||:|:
Db        646 CTTTCTGTGCCTGGCTTATTTCACTTAGCATAATGTCCTCCAGGTTCATCCATGTTGYTR 587

Qy       2410 CAAAAGGCATGATTTCATTCTTTTTTATGGCTATGTTCTTTCCCAATTTAGATAAAGAAC 2469
             |||| | || ||||||||||||||||||||||||    | | | ||| | |  ||| | | |
Db        586 CAAATGACAGGATTTCATTCTTTTTTATGGCTGAATAGTATTCCATTGTGTATATATACC 527

Qy       2470 ACTCGCACTTGCTC 2483
             ||     |||  ||
Db        526 ACATTTTCTTTATC 513


RESULT 91
US-09-949-654-3
; Sequence 3, Application US/09949654
; Patent No. US20020127644A1
; GENERAL INFORMATION:
;  APPLICANT: YAN, Chunhua et al.
;  TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
;  TITLE OF INVENTION:  NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
;  TITLE OF INVENTION:  AND USES THEREOF
;  FILE REFERENCE: CL000817
;  CURRENT APPLICATION NUMBER: US/09/949,654
;  CURRENT FILING DATE:  2001-09-12
;  PRIOR APPLICATION NUMBER: 60/231,572
;  PRIOR FILING DATE: 2000-09-11
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 368004
```

```
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(368004)
;   OTHER INFORMATION: n = A,T,C or G
US-09-949-654-3

  Query Match             2.5%;  Score 153.8;  DB 9;  Length 368004;
  Best Local Similarity   74.7%;  Pred. No. 3.4e-25;
  Matches  219;  Conservative    0;  Mismatches   72;  Indels    2;  Gaps    2;

Qy       2177 GGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCT 2236
              | |||| | | |||||||| | |||||||| || | |||||||||||| ||| ||||| |
Db      13776 GAGTATGCTTTACCTGAATAACGTGCATTGCACTCATTAAGTAATTTATCATCATCCACC 13835

Qy       2237 GACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTC-TACATC 2295
                | |||  ||||| |||  ||| ||| || || |||| || | |||| ||||     ||
Db      13836 CTCCTCTAACCCCCCCATTATTCAGAGTCTACAATGTCTATTAGTCCAAACTCAGTTGTC 13895

Qy       2296 TATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTG-TCTTTC 2354
                |||  |||||||||  |||||||| |||||||| | |  ||||| |||||||| | ||||
Db      13896 CATGAGTACACATTGCTTAGCTCCCACTTATAAATAAGCACATGTAATATTTGATTTTTC 13955

Qy       2355 TGTGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAA 2414
                |||||||| ||||||| |||||  |||||||| | ||||| ||||  | ||||| |||
Db      13956 TGTGTCTGACTTGTTTCACTTAAAATAATGGCCTCTAGTTCCATCTGTGTTGCTGAAAAC 14015

Qy       2415 GGCATGATTTCATTCTTTTTTATGGCTATGTTCTTTCCCAATTTAGATAAAGA 2467
                ||  |||||||||||||||||||| | ||  | | ||| |  | ||| | |
Db      14016 AACACAATTTCATTCTTTTTTATGGCCAAGTAGTATTCCATTGGATATATATA 14068


RESULT 92
US-10-027-632-166950
; Sequence 166950, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 166950
;   LENGTH: 766
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-166950

  Query Match             2.5%;  Score 153.4;  DB 13;  Length 766;
  Best Local Similarity   71.8%;  Pred. No. 1e-26;
  Matches  262;  Conservative    1;  Mismatches   82;  Indels   20;  Gaps    4;

Qy        334 CATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTCCTATAGTAAAAGACAGGTTAAT 393
              || || |||  ||||  |||| |:  |     |  |||||||| ||||||| |||||||
Db        422 CACTTACTCACCCATTGTAAGTGYTACGTTTGATCCTCCTATAATAAAAGATAGGTTAAC 481

Qy        394 AAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTACATGACATGGGAGTCTTCAGA 453
              ||||| ||| | | |||||||| ||||||||||||||||  |||||  |||| |||||||
Db        482 AAGAGAAAAGCATGACAAATTTGTTTAATCAAAGTTTTATGTGACACAGGAGCCTTCAGA 541

Qy        454 AATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTTTTTTGCTGAGGTTCGATGAAGA 513
              |||||||||| |||||||||||| |||||       |||| |||| ||||| |||||||||
Db        542 AATGAAGACCAAAAGACCCAGGGAAAACT-----ACTTTTATGCTTAGGTTTGATGAAGA 596

Qy        514 ATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGATATGATCTAGTGGTAAAGGACTC 573
              |||| | |         |    |||||||||||| |   ||||  |||||||| || |
Db        597 GTGGACATCCATGTTGAAATGTGATTAGACAAAAAG-GGTGATTGAGTGGTAAAAGA--C 653

Qy        574 AGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTTCTTGATCTCTCTCTCTCTATGT 633
              || || |||| ||| ||| ||||||| |||||||||||||| | | ||           |
Db        654 AGAGGAAAACCCAGTAAGTCCTGTCTGTTCAGATTCTTCTGGGTTTC-----------T 701

Qy        634 ATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTTCTTCAATGAGGGTCTTCAAGGG 693
              ||| |||| ||||| ||   |||||| |||||| | |||   |||| | |||| |||||
Db        702 GTAGAATTCCTTCCTTCTAGATATGGGCCAGGACCCCTCTGATATGAAGTTCTTTAAGGG 761

Qy        694 AGAAG 698
              |||||
Db        762 AGAAG 766


RESULT 93
US-10-027-632-166951
; Sequence 166951, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 166951
;   LENGTH: 766
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-166951

  Query Match             2.5%;  Score 153.4;  DB 13;  Length 766;
  Best Local Similarity   71.8%;  Pred. No. 1e-26;
  Matches  262;  Conservative    1;  Mismatches   82;  Indels   20;  Gaps    4;

Qy        334 CATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTCCTATAGTAAAAGACAGGTTAAT 393
              || || |||  ||||  |||| |:  |     |  |||||||| ||||||| |||||||
Db        422 CACTTACTCACCCATTGTAAGTGYTACGTTTGATCCTCCTATAATAAAAGATAGGTTAAC 481

Qy        394 AAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTACATGACATGGGAGTCTTCAGA 453
              ||||| ||| | | |||||||| ||||||||||||||||  |||||  |||| |||||||
Db        482 AAGAGAAAAGCATGACAAATTTGTTTAATCAAAGTTTTATGTGACACAGGAGCCTTCAGA 541

Qy        454 AATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTTTTTTGCTGAGGTTCGATGAAGA 513
              |||||||||| |||||||||||| |||||       |||| |||| ||||| |||||||||
```

```
Db        542 AATGAAGACCAAAAGACCCAGGGAAAACT-----ACTTTTATGCTTAGGTTTGATGAAGA 596

Qy        514 ATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGATATGATCTAGTGGTAAAGGACTC 573
              |||| | |         |    |||||||||||| |   ||||  ||||||||| ||  |
Db        597 GTGGACATCCATGTTGAAATGTGATTAGACAAAAAG-GGTGATTGAGTGGTAAAAGA--C 653

Qy        574 AGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTTCTTGATCTCTCTCTCTCTATGT 633
              || || |||| ||| ||| ||||||| |||||||||||||| | | ||            |
Db        654 AGAGGAAAACCCAGTAAGTCCTGTCTGTTCAGATTCTTCTGGGTTTC------------T 701

Qy        634 ATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTTCTTCAATGAGGGTCTTCAAGGG 693
              ||| |||| ||||| ||   |||||| |||||| | |||   |||| | |||| |||||
Db        702 GTAGAATTCCTTCCTTCTAGATATGGGCCAGGACCCCTCTGATATGAAGTTCTTTAAGGG 761

Qy        694 AGAAG 698
              |||||
Db        762 AGAAG 766


RESULT 94
US-10-027-632-166950
; Sequence 166950, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 166950
;   LENGTH: 766
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-166950

  Query Match             2.5%;  Score 153.4;  DB 17;  Length 766;
  Best Local Similarity   71.8%;  Pred. No. 1e-26;
  Matches  262;  Conservative    1;  Mismatches   82;  Indels   20;  Gaps    4;

Qy        334 CATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTCCTATAGTAAAAGACAGGTTAAT 393
              || || |||  ||||  |||| |:  |     |  |||||||| ||||||| |||||||
Db        422 CACTTACTCACCCATTGTAAGTGYTACGTTTGATCCTCCTATAATAAAAGATAGGTTAAC 481

Qy        394 AAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTACATGACATGGGAGTCTTCAGA 453
              ||||| ||| | | |||||||| ||||||||||||||||  |||||  |||| |||||||
Db        482 AAGAGAAAAGCATGACAAATTTGTTTAATCAAAGTTTTATGTGACACAGGAGCCTTCAGA 541

Qy        454 AATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTTTTTTTGCTGAGGTTCGATGAAGA 513
              |||||||||| |||||||||||| |||||        |||| |||| ||||| ||||||||
Db        542 AATGAAGACCAAAAGACCCAGGGAAAACT-----ACTTTTATGCTTAGGTTTGATGAAGA 596

Qy        514 ATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGATATGATCTAGTGGTAAAGGACTC 573
              |||| | |         |    |||||||||||| |   ||||  ||||||||| ||  |
Db        597 GTGGACATCCATGTTGAAATGTGATTAGACAAAAAG-GGTGATTGAGTGGTAAAAGA--C 653

Qy        574 AGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTTCTTGATCTCTCTCTCTCTATGT 633
              || || |||| ||| ||| ||||||| |||||||||||||| | | ||            |
Db        654 AGAGGAAAACCCAGTAAGTCCTGTCTGTTCAGATTCTTCTGGGTTTC------------T 701

Qy        634 ATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTTCTTCAATGAGGGTCTTCAAGGG 693
              ||| |||| ||||| ||   |||||| |||||| | |||   |||| | |||| |||||
Db        702 GTAGAATTCCTTCCTTCTAGATATGGGCCAGGACCCCTCTGATATGAAGTTCTTTAAGGG 761

Qy        694 AGAAG 698
              |||||
Db        762 AGAAG 766


RESULT 95
US-10-027-632-166951
; Sequence 166951, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 166951
;   LENGTH: 766
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-166951

  Query Match             2.5%;  Score 153.4;  DB 17;  Length 766;
  Best Local Similarity   71.8%;  Pred. No. 1e-26;
  Matches  262;  Conservative    1;  Mismatches   82;  Indels   20;  Gaps    4;

Qy        334 CATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTCCTATAGTAAAAGACAGGTTAAT 393
              || || |||  ||||  |||| |:  |     |  |||||||| ||||||| |||||||
Db        422 CACTTACTCACCCATTGTAAGTGYTACGTTTGATCCTCCTATAATAAAAGATAGGTTAAC 481

Qy        394 AAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTACATGACATGGGAGTCTTCAGA 453
              ||||| ||| | | |||||||| ||||||||||||||||  |||||  |||| |||||||
Db        482 AAGAGAAAAGCATGACAAATTTGTTTAATCAAAGTTTTATGTGACACAGGAGCCTTCAGA 541

Qy        454 AATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTTTTTTTGCTGAGGTTCGATGAAGA 513
              |||||||||| |||||||||||| |||||        |||| |||| ||||| ||||||||
Db        542 AATGAAGACCAAAAGACCCAGGGAAAACT-----ACTTTTATGCTTAGGTTTGATGAAGA 596

Qy        514 ATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGATATGATCTAGTGGTAAAGGACTC 573
              |||| | |         |    |||||||||||| |   ||||  ||||||||| ||  |
Db        597 GTGGACATCCATGTTGAAATGTGATTAGACAAAAAG-GGTGATTGAGTGGTAAAAGA--C 653

Qy        574 AGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTTCTTGATCTCTCTCTCTCTATGT 633
              || || |||| ||| ||| ||||||| |||||||||||||| | | ||            |
Db        654 AGAGGAAAACCCAGTAAGTCCTGTCTGTTCAGATTCTTCTGGGTTTC------------T 701

Qy        634 ATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTTCTTCAATGAGGGTCTTCAAGGG 693
              ||| |||| ||||| ||   |||||| |||||| | |||   |||| | |||| |||||
Db        702 GTAGAATTCCTTCCTTCTAGATATGGGCCAGGACCCCTCTGATATGAAGTTCTTTAAGGG 761
```

```
Qy        694 AGAAG 698
              |||||
Db        762 AGAAG 766


RESULT 96
US-10-027-632-260674/c
; Sequence 260674, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 260674
;   LENGTH: 654
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-260674

  Query Match             2.5%;  Score 152.4;  DB 13;  Length 654;
  Best Local Similarity   74.1%;  Pred. No. 1.6e-26;
  Matches  206;  Conservative    0;  Mismatches   71;  Indels    1;  Gaps    1;

Qy       2177 GGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCT 2236
              | ||| ||||||||  |||  ||  ||||||||||   | |||||| |||| | | | |
Db        280 GTGTAATCATCACCCAAATAGTGCACATTGTACCCGAGAGGTAATTGCTCATCCTTCACC 221

Qy       2237 GACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTCTACATCT 2296
                ||   ||| | || | |||  || ||||| |||| || ||||||| |||||  |
Db        220 CCACTCCCACCCTCCCACCTTTCCAAGCCTCCAATGTCTATTATTCCACTCTCTATGTTC 161

Qy       2297 ATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGTCTTTCTG 2356
              |||||||||||| |||||||| |||||||||||| || ||||  |||||| ||||| |
Db        160 ATGTGTACACATTGTTTAGCTCCCACTTATAAGTGAGAATATGCCGTATTTGACTTTC-G 102

Qy       2357 TGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAAGG 2416
              | ||||  ||  |||||||| ||||| |||| ||||||| |  || | | ||| |||||
Db        101 TTTCTGAGTTACTTTACTTAAGATAAAGGCCTCCAGTTCCACTTATGTTTCTGAAAAAGA 42

Qy       2417 CATGATTTCATTCTTTTTTATGGCTATGTTCTTTCCCA 2454
              ||| |||||||||||||||| |||  ||  |||  ||
Db         41 CATAATTTCATTCTTTTTTATAGCTGAGTATTTTTTCA 4


RESULT 97
US-10-027-632-260674/c
; Sequence 260674, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 260674
;   LENGTH: 654
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-260674

  Query Match             2.5%;  Score 152.4;  DB 17;  Length 654;
  Best Local Similarity   74.1%;  Pred. No. 1.6e-26;
  Matches  206;  Conservative    0;  Mismatches   71;  Indels    1;  Gaps    1;

Qy       2177 GGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCT 2236
              | ||| ||||||||  |||  ||  ||||||||||   | |||||| |||| | | | |
Db        280 GTGTAATCATCACCCAAATAGTGCACATTGTACCCGAGAGGTAATTGCTCATCCTTCACC 221

Qy       2237 GACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTCTACATCT 2296
                ||   ||| | || | |||  || ||||| |||| || ||||||| |||||  |
Db        220 CCACTCCCACCCTCCCACCTTTCCAAGCCTCCAATGTCTATTATTCCACTCTCTATGTTC 161

Qy       2297 ATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGTCTTTCTG 2356
              |||||||||||| |||||||| |||||||||||| || ||||  |||||| ||||| |
Db        160 ATGTGTACACATTGTTTAGCTCCCACTTATAAGTGAGAATATGCCGTATTTGACTTTC-G 102

Qy       2357 TGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAAGG 2416
              | ||||  ||  |||||||| ||||| |||| ||||||| |  || | | ||| |||||
Db        101 TTTCTGAGTTACTTTACTTAAGATAAAGGCCTCCAGTTCCACTTATGTTTCTGAAAAAGA 42

Qy       2417 CATGATTTCATTCTTTTTTATGGCTATGTTCTTTCCCA 2454
              ||| |||||||||||||||| |||  ||  |||  ||
Db         41 CATAATTTCATTCTTTTTTATAGCTGAGTATTTTTTCA 4


RESULT 98
US-09-854-867-358/c
; Sequence 358, Application US/09854867
; Publication No. US20030224356A1
; GENERAL INFORMATION:
;  APPLICANT: JOAN, KNOLL H
;  APPLICANT:  ROGAN, PETER K
;  TITLE OF INVENTION: SINGLE COPY GENOMIC HYBRIDIZATION PROBES AND METHOD OF GENERATING
;  FILE REFERENCE: 30307
;  CURRENT APPLICATION NUMBER: US/09/854,867
;  CURRENT FILING DATE:  2003-05-08
;  NUMBER OF SEQ ID NOS: 613
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 358
;   LENGTH: 897
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: repeat_region
;   LOCATION: (1)..(897)
;   OTHER INFORMATION: l1pb3
US-09-854-867-358

  Query Match             2.5%;  Score 151.4;  DB 10;  Length 897;
  Best Local Similarity   69.1%;  Pred. No. 3.4e-26;
```

```
  Matches  212;  Conservative    6;  Mismatches   87;  Indels    2;  Gaps    1;

Qy       2177 GGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCT 2236
              | |||  |||||| ||||  ||: ||||||||||  || |||||| ||| |  || |
Db        817 GTGTACCCATCACCCGAATAGTGWACATTGTACCCAATAGGTAATTTTTCATC--CCTCA 760

Qy       2237 GACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTCTACATCT 2296
              | |||    ||||| |||| |||| ||||| |||| || || |||| || ||   |
Db        759 CCCCTCTCCTACCCTCCACCTTTTGAGTCTCCARTGTCTATTATACCACTCTATATGCCC 700

Qy       2297 ATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGTCTTTCTG 2356
              | | ||| |||   |||||||| |||||||||||| ||||| ::|| ||||  ||||
Db        699 GTATATACCCATAGCTTAGCTCCCACTTATAAGTGAGAACATAYRATGTTTGGTTTTCCA 640

Qy       2357 TGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAAGG 2416
              |  |||  ||  || |||||  |||||:| | |||::|| ||| ||| |||||||||| |
Db        639 TTCCTGAGTTACTTCACTTAGAATAATRGTCTCCARYTCCATCCAGGTTGCTGCAAATGC 580

Qy       2417 CATGATTTCATTCTTTTTTATGGCTATGTTCTTTCCCAATTTAGATAAAGAACACTCGCA 2476
              ||| ||||||||||||||||||||  ||  | | |||  || ||| | | |||
Db        579 CATTATTTCATTCTTTTTTATGGCTGMGTAGTATTCCATGGTATATATATAYCACATTTT 520

Qy       2477 CTTGCTC 2483
              |||  ||
Db        519 CTTTATC 513


RESULT 99
US-10-786-970A-358/c
; Sequence 358, Application US/10786970A
; Publication No. US20050064449A1
; GENERAL INFORMATION:
;  APPLICANT: JOAN, KNOLL
;  APPLICANT:  ROGAN, PETER
;  TITLE OF INVENTION: SINGLE COPY GENOMIC HYBRIDIZATION PROBES AND METHOD OF GENERATING
;  FILE REFERENCE: 30307
;  CURRENT APPLICATION NUMBER: US/10/786,970A
;  CURRENT FILING DATE:  2004-02-24
;  PRIOR APPLICATION NUMBER: US/09/573,080
;  PRIOR FILING DATE: 2000-05-16
;  NUMBER OF SEQ ID NOS: 479
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 358
;   LENGTH: 897
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: repeat_region
;   LOCATION: (1)..(897)
;   OTHER INFORMATION: l1pb3
;  PUBLICATION INFORMATION:
;  PUBLICATION INFORMATION:
;   AUTHORS: Jurka, J; Walichiewicz, J; Milosavljevic, A
;   TITLE: Prototypic sequences for human repetitive DNA
;   JOURNAL: Journal of Molecular Evolution
;   VOLUME: 35
;   ISSUE: 4
;   PAGES: 286-291
;   DATE: 1992-10-__
;  PUBLICATION INFORMATION:
;  PUBLICATION INFORMATION:
;   AUTHORS: Jurka, J; Walichiewicz, J; Milosavljevic, A
;   TITLE: Prototypic sequences for human repetitive DNA
;   JOURNAL: Journal of Molecular Evolution
;   VOLUME: 35
;   ISSUE: 4
;   PAGES: 286-291
;   DATE: 1992-10-__
;   DATABASE ACCESSION NUMBER: Database of repetitive elements (repbase)
;   DATABASE ENTRY DATE: ____-__-__
;   DATABASE ACCESSION NUMBER: Database of repetitive elements (repbase)
;   DATABASE ENTRY DATE: 1996-01-26
;   DATABASE ENTRY DATE: ____-__-__
;   DATABASE ENTRY DATE: 1996-01-26
US-10-786-970A-358

  Query Match             2.5%;  Score 151.4;  DB 19;  Length 897;
  Best Local Similarity   69.1%;  Pred. No. 3.4e-26;
  Matches  212;  Conservative    6;  Mismatches   87;  Indels    2;  Gaps    1;

Qy       2177 GGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCT 2236
              | |||  |||||| ||||  ||: ||||||||||  || |||||| ||| |  || |
Db        817 GTGTACCCATCACCCGAATAGTGWACATTGTACCCAATAGGTAATTTTTCATC--CCTCA 760

Qy       2237 GACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTCTACATCT 2296
              | |||    ||||| |||| |||| ||||| |||| || || |||| || ||   |
Db        759 CCCCTCTCCTACCCTCCACCTTTTGAGTCTCCARTGTCTATTATACCACTCTATATGCCC 700

Qy       2297 ATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGTCTTTCTG 2356
              | | ||| |||   |||||||| |||||||||||| ||||| ::|| ||||  ||||
Db        699 GTATATACCCATAGCTTAGCTCCCACTTATAAGTGAGAACATAYRATGTTTGGTTTTCCA 640

Qy       2357 TGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAAGG 2416
              |  |||  ||  || |||||  |||||:| | |||::|| ||| ||| |||||||||| |
Db        639 TTCCTGAGTTACTTCACTTAGAATAATRGTCTCCARYTCCATCCAGGTTGCTGCAAATGC 580

Qy       2417 CATGATTTCATTCTTTTTTATGGCTATGTTCTTTCCCAATTTAGATAAAGAACACTCGCA 2476
              ||| ||||||||||||||||||||  ||  | | |||  || ||| | | |||
Db        579 CATTATTTCATTCTTTTTTATGGCTGMGTAGTATTCCATGGTATATATATAYCACATTTT 520

Qy       2477 CTTGCTC 2483
              |||  ||
Db        519 CTTTATC 513


RESULT 100
US-10-027-632-233626/c
; Sequence 233626, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 233626
;   LENGTH: 585
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-233626

  Query Match             2.5%;  Score 150.6;  DB 13;  Length 585;
  Best Local Similarity   75.5%;  Pred. No. 4.2e-26;
  Matches  200;  Conservative    0;  Mismatches   64;  Indels    1;  Gaps    1;

Qy       2177 GGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCT 2236
              | |||  ||||| |  |||||||| |||||||||| ||| |||||||    || ||  |
Db        458 GTGTAAACATCAACCAAATGATGTACATTGTACCCATTATGTAATTTTCATCCCTCTCCC 399
```

```
Qy       2237 GACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTCTACATCT 2296
              |    |    ||  | || | ||||||| ||  |||||| ||  ||||||  ||||  |
Db        398 CAAAACCCACCATCCCACCTTTCTGAGTCTTTATTGTCTAT-TTTCCACGTTCTATGTTC 340

Qy       2297 ATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGTCTTTCTG 2356
              |||||||| |||| ||||||||| |||||||||||| | |||| |||||||| |||||||
Db        339 ATGTGTACGCATTCTTTAGCTCCCACTTATAAGTGAGAGCATGTAATATTTGACTTTCTG 280

Qy       2357 TGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAAGG 2416
              | ||||   ||||| |||||  |||||||||| ||||||| ||| | | |||||| ||||
Db        279 TTTCTGAACTGTTTCACTTAAAATAATGGCCTCCAGTTCCATCCAYGTTGCTGCCAAAGA 220

Qy       2417 CATGATTTCATTCTTTTTTATGGCT 2441
              |||||||| ||||||||| || |||
Db        219 CATGATTTAATTCTTTTTCATTGCT 195
```

Search completed: April 22, 2005, 07:12:29
Job time : 2688.72 secs

This Page Blank (uspto)

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 20, 2005, 22:52:45 ; Search time 2165.11 Seconds
(without alignments)
13936.065 Million cell updates/sec

Title: US-09-939-209A-3_COPY_1_4970
Perfect score: 4970
Sequence: 1 agttcaagaccagcctgagc.....gtggctgaatgtctgccccc 4970

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5633728 seqs, 3035525691 residues

Total number of hits satisfying chosen parameters: 11267456

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq:*
19: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
20: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:*
21: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
22: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4970	100.0	20300	10	US-09-939-209A-3	Sequence 3, Appli
2	467.4	9.4	615	13	US-10-027-632-68208	Sequence 68208, A
3	467.4	9.4	615	13	US-10-027-632-68209	Sequence 68209, A
4	467.4	9.4	615	13	US-10-027-632-68210	Sequence 68210, A
5	467.4	9.4	615	13	US-10-027-632-68211	Sequence 68211, A
6	467.4	9.4	615	13	US-10-027-632-68742	Sequence 68742, A
7	467.4	9.4	615	13	US-10-027-632-68743	Sequence 68743, A
8	467.4	9.4	615	13	US-10-027-632-68744	Sequence 68744, A
9	467.4	9.4	615	13	US-10-027-632-70068	Sequence 70068, A
10	467.4	9.4	615	13	US-10-027-632-70069	Sequence 70069, A
11	467.4	9.4	615	13	US-10-027-632-70070	Sequence 70070, A
12	467.4	9.4	615	13	US-10-027-632-70071	Sequence 70071, A
13	467.4	9.4	615	13	US-10-027-632-295009	Sequence 295009,
14	467.4	9.4	615	13	US-10-027-632-295010	Sequence 295010,
15	467.4	9.4	615	13	US-10-027-632-311585	Sequence 311585,
16	467.4	9.4	615	13	US-10-027-632-311586	Sequence 311586,
17	467.4	9.4	615	17	US-10-027-632-68208	Sequence 68208, A
18	467.4	9.4	615	17	US-10-027-632-68209	Sequence 68209, A
19	467.4	9.4	615	17	US-10-027-632-68210	Sequence 68210, A
20	467.4	9.4	615	17	US-10-027-632-68211	Sequence 68211, A
21	467.4	9.4	615	17	US-10-027-632-68742	Sequence 68742, A
22	467.4	9.4	615	17	US-10-027-632-68743	Sequence 68743, A
23	467.4	9.4	615	17	US-10-027-632-68744	Sequence 68744, A
24	467.4	9.4	615	17	US-10-027-632-70068	Sequence 70068, A
25	467.4	9.4	615	17	US-10-027-632-70069	Sequence 70069, A
26	467.4	9.4	615	17	US-10-027-632-70070	Sequence 70070, A
27	467.4	9.4	615	17	US-10-027-632-70071	Sequence 70071, A
28	467.4	9.4	615	17	US-10-027-632-295009	Sequence 295009,
29	467.4	9.4	615	17	US-10-027-632-295010	Sequence 295010,
30	467.4	9.4	615	17	US-10-027-632-311585	Sequence 311585,
31	467.4	9.4	615	17	US-10-027-632-311586	Sequence 311586,
32	467.4	9.4	666	13	US-10-027-632-68745	Sequence 68745, A
33	467.4	9.4	666	17	US-10-027-632-68745	Sequence 68745, A
c 34	356.6	7.2	121600	18	US-10-723-860-1125	Sequence 1125, Ap
35	345.8	7.0	109730	19	US-10-741-600-17809	Sequence 17809, A
c 36	327.4	6.6	188971	17	US-10-235-192A-27	Sequence 27, Appl
37	323.6	6.5	163382	18	US-10-367-094-179	Sequence 179, App
c 38	322.2	6.5	26997	17	US-10-221-714A-474	Sequence 474, App
39	320.6	6.5	3310	13	US-10-027-632-113621	Sequence 113621,
40	320.6	6.5	3310	13	US-10-027-632-113622	Sequence 113622,
41	320.6	6.5	3310	17	US-10-027-632-113621	Sequence 113621,
42	320.6	6.5	3310	17	US-10-027-632-113622	Sequence 113622,
c 43	311.2	6.3	26997	17	US-10-221-714A-473	Sequence 473, App
44	308	6.2	1990	15	US-10-017-161-2383	Sequence 2383, Ap
45	308	6.2	1990	17	US-10-292-798-2025	Sequence 2025, Ap

ALIGNMENTS

RESULT 1
US-09-939-209A-3
; Sequence 3, Application US/09939209A
; Publication No. US20030113721A1
; GENERAL INFORMATION:
; APPLICANT: LEVITT, PAT R.
; APPLICANT: MIRNICS, KAROLY
; APPLICANT: KODAVALI, MENKATA CHOWDARI
; APPLICANT: NIMGAONKAR, VISHWAJIT L.
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR FACILITATING THE DIAGNOSIS AND TREATMENT
; TITLE OF INVENTION: SCHIZOPHRENIA
; FILE REFERENCE: 00-539-US
; CURRENT APPLICATION NUMBER: US/09/939,209A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: 60/228,021
; PRIOR FILING DATE: 2000-08-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 20300
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: A genomic sequence containing RGS4 nucleic acid sequence and sequ
; OTHER INFORMATION: ences upstream and downstream to the RGS4 nucleic acid sequence
US-09-939-209A-3

Query Match 100.0%; Score 4970; DB 10; Length 20300;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4970; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAATTA 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

```
Db          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAATTA 60

Qy         61 GACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAAT 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAAT 120

Qy        121 CACTTGAACCTGCTGGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAGC 180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACTTGAACCTGCTGGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAGC 180

Qy        181 CCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATA 240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATA 240

Qy        241 TATATATATATACACACACACACATATATATATACACATATATATACACATATATACA 300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TATATATATATACACACACACACATATATATATACACATATATATACACATATATACA 300

Qy        301 TATATACACATATATACACATATATATACACATACATATGTACACATATATATACACATA 360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TATATACACATATATACACATATATATACACATACATATGTACACATATATATACACATA 360

Qy        361 TGTATACACATATATACACATATATACACACATATATACACATATATACACACATATATA 420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGTATACACATATATACACATATATACACACATATATACACATATATACACACATATATA 420

Qy        421 CACATATATACACATATATACACATATACACATATATACACATATATACATATATACACA 480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CACATATATACACATATATACACATATACACATATATACACATATATACATATATACACA 480

Qy        481 TATATATAATATACACACATATATATACACATATATACACACATATATACACATATATAC 540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TATATATAATATACACACATATATATACACATATATACACACATATATACACATATATAC 540

Qy        541 ACATATATATACACATATATACACATATATACATATATACACATATATATACATATATAC 600
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ACATATATATACACATATATACACATATATACATATATACACATATATATACATATATAC 600

Qy        601 ACATATATACATATATACACATATATACATATATACACACATATATACACATACATATAC 660
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ACATATATACATATATACACATATATACATATATACACACATATATACACATACATATAC 660

Qy        661 ACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATAT 720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATAT 720

Qy        721 GCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTT 780
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTT 780

Qy        781 TGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATAC 840
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATAC 840

Qy        841 ATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATATATATA 900
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 ATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATATATATA 900

Qy        901 TAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTATAGAAGAATGG 960
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTATAGAAGAATGG 960

Qy        961 CTAAGATGCCAAAGCTCAGTGTATGTGGCAGAGGCATGGTGTAGGGTGTGTCCAGGTTCA 1020
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTAAGATGCCAAAGCTCAGTGTATGTGGCAGAGGCATGGTGTAGGGTGTGTCCAGGTTCA 1020

Qy       1021 TATATTGCATTAAGTGTGAGAACACCCTGGAGTATGAACCAAGAAAATGCAAAAGCCAGA 1080
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TATATTGCATTAAGTGTGAGAACACCCTGGAGTATGAACCAAGAAAATGCAAAAGCCAGA 1080

Qy       1081 AGTGATGGAGGAAATGAGACACAATAATGAAGATATTGAGAGGAGGGTGTGGGCCTAGAG 1140
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AGTGATGGAGGAAATGAGACACAATAATGAAGATATTGAGAGGAGGGTGTGGGCCTAGAG 1140

Qy       1141 TGAAGCTTTTCGTGCCAGTACTTCTTTTGAAGGCCCAGTTCTCTTCTCTCTCGGGGGCTC 1200
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 TGAAGCTTTTCGTGCCAGTACTTCTTTTGAAGGCCCAGTTCTCTTCTCTCTCGGGGGCTC 1200

Qy       1201 CTTCATCTCTCATAGAGTCCACAGCTTTTAAGGGCCAACACTTGAGGTCAGCCTGGCTCT 1260
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 CTTCATCTCTCATAGAGTCCACAGCTTTTAAGGGCCAACACTTGAGGTCAGCCTGGCTCT 1260

Qy       1261 CTCATTTGAGCTGGATAGAACATTTTAGAGCACCATCTATTCTTCAAGAGGAAGTTTAAA 1320
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CTCATTTGAGCTGGATAGAACATTTTAGAGCACCATCTATTCTTCAAGAGGAAGTTTAAA 1320

Qy       1321 AATAAAAGAACCTTGAAGAGGAAAAAATGTAGACATTCAATCTAACCTTTTCATTTTACT 1380
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AATAAAAGAACCTTGAAGAGGAAAAAATGTAGACATTCAATCTAACCTTTTCATTTTACT 1380

Qy       1381 AGCCAAAGCTAAATAGAATGCAGGTTACCTGTTTTTCAGCCAGGCACCATCATTTCCTAA 1440
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 AGCCAAAGCTAAATAGAATGCAGGTTACCTGTTTTTCAGCCAGGCACCATCATTTCCTAA 1440

Qy       1441 TTGTTATAAAATTTATTATTATTGTTGTTATTATTATTATTTGCCATAAGAAGTTTCCCA 1500
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 TTGTTATAAAATTTATTATTATTGTTGTTATTATTATTATTTGCCATAAGAAGTTTCCCA 1500

Qy       1501 TATCCTTTTAGTATAACAAAAACACAATTCACAAGCATTATAAAACCCATGGTGTCTAAC 1560
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 TATCCTTTTAGTATAACAAAAACACAATTCACAAGCATTATAAAACCCATGGTGTCTAAC 1560

Qy       1561 TATTAAAAAAAATTAAGTGGAACACACTTGTCCCAGCTACTGGGGAGGCTGAGGAGGGAGG 1620
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 TATTAAAAAAAATTAAGTGGAACACACTTGTCCCAGCTACTGGGGAGGCTGAGGAGGGAGG 1620

Qy       1621 ATCACGTGATCCCAGGGGGTCAAGGTTATGGAGAGCTATGATTGTGCCACTGCACTCCAG 1680
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 ATCACGTGATCCCAGGGGGTCAAGGTTATGGAGAGCTATGATTGTGCCACTGCACTCCAG 1680

Qy       1681 CCTGGGTGACAGGGAAAGACCCTGTCTCTAAAATTTTTTTTAAAAAAACTAAACTGGTTT 1740
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 CCTGGGTGACAGGGAAAGACCCTGTCTCTAAAATTTTTTTTAAAAAAACTAAACTGGTTT 1740

Qy       1741 TATTACAGAGATTCTGGAGACAGCTACACATAAAAGGGTGGTATGCCTCATATTAGCTAC 1800
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 TATTACAGAGATTCTGGAGACAGCTACACATAAAAGGGTGGTATGCCTCATATTAGCTAC 1800

Qy       1801 CCAGGGAGGTGGAATGCCAACTTAGGTGGTGTCACCACTATTAAAAATGCCCCAAAGCAA 1860
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 CCAGGGAGGTGGAATGCCAACTTAGGTGGTGTCACCACTATTAAAAATGCCCCAAAGCAA 1860

Qy       1861 TCAAAACTGAGAACTTCCTGGGAGCTTAGCATTGTGCAAAAGCAGCACAAAACACTTAAA 1920
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 TCAAAACTGAGAACTTCCTGGGAGCTTAGCATTGTGCAAAAGCAGCACAAAACACTTAAA 1920

Qy       1921 CAATTCACAGTTGTGTTGGAATGGGAAGGCCTGGAAATATAAACCAAAGAGTATATTGTC 1980
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 CAATTCACAGTTGTGTTGGAATGGGAAGGCCTGGAAATATAAACCAAAGAGTATATTGTC 1980

Qy       1981 TAAATTGATAGAGATTACAATTGCCTGAAAGAAAAAGTTGACTTTTAACTAGAATGTTCA 2040
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 TAAATTGATAGAGATTACAATTGCCTGAAAGAAAAAGTTGACTTTTAACTAGAATGTTCA 2040

Qy       2041 GAGTAGGTTTACAGAAGAAGCTCTTAAACTGGGCTCCAGTGGATTTGTCAATGCTTTGGA 2100
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 GAGTAGGTTTACAGAAGAAGCTCTTAAACTGGGCTCCAGTGGATTTGTCAATGCTTTGGA 2100

Qy       2101 AGCTGGTGGGGTGGGAGGGTTGGAGGGGGCATAAAAAGTCATGTTGGTATGCTCTGCTCA 2160
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 AGCTGGTGGGGTGGGAGGGTTGGAGGGGGCATAAAAAGTCATGTTGGTATGCTCTGCTCA 2160

Qy       2161 AGTCTCCATTCTGTTTCCTTTTCCTCTTTTCAATGTCATGTCCCATTATTTCATTATGGG 2220
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 AGTCTCCATTCTGTTTCCTTTTCCTCTTTTCAATGTCATGTCCCATTATTTCATTATGGG 2220
```

```
Qy      2221 CTTCCCTTTATCCAGGATCAATATGCCACCTCTTGGTTGTCTTTTACCTACTTCTCCACC 2280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2221 CTTCCCTTTATCCAGGATCAATATGCCACCTCTTGGTTGTCTTTTACCTACTTCTCCACC 2280

Qy      2281 TCACTATGGAATCGTCCTTGGGTAGCTCCTGTGCTTGGGAACCTGCACGGGCACTTTTCT 2340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2281 TCACTATGGAATCGTCCTTGGGTAGCTCCTGTGCTTGGGAACCTGCACGGGCACTTTTCT 2340

Qy      2341 GATGTCTTGATTCCAGCTTTACTCCTAAAACTTAAATGCTGAGGGGCCAACACCATGGCA 2400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2341 GATGTCTTGATTCCAGCTTTACTCCTAAAACTTAAATGCTGAGGGGCCAACACCATGGCA 2400

Qy      2401 GTGGTAGGGATGGGAATGGGGGTCTTGTAACACACTACATAAACTACACGAAATAAACTA 2460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2401 GTGGTAGGGATGGGAATGGGGGTCTTGTAACACACTACATAAACTACACGAAATAAACTA 2460

Qy      2461 CATGAAACTCAACATGTTTGCAAGACTCAGTTCACATCCATGAGGAGCTCATGCTTCTCC 2520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2461 CATGAAACTCAACATGTTTGCAAGACTCAGTTCACATCCATGAGGAGCTCATGCTTCTCC 2520

Qy      2521 CTCCTGCTCCCCTAGCACACATGATTATCTCTATTTGGAAATGTTTGGCATTTTTGGTGA 2580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2521 CTCCTGCTCCCCTAGCACACATGATTATCTCTATTTGGAAATGTTTGGCATTTTTGGTGA 2580

Qy      2581 AGTGAATGGTTCAATAACTTTCTCCACCATCAGAACAAAAGCTCTTTAAGGTTAGGGATG 2640
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2581 AGTGAATGGTTCAATAACTTTCTCCACCATCAGAACAAAAGCTCTTTAAGGTTAGGGATG 2640

Qy      2641 GGATCATACACACTTCCCTTGTCCAAGTCCCCATCACCCCTTATCTAGACAATTGCTACA 2700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2641 GGATCATACACACTTCCCTTGTCCAAGTCCCCATCACCCCTTATCTAGACAATTGCTACA 2700

Qy      2701 GTTTCCTACACACTCTTCTAACCTCTTGCAGTCTATTTTCATAAAACAGCTAGAGAACTT 2760
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2701 GTTTCCTACACACTCTTCTAACCTCTTGCAGTCTATTTTCATAAAACAGCTAGAGAACTT 2760

Qy      2761 TGAGATGTAAGTCAAAAAATAGAACATGTCGCTCTTTCCCATTGTTTTTGAAATAAAGTT 2820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2761 TGAGATGTAAGTCAAAAAATAGAACATGTCGCTCTTTCCCATTGTTTTTGAAATAAAGTT 2820

Qy      2821 CAACCCCCTTACCAGGGTCAACAAGGCCCTGCAATGATTTGGTCCTGTTAAAAATTCTTT 2880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2821 CAACCCCCTTACCAGGGTCAACAAGGCCCTGCAATGATTTGGTCCTGTTAAAAATTCTTT 2880

Qy      2881 AGCCTTAACTCATGCTGTTCTTCCTTACACTCACTGCATTCTAGCCATTGAGGTTTCTAT 2940
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2881 AGCCTTAACTCATGCTGTTCTTCCTTACACTCACTGCATTCTAGCCATTGAGGTTTCTAT 2940

Qy      2941 GCATCAAACTTTTTTTGGTCCCAGCACTGTGCACATCCTTCTGGGTAGAATGCCCCTTGA 3000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2941 GCATCAAACTTTTTTTGGTCCCAGCACTGTGCACATCCTTCTGGGTAGAATGCCCCTTGA 3000

Qy      3001 TTTGTATAATTAGCACCTCCTTCATCATTTAGGTCTTAGTATAACTACTACCTTCTTAGA 3060
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3001 TTTGTATAATTAGCACCTCCTTCATCATTTAGGTCTTAGTATAACTACTACCTTCTTAGA 3060

Qy      3061 GAAGCTCTGCTTCTTCATCCTATAAAAAAGTAAAATTCCTTACCCTGTTATTTTTTAAGT 3120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3061 GAAGCTCTGCTTCTTCATCCTATAAAAAAGTAAAATTCCTTACCCTGTTATTTTTTAAGT 3120

Qy      3121 CATCCGTGTTTCATTCTGTTAAAGTTCTTATCACAATTTATCATTATTTTATTTACAGTC 3180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3121 CATCCGTGTTTCATTCTGTTAAAGTTCTTATCACAATTTATCATTATTTTATTTACAGTC 3180

Qy      3181 ATGTGCCACATAACAATGTTTCAGTCAGGGATAGAACACAAATGTATCTGGCCCCATAAT 3240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3181 ATGTGCCACATAACAATGTTTCAGTCAGGGATAGAACACAAATGTATCTGGCCCCATAAT 3240

Qy      3241 ATTATAAGCTGAGAAATTTCTATTAACTAGTGATATCGCAGCCATCATAAGTGTAATGCA 3300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3241 ATTATAAGCTGAGAAATTTCTATTAACTAGTGATATCGCAGCCATCATAAGTGTAATGCA 3300

Qy      3301 GGACATTACCTTTTCTATGTTTAGATATGTTAGATACACAAATATATTTCATTGTGTTAT 3360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3301 GGACATTACCTTTTCTATGTTTAGATATGTTAGATACACAAATATATTTCATTGTGTTAT 3360

Qy      3361 AATTTCCTACAGTATTCAGTACAGTAACATGCTGTACAGGTTTGTAACCTAGGAGTAATA 3420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3361 AATTTCCTACAGTATTCAGTACAGTAACATGCTGTACAGGTTTGTAACCTAGGAGTAATA 3420

Qy      3421 GGCTATACCATACAGCTTAGGTGTGTAGTAGGCTATAACCATCTAGGTTTGTGTAAGTAC 3480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3421 GGCTATACCATACAGCTTAGGTGTGTAGTAGGCTATAACCATCTAGGTTTGTGTAAGTAC 3480

Qy      3481 ATTCTATGATATTCCCACAATGATGAAATCACCTAACTACACATTTCTCAGAATGTTTCA 3540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3481 ATTCTATGATATTCCCACAATGATGAAATCACCTAACTACACATTTCTCAGAATGTTTCA 3540

Qy      3541 CTGTTGTGAAGTGACCCATGACTATATTTTCCTATATACTTGATATTTTTGTGCATCTGC 3600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3541 CTGTTGTGAAGTGACCCATGACTATATTTTCCTATATACTTGATATTTTTGTGCATCTGC 3600

Qy      3601 CCATGAGAATGTAGTGTAAGATCAAAGGATGCAAGAATGGGTTCTATCCAGTATAGTACC 3660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3601 CCATGAGAATGTAGTGTAAGATCAAAGGATGCAAGAATGGGTTCTATCCAGTATAGTACC 3660

Qy      3661 CACTACACTGGTGGATGTCAATATGTATTTGTTAGATTAATATCTCAAGAATGAGCACCT 3720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3661 CACTACACTGGTGGATGTCAATATGTATTTGTTAGATTAATATCTCAAGAATGAGCACCT 3720

Qy      3721 TTCTCAGACACATAAAAGATGCTCAATATAAAAGTTTGTTGAACTGAACGTTATTGGCAA 3780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3721 TTCTCAGACACATAAAAGATGCTCAATATAAAAGTTTGTTGAACTGAACGTTATTGGCAA 3780

Qy      3781 ATGTAACATGATCGGATTTAAAGAGGAGCGAAACAGAGGTCTGGCTCAAACACCATACTT 3840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3781 ATGTAACATGATCGGATTTAAAGAGGAGCGAAACAGAGGTCTGGCTCAAACACCATACTT 3840

Qy      3841 CTAGAGTGCATAAGAGGTAGCAGTTGATTACCACTGGCGACAGGAGAAAAAAGAGCTTGA 3900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3841 CTAGAGTGCATAAGAGGTAGCAGTTGATTACCACTGGCGACAGGAGAAAAAAGAGCTTGA 3900

Qy      3901 CCGCAGGGTACTGTGAAGACATTTCAGGTTGATGGCACAGAACAGGGGAAATACATAAAT 3960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3901 CCGCAGGGTACTGTGAAGACATTTCAGGTTGATGGCACAGAACAGGGGAAATACATAAAT 3960

Qy      3961 GTGTGGGAATATTCAGTGGTCTGGGATGACTACATAGTAGAATATAATGAAGAAAAGAGT 4020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3961 GTGTGGGAATATTCAGTGGTCTGGGATGACTACATAGTAGAATATAATGAAGAAAAGAGT 4020

Qy      4021 GGAAGGGAAAGATGAAAAGTTGGAATGGGGATGAATTATGAAAGTACCAGAATGTTATGC 4080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4021 GGAAGGGAAAGATGAAAAGTTGGAATGGGGATGAATTATGAAAGTACCAGAATGTTATGC 4080

Qy      4081 TAAGGAATCTAGATTTTAAAATGTGAGGGCAAATTGAAGTCCTGGGCACGTTACAAAACT 4140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4081 TAAGGAATCTAGATTTTAAAATGTGAGGGCAAATTGAAGTCCTGGGCACGTTACAAAACT 4140

Qy      4141 AGAGGTCATAAAGTTTACCCTAATTTACCAAGATTTCCTAGAGGATCTATAATTGGAATC 4200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4141 AGAGGTCATAAAGTTTACCCTAATTTACCAAGATTTCCTAGAGGATCTATAATTGGAATC 4200

Qy      4201 CAGATCTGCCTCTCTGTAAAGTTCAAGCACTTTCCATGACACCATACTGTTTCTTTCCAC 4260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4201 CAGATCTGCCTCTCTGTAAAGTTCAAGCACTTTCCATGACACCATACTGTTTCTTTCCAC 4260

Qy      4261 CTGCACAATGCAAATGAACTCTTATGAAACTGCTGTTTCTATCCTGGGCTAAATGTTGCA 4320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4261 CTGCACAATGCAAATGAACTCTTATGAAACTGCTGTTTCTATCCTGGGCTAAATGTTGCA 4320

Qy      4321 GAAAAAAGATTTAATCTTTGGGATAAGGCTATTTTGGGTTTTCTCCTACTTCTTGGGAAA 4380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4321 GAAAAAAGATTTAATCTTTGGGATAAGGCTATTTTGGGTTTTCTCCTACTTCTTGGGAAA 4380

Qy      4381 CAAGGTTTTCTTCCCCTGGCTAATTAAGTGTGGTATTGTTCTTCCAGGGAAATCAGTGAT 4440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       4381 CAAGGTTTTCTTCCCCTGGCTAATTAAGTGTGGTATTGTTCTTCCAGGGAAATCAGTGAT 4440

Qy       4441 GCATCACCTGCTGCTATCAAATGTCAGGGTTGGAGTTCCTGATTTATTGCATGTGCCCAC 4500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4441 GCATCACCTGCTGCTATCAAATGTCAGGGTTGGAGTTCCTGATTTATTGCATGTGCCCAC 4500

Qy       4501 AAAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCT 4560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4501 AAAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCT 4560

Qy       4561 GTTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAA 4620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4561 GTTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAA 4620

Qy       4621 ATCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGG 4680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4621 ATCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGG 4680

Qy       4681 AGGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCT 4740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4681 AGGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCT 4740

Qy       4741 GAGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAACATCTGGTAGA 4800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4741 GAGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAACATCTGGTAGA 4800

Qy       4801 ACTAAAGAACGCAACGCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCC 4860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4801 ACTAAAGAACGCAACGCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCC 4860

Qy       4861 TTGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCT 4920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4861 TTGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCT 4920

Qy       4921 AGAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCC 4970
             |||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4921 AGAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCC 4970
```

RESULT 2
US-10-027-632-68208
; Sequence 68208, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 68208
; LENGTH: 615
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-68208

Query Match 9.4%; Score 467.4; DB 13; Length 615;
Best Local Similarity 99.1%; Pred. No. 5.3e-87;
Matches 465; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

```
Qy       4502 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 4561
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 60

Qy       4562 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAA 4621
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db         61 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAM 120

Qy       4622 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 4681
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 180

Qy       4682 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 4741
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 240

Qy       4742 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAACATCTGGTAGAA 4801
             |||||||||||||||||||||||||||||||||||||||||||||||:||||||||||||
Db        241 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAAYATCTGGTAGAA 300

Qy       4802 CTAAAGAACGCAACGCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 4861
             |||||||||||||:|||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTAAAGAACGCAACKCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 360

Qy       4862 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 4921
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 420

Qy       4922 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCC 4970
             |||||||||||||||||||||||||||||||||||||||||||||||:
Db        421 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCY 469
```

RESULT 3
US-10-027-632-68209
; Sequence 68209, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 68209
; LENGTH: 615
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-68209

Query Match 9.4%; Score 467.4; DB 13; Length 615;
Best Local Similarity 99.1%; Pred. No. 5.3e-87;
Matches 465; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

```
Qy       4502 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 4561
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 60

Qy       4562 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAA 4621
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db         61 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAM 120

Qy       4622 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 4681
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 180

Qy       4682 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 4741
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 240

Qy       4742 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAACATCTGGTAGAA 4801
             ||||||||||||||||||||||||||||||||||||||||||||||:||||||||||||||
Db         241 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAAYATCTGGTAGAA 300

Qy       4802 CTAAAGAACGCAACGCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 4861
             ||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||||||
Db         301 CTAAAGAACGCAACKCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 360

Qy       4862 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 4921
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         361 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 420

Qy       4922 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCC 4970
             ||||||||||||||||||||||||||||||||||||||||||||||||:
Db         421 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCY 469
```

RESULT 4
US-10-027-632-68210
; Sequence 68210, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 68210
; LENGTH: 615
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-68210

Query Match 9.4%; Score 467.4; DB 13; Length 615;
Best Local Similarity 99.1%; Pred. No. 5.3e-87;
Matches 465; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

```
Qy       4502 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 4561
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 60

Qy       4562 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAA 4621
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db         61 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAM 120

Qy       4622 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 4681
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 180

Qy       4682 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 4741
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 240

Qy       4742 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAACATCTGGTAGAA 4801
             ||||||||||||||||||||||||||||||||||||||||||||||:||||||||||||||
Db         241 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAAYATCTGGTAGAA 300

Qy       4802 CTAAAGAACGCAACGCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 4861
             ||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||||||
Db         301 CTAAAGAACGCAACKCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 360

Qy       4862 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 4921
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         361 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 420

Qy       4922 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCC 4970
             ||||||||||||||||||||||||||||||||||||||||||||||||:
Db         421 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCY 469
```

RESULT 5
US-10-027-632-68211
; Sequence 68211, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 68211
; LENGTH: 615
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-68211

Query Match 9.4%; Score 467.4; DB 13; Length 615;
Best Local Similarity 99.1%; Pred. No. 5.3e-87;
Matches 465; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

```
Qy       4502 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 4561
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 60

Qy       4562 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAA 4621
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
```

```
Db         61 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAM 120

Qy       4622 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 4681
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 180

Qy       4682 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 4741
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 240

Qy       4742 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAACATCTGGTAGAA 4801
             ||||||||||||||||||||||||||||||||||||||||||||||:||||||||||||
Db        241 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAAYATCTGGTAGAA 300

Qy       4802 CTAAAGAACGCAACGCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 4861
             ||||||||||||||:||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTAAAGAACGCAACKCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 360

Qy       4862 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 4921
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 420

Qy       4922 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCC 4970
             |||||||||||||||||||||||||||||||||||||||||||||||||:
Db        421 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCY 469
```

RESULT 6
US-10-027-632-68742
; Sequence 68742, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 68742
; LENGTH: 615
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-68742

Query Match 9.4%; Score 467.4; DB 13; Length 615;
Best Local Similarity 99.1%; Pred. No. 5.3e-87;
Matches 465; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

```
Qy       4502 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 4561
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 60

Qy       4562 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAA 4621
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db         61 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAM 120

Qy       4622 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 4681
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 180

Qy       4682 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 4741
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 240

Qy       4742 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAACATCTGGTAGAA 4801
             ||||||||||||||||||||||||||||||||||||||||||||||:||||||||||||
Db        241 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAAYATCTGGTAGAA 300

Qy       4802 CTAAAGAACGCAACGCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 4861
             ||||||||||||||:||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTAAAGAACGCAACKCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 360

Qy       4862 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 4921
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 420

Qy       4922 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCC 4970
             |||||||||||||||||||||||||||||||||||||||||||||||||:
Db        421 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCY 469
```

RESULT 7
US-10-027-632-68743
; Sequence 68743, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 68743
; LENGTH: 615
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-68743

Query Match 9.4%; Score 467.4; DB 13; Length 615;
Best Local Similarity 99.1%; Pred. No. 5.3e-87;
Matches 465; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

```
Qy       4502 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 4561
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 60

Qy       4562 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAA 4621
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db         61 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAM 120

Qy       4622 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 4681
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 180
```

```
Qy       4682 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 4741
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 240

Qy       4742 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAACATCTGGTAGAA 4801
             ||||||||||||||||||||||||||||||||||||||||||||||||||:||||||||||||
Db        241 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAAYATCTGGTAGAA 300

Qy       4802 CTAAAGAACGCAACGCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 4861
             ||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTAAAGAACGCAACKCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 360

Qy       4862 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 4921
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 420

Qy       4922 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCC 4970
             |||||||||||||||||||||||||||||||||||||||||||||||||:
Db        421 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCY 469
```

RESULT 8
US-10-027-632-68744
; Sequence 68744, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 68744
; LENGTH: 615
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-68744

Query Match 9.4%; Score 467.4; DB 13; Length 615;
Best Local Similarity 99.1%; Pred. No. 5.3e-87;
Matches 465; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

```
Qy       4502 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 4561
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 60

Qy       4562 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAA 4621
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db         61 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAM 120

Qy       4622 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 4681
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 180

Qy       4682 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 4741
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 240
```

```
Qy       4742 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAACATCTGGTAGAA 4801
             ||||||||||||||||||||||||||||||||||||||||||||||||||:||||||||||||
Db        241 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAAYATCTGGTAGAA 300

Qy       4802 CTAAAGAACGCAACGCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 4861
             ||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTAAAGAACGCAACKCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 360

Qy       4862 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 4921
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 420

Qy       4922 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCC 4970
             |||||||||||||||||||||||||||||||||||||||||||||||||:
Db        421 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCY 469
```

RESULT 9
US-10-027-632-70068
; Sequence 70068, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 70068
; LENGTH: 615
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-70068

Query Match 9.4%; Score 467.4; DB 13; Length 615;
Best Local Similarity 99.1%; Pred. No. 5.3e-87;
Matches 465; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

```
Qy       4502 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 4561
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 60

Qy       4562 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAA 4621
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db         61 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAM 120

Qy       4622 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 4681
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 180

Qy       4682 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 4741
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 240

Qy       4742 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAACATCTGGTAGAA 4801
             ||||||||||||||||||||||||||||||||||||||||||||||||||:||||||||||||
```

```
Db        241 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAAYATCTGGTAGAA 300

Qy       4802 CTAAAGAACGCAACGCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 4861
              ||||||||||||||:||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTAAAGAACGCAACKCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 360

Qy       4862 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 4921
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 420

Qy       4922 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCC 4970
              ||||||||||||||||||||||||||||||||||||||||||||||||:
Db        421 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCY 469


RESULT 10
US-10-027-632-70069
; Sequence 70069, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 70069
;   LENGTH: 615
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-70069

  Query Match             9.4%;  Score 467.4;  DB 13;  Length 615;
  Best Local Similarity   99.1%;  Pred. No. 5.3e-87;
  Matches  465;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4502 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 4561
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 60

Qy       4562 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAM 4621
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db         61 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAM 120

Qy       4622 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 4681
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 180

Qy       4682 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 4741
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 240

Qy       4742 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAACATCTGGTAGAA 4801
              ||||||||||||||||||||||||||||||||||||||||||||||||:|||||||||||
Db        241 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAAYATCTGGTAGAA 300

Qy       4802 CTAAAGAACGCAACGCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 4861
              ||||||||||||||:||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTAAAGAACGCAACKCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 360

Qy       4862 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 4921
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 420

Qy       4922 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCC 4970
              ||||||||||||||||||||||||||||||||||||||||||||||||:
Db        421 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCY 469


RESULT 11
US-10-027-632-70070
; Sequence 70070, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 70070
;   LENGTH: 615
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-70070

  Query Match             9.4%;  Score 467.4;  DB 13;  Length 615;
  Best Local Similarity   99.1%;  Pred. No. 5.3e-87;
  Matches  465;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4502 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 4561
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 60

Qy       4562 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAA 4621
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db         61 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAM 120

Qy       4622 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 4681
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 180

Qy       4682 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 4741
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 240

Qy       4742 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAACATCTGGTAGAA 4801
              ||||||||||||||||||||||||||||||||||||||||||||||||:|||||||||||
Db        241 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAAYATCTGGTAGAA 300

Qy       4802 CTAAAGAACGCAACGCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 4861
              ||||||||||||||:||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTAAAGAACGCAACKCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 360
```

```
Qy       4862 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 4921
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 420

Qy       4922 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCC 4970
             ||||||||||||||||||||||||||||||||||||||||||||||||||:
Db        421 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCY 469


RESULT 12
US-10-027-632-70071
; Sequence 70071, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 70071
;   LENGTH: 615
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-70071

  Query Match             9.4%;  Score 467.4;  DB 13;  Length 615;
  Best Local Similarity   99.1%;  Pred. No. 5.3e-87;
  Matches  465;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4502 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 4561
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 60

Qy       4562 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAA 4621
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db         61 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAM 120

Qy       4622 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 4681
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 180

Qy       4682 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 4741
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 240

Qy       4742 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAACATCTGGTAGAA 4801
             ||||||||||||||||||||||||||||||||||||||||||||||:||||||||||||||
Db        241 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAAYATCTGGTAGAA 300

Qy       4802 CTAAAGAACGCAACGCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 4861
             |||||||||||||:||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTAAAGAACGCAACKCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 360

Qy       4862 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 4921
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 420

Qy       4922 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCC 4970
             ||||||||||||||||||||||||||||||||||||||||||||||||||:
Db        421 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCY 469


RESULT 13
US-10-027-632-295009
; Sequence 295009, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 295009
;   LENGTH: 615
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-295009

  Query Match             9.4%;  Score 467.4;  DB 13;  Length 615;
  Best Local Similarity   99.1%;  Pred. No. 5.3e-87;
  Matches  465;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4502 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 4561
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 60

Qy       4562 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAA 4621
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db         61 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAM 120

Qy       4622 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 4681
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 180

Qy       4682 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 4741
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 240

Qy       4742 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAACATCTGGTAGAA 4801
             ||||||||||||||||||||||||||||||||||||||||||||||:||||||||||||||
Db        241 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAAYATCTGGTAGAA 300

Qy       4802 CTAAAGAACGCAACGCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 4861
             |||||||||||||:||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTAAAGAACGCAACKCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 360

Qy       4862 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 4921
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 420

Qy       4922 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCC 4970
             ||||||||||||||||||||||||||||||||||||||||||||||||||:
```

```
Db        421 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCY 469
```

RESULT 14
US-10-027-632-295010
; Sequence 295010, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 295010
; LENGTH: 615
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-295010

Query Match 9.4%; Score 467.4; DB 13; Length 615;
Best Local Similarity 99.1%; Pred. No. 5.3e-87;
Matches 465; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

```
Qy       4502 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 4561
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 60

Qy       4562 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAA 4621
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db         61 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAM 120

Qy       4622 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 4681
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 180

Qy       4682 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 4741
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 240

Qy       4742 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAACATCTGGTAGAA 4801
             ||||||||||||||||||||||||||||||||||||||||||||||||:||||||||||||
Db        241 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAAYATCTGGTAGAA 300

Qy       4802 CTAAAGAACGCAACGCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 4861
             |||||||||||||:||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTAAAGAACGCAACKCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 360

Qy       4862 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 4921
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 420

Qy       4922 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCC 4970
             |||||||||||||||||||||||||||||||||||||||||||||||||:
Db        421 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCY 469
```

RESULT 15
US-10-027-632-311585
; Sequence 311585, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 311585
; LENGTH: 615
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-311585

Query Match 9.4%; Score 467.4; DB 13; Length 615;
Best Local Similarity 99.1%; Pred. No. 5.3e-87;
Matches 465; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

```
Qy       4502 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 4561
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 60

Qy       4562 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAA 4621
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db         61 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAM 120

Qy       4622 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 4681
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 180

Qy       4682 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 4741
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 240

Qy       4742 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAACATCTGGTAGAA 4801
             ||||||||||||||||||||||||||||||||||||||||||||||||:||||||||||||
Db        241 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAAYATCTGGTAGAA 300

Qy       4802 CTAAAGAACGCAACGCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 4861
             |||||||||||||:||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTAAAGAACGCAACKCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 360

Qy       4862 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 4921
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 420

Qy       4922 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCC 4970
             |||||||||||||||||||||||||||||||||||||||||||||||||:
Db        421 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCY 469
```

RESULT 16
US-10-027-632-311586
; Sequence 311586, Application US/10027632

```
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 311586
;   LENGTH: 615
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-311586

  Query Match             9.4%;  Score 467.4;  DB 13;  Length 615;
  Best Local Similarity   99.1%;  Pred. No. 5.3e-87;
  Matches  465;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4502 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 4561
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 60

Qy       4562 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAA 4621
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db         61 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAM 120

Qy       4622 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 4681
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 180

Qy       4682 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 4741
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 240

Qy       4742 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAACATCTGGTAGAA 4801
             ||||||||||||||||||||||||||||||||||||||||||||||||:|||||||||||||||
Db        241 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAAYATCTGGTAGAA 300

Qy       4802 CTAAAGAACGCAACGCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 4861
             |||||||||||||:||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTAAAGAACGCAACKCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 360

Qy       4862 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 4921
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 420

Qy       4922 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCC 4970
             ||||||||||||||||||||||||||||||||||||||||||||||||:
Db        421 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCY 469


RESULT 17
US-10-027-632-68208
; Sequence 68208, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 68208
;   LENGTH: 615
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-68208

  Query Match             9.4%;  Score 467.4;  DB 17;  Length 615;
  Best Local Similarity   99.1%;  Pred. No. 5.3e-87;
  Matches  465;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4502 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 4561
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 60

Qy       4562 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAA 4621
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db         61 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAM 120

Qy       4622 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 4681
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 180

Qy       4682 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 4741
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 240

Qy       4742 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAACATCTGGTAGAA 4801
             ||||||||||||||||||||||||||||||||||||||||||||||||:|||||||||||||||
Db        241 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAAYATCTGGTAGAA 300

Qy       4802 CTAAAGAACGCAACGCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 4861
             |||||||||||||:||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTAAAGAACGCAACKCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 360

Qy       4862 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 4921
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 420

Qy       4922 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCC 4970
             ||||||||||||||||||||||||||||||||||||||||||||||||:
Db        421 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCY 469


RESULT 18
US-10-027-632-68209
; Sequence 68209, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
```

```
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 68209
;   LENGTH: 615
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-68209

  Query Match             9.4%;  Score 467.4;  DB 17;  Length 615;
  Best Local Similarity   99.1%;  Pred. No. 5.3e-87;
  Matches  465;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4502 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 4561
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 60

Qy       4562 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAA 4621
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db         61 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAM 120

Qy       4622 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 4681
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 180

Qy       4682 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 4741
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 240

Qy       4742 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAACATCTGGTAGAA 4801
             ||||||||||||||||||||||||||||||||||||||||||||||||||:||||||||||
Db        241 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAAYATCTGGTAGAA 300

Qy       4802 CTAAAGAACGCAACGCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 4861
             ||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTAAAGAACGCAACKCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 360

Qy       4862 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 4921
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 420

Qy       4922 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCC 4970
             ||||||||||||||||||||||||||||||||||||||||||||||||:
Db        421 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCY 469


RESULT 19
US-10-027-632-68210
; Sequence 68210, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 68210
;   LENGTH: 615
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-68210

  Query Match             9.4%;  Score 467.4;  DB 17;  Length 615;
  Best Local Similarity   99.1%;  Pred. No. 5.3e-87;
  Matches  465;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4502 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 4561
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 60

Qy       4562 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAA 4621
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db         61 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAM 120

Qy       4622 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 4681
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 180

Qy       4682 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 4741
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 240

Qy       4742 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAACATCTGGTAGAA 4801
             ||||||||||||||||||||||||||||||||||||||||||||||||||:||||||||||
Db        241 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAAYATCTGGTAGAA 300

Qy       4802 CTAAAGAACGCAACGCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 4861
             ||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTAAAGAACGCAACKCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 360

Qy       4862 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 4921
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 420

Qy       4922 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCC 4970
             ||||||||||||||||||||||||||||||||||||||||||||||||:
Db        421 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCY 469


RESULT 20
US-10-027-632-68211
; Sequence 68211, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE:.108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
```

```
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 68211
;   LENGTH: 615
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-68211

  Query Match             9.4%;  Score 467.4;  DB 17;  Length 615;
  Best Local Similarity   99.1%;  Pred. No. 5.3e-87;
  Matches  465;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4502 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 4561
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 60

Qy       4562 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAA 4621
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db         61 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAM 120

Qy       4622 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 4681
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 180

Qy       4682 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 4741
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 240

Qy       4742 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAACATCTGGTAGAA 4801
             ||||||||||||||||||||||||||||||||||||||||||||||||||||:||||||||||||
Db        241 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAAYATCTGGTAGAA 300

Qy       4802 CTAAAGAACGCAACGCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 4861
             |||||||||||||:||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTAAAGAACGCAACKCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 360

Qy       4862 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 4921
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 420

Qy       4922 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCC 4970
             ||||||||||||||||||||||||||||||||||||||||||||||||||:
Db        421 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCY 469


RESULT 21
US-10-027-632-68742
; Sequence 68742, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 68742
;   LENGTH: 615
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-68742

  Query Match             9.4%;  Score 467.4;  DB 17;  Length 615;
  Best Local Similarity   99.1%;  Pred. No. 5.3e-87;
  Matches  465;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4502 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 4561
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 60

Qy       4562 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAA 4621
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db         61 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAM 120

Qy       4622 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 4681
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 180

Qy       4682 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 4741
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 240

Qy       4742 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAACATCTGGTAGAA 4801
             ||||||||||||||||||||||||||||||||||||||||||||||||||||:||||||||||||
Db        241 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAAYATCTGGTAGAA 300

Qy       4802 CTAAAGAACGCAACGCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 4861
             |||||||||||||:||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTAAAGAACGCAACKCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 360

Qy       4862 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 4921
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 420

Qy       4922 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCC 4970
             ||||||||||||||||||||||||||||||||||||||||||||||||||:
Db        421 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCY 469


RESULT 22
US-10-027-632-68743
; Sequence 68743, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
```

```
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 68743
;   LENGTH: 615
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-68743

  Query Match             9.4%;  Score 467.4;  DB 17;  Length 615;
  Best Local Similarity   99.1%;  Pred. No. 5.3e-87;
  Matches  465;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4502 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 4561
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 60

Qy       4562 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAA 4621
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db         61 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAM 120

Qy       4622 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 4681
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 180

Qy       4682 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 4741
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 240

Qy       4742 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAACATCTGGTAGAA 4801
             ||||||||||||||||||||||||||||||||||||||||||||||||:|||||||||||
Db        241 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAAYATCTGGTAGAA 300

Qy       4802 CTAAAGAACGCAACGCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 4861
             |||||||||||||:|||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTAAAGAACGCAACKCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 360

Qy       4862 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 4921
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 420

Qy       4922 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCC 4970
             ||||||||||||||||||||||||||||||||||||||||||||||||:
Db        421 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCY 469


RESULT 23
US-10-027-632-68744
; Sequence 68744, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 68744
;   LENGTH: 615
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-68744

  Query Match             9.4%;  Score 467.4;  DB 17;  Length 615;
  Best Local Similarity   99.1%;  Pred. No. 5.3e-87;
  Matches  465;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4502 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 4561
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 60

Qy       4562 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAA 4621
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db         61 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAM 120

Qy       4622 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 4681
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 180

Qy       4682 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 4741
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 240

Qy       4742 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAACATCTGGTAGAA 4801
             ||||||||||||||||||||||||||||||||||||||||||||||||:|||||||||||
Db        241 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAAYATCTGGTAGAA 300

Qy       4802 CTAAAGAACGCAACGCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 4861
             |||||||||||||:|||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTAAAGAACGCAACKCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 360

Qy       4862 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 4921
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 420

Qy       4922 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCC 4970
             ||||||||||||||||||||||||||||||||||||||||||||||||:
Db        421 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCY 469


RESULT 24
US-10-027-632-70068
; Sequence 70068, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
```

```
; SEQ ID NO 70068
;   LENGTH: 615
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-70068

  Query Match             9.4%;  Score 467.4;  DB 17;  Length 615;
  Best Local Similarity   99.1%;  Pred. No. 5.3e-87;
  Matches  465;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4502 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 4561
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 60

Qy       4562 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAA 4621
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db         61 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAM 120

Qy       4622 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 4681
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 180

Qy       4682 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 4741
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 240

Qy       4742 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAACATCTGGTAGAA 4801
             ||||||||||||||||||||||||||||||||||||||||||||||||||:||||||||||
Db        241 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAAYATCTGGTAGAA 300

Qy       4802 CTAAAGAACGCAACGCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 4861
             |||||||||||||:||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTAAAGAACGCAACKCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 360

Qy       4862 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 4921
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 420

Qy       4922 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCC 4970
             ||||||||||||||||||||||||||||||||||||||||||||||||||:
Db        421 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCY 469


RESULT 25
US-10-027-632-70069
; Sequence 70069, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 70069
;   LENGTH: 615
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-70069

  Query Match             9.4%;  Score 467.4;  DB 17;  Length 615;
  Best Local Similarity   99.1%;  Pred. No. 5.3e-87;
  Matches  465;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4502 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 4561
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 60

Qy       4562 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAA 4621
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db         61 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAM 120

Qy       4622 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 4681
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 180

Qy       4682 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 4741
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 240

Qy       4742 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAACATCTGGTAGAA 4801
             ||||||||||||||||||||||||||||||||||||||||||||||||||:||||||||||
Db        241 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAAYATCTGGTAGAA 300

Qy       4802 CTAAAGAACGCAACGCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 4861
             |||||||||||||:||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTAAAGAACGCAACKCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 360

Qy       4862 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 4921
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 420

Qy       4922 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCC 4970
             ||||||||||||||||||||||||||||||||||||||||||||||||||:
Db        421 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCY 469


RESULT 26
US-10-027-632-70070
; Sequence 70070, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 70070
;   LENGTH: 615
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-70070
```

```
  Query Match             9.4%;  Score 467.4;  DB 17;  Length 615;
  Best Local Similarity   99.1%;  Pred. No. 5.3e-87;
  Matches  465;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4502 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 4561
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 60

Qy       4562 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAA 4621
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db         61 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAM 120

Qy       4622 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 4681
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 180

Qy       4682 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 4741
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 240

Qy       4742 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAACATCTGGTAGAA 4801
             ||||||||||||||||||||||||||||||||||||||||||||||||||:||||||||||||
Db        241 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAAYATCTGGTAGAA 300

Qy       4802 CTAAAGAACGCAACGCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 4861
             |||||||||||||:||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTAAAGAACGCAACKCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 360

Qy       4862 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 4921
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 420

Qy       4922 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCC 4970
             ||||||||||||||||||||||||||||||||||||||||||||||||:
Db        421 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCY 469


RESULT 27
US-10-027-632-70071
; Sequence 70071, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 70071
;   LENGTH: 615
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-70071

  Query Match             9.4%;  Score 467.4;  DB 17;  Length 615;
  Best Local Similarity   99.1%;  Pred. No. 5.3e-87;
  Matches  465;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4502 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 4561
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 60

Qy       4562 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAA 4621
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db         61 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAM 120

Qy       4622 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 4681
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 180

Qy       4682 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 4741
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 240

Qy       4742 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAACATCTGGTAGAA 4801
             ||||||||||||||||||||||||||||||||||||||||||||||||||:||||||||||||
Db        241 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAAYATCTGGTAGAA 300

Qy       4802 CTAAAGAACGCAACGCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 4861
             |||||||||||||:||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTAAAGAACGCAACKCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 360

Qy       4862 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 4921
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 420

Qy       4922 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCC 4970
             ||||||||||||||||||||||||||||||||||||||||||||||||:
Db        421 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCY 469


RESULT 28
US-10-027-632-295009
; Sequence 295009, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 295009
;   LENGTH: 615
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-295009

  Query Match             9.4%;  Score 467.4;  DB 17;  Length 615;
  Best Local Similarity   99.1%;  Pred. No. 5.3e-87;
  Matches  465;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4502 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 4561
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db          1 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 60

Qy       4562 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAA 4621
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db         61 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAM 120

Qy       4622 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 4681
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 180

Qy       4682 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 4741
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 240

Qy       4742 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAACATCTGGTAGAA 4801
             ||||||||||||||||||||||||||||||||||||||||||||||||||||:||||||||
Db        241 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAAYATCTGGTAGAA 300

Qy       4802 CTAAAGAACGCAACGCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 4861
             |||||||||||||:||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTAAAGAACGCAACKCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 360

Qy       4862 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 4921
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 420

Qy       4922 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCC 4970
             ||||||||||||||||||||||||||||||||||||||||||||||||:
Db        421 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCY 469


RESULT 29
US-10-027-632-295010
; Sequence 295010, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 295010
;   LENGTH: 615
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-295010

  Query Match             9.4%;  Score 467.4;  DB 17;  Length 615;
  Best Local Similarity   99.1%;  Pred. No. 5.3e-87;
  Matches  465;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4502 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 4561
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 60

Qy       4562 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAA 4621
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db         61 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAM 120

Qy       4622 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 4681
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 180

Qy       4682 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 4741
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 240

Qy       4742 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAACATCTGGTAGAA 4801
             ||||||||||||||||||||||||||||||||||||||||||||||||||||:||||||||
Db        241 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAAYATCTGGTAGAA 300

Qy       4802 CTAAAGAACGCAACGCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 4861
             |||||||||||||:||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTAAAGAACGCAACKCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 360

Qy       4862 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 4921
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 420

Qy       4922 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCC 4970
             ||||||||||||||||||||||||||||||||||||||||||||||||:
Db        421 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCY 469


RESULT 30
US-10-027-632-311585
; Sequence 311585, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 311585
;   LENGTH: 615
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-311585

  Query Match             9.4%;  Score 467.4;  DB 17;  Length 615;
  Best Local Similarity   99.1%;  Pred. No. 5.3e-87;
  Matches  465;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4502 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 4561
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 60

Qy       4562 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAA 4621
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db         61 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAM 120
```

```
Qy       4622 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 4681
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 180

Qy       4682 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 4741
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 240

Qy       4742 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAACATCTGGTAGAA 4801
             ||||||||||||||||||||||||||||||||||||||||||||||||||:||||||||||||
Db        241 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAAYATCTGGTAGAA 300

Qy       4802 CTAAAGAACGCAACGCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 4861
             ||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTAAAGAACGCAACKCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 360

Qy       4862 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 4921
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 420

Qy       4922 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCC 4970
             ||||||||||||||||||||||||||||||||||||||||||||||||:
Db        421 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCY 469
```

RESULT 31
US-10-027-632-311586
; Sequence 311586, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 311586
; LENGTH: 615
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-311586

Query Match 9.4%; Score 467.4; DB 17; Length 615;
Best Local Similarity 99.1%; Pred. No. 5.3e-87;
Matches 465; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

```
Qy       4502 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 4561
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 60

Qy       4562 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAA 4621
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db         61 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAM 120

Qy       4622 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 4681
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 180

Qy       4682 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 4741
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 240

Qy       4742 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAACATCTGGTAGAA 4801
             ||||||||||||||||||||||||||||||||||||||||||||||||||:||||||||||||
Db        241 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAAYATCTGGTAGAA 300

Qy       4802 CTAAAGAACGCAACGCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 4861
             ||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTAAAGAACGCAACKCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 360

Qy       4862 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 4921
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 420

Qy       4922 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCC 4970
             ||||||||||||||||||||||||||||||||||||||||||||||||:
Db        421 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCY 469
```

RESULT 32
US-10-027-632-68745
; Sequence 68745, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 68745
; LENGTH: 666
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-68745

Query Match 9.4%; Score 467.4; DB 13; Length 666;
Best Local Similarity 99.1%; Pred. No. 5.5e-87;
Matches 465; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

```
Qy       4502 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 4561
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 60

Qy       4562 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAA 4621
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db         61 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAM 120

Qy       4622 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 4681
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 180

Qy       4682 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 4741
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        181 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 240

Qy       4742 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAACATCTGGTAGAA 4801
             ||||||||||||||||||||||||||||||||||||||||||||||||||||:||||||||||
Db        241 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAAYATCTGGTAGAA 300

Qy       4802 CTAAAGAACGCAACGCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 4861
             |||||||||||||:|||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTAAAGAACGCAACKCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 360

Qy       4862 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 4921
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 420

Qy       4922 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCC 4970
             |||||||||||||||||||||||||||||||||||||||||||||||||:
Db        421 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCY 469


RESULT 33
US-10-027-632-68745
; Sequence 68745, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 68745
;   LENGTH: 666
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-68745

  Query Match             9.4%;  Score 467.4;  DB 17;  Length 666;
  Best Local Similarity   99.1%;  Pred. No. 5.5e-87;
  Matches  465;  Conservative    4;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4502 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 4561
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTG 60

Qy       4562 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAA 4621
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db         61 TTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAM 120

Qy       4622 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 4681
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGA 180

Qy       4682 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 4741
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTG 240

Qy       4742 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAACATCTGGTAGAA 4801
             ||||||||||||||||||||||||||||||||||||||||||||||||||||:||||||||||
Db        241 AGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAAYATCTGGTAGAA 300

Qy       4802 CTAAAGAACGCAACGCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 4861
             |||||||||||||:|||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTAAAGAACGCAACKCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCT 360

Qy       4862 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 4921
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTA 420

Qy       4922 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCC 4970
             |||||||||||||||||||||||||||||||||||||||||||||||||:
Db        421 GAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCY 469


RESULT 34
US-10-723-860-1125/c
; Sequence 1125, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
;  APPLICANT: Aziz, Natasha
;  APPLICANT:  Ginsburg, Wendy M.
;  APPLICANT:  Zlotnik, Albert
;  TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
;  TITLE OF INVENTION:  Methods for Screening for Soft Tissue Sarcoma Modulators
;  FILE REFERENCE: 05882.0193.NPUS01
;  CURRENT APPLICATION NUMBER: US/10/723,860
;  CURRENT FILING DATE:  2003-11-26
;  PRIOR APPLICATION NUMBER: 60/429,739
;  PRIOR FILING DATE: 2002-11-26
;  NUMBER OF SEQ ID NOS: 8393
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1125
;   LENGTH: 121600
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-723-860-1125

  Query Match             7.2%;  Score 356.6;  DB 18;  Length 121600;
  Best Local Similarity   74.7%;  Pred. No. 6.4e-63;
  Matches  543;  Conservative    0;  Mismatches  169;  Indels   15;  Gaps    7;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAA--- 57
              || || |||||| ||||    ||||  |||||||||| |||||||||||||| |||||
Db      44776 AGATCGAGACCATCCTGGCTAACACAGTGAAACCCCGTCTCTACTAAAAATCCAAAAAAA 44717

Qy         58 --TTAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGG 115
                |||| | ||| |||||      | |||||| |||||||||||  |||||| ||||||||
Db      44716 AGTTAGCCGGGCGTGGTGGCGGGCACCTGTAGTCCCAGCTACTCTGGAGGCTGAGGCAGG 44657

Qy        116 AGAATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCAC 174
              |||||  | ||||| ||  || || |||| |||| | ||| | ||||   |||| |||||
Db      44656 AGAATGGCGTGAACTTG--GGAGGCGGAGCTTGCAGTGAGCCGAGATAGCGCCACTGCAC 44599

Qy        175 TCCAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATA 234
              ||||| |  ||| || ||||||| |   ||||||| |||||||| |      | ||||||
Db      44598 TCCAGTCTGGGCGAC-AGAGCGAGACTCCATCTCAAAAAAAAAATA----TATATATATA 44544

Qy        235 TATATATATATATATATACACACACACACATATATATATACACATATATATACACATA 294
              || ||||||||| ||||||| | | | | | ||||||| ||| | | ||||||| | |||
Db      44543 TAGATATATATAGATATACAGATATATAGATATATATAGATATATAGATATATATAGATA 44484

Qy        295 TATACATATATACACATATATACACATATATATACACATACATATGTACACATATATATA 354
              |||| ||||||| | ||||||| | ||||||| | | ||| |||| || | |||||||
Db      44483 TATAGATATATATAGATATATAGATATATATAGATATATAGATATATATAGATATATAGA 44424

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              | ||||  ||| | | ||||| | | ||||| | | ||||| | | ||||| | | | |
Db      44423 TATATATAGATATATAGATATATATAGATATATAGATATATAGATATATATAGATATATA 44364

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              ||||| | | ||||| | ||||||| | ||||| | ||||||| | ||||||| |||||
```

```
Db      44363 GATATATATAGATATATAGATATATATAGATATATAGATATATAGATATATATAGATATA 44304

Qy        475 TACACATATATATAATATACACACATATATATACACATATATACACACATATATACACAT 534
              || |  |||| || ||||| | | ||||| ||| | ||| ||| | | ||||||| | ||
Db      44303 TAGATGTATAGAT-ATATATAGATATATAGATATATATAGATATATAGATATATAGATAT 44245

Qy        535 ATATACACATATATATACACATATATACACATATATACATATATACACATATATATACAT 594
              ||||| | ||||| ||| | ||| ||| | ||||| | ||||| | | ||| |||||  |
Db      44244 ATATAGATATATAGATATATATAGATATATATATAGATATATAGATATATAGATATATAG 44185

Qy        595 ATATACACATATATACATATATACA-CATATATACATATATACACACATATATACACATA 653
              ||||| | ||||||| ||||||| |  ||||||| ||||||| | | ||||| | | |||
Db      44184 ATATATAGATATATAGATATATAGATGATATATAGATATATATAGATATATAGATATATA 44125

Qy        654 CATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATA 713
              ||||| | | | ||||||||| ||| ||||| | ||||||| |  |||| |||| ||||
Db      44124 GATATATAGATATATAGATATATATAGATATATAGATATATAGATATATAGATATATATA 44065

Qy        714 TATATAT 720
              |||||||
Db      44064 GATATAT 44058


RESULT 35
US-10-741-600-17809
; Sequence 17809, Application US/10741600
; Publication No. US20050026169A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001499
;  CURRENT APPLICATION NUMBER: US/10/741,600
;  CURRENT FILING DATE:  2003-12-22
;  NUMBER OF SEQ ID NOS: 73997
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17809
;   LENGTH: 109730
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-741-600-17809

  Query Match             7.0%;  Score 345.8;  DB 19;  Length 109730;
  Best Local Similarity   73.7%;  Pred. No. 1.1e-60;
  Matches  537;  Conservative    0;  Mismatches  177;  Indels   15;  Gaps    7;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAATTA 60
              ||||  |||||||||||| |||||||| |||||||| |||| ||||||||||||||||||
Db     107515 AGTTTGAGACCAGCCTGACCAACATGGAGAAACCCCGTCTCCACTAAAAATACAAAATTA 107574

Qy         61 GACAGGCATGGTGATACACGCCTGT-AATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              |   |||||||||    ||  ||||| |||||||||||||| |||||| ||||||||||||
Db     107575 GCGGGGCATGGTGGCGCATCCCTGTAAATCCCAGCTACTTGGGAGGCTGAGGCAGGAGAA 107634

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCCA 178
              |||||||||  | | || |||||||||||| || ||| | ||||   |||||| |||||||
Db     107635 TCACTTGAA--TCCAGGAGGTGGAGGTTGTGGTGAGCCGAGATGGAGCCATTACACTCCA 107692

Qy        179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAA--------AAGGCATTTTA 230
              |||  |||||||||||| |||   ||||||| ||||| | |         | | || |||
Db     107693 GCCTGGGCAACAAGAGCAAAACTCCATCTCAAAAAAATATATATATATGTATGTATATTA 107752

Qy        231 TATATATATATATATATATATACACACACACACATATATATATACACATATATATACA 290
              |||| ||||| |||||||||| | | | | | |||| |||||||| ||| ||||| |
Db     107753 TATAAATATAAATATATATATTTATATAAATATATATATTTATATACATATAAATATAAA 107812

Qy        291 CATATATACATATATACACATATATACACATATATATACACATACATATGTACACATATA 350
              |||||||  |||||||| |||||||    |||| |||| ||||  ||| || | ||||||
Db     107813 TATATATATTTATATACATATATATATTTATATTTATATACATATGTATATAAATATATA 107872

Qy        351 TATACACATATGTATACACATATATACACATATATACACACATATATACACATATATACA 410
              |||  |  ||| || | | ||||||  | || |||| | | |||||||   || |||| |
Db     107873 TATTTATGTATATAAATATATATATTTATATGTATATAAATATATATATTTATGTATATA 107932

Qy        411 CACATATATACACATATATACACATATATACACATATACACATATATACACATATATACA 470
              | |||||||   |||| || | | ||||| | || || |  |||||| | ||||||
Db     107933 AATATATATATTTATATGTATATAAATATATATATCTATATGTATATAAATATATATTTT 107992

Qy        471 TATATACACATATATATAATATACACACATATATATACACATATATACACACATATATAC 530
              |||||  | ||| |||| ||||    | || |||||| | ||| ||    | || ||||
Db     107993 TATATGTATATAAATAT-ATATTTTTATATGTATATAAATATAGATTTTTATATGTATAT 108051

Qy        531 ACATATATACACATATATATACACATATATACACATATATACATATATACACATATATAT 590
              | |||||||   |||| |||| | |||||||   |||| || ||| ||| | || |||||
Db     108052 AAATATATATTTATATGTATATAAATATATATTTATATGTATATAAATATATATTTATAT 108111

Qy        591 ACATATATACACACATAT-ATACATATATACACATATATACATATATACACACATATATACA 649
              |||||  | | |||| | | |||| || | ||| ||| |||||||     ||||||  |
Db     108112 GTATATAAATATATATAACATATATTTATATATAAATATATATATATTTTTATATATGTA 108171

Qy        650 CATACATATACACACACATAGATATACATATATATACACATATATAT-ACGTATATATAT 708
              |||  |||||  | |   |||  |||| | ||||||| | |||||||| || |||||  |||
Db     108172 TATAAATATATATATTTATATGTATATAAATATATATAAATATATATAACATATATTTAT 108231

Qy        709 GTATATATA 717
              |||||||||
Db     108232 GTATATATA 108240


RESULT 36
US-10-235-192A-27/c
; Sequence 27, Application US/10235192A
; Publication No. US20040043389A1
; GENERAL INFORMATION:
;  APPLICANT: McCarthy, Jeanette
;  TITLE OF INVENTION: Methods and Compositions for Identifying
;  TITLE OF INVENTION:  Risk Factors for Abnormal Lipid Levels and the Diseases
;  TITLE OF INVENTION:  and Disorders Associated Therewith
;  FILE REFERENCE: MMI-011
;  CURRENT APPLICATION NUMBER: US/10/235,192A
;  CURRENT FILING DATE:  2002-09-04
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 27
;   LENGTH: 188971
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 145329-145428, 187896-187995, 195894-195993
;   OTHER INFORMATION: N = any nucleotide
US-10-235-192A-27

  Query Match             6.6%;  Score 327.4;  DB 17;  Length 188971;
  Best Local Similarity   67.6%;  Pred. No. 9.3e-57;
  Matches  601;  Conservative    0;  Mismatches  261;  Indels   27;  Gaps    9;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC--AAAAT 58
              || ||||||||| ||||  |||||||||||||| || ||||||||||||||  |||||
Db     104585 AGATCAAGACCATCCTGGCCAACATGGTGAAACTCCGTCTCTACTAAAAATACAAAAAAT 104526

Qy         59 TAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGA 118
              |||   ||| ||||| |    | |||||||||||||||||| |||||| || ||||||||
Db     104525 TAGCTGGGCGTGGTGGTGGGCTCCTGTAATCCCAGCTACTTGGGAGGCTGAAGCAGGAGA 104466

Qy        119 ATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTG---CAC 174
              ||| ||||||||  | || || |||||||| | ||| |||||||  |||||||   |||
Db     104465 ATCGCTTGAACC--CAGGAGGCGGAGGTTGCAGTGAGCCAAGATCGCGCCATTGCACCAC 104408

Qy        175 TCCAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATA 234
              |||||||  |||||  ||||| ||||||||||  |     | | |
Db     104407 TCCAGCCTGGGCAAAAAGAGAGAAACTCTGTCTCAAAAAAAAAAAAAAAAAAAAAAAAA 104348

Qy        235 TATATATATATATATATACACACACACACATATATATATACACATATATATACACATA 294
              |||||||||||||||| | | | |||||  | |||||||  ||||||||        |
Db     104347 AATATATATATATATATATATATATACACACACATATATATTCATATATA-------A 104295
```

```
Qy        295 TATACATATATACACATATATACACATATATATACACATACATATGTACACATATATATA 354
              |  | |||||||  ||||||||   ||  |||||  | ||  |||||  ||  ||||||||||
Db     104294 TTCATATATATATTCATATATA---ATTCATATATATATTCATATATATTCATATATATA 104238

Qy        355 CACATATGT-ATACACATATATACACATATATACACACATATATACACATATATACACAC 413
              |||||  |  ||  ||  |||||||  ||||||||  ||  |||||||  |||||||| |  |
Db     104237 TTCATATATAATTCATATATATATTCATATATATTCATATATATATTCATATATATATTC 104178

Qy        414 ATATATACACATATATACACATATATACACATATACACATATATACACATATATACATAT 473
              |||||||      ||  | |  ||||||  ||  |||||  ||||||||  |    |||||  ||||
Db     104177 ATATATA-----ATTCATATATATATTCATATATATTCATATATATATTCATATATATAT 104123

Qy        474 ATACACATATATATAATATACACACATATATATACACATATATACACACATATATACACA 533
              | | |||  |   |||||  |  |||||||| ||  |||||||  ||  ||||||  || |
Db     104122 TCATATATAATTCATATATATATTCATATATATTCATATATATATTCATATATATTCATA 104063

Qy        534 TATATACACATATATATACACATATATACACATATATACATATATACACATATATATACA 593
              ||||||  ||||||||||  ||||||||  ||||||||  ||  |||  | ||||  ||| |
Db     104062 TATATATTCATATATATATTCATATATATTCATATATATATTCATATATATATTCATATA 104003

Qy        594 TATATACACATATATACATATATACACATATATACATATATACACACATATATACACATA 653
              |||||  ||  |||||||   |||||  |  ||  |||  ||||||  ||  |  ||||  |  | |||
Db     104002 TATATTCATATATATATTCATATATATATTCATATATATATTCATATATATTCATATATA 103943

Qy        654 CATATACACACACATAGATATACA---TATATATACACATATATATACGTATATATATGT 710
              ||  | | |  |  ||  |||||  |    |||||||  |   |||||||  |  ||||||
Db     103942 TATTCATATATATATTCATATATAATTCATATATATATTCATATATATATTCATATATAA 103883

Qy        711 ATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGA 770
              ||  ||||||  | ||  |  |       |  |||   ||  ||    |  |    |||||
Db     103882 ATTCATATATAATTCATATATATATTCATATAGATATTCATATATAATTCATATAGATAT 103823

Qy        771 AAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTA 830
              |  |||  ||  |  |    |  ||   ||  |   | ||  |   |      ||  |||  |
Db     103822 TCATAGATATTCATATATATTCATATAGATATTCATATATTCATATATATTCATATATTC 103763

Qy        831 GGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGA 879
              | | ||  ||||  |  | |    | |   ||  |  |   |  ||  |  |
Db     103762 ATGTATATTCATATAGATATTCATATATAGATATTCATAGATATTCATA 103714


RESULT 37
US-10-367-094-179
; Sequence 179, Application US/10367094
; Publication No. US20040170982A1
; GENERAL INFORMATION:
;  APPLICANT: David W. Morris
;  APPLICANT:  Marc Malandro
;  TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
;  FILE REFERENCE: 529452001500
;  CURRENT APPLICATION NUMBER: US/10/367,094
;  CURRENT FILING DATE:  2003-02-14
;  NUMBER OF SEQ ID NOS: 203
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 179
;   LENGTH: 163382
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(163382)
;   OTHER INFORMATION: n = A,T,C or G
US-10-367-094-179

  Query Match             6.5%;  Score 323.6;  DB 18;  Length 163382;
  Best Local Similarity   68.2%;  Pred. No. 5.4e-56;
  Matches  495;  Conservative    0;  Mismatches  224;  Indels    7;  Gaps    3;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              |||||||||||||||||  |||||||||||||||||||||||||||||||||| ||||||
Db     144405 AGTTCAAGACCAGCCTGGCCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAAATT 144464

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              || |||||  |||||  ||  |||||||||||||||||    ||||||  ||||||||||||
Db     144465 AGCCAGGCGTGGTGGCACGTGCCTGTAATCCCAGCTACCCAGGAGGCTGAGGCAGGAGAA 144524

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              |  ||||||  |    | ||  | ||  |  |  |  |||||||||||||||||||||||
Db     144525 TTGCTTGAAGC-CAAGAGGCAGAAGTTGCAGTGAGCCGAGATCATGCCATTGCACTCCAG 144583

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  ||  ||||||  |||  |  ||||  |||  ||||||||||        |  |    |  |
Db     144584 CCTGGGGGACAAGAACGAGACTTCATGTCAAAAAAAAAAAATACACAGTCTTCCTTGGCT 144643

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
                          | | | | | | | || ||| |||| | |||| |||| | || |||
Db     144644 TAGCTGTATGGGAAATATATATATATATTTATTTATATATATATTTATATATATTTATTT 144703

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACAT 359
              |||||||   || |||| | || ||| || | |||  ||| || | || ||| || | |
Db     144704 ATATATATTTATTTATATATATTTATTTATATATATTTATTTATATATTTATTTATATTT 144763

Qy        360 ATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATAT 419
              || | || | || |||  | ||||||     | ||||  ||   ||||||  |   ||||||
Db     144764 ATTTTTATATATTTATTTATATATATTTGTATATATTTATTTATATATTTATTTATATAT 144823

Qy        420 ACACATATATACACATATATACACATATACACATATATACACATATATACATA-----TA 474
              |  ||||||  |||||||    |  | | |||| || | || |||| |||     ||
Db     144824 TTATTTATATATTTATATATATTAGTTCATATATATTTATATATTTATATATATTAGTTA 144883

Qy        475 TACACATATATATAATATACACACATATATATACACATATATACACACATATATACACAT 534
              || | || ||||||  ||   | | |||| |||| | || | || | |  ||| || | ||
Db     144884 TATATATTTATATATTAGTTATATATATTTATATATATTTTTATATATTTATTTATATAT 144943

Qy        535 ATATACACATATATATACACATATATACACATATATACATATATACACATATATATACAT 594
              || || | ||  | |||   |||||||  |||||||  |||||||  | |||||||  ||
Db     144944 ATGTATATATTAACATATTTATATATATTCATATATTTATATATAAATATATATATTAAT 145003

Qy        595 ATATACACATATATACATATATACACATATATACATATATACACACATATATACACATAC 654
              ||||  | ||||||  ||||||  | |||| | ||||||  | ||||||  ||   |||
Db     145004 ATATTTATATATATTTATATATTAATATATGTTTATATATTTATACATATTTATGTATAT 145063

Qy        655 ATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATAT 714
              |||  | | |  || |||||  |||||||  |  ||||||||  |||||||  | |||  |
Db     145064 TAATATATATATTTATATATATTTATATATGTATTTATATATATTTATATATTTATATTT 145123

Qy        715 ATATAT 720
              || |||
Db     145124 ATTTAT 145129


RESULT 38
US-10-221-714A-474/c
; Sequence 474, Application US/10221714A
; Publication No. US20040048254A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with
;  TITLE OF INVENTION:  tumor suppressor genes and oncogenes
;  FILE REFERENCE: 5013.1005
;  CURRENT APPLICATION NUMBER: US/10/221,714A
;  CURRENT FILING DATE:  2003-01-21
;  PRIOR APPLICATION NUMBER: PCT/EP01/02955
;  PRIOR FILING DATE: 2001-03-15
;  PRIOR APPLICATION NUMBER: DE 10013847.0
;  PRIOR FILING DATE: 2000-03-15
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
```

```
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 540
; SEQ ID NO 474
;   LENGTH: 26997
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-714A-474

  Query Match             6.5%;  Score 322.2;  DB 17;  Length 26997;
  Best Local Similarity   66.0%;  Pred. No. 4.7e-56;
  Matches  498;  Conservative    0;  Mismatches  253;  Indels    4;  Gaps    2;

Qy        207 TCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATATACACACACACACACA 266
              | |  |||||||| |     | ||||||||||||||||||||||||||||||| | | | |
Db       6221 TAAATAAAAAAAATATAATATATATATATATATATATATATATATACACACATATATATA 6162

Qy        267 TATATATATACACATATATATACACATATATACATATATACACATATATACACATATATA 326
              | ||||||||||| | ||||||| ||||| ||| ||||| ||||||||||||||||||
Db       6161 AACATATATACACACACATATACATATATACACACATATATACATATATACACATATATA 6102

Qy        327 TACACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATA 386
              || | ||| | || || | | ||||||| | | |  |||| | ||||||||| |||||||
Db       6101 TATATATATACATATATATACATATATATATACACATATATATATATATACATATATATA 6042

Qy        387 CACACATATATACACATATATACACACATATATACACATATATACACATATATACACATA 446
              || | | |||||||||||||||| | | ||| ||| | ||| ||| ||||||| ||| |||
Db       6041 CATATACATATACACATATATATATATATACATATATATACATATACATATACACATATA 5982

Qy        447 TACACATATATACACATATATACATATATACACATATATATA-ATATACACACATATATA 505
              || | ||||| | | ||| | |||||||||||||||||||| ||||||| | | ||| |
Db       5981 TATATATATACATATACATACATATATATACACATATATATATATATACATATACATACA 5922

Qy        506 TACACATATATACACACATATATACACATATATACACATATATATACACATATATACACA 565
              || | ||| | | | | ||||||| |||||||||||| |||||| | | ||| | ||| |
Db       5921 TATATATACACATATATATATATATACATATATACACGTATATACATATATACACACATA 5862

Qy        566 TATATACATATATACACATATATATACATATATACACATATATACATATATACACA---T 622
              | |||| |||  || | | ||||||| |||||||| ||||||| ||||||| |||   |
Db       5861 TTTATATATACGTATATACATATATATATATATACATATATATATATATATACATATT 5802

Qy        623 ATATACATATATACACACATATATACACATACATATACACACACATAGATATACATATAT 682
              ||||| ||||||| | |||||||| | |||||||| | | | ||| | ||| ||| ||
Db       5801 ATATATATATATATATACATATATAAATATACATATATATATATATACATATATACAT 5742

Qy        683 ATACACATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTA 742
              ||| | ||||||| |  |||||||| |||| ||||          || |     ||
Db       5741 ATATATATATATAAATATATATATATATACGTATATATATATACACACACACCATA 5682

Qy        743 GCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAA 802
                |   ||  | |    | | |  |||| |  |||  | | |  | |   | |  | |
Db       5681 AAATACTATACACCCATAAAAAAAAAAAAAAATAACATTTACAACAACTTAAATAAAATTA 5622

Qy        803 GGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTAT 862
                |    | |  | |  | ||  | |      | | | |   | | ||| |  ||
Db       5621 AAAACCATTATTCTAACCGAAATAACTCAAAAATAAAAAACCAAATATTATATATTCTCA 5562

Qy        863 CTGTAAGGTAGTTAAGAAGGTAACACTATATATATATATAGTGAAAGCAGTGTAAACCTA 922
              ||  |  ||  |  ||  ||  | | | | | | |||  | |  |  |||| |
Db       5561 CTTATAACTAAAAACTAAACTATAAAAACACAAAAACATAAAAATAATACAATAAACTTT 5502

Qy        923 AAGGATGGGCCAAGGATTTAAATGTTATAGAAGAA 957
              ||  |    ||  |   ||  || | || ||
Db       5501 AAAAACTCTAAAAAAAACATAAAAATAAATAAAAA 5467


RESULT 39
US-10-027-632-113621
; Sequence 113621, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 113621
;   LENGTH: 3310
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-113621

  Query Match             6.5%;  Score 320.6;  DB 13;  Length 3310;
  Best Local Similarity   65.7%;  Pred. No. 3.9e-56;
  Matches  559;  Conservative    0;  Mismatches  284;  Indels    8;  Gaps    6;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACA-AAATT 59
              |||||||.| || |||| ||| |||||||||||||| ||||||||  ||||||| |||||
Db       1488 AGTTCAAGTCTAGTCTGACCAATATGGTGAAACCCCGTCTCTACTGGAAATACAGAAATT 1547

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              || ||||  ||||  | | |||||||| ||| |||||||  |||||| ||||||||||||
Db       1548 AGCCAGGTGTGGTAGTGCTCGCCTGTAGTCCTAGCTACTCGGGAGGCTGAGGCAGGAGAA 1607

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCCA 178
              || |||||||||  || ||  |||||||| | ||| | ||||| |||| |||||||||
Db       1608 TCGCTTGAACCTG--GGAGGCAGAGGTTGCAGTGAGCCGAGATCCTGCCCCTGCACTCCA 1665

Qy        179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATA 238
              |||   |  | ||||||| ||    ||||| ||||||    | | | | || ||||||||
Db       1666 GCC--TGGCAACAGAGCGAGATTCTGTCTCAAAAAAAACCCCATGTA-TATACATATATA 1722

Qy        239 TATATATATATATACACACACACACACATATATATATACACATATATATACACATATATA 298
              ||||||||  | | | | |   | |||||||| |||| ||| ||||||||  |||||||
Db       1723 CGTATATATACACATATATATGTATACATATATGTATATACACATATATACGTATATATA 1782

Qy        299 CATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACA 358
              | |||| |||  |||| |  |||| ||| |  || |||   | | | | || ||   |
Db       1783 TGTGTATATACACGTATATATGTATACATATATGTATATACACATATACACATGTATGTA 1842

Qy        359 TATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATA 418
              ||  |||| |  |||||||   ||||||   | | ||||    ||||||||||  |||||
Db       1843 TACATATATATGTATATACGTGTATATATGTATACATATGTGTATATATACACGTATATA 1902

Qy        419 TACACATATATACACATATATACACATATACACATATATACACATATATACATATATACA 478
              ||||||||| || || | | || | |||     |||| |||||  ||||||||  || |
Db       1903 TACACATATGTATACGTGTGTATATATACGTGTATATACACACACGTATACATACGTATA 1962

Qy        479 CATATATAT-AATATACACACATATATATACACATATATACACACATATATACACATATA 537
              |||||  |  ||||| ||||| || |||||| || | || |    |||||| ||| | |
Db       1963 TATATACGTGTATATATACACACATGTATACATATGTGTATATGTGTATATATACACACA 2022

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATACATATA 597
              |  | | ||||||    |||||||||||||| || | ||| ||   |||||| ||| |||
Db       2023 TGTATACATATATGTGTATATATACACATATGTATACATACACGTGTATATACACACATA 2082

Qy        598 TACACATATATACATATATACACATATATACATATATACACACATATATACACATACATA 657
              |  | | ||| || | |||| | | | |||  |||| | | |  | |||||| |  ||
```

```
Db       2083 TGTATACATACACGTGTATATATACACATATGTATACATATATGTGTATACATACGCACG 2142

Qy        658 TACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATA 717
              || |||||  ||  |||||||||| | || | |||   | |  || | |||||||||| |||
Db       2143 TATACACATGTATGTATACATATTTGTATATATACTCACATATACACATATGTATACATA 2202

Qy        718 TATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGA 777
               |||          |  |         ||  |   | |   |  | | | |   | | | |
Db       2203 CATGTGTATATACACGCACACATACACATATATGTGCATATACGCACATATATACATATA 2262

Qy        778 GTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAA 837
              |   |||    || |||     | |   | ||       ||    ||  ||       | |
Db       2263 TGTACACATATGTATATATACACACATATATACATATATGTACACATGTACGTACACATA 2322

Qy        838 TACATAAATGT 848
              |||||| ||||
Db       2323 TACATATATGT 2333


RESULT 40
US-10-027-632-113622
; Sequence 113622, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 113622
;   LENGTH: 3310
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-113622

  Query Match             6.5%;  Score 320.6;  DB 13;  Length 3310;
  Best Local Similarity   65.7%;  Pred. No. 3.9e-56;
  Matches  559;  Conservative    0;  Mismatches  284;  Indels    8;  Gaps    6;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACA-AAATT 59
              |||||||| | || |||| ||| |||||||||||||| ||||||||  |||||| |||||
Db       1488 AGTTCAAGTCTAGTCTGACCAATATGGTGAAACCCCGTCTCTACTGGAAATACAGAAATT 1547

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              || ||||  ||||  | | |||||||| ||| |||||||  ||||||.||||||||||||
Db       1548 AGCCAGGTGTGGTAGTGCTCGCCTGTAGTCCTAGCTACTCGGGAGGCTGAGGCAGGAGAA 1607

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCCA 178
              || ||||||||||  || ||  |||||||| | ||| | |||||  ||||  |||||||||
Db       1608 TCGCTTGAACCTG--GGAGGCAGAGGTTGCAGTGAGCCGAGATCCTGCCCCTGCACTCCA 1665

Qy        179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATA 238
              |||   |  |  ||||||| ||    ||||| ||||||     | | | | || |||||||
Db       1666 GCC--TGGCAACAGAGCGAGATTCTGTCTCAAAAAAAACCCCATGTA-TATACATATATA 1722

Qy        239 TATATATATATATACACACACACACACATATATATATACACATATATATACACATATATA 298
```

```
              |||||||| | | | | |   | |||||||| |||| ||| ||||||||  |||||||
Db       1723 CGTATATATACACATATATATGTATACATATATGTATATACACATATATACGTATATATA 1782

Qy        299 CATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACA 358
              | |||| |||  |||| |  |||| ||| |  || |||   | | | | || ||   |
Db       1783 TGTGTATATACACGTATATATGTATACATATATGTATATACACATATACACATGTATGTA 1842

Qy        359 TATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATA 418
              ||  |||| |  |||||||  ||||||  | | ||||    ||||||||||  |||||
Db       1843 TACATATATATGTATATACGTGTATATATGTATACATATGTGTATATATACACGTATATA 1902

Qy        419 TACACATATATACACATATATACACATATACACATATATACACATATATACATATATACA 478
              ||||||||| || || | | || | |||      |||||  |||||  ||||||||  || |
Db       1903 TACACATATGTATACGTGTGTATATATACGTGTATATACACACACGTATACATACGTATA 1962

Qy        479 CATATATAT-AATATACACACATATATATACACATATATACACACATATATACACATATA 537
              |||||  |  |||||  |||||  || ||||||  || | || |    ||||||  ||| | |
Db       1963 TATATACGTGTATATATACACACATGTATACATATGTGTATATGTGTATATATACACACA 2022

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATACATATA 597
              | | | ||||||    |||||||||||||| || | ||| ||   |||||| ||| |||
Db       2023 TGTATACATATATGTGTATATATACACATATGTATACATACACGTGTATATACACACATA 2082

Qy        598 TACACATATATACATATATACACATATATACATATATACACACATATATACACATACATA 657
              | | | ||| || | |||| | | | |||  |||| | | |  | |||||| |  ||
Db       2083 TGTATACATACACGTGTATATATACACATATGTATACATATATGTGTATACATACGCACG 2142

Qy        658 TACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATA 717
              || |||||  ||  |||||||||| | || | |||   | |  || | |||||||||| |||
Db       2143 TATACACATGTATGTATACATATTTGTATATATACTCACATATACACATATGTATACATA 2202

Qy        718 TATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGA 777
               |||          |  |         ||  |   | |   |  | | | |   | | | |
Db       2203 CATGTGTATATACACGCACACATACACATATATGTGCATATACGCACATATATACATATA 2262

Qy        778 GTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAA 837
              |   |||    || |||     | |   | ||       ||    ||  ||       | |
Db       2263 TGTACACATATGTATATATACACACATATATACATATATGTACACATGTACGTACACATA 2322

Qy        838 TACATAAATGT 848
              |||||| ||||
Db       2323 TACATATATGT 2333


RESULT 41
US-10-027-632-113621
; Sequence 113621, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 113621
;   LENGTH: 3310
```

```
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-113621

  Query Match             6.5%;  Score 320.6;  DB 17;  Length 3310;
  Best Local Similarity   65.7%;  Pred. No. 3.9e-56;
  Matches  559;  Conservative    0;  Mismatches  284;  Indels    8;  Gaps    6;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACA-AAATT 59
              |||||||| | || |||| ||| |||||||||||||| |||||||  ||||||| |||||
Db       1488 AGTTCAAGTCTAGTCTGACCAATATGGTGAAACCCCGTCTCTACTGGAAATACAGAAATT 1547

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              || ||||  ||||  | | |||||||| ||| |||||||  |||||| ||||||||||||
Db       1548 AGCCAGGTGTGGTAGTGCTCGCCTGTAGTCCTAGCTACTCGGGAGGCTGAGGCAGGAGAA 1607

Qy        120 TCACTTGAACCTGCTGGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCCA 178
              || |||||||||  || ||  |||||||| | ||| | ||||| ||||  |||||||||
Db       1608 TCGCTTGAACCTG--GGAGGCAGAGGTTGCAGTGAGCCGAGATCCTGCCCCTGCACTCCA 1665

Qy        179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATA 238
              |||   |  | ||||||| ||     ||||| ||||||     | | | | || |||||||
Db       1666 GCC--TGGCAACAGAGCGAGATTCTGTCTCAAAAAAAACCCCATGTA-TATACATATATA 1722

Qy        239 TATATATATATATACACACACACACATATATATATACACATATATATACACATATATA 298
              ||||||||| | | | |   | ||||||||| |||| ||| ||||||||  |||||||
Db       1723 CGTATATATACACATATATATGTATACATATATGTATATACACATATATACGTATATATA 1782

Qy        299 CATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACA 358
              | |||| |||  |||| |  |||| ||| |  || |||   | | | | || ||   |
Db       1783 TGTGTATATACACGTATATATGTATACATATATGTATATACACATATACACATGTATGTA 1842

Qy        359 TATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATA 418
              ||  |||| |  |||||||  ||||||  | | ||||   |||||||||  |||||
Db       1843 TACATATATATGTATATACGTGTATATATGTATACATATGTGTATATATACACGTATATA 1902

Qy        419 TACACATATATACACATATATACACATATACACATATATACACATATATACATATATACA 478
              |||||||||| || || | | || | |||     ||||| |||||  ||||||||  || |
Db       1903 TACACATATGTATACGTGTGTATATATACGTGTATATACACACACGTATACATACGTATA 1962

Qy        479 CATATATAT-AATATACACACATATATATACACATATATACACACATATATACACATATA 537
              |||||  |  ||||| ||||| || |||||| || | || |    ||||| ||| | |
Db       1963 TATATACGTGTATATATACACACATGTATACATATGTGTATATGTGTATATATACACACA 2022

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATACATATA 597
              |  | | ||||||    |||||||||||||| || | ||| ||   |||||| ||| |||
Db       2023 TGTATACATATATGTGTATATATACACATATGTATACATACACGTGTATATACACACATA 2082

Qy        598 TACACATATATACATATATACACATATATACATATATACACACATATATACACATACATA 657
              |  | | ||| || | |||| | | | |||  |||| | | |  | |||||| | ||
Db       2083 TGTATACATACACGTGTATATATACACATATGTATACATATATGTGTATACATACGCACG 2142

Qy        658 TACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATA 717
              || |||||  ||  |||||||||| | || | |||   | |  || | |||||||||| |||
Db       2143 TATACACATGTATGTATACATATTTGTATATATACTCACATATACACATATGTATACATA 2202

Qy        718 TATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGA 777
              |||             |  |      ||  |  | |  |  | | | |  | | | |
Db       2203 CATGTGTATATACACGCACACATACACATATATGTGCATATACGCACATATATACATATA 2262

Qy        778 GTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAA 837
              |  |||   || |||    | |   | ||       ||  ||  ||||||  |  || |
Db       2263 TGTACACATATGTATATATACACACATATATACATATATGTACACATGTACGTACACATA 2322

Qy        838 TACATAAATGT 848
              |||||| ||||
Db       2323 TACATATATGT 2333


RESULT 42
US-10-027-632-113622
; Sequence 113622, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 113622
;   LENGTH: 3310
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-113622

  Query Match             6.5%;  Score 320.6;  DB 17;  Length 3310;
  Best Local Similarity   65.7%;  Pred. No. 3.9e-56;
  Matches  559;  Conservative    0;  Mismatches  284;  Indels    8;  Gaps    6;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACA-AAATT 59
              |||||||| | || |||| ||| |||||||||||||| |||||||  ||||||| |||||
Db       1488 AGTTCAAGTCTAGTCTGACCAATATGGTGAAACCCCGTCTCTACTGGAAATACAGAAATT 1547

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              || ||||  ||||  | | |||||||| ||| |||||||  |||||| ||||||||||||
Db       1548 AGCCAGGTGTGGTAGTGCTCGCCTGTAGTCCTAGCTACTCGGGAGGCTGAGGCAGGAGAA 1607

Qy        120 TCACTTGAACCTGCTGGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCCA 178
              || |||||||||  || ||  |||||||| | ||| | ||||| ||||  |||||||||
Db       1608 TCGCTTGAACCTG--GGAGGCAGAGGTTGCAGTGAGCCGAGATCCTGCCCCTGCACTCCA 1665

Qy        179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATA 238
              |||   |  | ||||||| ||     ||||| ||||||     | | | | || |||||||
Db       1666 GCC--TGGCAACAGAGCGAGATTCTGTCTCAAAAAAAACCCCATGTA-TATACATATATA 1722

Qy        239 TATATATATATATACACACACACACATATATATATACACATATATATACACATATATA 298
              ||||||||| | | | |   | ||||||||| |||| ||| ||||||||  |||||||
Db       1723 CGTATATATACACATATATATGTATACATATATGTATATACACATATATACGTATATATA 1782

Qy        299 CATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACA 358
              | |||| |||  |||| |  |||| ||| |  || |||   | | | | || ||   |
Db       1783 TGTGTATATACACGTATATATGTATACATATATGTATATACACATATACACATGTATGTA 1842

Qy        359 TATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATA 418
              ||  |||| |  |||||||  ||||||  | | ||||   |||||||||  |||||
Db       1843 TACATATATATGTATATACGTGTATATATGTATACATATGTGTATATATACACGTATATA 1902

Qy        419 TACACATATATACACATATATACACATATACACATATATACACATATATACATATATACA 478
              |||||||||| || || | | || | |||     ||||| |||||  ||||||||  || |
Db       1903 TACACATATGTATACGTGTGTATATATACGTGTATATACACACACGTATACATACGTATA 1962

Qy        479 CATATATAT-AATATACACACATATATATACACATATATACACACATATATACACATATA 537
              |||||  |  ||||| ||||| || |||||| || | || |    ||||| ||| | |
Db       1963 TATATACGTGTATATATACACACATGTATACATATGTGTATATGTGTATATATACACACA 2022

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATACATATA 597
              |  | | ||| || | |||| | | | |||  |||| | | |  | |||||| | |||
Db       2023 TGTATACATATATGTGTATATATACACATATGTATACATACACGTGTATATACACACATA 2082
```

```
Qy        598 TACACATATATACATATATACACATATATACATATATACACACATATATACACATACATA 657
              |  | | ||| || | |||| | | | |||  |||| | | |  | |||||| |  ||
Db       2083 TGTATACATACACGTGTATATATACACATATGTATACATATATGTGTATACATACGCACG 2142

Qy        658 TACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATA 717
              || |||||  ||  |||||||||| | || | |||   | |  || | |||||||||| |||
Db       2143 TATACACATGTATGTATACATATTTGTATATATACTCACATATACACATATGTATACATA 2202

Qy        718 TATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGA 777
              |||           |  |          ||  |   | |  |  | | | |   | | | |
Db       2203 CATGTGTATATACACGCACACATACACATATATGTGCATATACGCACATATATACATATA 2262

Qy        778 GTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAA 837
                |  |||   || |||    | |   | ||       ||   ||  ||        | |
Db       2263 TGTACACATATGTATATATACACACATATATACATATATGTACACATGTACGTACACATA 2322

Qy        838 TACATAAATGT 848
              |||||| ||||
Db       2323 TACATATATGT 2333


RESULT 43
US-10-221-714A-473/c
; Sequence 473, Application US/10221714A
; Publication No. US20040048254A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with
;  TITLE OF INVENTION:  tumor suppressor genes and oncogenes
;  FILE REFERENCE: 5013.1005
;  CURRENT APPLICATION NUMBER: US/10/221,714A
;  CURRENT FILING DATE:  2003-01-21
;  PRIOR APPLICATION NUMBER: PCT/EP01/02955
;  PRIOR FILING DATE: 2001-03-15
;  PRIOR APPLICATION NUMBER: DE 10013847.0
;  PRIOR FILING DATE: 2000-03-15
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 540
; SEQ ID NO 473
;   LENGTH: 26997
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-714A-473

  Query Match             6.3%;  Score 311.2;  DB 17;  Length 26997;
  Best Local Similarity   76.9%;  Pred. No. 9.1e-54;
  Matches  409;  Conservative    0;  Mismatches  113;  Indels   10;  Gaps    2;

Qy        203 CATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACA 262
              | | ||| ||| | | ||        | ||||||||||||||||| | | || || | | |
Db      21339 CTTTTCATAAATACATAATATTCCATAATATATATATATATATACACACACGCATATATA 21280

Qy        263 CACATATATATATACACATATATATACACATATATACATATATACACATATATACACATA 322
              ||||||||   | ||| ||| ||||| | |||||||||||| ||| |||||||  |  ||
Db      21279 CACATATACCCACACATATACATATATATATATACATATACACATATATATATCTA 21220

Qy        323 TATATACACATACATATGT-----ACACATATATATACACATATGTATACACATATATAC 377
              |||||||||  | |||  |     | | ||| | ||| |||||| |||| ||| ||||||
Db      21219 CATATACACACATATACATATATAATATATACACATATACATATATATATACACATATAC 21160

Qy        378 ACATATATACACACATATATACACATATATACACACATATATACACATATATACACATAT 437
              |||||||  | || || ||||| |||||||| | ||||||||||||| |||| ||| |||
Db      21159 ACATATACATACGCACATATAAACATATATATATACATATATACACGTATACACATATAC 21100

Qy        438 ATACACATATACACATATATACACATATATACATATATACACATATATATA-----ATAT 492
              ||| ||| ||| | |||||||||| ||||| ||||||| | | ||| |||||      |||
Db      21099 ATATACACATATATATATATACATATATACACATATACATATATACATATATATATAC 21040

Qy        493 ACACACATATATATACACATATATACACACATATATACACATATATACACATATATATAC 552
              | | | | | ||||||| ||||| | | | ||||||| ||||||||||||||||| |||
Db      21039 ATATATACACATATACATATATACATATATATATATATACATATATACACATATACATAT 20980

Qy        553 ACATATATACACATATATACATATATACACATATATATACATATATACACATATATACAT 612
              | | ||||| | |||||||||||||||||||||||| ||| ||| ||| | |||||||| |
Db      20979 ATACATATATATATATATACATATATACACATATACATATATACATATATATATATAC 20920

Qy        613 ATATACACATATATACATATATACACACATATATACACATACATATACACACACATAGAT 672
              ||||| ||| |||||||||||||| | ||||||| | ||| ||||| | | | ||| ||
Db      20919 ATATATACACATATACATATATACATATATATATATATATATATATATATATACAT 20860

Qy        673 ATACATATATATACACATATATATACGTATATATATGTATATATATATGCTC 724
              ||| |||||||||| |||||| ||  ||||||||| |||||||||| | |
Db      20859 ATATATATATATACATATATATCTATATATATATATATATATATATATACAC 20808


RESULT 44
US-10-017-161-2383
; Sequence 2383, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
;  FILE REFERENCE: 084335/0152
;  CURRENT APPLICATION NUMBER: US/10/017,161
;  CURRENT FILING DATE:  2002-12-18
;  PRIOR APPLICATION NUMBER: JP 2001/246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2430
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2383
;   LENGTH: 1990
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: source
;   LOCATION: (1)..(1990)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(1790)
US-10-017-161-2383

  Query Match             6.2%;  Score 308;  DB 15;  Length 1990;
  Best Local Similarity   80.2%;  Pred. No. 1.3e-53;
  Matches  401;  Conservative    0;  Mismatches   90;  Indels    9;  Gaps    3;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | || | | |||| |||||||| | | | | ||||||| || ||||||||||||||||||
Db        826 TATACACACATATGTATATATACACATATATACACACATATGTATATATACACATATATA 885

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              |||||||| || ||||| ||| ||||| |||||||| |||| | | |||||| ||||||
Db        886 CACACATATGTATATATACACATATATACACACATATGTATATATACACATATATACACA 945

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||| || | |||   ||| | | |  || |  |||||| |||||||||||||||||| |
Db        946 CATATGTATATATACACATATATACACGTATATGTATATATACACATATATACACATACA 1005

Qy        407 TACACA----CATATATACACATATATACACATATATACACATATACACATATATACACA 462
              ||||||    ||||| |||||||||||||||| ||||||||||||  ||||||||||
Db       1006 TACACATATGTATATACACACATATATACACATACATACACATATACGTATATATACACA 1065

Qy        463 TATATACATATATACACATATATATAATATACACACATATATATACACATATATACACAC 522
```

```
Db       1066 TATACACATACATACACATATACGT-ATATATACACATATATACACATATGCGTATATAC 1124

Qy        523 ATATATACACATATATACACATATATATACACATATATACACATATATACATATATACAC 582
              |||||||||||||||||||    ||||||  ||||||||||||||||||| ||||||
Db       1125 ATATATACACATATATACAC----ATATACGTATATATACACATATATACACATATACGT 1180

Qy        583 ATATATATACATATATACACATATATACATATATACACATATATACATATATACACACAT 642
              ||||||| |||||||||||||||||||| |||||   ||||| |||||||||||||| ||
Db       1181 ATATATACACATATATACACATATATACACATATATGTATATACACATATATACACATAT 1240

Qy        643 ATATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTAT 702
              |||| | | ||||||| ||||| |||  |||| ||| ||||| |||  ||| ||  |||
Db       1241 GTATATATACACATATATACACATATATGTATATATACATATATACACGTATGTATATAT 1300

Qy        703 ATATATGTATATATATATGC 722
              | | || ||| ||||||| |
Db       1301 ACACATATATGTATATATAC 1320


RESULT 45
US-10-292-798-2025
; Sequence 2025, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
;  FILE REFERENCE: 084335/166
;  CURRENT APPLICATION NUMBER: US/10/292,798
;  CURRENT FILING DATE:  2002-11-13
;  PRIOR APPLICATION NUMBER: 10/017,161
;  PRIOR FILING DATE: 2001-12-18
;  PRIOR APPLICATION NUMBER: JP 2001-246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2070
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2025
;   LENGTH: 1990
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   LOCATION: source
;   FEATURE:
;   LOCATION: (1)..(1990)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(1790)
US-10-292-798-2025

  Query Match             6.2%;  Score 308;  DB 17;  Length 1990;
  Best Local Similarity   80.2%;  Pred. No. 1.3e-53;
  Matches  401;  Conservative    0;  Mismatches   90;  Indels    9;  Gaps    3;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | || | | |||| |||||||| | | | | ||||||| || ||||||||||||||||||
Db        826 TATACACACATATGTATATATACACATATATACACACATATGTATATATACACATATATA 885

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              |||||||| || ||||| ||| ||||| |||||||| |||| | | |||||| ||||||
Db        886 CACACATATGTATATATACACATATATACACACATATGTATATATACACATATATACACA 945

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||| || | |||   ||| | | |  || |  |||||| |||||||||||||||||| |
Db        946 CATATGTATATATACACATATATACACGTATATGTATATATACACATATATACACATACA 1005

Qy        407 TACACA----CATATATACACATATATACACATATATACACATATACACATATATACACA 462
              ||||||     ||||| |||||||||||||||| ||||||||||||  ||||||||||
Db       1006 TACACATATGTATATACACACATATATACACATACATACACATATACGTATATATACACA 1065

Qy        463 TATATACATATATACACATATATATAATATACACACATATATATACACATATATACACAC 522
```

```
Db       1066 TATACACATACATACACATATACGT-ATATATACACATATATACACATATGCGTATATAC 1124

Qy        523 ATATATACACATATATACACATATATATACACATATATACACATATATACATATATACAC 582
              |||||||||||||||||||    ||||||  ||||||||||||||||||| ||||||
Db       1125 ATATATACACATATATACAC----ATATACGTATATATACACATATATACACATATACGT 1180

Qy        583 ATATATATACATATATACACATATATACATATATACACATATATACATATATACACACAT 642
              ||||||| |||||||||||||||||||| |||||   ||||| |||||||||||||| ||
Db       1181 ATATATACACATATATACACATATATACACATATATGTATATACACATATATACACATAT 1240

Qy        643 ATATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTAT 702
              |||| | | ||||||| ||||| |||  |||| ||| ||||| |||  ||| ||  |||
Db       1241 GTATATATACACATATATACACATATATGTATATATACATATATACACGTATGTATATAT 1300

Qy        703 ATATATGTATATATATATGC 722
              | | || ||| ||||||| |
Db       1301 ACACATATATGTATATATAC 1320


RESULT 46
US-10-437-963-12666
; Sequence 12666, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 12666
;   LENGTH: 1800
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_18773C.1
US-10-437-963-12666

  Query Match             6.0%;  Score 300.6;  DB 18;  Length 1800;
  Best Local Similarity   78.2%;  Pred. No. 4.3e-52;
  Matches  387;  Conservative    0;  Mismatches  104;  Indels    4;  Gaps    2;

Qy        229 TATATATATATATATATATATATACACACACACACACATATATATATACACATATATATA 288
              |||||||||||| |||||||||| | | ||| | ||| ||||||||| ||||| |||
Db       1298 TATATATATATACATATATATATATATATATACATATATACATATATACATATATACATA 1357

Qy        289 CACATATATACATATATACACATATATAC---ACATATATATACACATACATATGTACAC 345
              ||||||| | ||||| ||||| ||||||    ||||||| ||| | |||||||| ||
Db       1358 TACATATACATATATACACACACATATACATAACATATACATATATATACATATATATGT 1417

Qy        346 ATATATATACACATATGTATACACATATATACACATATATACACACATATATACACATAT 405
              ||||||| | | |||| || | | ||| ||| | || |||| | ||||||||| ||||||
Db       1418 ATATATACATATATATATACATATATACATATATATGTATATATACATATATATACATAT 1477

Qy        406 ATACACACATATATACACATATATACACATATATACACATATACACATATATACACATAT 465
              ||| ||  |||||||| |||||||| |||||||| ||||| |||||  || | | |
Db       1478 ATATACGTATATATACATATATATACGTATATATACATATATATACATACGTATATACGT 1537

Qy        466 ATACATATATACACATATATATAATATACACACATATATATACACATATATACACACATA 525
              |||||||  || | ||| |||| ||||||| || |||||||||| ||||||||| || ||
Db       1538 ATACATACGTATATATACATAT-ATATACATACGTATATATACATATATATACATACGTA 1596

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| ||||||||| |||||||||| | ||||| | | ||||| | ||||| |
```

```
Db       1597 TATATACATATATATACATATATATATATATATACATATATATACATATATACATTTTCG 1656

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| |||||||   ||||||| |||| || | | |||||
Db       1657 TATATATATATATATATATATATATATATATATGTATATATATATATGTATATATATATA 1716

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | |   ||| ||||| |||| |||| | |||||| ||  ||||||
Db       1717 TATATATATATATATATATGTATATATATATATATGTATATATATATATGTATATATATA 1776

Qy        706 TATGTATATATATAT 720
              ||||||||||||||
Db       1777 TATGTATATATATAT 1791


RESULT 47
US-09-783-203-1
; Sequence 1, Application US/09783203
; Patent No. US20020098582A1
; GENERAL INFORMATION:
;  APPLICANT: Geron Corporation
;  APPLICANT:  Gold, Joseph
;  APPLICANT:  Lebkowski, Jane
;  TITLE OF INVENTION: Tpacked stem cells
;  FILE REFERENCE: 096/003
;  CURRENT APPLICATION NUMBER: US/09/783,203
;  CURRENT FILING DATE:  2001-02-13
;  PRIOR APPLICATION NUMBER: 60/253,443
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/253,357
;  PRIOR FILING DATE: 2000-11-27
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
;   LENGTH: 15418
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-783-203-1

  Query Match             6.0%;  Score 296;  DB 9;  Length 15418;
  Best Local Similarity   69.2%;  Pred. No. 1e-50;
  Matches  479;  Conservative    0;  Mismatches  200;  Indels   13;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||| |||||||||||  ||| || | ||||||||||||| ||||||||| || ||||||
Db        646 AGTTCGAGACCAGCCTGGCCAATATAGCGAAACCCCATCTGTACTAAAAACACAAAAATT 705

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   ||  |||||| ||| ||||||||||||| ||||  ||||||  ||||||||||
Db        706 AGCTGGGTGTGGTGACACATGCCTGTAATCCCAGGTACTCAGGAGGCTAAGGCAGGAGAA 765

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              |  |||||||  |  ||  | ||  |||      |   ||||||    |||  ||||||||||
Db        766 TTGCTTGAACTGGGAGGCAGAGGTGGT-----GAGCCAAGATTGCACCAGTGCACTCCAG 820

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  ||  || |||| ||||   ||||||| ||| ||| |||   | |  ||||| |||
Db        821 CCTTGGTGAC-AGAGTGAAACTCCATCTCAAAAACAAACAAACAAAATACATATACATAA 879

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              ||||||||  | | | | | ||| | ||| ||||||| ||| ||||||| |  |||||||
Db        880 ATATATATGCACATATATATACATATATAAATATATATACACATATATAAATCTATATAC 939

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACAT-----ATATATA 354
              ||||||||| ||||| ||| ||||| ||  | |||||||| || ||||     ||||||
Db        940 ATATATACATATATACACATATATAAATCTATATACATATATATACATATATAATATATT 999

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              ||||||  || | | ||| ||| | | |||||||  | |||  ||||| ||||| | | |||
Db       1000 TACATATATAAATATATACATATATAAATATACATATATAAATACATATATAAATATACA 1059

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              |||||| | ||| ||| | | ||||||| ||||| | ||||| | | ||| ||| | |||
Db       1060 TATATAAATATACATATATAAATATACATATATAAATATATACATATATAAATATACATA 1119

Qy        475 TACACATATATATA-ATATACACACATATATATACACATATATACACACATATATACACA 533
              || | ||||||||| ||||| | | ||||| ||| |||   |||| ||| |||||||||| |
Db       1120 TATAAATATATATACATATATAAATATATAAATATACAAGTATATACAAATATATACATA 1179

Qy        534 TATATACACATATATATACACATATATACACATATATACATATATACACATATATATACA 593
              |||| |   |||||  || | | ||||| | ||| ||| ||| | | || | |   |
Db       1180 TATAAATGTATATACGTATATACATATATATATAAATATATAAAAAAACTTTTGGCTGGG 1239

Qy        594 TATATACACATATATACATATATACACATATATACATATATACACACATATATACACATA 653
              |  |   || || |      | | | |  | |     ||  | || ||| | | |||||
Db       1240 CACCTTTCCAAATCTCATGGCACATATAAGTCTCATGGTAACCTCAAATAAAAAAACATA 1299

Qy        654 CATATACACACACATAGATATACATATATATA 685
              |           || | ||| | |   ||| |
Db       1300 TAACAGATACACCAAAAATAAAAACCAATAAA 1331


RESULT 48
US-09-994-427A-1
; Sequence 1, Application US/09994427A
; Patent No. US20020128221A1
; GENERAL INFORMATION:
;  APPLICANT: Geron Corporation
;  APPLICANT:  Schiff, J. Michael
;  TITLE OF INVENTION: GLYCOSYLTRANSFERASE VECTORS FOR TREATING CANCER
;  FILE REFERENCE: 083,002
;  CURRENT APPLICATION NUMBER: US/09/994,427A
;  CURRENT FILING DATE:  2002-02-26
;  PRIOR APPLICATION NUMBER: 60/253,395
;  PRIOR FILING DATE: 2000-11-27
;  NUMBER OF SEQ ID NOS: 17
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 15418
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-994-427A-1

  Query Match             6.0%;  Score 296;  DB 9;  Length 15418;
  Best Local Similarity   69.2%;  Pred. No. 1e-50;
  Matches  479;  Conservative    0;  Mismatches  200;  Indels   13;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||| |||||||||||  ||| || | ||||||||||||| ||||||||| || ||||||
Db        646 AGTTCGAGACCAGCCTGGCCAATATAGCGAAACCCCATCTGTACTAAAAACACAAAAATT 705

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   ||  |||||| ||| ||||||||||||| ||||  ||||||  ||||||||||
Db        706 AGCTGGGTGTGGTGACACATGCCTGTAATCCCAGGTACTCAGGAGGCTAAGGCAGGAGAA 765

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              |  |||||||  |  ||  | ||  |||      |   ||||||    |||  ||||||||||
Db        766 TTGCTTGAACTGGGAGGCAGAGGTGGT-----GAGCCAAGATTGCACCAGTGCACTCCAG 820

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  ||  || |||| ||||   ||||||| ||| ||| |||   | |  ||||| |||
Db        821 CCTTGGTGAC-AGAGTGAAACTCCATCTCAAAAACAAACAAACAAAATACATATACATAA 879

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              ||||||||  | | | | | ||| | ||| ||||||| ||| ||||||| |  |||||||
Db        880 ATATATATGCACATATATATACATATATAAATATATATACACATATATAAATCTATATAC 939

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACAT-----ATATATA 354
              ||||||||| ||||| ||| ||||| ||  | |||||||| || ||||     ||||||
Db        940 ATATATACATATATACACATATATAAATCTATATACATATATATACATATATAATATATT 999

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              ||||||  || | | ||| ||| | | |||||||  | |||  ||||| ||||| | | |||
Db       1000 TACATATATAAATATATACATATATAAATATACATATATAAATACATATATAAATATACA 1059

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
```

```
Db       1060 TATATAAATATACATATATAAATATACATATATAAATATATACATATATAAATATACATA 1119

Qy        475 TACACATATATATA-ATATACACACATATATATACACATATATACACACATATATACACA 533
Db       1120 TATAAATATATATACATATATAAATATATAAATATACAAGTATATACAAATATATACATA 1179

Qy        534 TATATACACATATATATACACATATATACACATATATACATATATACACATATATATACA 593
Db       1180 TATAAATGTATATACGTATATACATATATATATAAATATATAAAAAAACTTTTGGCTGGG 1239

Qy        594 TATATACACATATATACATATATACACATATATACATATATACACACATATATACACATA 653
Db       1240 CACCTTTCCAAATCTCATGGCACATATAAGTCTCATGGTAACCTCAAATAAAAAAACATA 1299

Qy        654 CATATACACACACATAGATATACATATATATA 685
Db       1300 TAACAGATACACCAAAAATAAAAACCAATAAA 1331
```

RESULT 49
US-09-995-419A-1
; Sequence 1, Application US/09995419A
; Publication No. US20030032187A1
; GENERAL INFORMATION:
; APPLICANT: Geron Corporation
; APPLICANT: McWhir, Jim
; APPLICANT: Gold, Joseph D.
; APPLICANT: Schiff, J. Michael
; TITLE OF INVENTION: 096,004 - SeqList
; FILE REFERENCE: 096,004 - SeqList
; CURRENT APPLICATION NUMBER: US/09/995,419A
; CURRENT FILING DATE: 2001-11-26
; PRIOR APPLICATION NUMBER: 60/253,357
; PRIOR FILING DATE: 2000-11-27
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 15418
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-995-419A-1

Query Match 6.0%; Score 296; DB 10; Length 15418;
Best Local Similarity 69.2%; Pred. No. 1e-50;
Matches 479; Conservative 0; Mismatches 200; Indels 13; Gaps 5;

```
Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
Db        646 AGTTCGAGACCAGCCTGGCCAATATAGCGAAACCCCATCTGTACTAAAAACACAAAAATT 705

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
Db        706 AGCTGGGTGTGGTGACACATGCCTGTAATCCCAGGTACTCAGGAGGCTAAGGCAGGAGAA 765

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
Db        766 TTGCTTGAACTGGGAGGCAGAGGTGGT-----GAGCCAAGATTGCACCAGTGCACTCCAG 820

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
Db        821 CCTTGGTGAC-AGAGTGAAACTCCATCTCAAAAACAAACAAACAAAATACATATACATAA 879

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
Db        880 ATATATATGCACATATATATACATATATAAATATATATACACATATATAAATCTATATAC 939

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACAT-----ATATATA 354
Db        940 ATATATACATATATACACATATATAAATCTATATACATATATATACATATATAATATATT 999

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
```

```
Db       1000 TACATATATAAATATATACATATATAAATATACATATATAAATACATATATAAATATACA 1059

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
Db       1060 TATATAAATATACATATATAAATATACATATATAAATATATACATATATAAATATACATA 1119

Qy        475 TACACATATATATA-ATATACACACATATATATACACATATATACACACATATATACACA 533
Db       1120 TATAAATATATATACATATATAAATATATAAATATACAAGTATATACAAATATATACATA 1179

Qy        534 TATATACACATATATATACACATATATACACATATATACATATATACACATATATATACA 593
Db       1180 TATAAATGTATATACGTATATACATATATATATAAATATATAAAAAAACTTTTGGCTGGG 1239

Qy        594 TATATACACATATATACATATATACACATATATACATATATACACACATATATACACATA 653
Db       1240 CACCTTTCCAAATCTCATGGCACATATAAGTCTCATGGTAACCTCAAATAAAAAAACATA 1299

Qy        654 CATATACACACACATAGATATACATATATATA 685
Db       1300 TAACAGATACACCAAAAATAAAAACCAATAAA 1331
```

RESULT 50
US-10-141-220-1
; Sequence 1, Application US/10141220
; Publication No. US20030040111A1
; GENERAL INFORMATION:
; APPLICANT: Geron Corporation
; APPLICANT: Gold, Joseph
; APPLICANT: Lebkowski, Jane
; TITLE OF INVENTION: Tpacked stem cells
; FILE REFERENCE: 096/003
; CURRENT APPLICATION NUMBER: US/10/141,220
; CURRENT FILING DATE: 2002-05-07
; PRIOR APPLICATION NUMBER: US/09/783,203
; PRIOR FILING DATE: 2001-02-13
; PRIOR APPLICATION NUMBER: 60/253,443
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/253;357
; PRIOR FILING DATE: 2000-11-27
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 15418
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-141-220-1

Query Match 6.0%; Score 296; DB 14; Length 15418;
Best Local Similarity 69.2%; Pred. No. 1e-50;
Matches 479; Conservative 0; Mismatches 200; Indels 13; Gaps 5;

```
Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
Db        646 AGTTCGAGACCAGCCTGGCCAATATAGCGAAACCCCATCTGTACTAAAAACACAAAAATT 705

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
Db        706 AGCTGGGTGTGGTGACACATGCCTGTAATCCCAGGTACTCAGGAGGCTAAGGCAGGAGAA 765

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
Db        766 TTGCTTGAACTGGGAGGCAGAGGTGGT-----GAGCCAAGATTGCACCAGTGCACTCCAG 820

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
Db        821 CCTTGGTGAC-AGAGTGAAACTCCATCTCAAAAACAAACAAACAAAATACATATACATAA 879

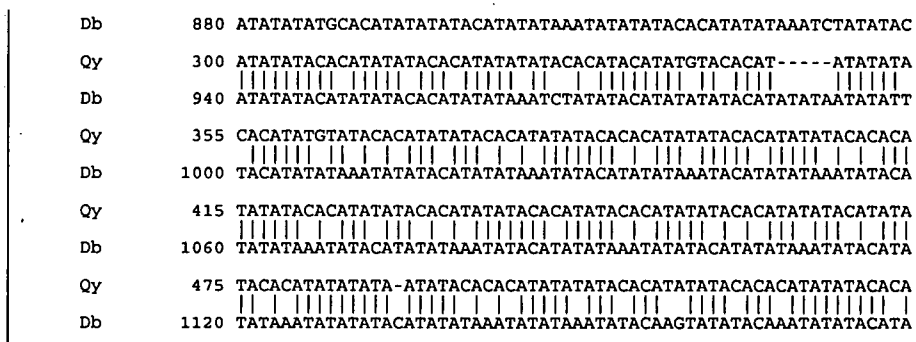
Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
Db        880 ATATATATGCACATATATATACATATATAAATATATATACACATATATAAATCTATATAC 939
```



RESULT 51
US-10-023-969-1
; Sequence 1, Application US/10023969
; Publication No. US20030095989A1
; GENERAL INFORMATION:
; APPLICANT: Geron Corporation
; APPLICANT: Irving, John
; APPLICANT: Lebkowski, Jane
; TITLE OF INVENTION: Chimeric Cytolytic Viruses for Cancer Treatment
; FILE REFERENCE: 084,002
; CURRENT APPLICATION NUMBER: US/10/023,969
; CURRENT FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: 60/256,418
; PRIOR FILING DATE: 2000-12-18
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 15418
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-023-969-1

Query Match 6.0%; Score 296; DB 14; Length 15418;
Best Local Similarity 69.2%; Pred. No. 1e-50;
Matches 479; Conservative 0; Mismatches 200; Indels 13; Gaps 5;

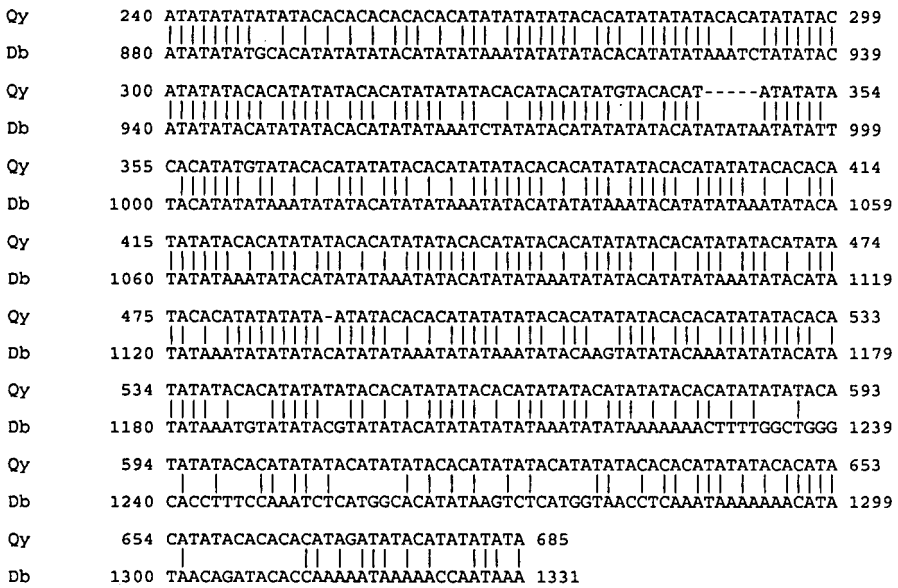




RESULT 52
US-10-206-447-1
; Sequence 1, Application US/10206447
; Publication No. US20030099616A1
; GENERAL INFORMATION:
; APPLICANT: Geron Corporation
; APPLICANT: Irving, John
; APPLICANT: Karpf, David
; APPLICANT: Schiff, Michael
; TITLE OF INVENTION: DUAL SPECIFICITY TUMOR KILLING VECTORS DRIVEN BY THE TELOMERASE P
; FILE REFERENCE: 085/002
; CURRENT APPLICATION NUMBER: US/10/206,447
; CURRENT FILING DATE: 2002-07-25
; PRIOR APPLICATION NUMBER: 60/308,029
; PRIOR FILING DATE: 2001-07-25
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 15418
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-206-447-1

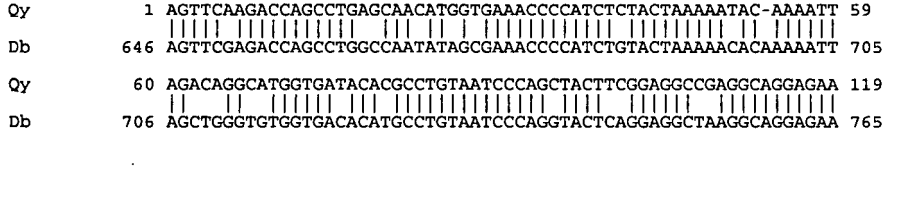
Query Match 6.0%; Score 296; DB 14; Length 15418;
Best Local Similarity 69.2%; Pred. No. 1e-50;
Matches 479; Conservative 0; Mismatches 200; Indels 13; Gaps 5;

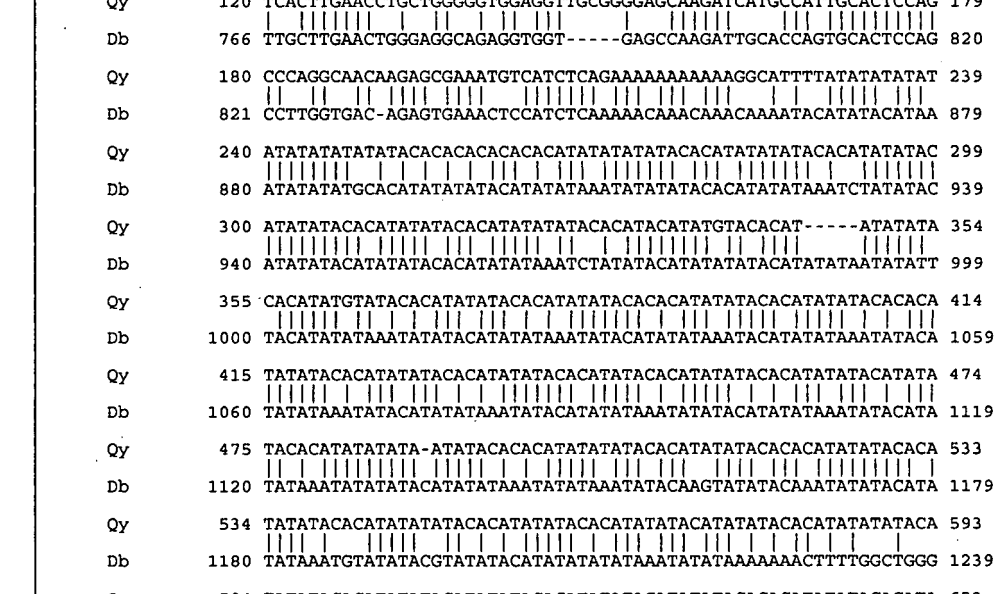




RESULT 53
US-10-674-836-1
; Sequence 1, Application US/10674836
; Publication No. US20040072787A1
; GENERAL INFORMATION:
; APPLICANT: Morin, Gregg B.
; APPLICANT: Lichtsteiner, Serge
; APPLICANT: Vasserot, Alain
; APPLICANT: Adams, Robert R.
; APPLICANT: Geron Corporation
; TITLE OF INVENTION: Telomerase Reverse Transcriptase Transcriptional
; TITLE OF INVENTION: Regulatory Sequences and Methods of Using
; FILE REFERENCE: 019/246P
; CURRENT APPLICATION NUMBER: US/10/674,836
; CURRENT FILING DATE: 2003-09-29
; PRIOR APPLICATION NUMBER: US/09/244,438
; PRIOR FILING DATE: 1999-02-04
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 15418
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Human TERT promoter
US-10-674-836-1

Query Match 6.0%; Score 296; DB 17; Length 15418;
Best Local Similarity 69.2%; Pred. No. 1e-50;
Matches 479; Conservative 0; Mismatches 200; Indels 13; Gaps 5;





RESULT 54
US-10-811-012-1
; Sequence 1, Application US/10811012
; Publication No. US20040152189A1
; GENERAL INFORMATION:
; APPLICANT: Geron Corporation
; APPLICANT: McWhir, Jim
; APPLICANT: Gold, Joseph D.
; APPLICANT: Schiff, J. Michael
; TITLE OF INVENTION: Selective Antibody Targeting of Undifferentiated Stem Cells
; FILE REFERENCE: 096,006D- SeqList
; CURRENT APPLICATION NUMBER: US/10/811,012
; CURRENT FILING DATE: 2004-03-26
; PRIOR APPLICATION NUMBER: 09/995,419
; PRIOR FILING DATE: 2001-11-26
; PRIOR APPLICATION NUMBER: 60/253,357
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/253,443
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/253,395
; PRIOR FILING DATE: 2000-11-27
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 15418
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-811-012-1

Query Match 6.0%; Score 296; DB 18; Length 15418;
Best Local Similarity 69.2%; Pred. No. 1e-50;
Matches 479; Conservative 0; Mismatches 200; Indels 13; Gaps 5;

```
Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||| |||||||||||  ||| || | ||||||||||| |||||||| || ||||||
Db        646 AGTTCGAGACCAGCCTGGCCAATATAGCGAAACCCCATCTGTACTAAAAACACAAAAATT 705

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   ||  |||||| ||| |||||||||||||| |||| ||||||  ||||||||||
Db        706 AGCTGGGTGTGGTGACACATGCCTGTAATCCCAGGTACTCAGGAGGCTAAGGCAGGAGAA 765

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              | |||||||  |  ||  | || |||     |   ||||||    ||| ||||||||||
Db        766 TTGCTTGAACTGGGAGGCAGAGGTGGT-----GAGCCAAGATTGCACCAGTGCACTCCAG 820

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  ||  || |||| ||||   ||||||| ||| ||| |||   | |  ||||| |||
Db        821 CCTTGGTGAC-AGAGTGAAACTCCATCTCAAAAACAAACAAACAAAATACATATACATAA 879

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              |||||||| | | | | | ||| | ||| |||||||| ||| |||||||| |   |||||||
Db        880 ATATATATGCACATATATATACATATATAAATATATATACACATATATAAATCTATATAC 939

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACAT-----ATATATA 354
              ||||||||| ||||| ||| ||||| ||  | |||||||| || ||||      ||||||
Db        940 ATATATACATATATACACATATATAAATCTATATACATATATATACATATATAATATATT 999

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              ||||||  || | | ||| ||| | | ||||||| | ||| ||||| ||||| | | |||
Db       1000 TACATATATAAATATATACATATATAAATATACATATATAAATACATATATAAATATACA 1059

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              ||||||  | ||| ||| | | ||||||| ||||| | ||||| | | ||| ||| | |||
Db       1060 TATATAAATATATACATATATAAATATACATATATAAATATATACATATATAAATATACATA 1119

Qy        475 TACACATATATATA-ATATACACACATATATATACACATATATACACACATATATACACA 533
              || | |||||||||  ||||| | | ||||| ||| |||  |||| ||| |||||||| |
Db       1120 TATAAATATATATACATATATAAATATATAAATATACAAGTATATACAAATATATACATA 1179

Qy        534 TATATACACATATATATACACATATATACACATATATACATATATACACATATATATACA 593
              ||||  |   |||||  || | | ||||| | ||| ||| ||| | | || | |   |
Db       1180 TATAAATGTATATACGTATATACATATATATATAAATATATAAAAAAACTTTGGCTGGG 1239

Qy        594 TATATACACATATATACATATATACACATATATACATATATACACACATATATACACATA 653
              | |   || || |       | | | |  |       || | || ||| | | |||||
Db       1240 CACCTTTCCAAATCTCATGGCACATATAAGTCTCATGGTAACCTCAAATAAAAAAACATA 1299

Qy        654 CATATACACACACATAGATATACATATATATA 685
              |       || | ||| | |   ||| |
Db       1300 TAACAGATACACCAAAAATAAAAACCAATAAA 1331


RESULT 55
US-10-292-798-2005/c
; Sequence 2005, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
;  FILE REFERENCE: 084335/166
;  CURRENT APPLICATION NUMBER: US/10/292,798
;  CURRENT FILING DATE:  2002-11-13
;  PRIOR APPLICATION NUMBER: 10/017,161
;  PRIOR FILING DATE: 2001-12-18
;  PRIOR APPLICATION NUMBER: JP 2001-246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2070
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2005
;   LENGTH: 1486
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   LOCATION: source
;   FEATURE:
;   LOCATION: (1)..(1486)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(1286)
US-10-292-798-2005

  Query Match             5.9%;  Score 294.8;  DB 17;  Length 1486;
  Best Local Similarity   74.9%;  Pred. No. 6.3e-51;
  Matches  397;  Conservative    0;  Mismatches  127;  Indels    6;  Gaps    2;

Qy        197 AAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACA 256
              | || |   | || | || | | |      || || | | ||  | ||||||| |||||||
Db        999 ATATATACACACATATAATATATACATATTTATACACACATGCACATATATAAATACACA 940

Qy        257 CACACACACATATATATATACACATATATATACACATATATACATATATACACATATATA 316
              | | ||| ||||| | | | || |||||| |||||||||||||||||||||||||| ||
Db        939 TATATACATATATACACACATATATGTATATATACATATATACATATATACACATATGTA 880

Qy        317 CACATATATATACACATACATATGTACACATATATATACACATATGTATACACATATATA 376
              | ||| ||||| ||||| |||   | |  |||||||||||||||||||| | | | |||
Db        879 TATATACATATATACATATATACACATATGTATATATACACATATGTATATATACACATA 820

Qy        377 CACATATATACACACATATATACACATATATACACACATATATACACATATATACACATA 436
              |||||||  || | | | |||||||||| || | | | | |||||||||| || | |||
Db        819 CACATATGTATATATACACATACACATATGTATATATACACATACACATATGTATATATA 760

Qy        437 TATACACATATACACATATATACACATATATACATATATACACATATATA-----ATA 491
              | | ||| |||    |||||||||||| ||  |||||||||||||||| ||     | |
Db        759 CACATACACATATGTATATATACACATACATGTATATATACACATATATGTATATATACA 700

Qy        492 TACACACATATATATACACATATATACACACATATATACACATATATACACATATATATA 551
              | |||||| |||| ||| | | ||| |   |||||||||||||| | | | ||| | |||
Db        699 CATACACATGTATACACATACACATATATGTATATATACACATACACATATATACACATA 640

Qy        552 CACATATATACACATATATACATATATACACATATATATACATATATACACATATATACA 611
              ||| |||||||||||||||| |||||||||||||||   |||||||||||||| |||
Db        639 TACACATATACACATATATACACATATACACATATATAAGTATATATACACATATACACA 580

Qy        612 TATATACACATATATACATATATACACACATATATACACATACATAT-ACACACACATAG 670
              ||||    | ||||| |||||||||||||| |||  || | |||| || | ||||| |||
Db        579 TATACGTATATATACACATATATACACATATACGTATATATACACATAATACACATATAC 520

Qy        671 ATATACATATATATACACATATATATACGTATATATATGTATATATATAT 720
              |||| ||| | ||| | | | |||||||||||||||| ||||||| | ||
Db        519 GTATAAATACACATATATACACATATACGTATATATACGTATATACACAT 470


RESULT 56
US-09-841-158-5
; Sequence 5, Application US/09841158
; Publication No. US20030166046A1
; GENERAL INFORMATION:
;  APPLICANT: BEASLEY, Ellen M.
;  TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
;  TITLE OF INVENTION:  NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND
;  TITLE OF INVENTION:  USES THEREOF
;  FILE REFERENCE: CL001229
;  CURRENT APPLICATION NUMBER: US/09/841,158
;  CURRENT FILING DATE:  2001-04-25
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
;   LENGTH: 27067
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(27067)
;   OTHER INFORMATION: n = A,T,C or G
US-09-841-158-5
```

```
  Query Match             5.9%;  Score 293.6;  DB 10;  Length 27067;
  Best Local Similarity   75.8%;  Pred. No. 4.2e-50;
  Matches  391;  Conservative    0;  Mismatches  119;  Indels    6;  Gaps    2;

Qy        230 ATATATATATATATATATATATACACACACACACACATATAT-ATATACACATATATATA 288
              || ||||  |  | |  |  ||| |     | ||| |||||| | ||| | ||||| | |
Db      26509 ATGTATAACTTAAAAATTTCATATATGTGTATACATATATATGAAATATATATATACAGA 26568

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              |||| ||||| || |||||| ||||||||||||||||||  | | |||| |||||||
Db      26569 CACACATATATATGTATACATATATATACACATATATATGTATACATATATACACACATA 26628

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              ||| || |||||| |||||||| ||||||||||||||||||||| ||| | | ||| | |
Db      26629 TATGTATACATATATATACACACATATACACATATATACACACACATACATATATACACA 26688

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              |||| ||||| |||||||||||||| ||  |||||||| |  |||| | | ||| | |
Db      26689 CACATATATACACACATATATACACACATGCACACATATATATGTATACATATATACACA 26748

Qy        469 CATATATACACATATATATA-----ATATACACACATATATATACACATATATACACACA 523
              ||| |||||  ||||||| |     ||||||||||||||||||||||| ||||||||||
Db      26749 CATGTATACGTATATATACACACATATATACACACATATATATACACACATATACACACA 26808

Qy        524 TATATACACATATATACACATATATATACACATATATACACATATATACATATATACACA 583
              || | |||||||||||||| |||||||||| ||||| ||| | ||| |||| || |||
Db      26809 TACACACACATATATACACACATATATACACACATATATACACACATATATATGTATACA 26868

Qy        584 TATATATACATATATACACATATATACATATATACACATATATACATATATACACACATA 643
              |||||||||| | ||| | | | ||||| | ||||| ||||| ||||||| ||| | | |
Db      26869 TATATATACACACATATATACACATACACACATACATATATACACATATACACATATACA 26928

Qy        644 TATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATA 703
              | ||| ||||| | |    | ||||| ||||||| | || || |||||| |||||   | |
Db      26929 CACACATATACACACATGTATACATATATATACACACATGTATACATATGTATACACACA 26988

Qy        704 TATATGTATATATATATGCTCCAGAGTTCATAAGAG 739
              |||| | |||| ||||    |      | |||| | |
Db      26989 CATATATGTATACATATATACACACATACATATGTG 27024


RESULT 57
US-10-801-837-5
; Sequence 5, Application US/10801837
; Publication No. US20040152167A1
; GENERAL INFORMATION:
;  APPLICANT: BEASLEY, Ellen M. et al.
;  TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
;  TITLE OF INVENTION:  NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND
;  TITLE OF INVENTION:  USES THEREOF
;  FILE REFERENCE: CL001229-DIV
;  CURRENT APPLICATION NUMBER: US/10/801,837
;  CURRENT FILING DATE:  2004-03-17
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
;   LENGTH: 27067
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(27067)
;   OTHER INFORMATION: n = A,T,C or G
US-10-801-837-5

  Query Match             5.9%;  Score 293.6;  DB 18;  Length 27067;
  Best Local Similarity   75.8%;  Pred. No. 4.2e-50;
  Matches  391;  Conservative    0;  Mismatches  119;  Indels    6;  Gaps    2;

Qy        230 ATATATATATATATATATATATACACACACACACACATATAT-ATATACACATATATATA 288
              || ||||  |  | |  |  ||| |     | ||| |||||| | ||| | ||||| | |
Db      26509 ATGTATAACTTAAAAATTTCATATATGTGTATACATATATATGAAATATATATATACAGA 26568

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              |||| ||||| || |||||| ||||||||||||||||||  | | |||| |||||||
Db      26569 CACACATATATATGTATACATATATATACACATATATATGTATACATATATACACACATA 26628

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              ||| || |||||| |||||||| ||||||||||||||||||||| ||| | | ||| | |
Db      26629 TATGTATACATATATATACACACATATACACATATATACACACACATACATATATACACA 26688

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              |||| ||||| |||||||||||||| ||  |||||||| |  |||| | | ||| | |
Db      26689 CACATATATACACACATATATACACACATGCACACATATATATGTATACATATATACACA 26748

Qy        469 CATATATACACATATATATA-----ATATACACACATATATATACACATATATACACACA 523
              ||| |||||  ||||||| |     ||||||||||||||||||||||| ||||||||||
Db      26749 CATGTATACGTATATATACACACATATATACACACATATATATACACACATATACACACA 26808

Qy        524 TATATACACATATATACACATATATATACACATATATACACATATATACATATATACACA 583
              || | |||||||||||||| |||||||||| ||||| ||| | ||| |||| || |||
Db      26809 TACACACACATATATACACACATATATACACACATATATACACACATATATATGTATACA 26868

Qy        584 TATATATACATATATACACATATATACATATATACACATATATACATATATACACACATA 643
              |||||||||| | ||| | | | ||||| | ||||| ||||| ||||||| ||| | | |
Db      26869 TATATATACACACATATATACACATACACACATACATATATACACATATACACATATACA 26928

Qy        644 TATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATA 703
              | ||| ||||| | |    | ||||| ||||||| | || || |||||| |||||   | |
Db      26929 CACACATATACACACATGTATACATATATATACACACATGTATACATATGTATACACACA 26988

Qy        704 TATATGTATATATATATGCTCCAGAGTTCATAAGAG 739
              |||| | |||| ||||    |      | |||| | |
Db      26989 CATATATGTATACATATATACACACATACATATGTG 27024


RESULT 58
US-10-017-161-2361/c
; Sequence 2361, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
;  FILE REFERENCE: 084335/0152
;  CURRENT APPLICATION NUMBER: US/10/017,161
;  CURRENT FILING DATE:  2002-12-18
;  PRIOR APPLICATION NUMBER: JP 2001/246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2430
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2361
;   LENGTH: 1486
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: source
;   LOCATION: (1)..(1486)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(1286)
US-10-017-161-2361

  Query Match             5.9%;  Score 293.2;  DB 15;  Length 1486;
  Best Local Similarity   74.7%;  Pred. No. 1.4e-50;
  Matches  396;  Conservative    0;  Mismatches  128;  Indels    6;  Gaps    2;

Qy        197 AAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACA 256
              | || |   | || | || | | |      || || | | ||  | || |||| |||||||
Db        999 ATATATACACACATATAATATATACATATTTATACACACATGCACATGTATAAATACACA 940

Qy        257 CACACACACATATATATATACACATATATATACACATATATACATATATACACATATATA 316
              | | ||| ||||| | | | | || |||||| |||||||||||||||||||||||| ||
```

```
Db       939 TATATACATATATACACACATATATGTATATATACATATATACATATATACACATATGTA 880

Qy       317 CACATATATATACACATACATATGTACACATATATATACACATATGTATACACATATATA 376
             | ||| ||||| ||||| |||    | |  ||||||||||||||||||| | | | |||
Db       879 TATATACATATATACATATATACACATATGTATATATACACATATGTATATATACACATA 820

Qy       377 CACATATATACACACATATATACACATATATACACACATATATACACATATATACACATA 436
             ||||||| || | | | | |||||||||| || | | | | |||||||||| || | |||
Db       819 CACATATGTATATATACACATACACATATGTATATATACACATACACATATGTATATATA 760

Qy       437 TATACACATATACACATATATACACATATATACATATATACACATATATATA-----ATA 491
             | | ||| |||   |||||||||||||| ||  |||||||||||||||| ||     | |
Db       759 CACATACACATATGTATATATACACATACATGTATATATACACATATATGTATATATACA 700

Qy       492 TACACACATATATATACACATATATACACACATATATACACATATATACACATATATATA 551
             | |||||| |||| ||| | | ||| |   |||||||||||||| | | | ||| | |||
Db       699 CATACACATGTATACACATACACATATATGTATATATACACATACACATATATACACATA 640

Qy       552 CACATATATACACATATATACATATATACACATATATATACATATATACACATATATACA 611
             ||| |||||||||||||||| ||||||||||||||   |||||||||||||| |||
Db       639 TACACATATACACATATATACACATATACACATATATAAGTATATATACACATATACACA 580

Qy       612 TATATACACATATATACATATATACACACATATATACACATACATAT-ACACACACATAG 670
             ||||   | ||||| |||||||||||||| |||  || | ||||| || | ||||| |||
Db       579 TATACGTATATATACACATATATACACATATACGTATATATACACATAATACACATATAC 520

Qy       671 ATATACATATATATACACATATATATACGTATATATATGTATATATATAT 720
             |||| ||| | ||| | | | |||||||||||||||| ||||||| | ||
Db       519 GTATAAATACACATATATACACATATACGTATATATACGTATATACACAT 470


RESULT 59
US-10-017-161-2381/c
; Sequence 2381, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
;  FILE REFERENCE: 084335/0152
;  CURRENT APPLICATION NUMBER: US/10/017,161
;  CURRENT FILING DATE:  2002-12-18
;  PRIOR APPLICATION NUMBER: JP 2001/246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2430
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2381
;   LENGTH: 1672
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: source
;   LOCATION: (1)..(1672)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(1472)
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (208)..(307)
;   OTHER INFORMATION: a, t, c, g, unknown or other
US-10-017-161-2381

  Query Match             5.9%;  Score 291.6;  DB 15;  Length 1672;
  Best Local Similarity   75.5%;  Pred. No. 3.1e-50;
  Matches  376;  Conservative    0;  Mismatches  119;  Indels    3;  Gaps    1;

Qy       227 TTTATATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
             | |||||||||||||||||||||||| | | ||||||||| | ||||||||||||||||||
Db      1390 TATATATATATATATATATATATATATATACACACACACACATATATACACATATATA 1331

Qy       287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
             |||||| | | | ||||| ||| ||||| |||||||||||||||||||| |||   | | |
Db      1330 TACACACACACATATATACACATATATACACACATATATATACACATATATACACATATA 1271

Qy       347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
             |||| | | | ||||  | | | |||| |||||||||||||||||| |||| || ||||
Db      1270 TATACACATATATATACACATATATATGTACACATATATACACACATGTATATACGTATA 1211

Qy       407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
             |    |||||||||||||| | ||| |  ||| |  | ||||||||||||||  | |||||
Db      1210 TGTGTACATATATACACACACATATATGTATGTGTATATATACACATATATGTATATATA 1151

Qy       467 TACATATATACACATATATATA---ATATACACACATATATATACACATATATACACACA 523
             |||| | | | |||| ||||||    |||| | ||| | ||| | | || |||| |
Db      1150 TACACACACAGACATGTATATATGTGTATATATACACACATAGATATATGTATATATGTG 1091

Qy       524 TATATACACATATATACACATATATATACACATATATACACATATATACATATATACACA 583
             |||||||||| | ||| |  | |||||||||| |||||    |||||||| | | | | |
Db      1090 TATATACACACACATATATGTGTATATACACACATATATGTGTATATACACACACATATA 1031

Qy       584 TATATATACATATATACACATATATACATATATACACATATATACATATATACACACATA 643
             | | |||| | | | ||| |||| |  |||| |||||  ||||  | |||| ||||| |
Db      1030 TGTGTATATACACACACATATATGTGTATATACACACACATATATGTGTATATACACACA 971

Qy       644 TATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATA 703
             ||| | |  ||||||||||||||  |||||  | ||||||||||||||| ||  |||
Db       970 CATATATGTGTATATACACACACATGTATATATGTGTATATACACATATATGTATATATG 911

Qy       704 TATATGTATATATATATG 721
             | ||| || | |||||||
Db       910 TGTATATACACATATATG 893


RESULT 60
US-10-292-798-2023/c
; Sequence 2023, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
;  FILE REFERENCE: 084335/166
;  CURRENT APPLICATION NUMBER: US/10/292,798
;  CURRENT FILING DATE:  2002-11-13
;  PRIOR APPLICATION NUMBER: 10/017,161
;  PRIOR FILING DATE: 2001-12-18
;  PRIOR APPLICATION NUMBER: JP 2001-246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2070
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2023
;   LENGTH: 1672
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   LOCATION: source
;   FEATURE:
;   LOCATION: (1)..(1672)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(1472)
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (208)..(307)
;   OTHER INFORMATION: a, t, c, g, unknown or other
US-10-292-798-2023

  Query Match             5.9%;  Score 291.6;  DB 17;  Length 1672;
  Best Local Similarity   75.5%;  Pred. No. 3.1e-50;
  Matches  376;  Conservative    0;  Mismatches  119;  Indels    3;  Gaps    1;

Qy       227 TTTATATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
```

```
Db       1390 TATATATATATATATATATATATATATATATACACACACACACATATATACACATATATA 1331

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              |||||| | | ||||| ||| ||||| ||||||||||||||||||| ||| |  | | |
Db       1330 TACACACACACATATATACACATATATACACACATATATATACACATATATACACATATA 1271

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||| | | | ||||  | | | |||| |||||||||||||||||||| |||| || ||||
Db       1270 TATACACATATATATACACATATATATGTACACATATATACACACATGTATATACGTATA 1211

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              |    |||||||||||||| | ||| |  ||| |  | ||||||||||||||  | |||||
Db       1210 TGTGTACATATATACACACACATATATGTATGTGTATATATACACATATATGTATATATA 1151

Qy        467 TACATATATACACATATATATA---ATATACACACATATATATACACATATATACACACA 523
              |||| | | | |||| ||||||    |||| | ||| | ||| | | || |||| |
Db       1150 TACACACACAGACATGTATATATGTGTATATATACACACATAGATATATGTATATATGTG 1091

Qy        524 TATATACACATATATACACATATATATACACATATATACACATATATACATATATACACA 583
              |||||||||| | ||| |  | |||||||||| |||||    |||||||| | | | | |
Db       1090 TATATACACACACATATATGTGTATATACACACATATATGTGTATATACACACACATATA 1031

Qy        584 TATATATACACATATATACACATATATACATATATACACATATATACATATATACACACATA 643
              | | |||| | | | ||| |||| |  ||||| ||||| |||||  | |||| ||||| |
Db       1030 TGTGTATATACACACACATATATGTGTATATACACACACATATATGTGTATATACACACA 971

Qy        644 TATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATA 703
              |||  |   |||||||||||||||   |||||  | ||||||||||||||||  ||  |||
Db        970 CATATATGTGTATATACACACACATGTATATATGTGTATATACACATATATGTATATATG 911

Qy        704 TATATGTATATATATATG 721
              | ||| || | |||||||
Db        910 TGTATATACACATATATG 893


RESULT 61
US-10-741-601-5636/c
; Sequence 5636, Application US/10741601
; Publication No. US20040166519A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  STENOSIS, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001500
;  CURRENT APPLICATION NUMBER: US/10/741,601
;  CURRENT FILING DATE:  2003-12-22
;  NUMBER OF SEQ ID NOS: 26415
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5636
;   LENGTH: 81924
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(81924)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-741-601-5636

  Query Match             5.5%;  Score 273.8;  DB 18;  Length 81924;
  Best Local Similarity   65.9%;  Pred. No. 9e-46;
  Matches  458;  Conservative    0;  Mismatches  232;  Indels    5;  Gaps    4;

Qy         42 TACTAAAAATACAAAATTAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCG 101
              ||  ||||| | ||| ||||   |||||||||  |||| ||||||||||||||||| || |
Db      38267 TAAAAAAAAAAAAAAGTTAGTGGGGCATGGTGGCACACACCTGTAATCCCAGCTATTTGG 38208

Qy        102 GAGGCCGAGGCAGGAGAATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGC-AAGA 160
              ||||| |||||||||||||  ||||||||  | || |||||||||||| | ||||  | |
Db      38207 GAGGCTGAGGCAGGAGAATTGCTTGAACC--CAGGAGGTGGAGGTTGCAGTGAGCTGAAA 38150

Qy        161 TCATGCCATTGCACTCCAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAA 220
              |  ||| ||||||||||||||   |  || |||| || |    ||||| || | | | |

Db      38149 TTGTGCTATTGCACTCCAGCCTGAGTGAC-AGAGTGAGACTCTGTCTCAAAATATATATA 38091

Qy        221 AGGCATTTTATATATATATATATATATATACACACACACACATATATATATACACA 280
                  | |||| | | | | |||||| | ||||||||| | || | | | || | |
Db      38090 TACACACACACATATGTGTGTGTGTATATACATACACACACATATATGTGTGTGTATATA 38031

Qy        281 TATATATACACATATATACATATATACACATATATACACATATATATACACATACATATG 340
              ||| ||| ||| ||| ||| | |||    ||||||||  |||| | | | ||   |||
Db      38030 CATACATATACACATACACACACATATGTGTATATACATGTATACACATATATGTGTATA 37971

Qy        341 TACACATATATATACACATATGTATACACATATATACACATATATACACACATATATACA 400
              ||||| | | | || |  | | |||||||||||| |     |||||    | ||| | | |
Db      37970 TACACGTGTGTGTATATGTGTATATACACATATGTGTGTGTATATGTGTATATACACATA 37911

Qy        401 CATATATACACACATATATACACATATATACACATATATACACATATACACATATATACA 460
              | | || |    |||||||||||||| |     |||||    |||||||||||| |
Db      37910 TGTGTGTATATGTGTATATACACATATGTGTGTGTATATGTGTATATACACATATGTGTG 37851

Qy        461 CATATATACATATATACACATATATATAATATACACACATATATATACACATATATACAC 520
              |||||   |||||||||||||| | |  | || |    ||||||||  ||||||||||
Db      37850 TGTATATGTGTATATACACATATGTGTG-TGTATATGTGTATATATATGTATATATACAC 37792

Qy        521 ACATATATACACATATATACACATATATATACACATATATACACATATATACATATATAC 580
              | | | | ||| |||| |     | ||| || | | | | | |  | | ||  |||||||
Db      37791 ATACACACACATATATGTTTGTGTGTATGTATATACACACATATGTGTGTATGTATATAC 37732

Qy        581 ACATATATATACATATATACACATATATACATATATACACATATATACATATATACACAC 640
              ||| |||| |   |||| ||| ||||  ||| ||||| |||  | |  |||||||| ||
Db      37731 ACACATATGTGTGTATACACATATATGCACACATATATACACGTGTGTATATATACCCAT 37672

Qy        641 ATATATACACATACATATACACACACATAGATATACATATATATACACATATATATACGT 700
              |||| || | ||   |||| | |||  |   ||||||||  | |  | ||||| ||  ||
Db      37671 ATATGTATATATGTGTATATATACATGTGTGTATACATACGTGTGTATATATACATGTGT 37612

Qy        701 ATATATATGTATATATATATGCTCCAGAGTTCATA 735
              |||| ||   | | |||||  |    | ||  | |
Db      37611 GTATACATACGTGTGTATATACATGTGTGTATACA 37577


RESULT 62
US-10-741-600-17799
; Sequence 17799, Application US/10741600
; Publication No. US20050026169A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001499
;  CURRENT APPLICATION NUMBER: US/10/741,600
;  CURRENT FILING DATE:  2003-12-22
;  NUMBER OF SEQ ID NOS: 73997
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17799
;   LENGTH: 14603
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-741-600-17799

  Query Match             5.4%;  Score 267.2;  DB 19;  Length 14603;
  Best Local Similarity   71.5%;  Pred. No. 9.8e-45;
  Matches  449;  Conservative    0;  Mismatches  168;  Indels   11;  Gaps    7;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC--AAAAT 58
              || ||||||||| ||||  |||||||||||||| || ||||||||||||||  |||||
Db      10438 AGATCAAGACCATCCTGGCCAACATGGTGAAACTCCGTCTCTACTAAAAATACAAAAAAT 10497

Qy         59 TAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGA 118
              |||   ||| ||||| |   | |||||||||||||||||||| |||||| || ||||||||
Db      10498 TAGCTGGGCGTGGTGGTGGGCTCCTGTAATCCCAGCTACTTGGGAGGCTGAAGCAGGAGA 10557

Qy        119 ATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTG---CAC 174
              ||| ||||||||  | || || ||||||||| | ||| |||||||  |||||||   |||
Db      10558 ATCGCTTGAACC--CAGGAGGCGGAGGTTGCAGTGAGCCAAGATCGCGCCATTGCACCAC 10615
```

```
Qy       175 TCCAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATA 234
Db     10616 TCCAGCCTGGGCAAAAAGAGAGAAACTCTGTCTCAAAAAAAAAAAAAAAAAAAAAAAAAA 10675

Qy       235 TATATATATATATATATACACACACACACATATATATATACACATATATATACACATA 294
Db     10676 AATATATATATATATATATATACACACACATATATATTCATATATA-ATTCATATATA 10734

Qy       295 TATACATATATACACATATATACACATATATATACACATACATATGTACACATATATATA 354
Db     10735 TATTCATATATAATTCATATATAATTCATATATATATTCATATAGATATTCATATATA-A 10793

Qy       355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
Db     10794 TTCATATAGATATTCATAGATATTCATATATATTCATATAGATATTCATATATTCATATA 10853

Qy       415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
Db     10854 TATTCATATATTCATGTATATTCATATAGATATTCATATATAGATATTCATAGATATTCA 10913

Qy       475 TACACATATATATAATATACACACATATATATACACATATATACACACATATATACACAT 534
Db     10914 TATATAGATAT-TCATATAGATATTCATATATAGATATTCATATAGATATTCATGTATAG 10972

Qy       535 ATATACACATATATATACACATATATACACATATATACATATATACACATATATATACAT 594
Db     10973 ATATTCATATATATATTCATATATAGATATTCATATAGATATATTCATATCTATATGAAT 11032

Qy       595 ATATACACATATATACATATATACACAT 622
Db     11033 ATATATATGAATATATCTATATATGAAT 11060
```

RESULT 63
US-10-311-455-1210/c
; Sequence 1210, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1210
; LENGTH: 6971
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 2741
; OTHER INFORMATION: n is a or g or c or t
US-10-311-455-1210

Query Match 5.4%; Score 266.4; DB 15; Length 6971;
Best Local Similarity 71.1%; Pred. No. 1e-44;
Matches 367; Conservative 0; Mismatches 146; Indels 3; Gaps 1;

```
Qy       212 AAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACATATAT 271
Db      6472 ATACATATATACCTATATATATATACCTATATATATAAATACACATATATATATATAT 6413

Qy       272 ATATACACATATATATACACATATATACATATATACACATATATACACATATATATACAC 331
Db      6412 ATACCAATATATATATATACACGTATATATATATACACACTTATATACATATATACCTAT 6353

Qy       332 ATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACACAC 391
Db      6352 ATATATAAATACACATATATATATATATACCTATATATATAAATACACATTTATATATAT 6293

Qy       392 ATATATACACATATATACACACATATATACACATATATACACATATATACACATATACAC 451
Db      6292 AAATACACATTTATATATATATACGCCTATATAAATACACATTTATATATATATATACCT 6233

Qy       452 ATATATACACATATATACATATATACACATATATATAATATACACACATATATATACACA 511
Db      6232 ATATATATACCTATATATATAAATACCTATATATATAAAACACATTTATATATATATATA 6173

Qy       512 TATATACACACATATATACACATATATACACATATATATACACATATATACACATATATA 571
Db      6172 CCTATATATAAATACACATTTATATATATACCTCTATATATATACCTATATATAAAAATA 6113

Qy       572 CATATATACACATATATATACATATATACACATATATACATATATACACATATATACATA 631
Db      6112 CACATTTATATATACGCCTATATATAAAAATACACATTTATATATACGCCTATATATATA 6053

Qy       632 TATACACACATATATACACATACATATACACACACATAGATATACATATATAT---ACAC 688
Db      6052 AATACACATTTATATATATATACGCCTATATATATAAATACCTATATATATATAAAACAC 5993

Qy       689 ATATATATACGTATATATATGTATATATATATGCTC 724
Db      5992 ATTTATATATATATATATACCTATATATAAATACAC 5957
```

RESULT 64
US-10-812-232-6/c
; Sequence 6, Application US/10812232
; Publication No. US20040265961A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Wei
; APPLICANT: Wu, Leeying
; APPLICANT: Channavajhala, Padma L,
; APPLICANT: Lin, Lih-Ling
; APPLICANT: Zhang, Yuhua
; TITLE OF INVENTION: Novel Proteins Homologous to Kinase Suppressor of Ras
; FILE REFERENCE: 01997.026700
; CURRENT APPLICATION NUMBER: US/10/812,232
; CURRENT FILING DATE: 2004-03-29
; PRIOR APPLICATION NUMBER: US 60/457,928
; PRIOR FILING DATE: 2003-03-28
; PRIOR APPLICATION NUMBER: US 60/491,283
; PRIOR FILING DATE: 2003-07-31
; NUMBER OF SEQ ID NOS: 254
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6
; LENGTH: 263852
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3814)..(4319)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (6552)..(6571)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (17206)..(17225)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (19680)..(19699)

```
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (24320)..(24339)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (26572)..(26591)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (32986)..(33005)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (56103)..(56724)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (59672)..(59691)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (61525)..(61544)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (71035)..(71054)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (79771)..(79807)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (91343)..(91801)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (141749)..(141768)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (144739)..(144758)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (164171)..(164340)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (173689)..(173708)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (175051)..(175070)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (192924)..(192993)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (202010)..(202029)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (247837)..(247856)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (251852)..(252421)
;   OTHER INFORMATION: n is a, c, g, or t
```

```
US-10-812-232-6

  Query Match             5.3%;  Score 265.6;  DB 18;  Length 263852;
  Best Local Similarity   73.7%;  Pred. No. 7.8e-44;
  Matches  367;  Conservative    0;  Mismatches  124;  Indels    7;  Gaps    2;

Qy        229 TATATATATATATATATATATACACACACACACATATATATATACACATATATATA 288
              || |||||||||| ||| ||| | ||| ||| ||| || |  | ||||||| ||  |||
Db      94187 TACATATATATATACATACATACATACATACATACATATGTCCACATACACACATGCATA 94128

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              | ||||||| | ||||| | ||||||| ||| | | | |  ||| ||||  |||| | |
Db      94127 TATATATATATACATATATATATATATATACACACACACATTCATGCATAAATACAAACA 94068

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACAT------ATATACACA 402
              |||| |||||||  |||| ||| ||| ||| | |||| |||||||      ||| |||||
Db      94067 TATACACACATACATATAAACACATACACAGAGATATGCACACATACATACATACACACA 94008

Qy        403 TATATACACACATATATACACATATATACACATATATACACATATACACATATATACACA 462
              || ||| | ||||  ||||||||| ||||||| | | ||||||  | || |||||| |||
Db      94007 TACATATATACATGCATACACATAAATACACACAGACACACATGCATACGTATATATACA 93948

Qy        463 TATATACATATATACACATATATATAATATACACACATATATATACACATATATACACAC 522
              |||||  ||| ||| ||||||| |  | ||||| | | | ||||||||| ||||||||||
Db      93947 TATATGTATACATATACATATACA-CACATACATAAACACATATACACAGATATACACAC 93889

Qy        523 ATATATACACATATATACACATATATATACACATATATACACATATATACATATATACAC 582
              ||| | | ||| | | | |||||| |||||| ||| | | ||||||  | | | |||  |
Db      93888 ATACACAAACACACAGAGACATATGTATACATATACACATACATATGCAAACACATAAGC 93829

Qy        583 ATATATATACATATATACACATATATACATATATACACATATATACATATATACACACAT 642
              | ||| |||||||||||||| | ||||||| ||| | ||||| ||||||| ||||| | |
Db      93828 ACATACATACATATATACATACATATACACATACATACATACATACATAAATACATAGAG 93769

Qy        643 ATATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTAT 702
              | | |||   ||||||| ||||   | | ||||| ||| ||||||| | | | |||| |
Db      93768 ACACACATGCACATATATACACATGCATACATACACATAAATACACACAGACACACGTGT 93709

Qy        703 ATATATGTATATATATAT 720
              |||||   | ||| | ||
Db      93708 ATATACACACATACACAT 93691


RESULT 65
US-10-108-260A-543/c
; Sequence 543, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
;  APPLICANT: HELIX RESEARCH INSTITUTE
;  TITLE OF INVENTION: No. US20040005560A1el full length cDNA
;  FILE REFERENCE: H1-A0106
;  CURRENT APPLICATION NUMBER: US/10/108,260A
;  CURRENT FILING DATE:  2002-03-27
;  NUMBER OF SEQ ID NOS: 5458
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 543
;   LENGTH: 2522
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-108-260A-543

  Query Match             5.3%;  Score 263.6;  DB 17;  Length 2522;
  Best Local Similarity   68.9%;  Pred. No. 2.5e-44;
  Matches  376;  Conservative    0;  Mismatches  169;  Indels    1;  Gaps    1;

Qy        236 ATATATATATATATATACACACACACACACATATATATATACACATATATATACACATAT 295
              ||  |  | | |||||| | | ||||||| |  |||||||    ||||||||||||  |
Db       2217 ATTCAACTGTGTATATATATATACACACATACGTATATATGTGTATATATATACACACGT 2158

Qy        296 ATACATATATACACATATATACACATATATATACACATACATATGTACACATATATATAC 355
              ||| ||| | | | || |||||| | | |   | | ||||| || ||   ||| ||| ||
Db       2157 ATATATACACATATATGTATACATACACACGCATATATACACATATATGTATACATACAC 2098
```

```
Qy        356 ACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACAT 415
              || ||| |||||||||||||  | | ||| |||| ||||||||||||||||||  | ||||
Db       2097 ACGTATATATACACATATATGTATACATACACACGCATATATACACATATATGTATACAT 2038

Qy        416 ATATACACATATATACACATATATACACATATACACATATATACACATATATACATATAT 475
              | | ||  |||||||||||||||| || | |||||||   ||| | | | ||| ||  ||
Db       2037 ACACACGTATATATACACATATATGCATACATACACACGCATATATACACATATATGCAT 1978

Qy        476 ACACATATATATAATATACACACATATATATACACATATATACACACATATATACACATA 535
              ||| | |||  | ||||| |||||||||| || ||||| ||||  | | ||| ||| |||
Db       1977 ACATACATACGT-ATATATACACATATATGTATACATACATACGTATACATACACATATA 1919

Qy        536 TATACACATATATATACACATATATACACATATATACATATATACACATATATATACATA 595
              | || |||||  ||   || ||||| ||| |||| || | ||| | || ||||||||| |
Db       1918 TGTATACATACATGCGCATATATACACATATATGTATACATACATACGTATATATACACA 1859

Qy        596 TATACACATATATACATATATACACATATATACATATATACACACATATATACACATACA 655
              ||||  ||| |||||||  || ||||| | | || |||||| |||||  || | |||||
Db       1858 TATATGTATACATACATACGTATACATACACATATGTATACATACATACGTATATATACA 1799

Qy        656 TATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATA 715
              ||| |   | | | | || | ||||||||||||||||| ||   ||| ||  |||||||
Db       1798 CATATATGTATACATACATGCGTATATATACACATATATGTATACATACATGCGTATATA 1739

Qy        716 TATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGA 775
              || |      |  | ||||      | ||  |   ||    || | ||| |   | | |
Db       1738 TACACATATATGCATACATACATATGAACACATATATATGTGTGTGTATATATACATATA 1679

Qy        776 GAGTTT 781
                 |||
Db       1678 AGATTT 1673
```

RESULT 66
US-10-723-860-1357/c
; Sequence 1357, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1357
; LENGTH: 227968
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-723-860-1357

Query Match 5.3%; Score 263; DB 18; Length 227968;
Best Local Similarity 71.6%; Pred. No. 2.5e-43;
Matches 373; Conservative 0; Mismatches 145; Indels 3; Gaps 2;

```
Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||  |||||||||||  ||||||||||||||||||||||||||||||||||| ||||||
Db       5103 AGTTTGAGACCAGCCTGGCCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAAATT 5044

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              || ||||||||||  ||| ||||||||||||||||||| | |||||| || | |||||||
Db       5043 AGCCAGGCATGGTGGCACATGCCTGTAATCCCAGCTACCTGGGAGGCTGAAGTAGGAGAA 4984

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              || || ||||| |  || ||  |||| ||| | |||| |         | | |   ||
Db       4983 TCTCTGGAACCAG--GGAGGCAGAGGCTGCAGTGAGCCAAGATCATTCCTGGGTGTCAGA 4926

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAGGCATTTTATATATATAT 239
```

```
              || |  ||     |      |  | || ||||||||||||   |    | | | | |
Db       4925 GAGAGACTCCATCCCTGCCCGTCCCCCGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4866

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              | | | ||||||| | | | | | ||||||||||| | ||||| | ||||| | |||
Db       4865 AAAAAAATATATATATATATATATATATATATATATATATATACACACACACACATAT 4806

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACAT 359
              ||||| | | ||||| |||||||||||||||| ||| |||  || | ||| |||| | ||
Db       4805 ATATACATATATATACACACATATATATACATATATATACATATATATATACATATATAT 4746

Qy        360 ATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATAT 419
              | ||||| | ||| ||| | ||||| ||| | ||||| ||| |||||||||| | ||||||
Db       4745 ACGTATATATATACATATATATATACACATATATATACACATATATATACATATATATAT 4686

Qy        420 ACACATATATACACATATATACACATATACACATATATACACATATATACATATATACAC 479
              |||||||||| ||| ||||| | | ||| | ||||| ||| ||||||||||||||| |
Db       4685 ACACATATATATACACATATATATACATATATATATACACATATATATACATATATATAT 4626

Qy        480 ATATATATAATATACACACATATATATACACATATATACAC 520
              || |||||  |||  | |       || |    || || |||
Db       4625 ATGTATATTTTATTGAGATCACATTAAATGTGTACATTCAC 4585
```

RESULT 67
US-10-741-600-17809/c
; Sequence 17809, Application US/10741600
; Publication No. US20050026169A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001499
; CURRENT APPLICATION NUMBER: US/10/741,600
; CURRENT FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 73997
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17809
; LENGTH: 109730
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-741-600-17809

Query Match 5.3%; Score 262.6; DB 19; Length 109730;
Best Local Similarity 72.0%; Pred. No. 2.2e-43;
Matches 384; Conservative 0; Mismatches 144; Indels 5; Gaps 3;

```
Qy        204 ATCTCAGAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACAC 263
              || | | | | | | |||   || |||||||||| |||||||| |||| ||| | | |
Db     108244 ATGTTATATATACATAAATATATGTTATATATATTTATATATATTTATATACATATAAAT 108185

Qy        264 ACATATATATATACACATATATATACACATATATACATATATACACATATATA-CACATA 322
              | |||||| |||| ||||||||| | | ||||||| || |||| | | |||||     |||
Db     108184 ATATATATTTATATACATATATAAAAATATATATATATTTATATATAAATATATGTTATA 108125

Qy        323 TATATACACATACATATGTACACATATATATACACATATGTATACACATATATACACATA 382
              |||||  | ||||||||  | | |||| |||| ||||||  ||| | || |||| |||||
Db     108124 TATATTTATATACATATAAATATATATTTATATACATATAAATATATATTTATATACATA 108065

Qy        383 TATACACACATATATACACATATATACACACATATATACACATATATACACATATATACA 442
              || | | | || |||| |||||| | |   |||| || | | ||||| | |||||||
Db     108064 TAAATATATATTTATATACATATAAAAATCTATATTTATATACATATAAAAATATATATT 108005

Qy        443 CATATACACATATATACACATATATACATATATACACATATATATA---ATATACACACA 499
              |||||| ||| | | | |||| || ||| ||| | |||||||||   |||||| | |
Db     108004 TATATACATATAAAAATATATATTTATATACATATAGATATATATATTTATATACATATA 107945

Qy        500 TATATATACACATATATACACACATATATACACATATATACACATATATATACACATATA 559
              |||||||  |  |||||||| | |||||||  |  |||||||| ||| ||||| | || ||
Db     107944 AATATATATATTTATATACATAAATATATATATTTATATACATATAAATATATATATTTA 107885

Qy        560 TACACATATATACATATATACACATATATATACATATATACACATATATACATATATACA 619
              || ||||| ||| ||||||  | ||| ||||  ||||| | | | ||||| |||| || |
```

```
Db     107884 TATACATAAATATATATATTTATATACATATGTATATAAATATAAATATATATATGTATA 107825

Qy        620 CATATATACATATATACACACATATATACACATACATATACACACACATAGATATACATA 679
              | ||||| |||| || |   |||| || | | | ||||| |   | ||| ||||| |||
Db     107824 TAAATATATATATTTATATTTATATGTATATAAATATATATATTTATATAAATATATATA 107765

Qy        680 TATATACACATAT-ATATACGTATATATATGTATATATATATGCTCCAGAGTT 731
              | ||||   |||| |||||| || |||||| ||||| | | ||     |||||
Db     107764 TTTATATTTATATAATATACATACATATATATATATTTTTTTGAGATGGAGTT 107712


RESULT 68
US-10-437-963-12666/c
; Sequence 12666, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 12666
;   LENGTH: 1800
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_18773C.1
US-10-437-963-12666

  Query Match             5.3%;  Score 262;  DB 18;  Length 1800;
  Best Local Similarity   72.0%;  Pred. No. 4.6e-44;
  Matches  370;  Conservative    0;  Mismatches  140;  Indels    4;  Gaps    2;

Qy        229 TATATATATATATATATATATATACACACACACACATATATATATACACATATATATA 288
              |||||||||||||||| |||||| | | ||| | | ||||||| ||| | |||||||||
Db       1796 TATATATATATATATACATATATATATATACATATATATATATACATATATATATATA 1737

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              || ||||||| ||||||| | ||||||| | | ||||||| | |||||||| || | |||
Db       1736 CATATATATATATATATATATATATATATATACATATATATATACATATATATATATA 1677

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              |||||| | |||| |||| ||  | ||  | |||| || | | || |||| | |||||||
Db       1676 TATATATATATATATATATACGAAAATGTATATATGTATATATATGTATATATATATATA 1617

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              |   |||||||   ||||||||  || |||| | || || | |||  ||   |||||||
Db       1616 TATGTATATATATGTATATATACGTATGTATATATATGTATATATACGTATGTATATATA 1557

Qy        469 CATATATACACATATATATAATATACACACATATATATACACATATATACACACATATAT 528
              ||||||  || ||| ||||   || | || ||| |||| | || |||| | || |||||
Db       1556 TGTATATATACGTATGTATACG-TATATACGTATGTATATATATGTATATATACGTATAT 1498

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACATATAT 588
              | |  |||||| || ||||||||   |||||||  ||||||||||||||   ||||||
Db       1497 ATATGTATATATACGTATATATATGTATATATATGTATATATACATATATATGTATATAT 1438

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATA---TACACACATATA 645
              || |||||||   ||||||||||||||   |||||||  ||||   ||    | || |
Db       1437 GTATATATATATGTATATATACATATATATGTATATATATGTATATGTTATGTATATGTG 1378

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              |    ||| || || |   | ||  |||||  |||||||  | |||| ||||  ||||||
Db       1377 TGTGTATATATGTATATGTATATGTATATATGTATATATGTATATATGTATATATATATA 1318

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAG 739
              ||| ||| |||||||      |  |  ||||| |
Db       1317 TATATATGTATATATATATATGTATATATAAGCG 1284


RESULT 69
US-10-719-993-6815
; Sequence 6815, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001496
;  CURRENT APPLICATION NUMBER: US/10/719,993
;  CURRENT FILING DATE:  2003-11-24
;  NUMBER OF SEQ ID NOS: 55342
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6815
;   LENGTH: 1980090
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1980090)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-2
US-10-719-993-6815

  Query Match             5.3%;  Score 262;  DB 18;  Length 1980090;
  Best Local Similarity   70.7%;  Pred. No. 1.1e-42;
  Matches  420;  Conservative    0;  Mismatches  165;  Indels    9;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC--AAAAT 58
              ||||| |||||||||||  |||||||| ||||||||||||||||||||||||  |||||
Db    1507113 AGTTCGAGACCAGCCTGGTCAACATGGTAAAACCCCATCTCTACTAAAAATACAAAAAAT 1507172

Qy         59 TAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGA 118
              ||| | ||  |||||      |||||||||||||||||||  |||||| ||| || ||||
Db    1507173 TAGCCGGGTGTGGTGGCGAGTGCCTGTAATCCCAGCTACTCTGGAGGCTGAGACATGAGA 1507232

Qy        119 ATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCC 177
              ||| ||||||||||  || || ||||||||| | ||| | |||||  |||| | ||||||
Db    1507233 ATCGCTTGAACCTG--GGAGGCGGAGGTTGCAGTGAGCCGAGATCGCGCCACTACACTCC 1507290

Qy        178 AGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATAT 237
              ||||  |||||||||||||||||      ||||| ||||||||||||    |   | | | ||
Db    1507291 AGCCTGGGCAACAAGAGCGAAACTCTGTCTCAAAAAAAAAAAAA---AAAAAAAAAAAAT 1507347

Qy        238 ATATATATATATATACACACACACACATATATATATACACATATATATACACATATAT 297
              |||||||||||||||| | | | | |   || ||||||||     |||||||  | |||| |
Db    1507348 ATATATATATATATATATATATATATGTATGTATATATATGTGTATATATGTATATATGT 1507407

Qy        298 ACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACAC 357
              ||||||||   |||||||    || |||||   ||| || |||| |  | ||||||
Db    1507408 GCATATATATGTATATATATGTGTAGATATATGTATATATGTGTATATGTGTATATATGT 1507467

Qy        358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417
              |||| | || | || | | || | || |||| |    | |||| |  ||||||  |    |||
Db    1507468 ATATATGTATATATGTGTATATATGTATATATGTGTGTATATATGTATATATATGTGTAT 1507527

Qy        418 ATACACATATATACACATATATACACATATACACATATATACACATATATACATATATAC 477
              |||   |||||||    ||||||   ||||      ||||||   |||||||  | | |
Db    1507528 ATATGTATATATATGTGTATATATGTATATGTGTGTATATATGTATATATATGTGTGTGT 1507587

Qy        478 ACATATATATAATATACACACATATATATACACATATATACACACATATATACACATATA 537
              | |||| ||| ||||      ||||||||  | ||||||  | | |||| |     ||||
Db    1507588 ATATATGTAT-ATATGTGTGTATATATATGTATATATATGTATATATATGTGTGTGTATA 1507646

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATA 591
              || |  ||||||||    |||||||     | |||| ||||||  | |||||||||
Db    1507647 TATATGTATATATATGTATATATATGTGTGTATATATATATGTATATATATATA 1507700
```

```
RESULT 70
US-10-741-600-17676
; Sequence 17676, Application US/10741600
; Publication No. US20050026169A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001499
;  CURRENT APPLICATION NUMBER: US/10/741,600
;  CURRENT FILING DATE:  2003-12-22
;  NUMBER OF SEQ ID NOS: 73997
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17676
;   LENGTH: 1980090
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1980090)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-741-600-17676

  Query Match             5.3%;  Score 262;  DB 19;  Length 1980090;
  Best Local Similarity   70.7%;  Pred. No. 1.1e-42;
  Matches  420;  Conservative    0;  Mismatches  165;  Indels    9;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC--AAAAT 58
              ||||| |||||||||||  ||||||||| |||||||||||||||||||||||  |||||
Db    1507113 AGTTCGAGACCAGCCTGGTCAACATGGTAAAACCCCATCTCTACTAAAAATACAAAAAT 1507172

Qy         59 TAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGA 118
              ||| | ||  |||||      |||||||||||||||||||  |||||| ||| || ||||
Db    1507173 TAGCCGGGTGTGGTGGCGAGTGCCTGTAATCCCAGCTACTCTGGAGGCTGAGACATGAGA 1507232

Qy        119 ATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCC 177
              ||| ||||||||||  || || |||||||| | ||| | |||||  |||| | ||||||
Db    1507233 ATCGCTTGAACCTG--GGAGGCGGAGGTTGCAGTGAGCCGAGATCGCGCCACTACACTCC 1507290

Qy        178 AGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATAT 237
              ||||  |||||||||||||||     ||||| ||||||||||   |     | | | ||
Db    1507291 AGCCTGGGCAACAAGAGCGAAACTCTGTCTCAAAAAAAAAAAA---AAAAAAAAAAAT 1507347

Qy        238 ATATATATATATATACACACACACACATATATATATACACATATATATACACATATAT 297
              |||||||||||||| | | | |   || ||||||||    |||||||  | |||| |
Db    1507348 ATATATATATATATATATATATATATGTATGTATATATATGTGTATATATGTATATATGT 1507407

Qy        298 ACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACAC 357
              ||||||||  |||||||    || |||||  ||| || |||| |  | ||||||
Db    1507408 GCATATATATGTATATATATGTGTAGATATATGTATATATGTGTATATGTGTATATATGT 1507467

Qy        358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417
              |||| | || | || | || | || |||| |    | |||| |  |||||| |    |||
Db    1507468 ATATATGTATATATGTGTATATATGTATATATGTGTGTATATATGTATATATATGTGTAT 1507527

Qy        418 ATACACATATATACACATATATACACATATACACATATATACACATATATACATATATAC 477
              |||   |||||||     ||||||   ||||     ||||||   |||||||  | | |
Db    1507528 ATATGTATATATATGTGTATATATGTATATGTGTGTATATATGTATATATATGTGTGTGT 1507587

Qy        478 ACATATATATAATATACACACATATATATACACATATATACACACATATATACACATATA 537
              | |||| ||| ||||     |||||||  | ||||||  | | |||| |     ||||
Db    1507588 ATATATGTAT-ATATGTGTGTATATATATGTATATATATGTATATATATGTGTGTGTATA 1507646

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATA 591
              || |  ||||||||   |||||||    | |||| ||||||  | |||||||||
Db    1507647 TATATGTATATATATGTATATATATGTGTGTATATATATATGTATATATATATA 1507700


RESULT 71
US-10-257-166-11/c
; Sequence 11, Application US/10257166
; Publication No. US20040023230A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Method and Nucleic Acids for Analysing the Methylation of
;  TITLE OF INVENTION:  Genes Implicated in Pharmacogenomics
;  FILE REFERENCE: 5013.1011
;  CURRENT APPLICATION NUMBER: US/10/257,166
;  CURRENT FILING DATE:  2002-10-07
;  PRIOR APPLICATION NUMBER: PCT/EP01/07470
;    DE 10032529.7
;    DE 10043826.1
;  PRIOR FILING DATE: 2001-06-29
;    2000-06-30
;    2000-09-01
;  NUMBER OF SEQ ID NOS: 178
; SEQ ID NO 11
;   LENGTH: 12069
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (390, 398)
US-10-257-166-11

  Query Match             5.3%;  Score 261.4;  DB 17;  Length 12069;
  Best Local Similarity   70.5%;  Pred. No. 1.4e-43;
  Matches  363;  Conservative    0;  Mismatches  151;  Indels    1;  Gaps    1;

Qy        204 ATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACAC 263
              || ||   | | | | ||      | |||||||||||| |||| |||||| | | | | |
Db       9846 ATATCTATATATATATAAATATATCTATATATATATAAATATCTATATATATATAAATAT 9787

Qy        264 ACATATATATATACACATATATATACACATATATACATATATACACATATATACACATAT 323
              |||||  |||| | || |||||| | ||| |||  ||||| ||| |||||  | ||||
Db       9786 CTATATACGTATAAATATCTATATATATATAAATATCTATATCCACGTATATCTATATAT 9727

Qy        324 ATATACACATACATATGTACACATATATATACACATATGTATACACATATATACACATAT 383
              ||||| | ||  |||| || | | ||||||| |   | |||||   ||||||| | | ||
Db       9726 ATATAAATATCTATATATATATAAATATATATATCCACGTATATCTATATATATATAAATAT 9667

Qy        384 ATACACACATATATACACATATATACACACATATATACACATATATACACATATATACAC 443
              ||  | | |||||||| | |||||||   ||| |||||  | |||||||  | |||| |||
Db       9666 ATCTATATATATATAAATATATATATCCACGTATATCTATATATATATAAATATCTACGT 9607

Qy        444 ATATACACATATATACACATATATACATATATACACATATATATAATATACACACATATA 503
              ||| | | |||||| ||| |||||  |||||||  | | |||| | |  || | | |||||
Db       9606 ATAAATATATATATCCACGTATATCTATATATATATAAATATCT-ACGTATAAATATATA 9548

Qy        504 TATACACATATATACACACATATATACACATATATACACATATATATACACATATATACA 563
              ||| ||| |||||  | | ||||||| | || |||| | |||||  ||| | ||||  ||
Db       9547 TATCCACGTATATCTATATATATATAAATATCTATATATATATAAATATATATATCCACG 9488

Qy        564 CATATATACATATATACACATATATATACATATATACACATATATACATATATACACATA 623
              |||| || ||||||| | |||| |||| ||||||| | || | || ||||||| | |||
Db       9487 TATATCTATATATATATAAATATCTATATATATATAAATATCTCTATATATATAAATATA 9428

Qy        624 TATACATATATACACACATATATACACATACATATACACACACATAGATATACATATATA 683
              ||||   | |||  |   ||||||| | | | || || | | | ||| |||| |||||||
Db       9427 TATATCCACATATATCTATATATATATAAATATCTATATATATATAAATATCTATATATA 9368

Qy        684 TACACATATATATACGTATATATATGTATATATAT 718
              || | | |||| ||  ||||||||  ||| |||||
Db       9367 TATATAAATATCTATATATATATAAATATCTATAT 9333


RESULT 72
US-10-719-993-6815/c
; Sequence 6815, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
```

```
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001496
;  CURRENT APPLICATION NUMBER: US/10/719,993
;  CURRENT FILING DATE:  2003-11-24
;  NUMBER OF SEQ ID NOS: 55342
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6815
;   LENGTH: 1980090
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1980090)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-719-993-6815

  Query Match             5.2%;  Score 259.6;  DB 18;  Length 1980090;
  Best Local Similarity   84.6%;  Pred. No. 3.4e-42;
  Matches  303;  Conservative    0;  Mismatches   54;  Indels    1;  Gaps    1;

Qy        270 ATATATACACATATATATACACATATATACATATATACACATATATACACATATATATAC 329
              ||||||| | ||| ||||| | ||| | ||||||||| ||||||||| ||||||||||||
Db    1507701 ATATATATATATACATATATATATACACACATATATATACATATATATACATATATATAC 1507642

Qy        330 ACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACAC 389
              ||| |||||| || ||||||||||||| |||| || ||||||||||||| ||||| ||||
Db    1507641 ACACACATATATATACATATATATACATATATATACACACATATATACATATATACACAC 1507582

Qy        390 ACATATATACACATATATACACACATATATACACATATATACACATATATACACATATAC 449
              ||||||||| |||||||||||||||||||| | | ||| | | | ||||| | | ||| ||
Db    1507581 ACATATATATACATATATACACACATATACATATATACACATATATATACATATATACAC 1507522

Qy        450 ACATATATACACATATATACATATATACACATATATATA-ATATACACACATATATATAC 508
              | ||||||||| ||||| ||| ||||| ||||||||| | ||||| ||| ||||| |||
Db    1507521 ATATATATACATATATACACACATATATACATATATACACATATATATACATATATACATAT 1507462

Qy        509 ACATATATACACACATATATACACATATATACACATATATATACACATATATACACATAT 568
              | | | ||| |||||||||||||| |||| ||||||||||||||||| |||||| |||||||
Db    1507461 ATACACATATACACATATATACATATATCTACACATATATATACATATATATGCACATAT 1507402

Qy        569 ATACATATATACACATATATATACATATATACACATATATACATATATACACATATAT 626
              |||||||||||||||||||||||||||| ||| | ||||||| ||||||| | |||| |
Db    1507401 ATACATATATACACATATATATACATACATATATATATATATATATATATATATTT 1507344


RESULT 73
US-10-741-600-17676/c
; Sequence 17676, Application US/10741600
; Publication No. US20050026169A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001499
;  CURRENT APPLICATION NUMBER: US/10/741,600
;  CURRENT FILING DATE:  2003-12-22
;  NUMBER OF SEQ ID NOS: 73997
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17676
;   LENGTH: 1980090
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1980090)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-741-600-17676

  Query Match             5.2%;  Score 259.6;  DB 19;  Length 1980090;
  Best Local Similarity   84.6%;  Pred. No. 3.4e-42;
  Matches  303;  Conservative    0;  Mismatches   54;  Indels    1;  Gaps    1;

Qy        270 ATATATACACATATATATACACATATATACATATATACACATATATACACATATATATAC 329
              ||||||| | ||| ||||| | ||| | ||||||||| ||||||||| ||||||||||||
Db    1507701 ATATATATATATACATATATATATACACACATATATATACATATATATACATATATATAC 1507642

Qy        330 ACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACAC 389
              ||| |||||| || ||||||||||||| |||| || ||||||||||||| ||||| ||||
Db    1507641 ACACACATATATATACATATATATACATATATATACACACATATATACATATATACACAC 1507582

Qy        390 ACATATATACACATATATACACACATATATACACATATATACACATATATACACATATAC 449
              ||||||||| |||||||||||||||||||| | | ||| | | | ||||| | | ||| ||
Db    1507581 ACATATATATACATATATACACACATATACATATATACACATATATATACATATATACAC 1507522

Qy        450 ACATATATACACATATATACATATATACACATATATATA-ATATACACACATATATATAC 508
              | ||||||||| ||||| ||| ||||| ||||||||| | ||||| ||| ||||| |||
Db    1507521 ATATATATACATATATACACACATATATACATATATACACATATATATACATATATACATAT 1507462

Qy        509 ACATATATACACACATATATACACATATATACACATATATATACACATATATACACATAT 568
              | | | ||| |||||||||||||| |||| ||||||||||||||||| |||||| |||||||
Db    1507461 ATACACATATACACATATATACATATATCTACACATATATATACATATATATGCACATAT 1507402

Qy        569 ATACATATATACACATATATATACATATATACACATATATACATATATACACATATAT 626
              |||||||||||||||||||||||||||| ||| | ||||||| ||||||| | |||| |
Db    1507401 ATACATATATACACATATATATACATACATATATATATATATATATATATATATTT 1507344


RESULT 74
US-10-017-161-2363
; Sequence 2363, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
;  FILE REFERENCE: 084335/0152
;  CURRENT APPLICATION NUMBER: US/10/017,161
;  CURRENT FILING DATE:  2002-12-18
;  PRIOR APPLICATION NUMBER: JP 2001/246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2430
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2363
;   LENGTH: 1489
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: source
;   LOCATION: (1)..(1489)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(1289)
US-10-017-161-2363

  Query Match             5.2%;  Score 258.8;  DB 15;  Length 1489;
  Best Local Similarity   72.9%;  Pred. No. 2e-43;
  Matches  360;  Conservative    0;  Mismatches  132;  Indels    2;  Gaps    2;

Qy        229 TATATATATATATATATATATACACACACACACATATATATATACACATATATATA 288
              | ||||||||| ||| |||||||| ||| | | |||||||| |||||| || ||   |||
Db        519 TGTATATATATGTATGTATATATATACATATATACACATATGTATATATACGTACC-ATA 577

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              |  |||||   |||||| || |||||| |||| |||||| || || ||||  ||| |||
Db        578 TATGTATATGTGTATATATACGTATATATACATGTATATATACGTATATATACACATATA 637

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              |||||  |||| | |||| || |||||| || |||| | | | ||| |    | |||||
Db        638 CATATATACATTTATATATACGTATATATACGTATACATATATATACACGTATATATACA 697

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              |||  |||||  ||   |||||| || |||| | | ||| ||   |||||| || |||| |
```

```
Db        698 TACATATATACACGTGTATATATACGTATACATATATACACGTGTATATATACGTATACA 757

Qy        469 CATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATATA 527
              |||||||||| | |||||| |  || ||| ||||| |      ||||||||  | | |||
Db        758 TATATATACACGTGTATATATACGTATACATATATACACGTGTATATATACGTATACATA 817

Qy        528 TACACATATATACACATATATATACACATATATACACATATATACATATATACACATATA 587
              || |||  | || | |||  |||||| ||||| ||   |||||| |  |||||| |||||
Db        818 TATACACGTGTATATATACGTATACATATATACACGTGTATATATACGTATACATATATA 877

Qy        588 TATACATATATACACATATATACATATATACACATATATACATATATACACACATATATA 647
              |    |||||||| | ||||| | |  | | ||| |||  |||||||||   ||||||
Db        878 CACGTGTATATACATACATATATACACGTGTATATACATATGTATATACACGTGTATATA 937

Qy        648 CACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATA 707
              |||  |||||| | || || |||||||| |||||||| |||||||||| ||||||
Db        938 TGTATATGTATACATATACGTATATATACATGTATATACACGTATATATACGTGTATATA 997

Qy        708 TGTATATATATATG 721
              |||||| ||||| |
Db        998 TGTATACATATACG 1011


RESULT 75
US-10-292-798-2007
; Sequence 2007, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
;  FILE REFERENCE: 084335/166
;  CURRENT APPLICATION NUMBER: US/10/292,798
;  CURRENT FILING DATE:  2002-11-13
;  PRIOR APPLICATION NUMBER: 10/017,161
;  PRIOR FILING DATE: 2001-12-18
;  PRIOR APPLICATION NUMBER: JP 2001-246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2070
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2007
;   LENGTH: 1489
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   LOCATION: source
;   FEATURE:
;   LOCATION: (1)..(1489)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(1289)
US-10-292-798-2007

  Query Match             5.2%;  Score 258.8;  DB 17;  Length 1489;
  Best Local Similarity   72.9%;  Pred. No. 2e-43;
  Matches  360;  Conservative    0;  Mismatches  132;  Indels    2;  Gaps    2;

Qy        229 TATATATATATATATATATATATACACACACACACATATATATATACACATATATATA 288
              | |||||||||| ||| |||||||| ||| | | |||||||| |||||| || ||   |||
Db        519 TGTATATATATGTATGTATATATATACATATATACACATATGTATATATACGTACC-ATA 577

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              |  |||||   |||||| || |||||| |||| |||||| || || ||||  ||| |||
Db        578 TATGTATATGTGTATATATACGTATATATACATGTATATATACGTATATATACACATATA 637

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              . ||||| |||| | |||| || |||||| || |||| | | | ||| |    | ||||| |
Db        638 CATATATACATTTATATATACGTATATATACGTATACATATATATACACGTATATATACA 697

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              |||  ||||| ||   |||||| || |||| | | ||| ||   |||||| || |||| |

Db        698 TACATATATACACGTGTATATATACGTATACATATATACACGTGTATATATACGTATACA 757

Qy        469 CATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATATA 527
              |||||||||| | |||||| |  || ||| ||||| |      ||||||||  | | |||
Db        758 TATATATACACGTGTATATATACGTATACATATATACACGTGTATATATACGTATACATA 817

Qy        528 TACACATATATACACATATATATACACATATATACACATATATACATATATACACATATA 587
              || |||  | || | |||  |||||| ||||| ||   |||||| |  |||||| |||||
Db        818 TATACACGTGTATATATACGTATACATATATACACGTGTATATATACGTATACATATATA 877

Qy        588 TATACATATATACACATATATACATATATACACATATATACATATATACACACATATATA 647
              |    |||||||| | ||||| | |  | | ||| |||  |||||||||   ||||||
Db        878 CACGTGTATATACATACATATATACACGTGTATATACATATGTATATACACGTGTATATA 937

Qy        648 CACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATA 707
              |||  |||||| | || || |||||||| |||||||| |||||||||| ||||||
Db        938 TGTATATGTATACATATACGTATATATACATGTATATACACGTATATATACGTGTATATA 997

Qy        708 TGTATATATATATG 721
              |||||| ||||| |
Db        998 TGTATACATATACG 1011


RESULT 76
US-10-367-094-179/c
; Sequence 179, Application US/10367094
; Publication No. US20040170982A1
; GENERAL INFORMATION:
;  APPLICANT: David W. Morris
;  APPLICANT:  Marc Malandro
;  TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
;  FILE REFERENCE: 5294520001500
;  CURRENT APPLICATION NUMBER: US/10/367,094
;  CURRENT FILING DATE:  2003-02-14
;  NUMBER OF SEQ ID NOS: 203
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 179
;   LENGTH: 163382
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(163382)
;   OTHER INFORMATION: n = A,T,C or G
US-10-367-094-179

  Query Match             5.2%;  Score 258.6;  DB 18;  Length 163382;
  Best Local Similarity   70.0%;  Pred. No. 1.8e-42;
  Matches  362;  Conservative    0;  Mismatches  154;  Indels    1;  Gaps    1;

Qy        204 ATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACAC 263
              || | |  | | | |||          |||| ||||||| |||| |||| | ||| | | |
Db     145185 ATATAAATATATATAAATAAATAAATATAAATATATAAATATGTATAAATACATATATAA 145126

Qy        264 ACATATATATATACACATATATATACACATATATACATATATACACATATATACACATAT 323
              | | ||||| ||| | | ||||||| | ||| ||| ||| ||| | ||| ||| | ||||
Db     145125 ATAAATATAAATATATAAATATATATAAATACATATATAAATATATATAAATATATATAT 145066

Qy        324 ATATACACATACATATGTACACATATATATACACATATGTATACACATATATACACATAT 383
              |||  |||||  |||||||  | |||||||  ||| ||||  |||  | | |||||  | | ||
Db     145065 TAATATACATAAATATGTATAAATATATAAACATATATTAATATATAAATATATATAAAT 145006

Qy        384 ATACACACATATATACACATATATACACACATATATACACATATATACACATATATACAC 443
              |||   | |||||||   |||||||  | | ||  |||  | |||  |||      ||||| ||
Db     145005 ATATTAATATATATATTTATATATAAATATATGAATATATATAAATATGTTAATATATAC 144946

Qy        444 ATATACACATATATACACATATATACATATATACACATATATATAATATACACACATATA 503
              |||||  | | |||  | | ||| | | |||||||  |  |||||||    ||| | | |||||
Db     144945 ATATATATAAATAAATATATAAAAATATATATAAATATATATAACTAATATATAAATATA 144886

Qy        504 TATACACATATATACACACATATATACACATATATACACATATATATACACATATATACA 563
              |||| ||  |||||  | | |||||||  | ||||||   ||  |||||||  | |||||||  |
Db     144885 TATA-ACTAATATATATAAATATATAAATATATATGAACTAATATATATAAATATATAAA 144827
```

```
Qy        564 CATATATACATATATACACATATATATACATATATACACATATATACATATATACACATA 623
              | ||||| | ||| | | ||| ||| | ||||| | | ||||||| ||| ||| | | |
Db     144826 TAAATATATAAATAAATATATAAATAAATATATACAAATATATATAAATAAATATATAAA 144767

Qy        624 TATACATATATACACACATATATACACATACATATACACACACATAGATATACATATATA 683
              ||| ||||| | | | ||||| | | ||| ||||| | | | ||| ||| | | |||||
Db     144766 AATAAATATAAATAAATATATAAATAAATATATATAAATAAATATATATAAATAAATATA 144707

Qy        684 TACACATATATATACGTATATATATGTATATATATAT 720
              || | ||| ||||| || |||||| |||| ||| ||
Db     144706 TATAAATAAATATATATAAATATATATATAAATAAAT 144670


RESULT 77
US-09-818-657-3
; Sequence 3, Application US/09818657
; Publication No. US20030077773A1
; GENERAL INFORMATION:
;  APPLICANT: BRANDON, Rhonda et al.
;  TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
;  TITLE OF INVENTION:  NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
;  TITLE OF INVENTION:  AND USES THEREOF
;  FILE REFERENCE: CL001006-CIP
;  CURRENT APPLICATION NUMBER: US/09/818,657
;  CURRENT FILING DATE:  2001-03-28
;  PRIOR APPLICATION NUMBER: 09/730,002
;  PRIOR FILING DATE: 2000-12-06
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 82938
;   TYPE: DNA
;   ORGANISM: Human
US-09-818-657-3

  Query Match             5.2%;  Score 258.4;  DB 10;  Length 82938;
  Best Local Similarity   70.5%;  Pred. No. 1.5e-42;
  Matches  361;  Conservative    0;  Mismatches  146;  Indels    5;  Gaps    1;

Qy        214 AAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACACATATATAT 273
              | | | | |    | |||| | | | |||| |||| |||||||||    | |||||| |||
Db       6946 ATACATATGTATATATATACACACACATATGTATACATACACACATGTATACATATGTAT 7005

Qy        274 ATACACATATATATACACATATATACATATATACACATATATACACATATATATACACAT 333
              |||   ||||| |||||||| |||||||| || | || |||| | ||| | ||  | |
Db       7006 ATATGTATATAGATACACATGTATACATATGTATATATGTATATAGATACACATGTATAC 7065

Qy        334 ACATATGTACACATATATATACACATATGTATACACATATATACACATATATACACACAT 393
              | || | || |   |||||| | ||| |||||||||| || |||| |  |||||| | |||
Db       7066 ATATGTATATATGTATATAGATACACATGTATACATATGTATATATGTATATAGATACAC 7125

Qy        394 ATATACACATATATACACACATATATACACATATATACACATATATACACATATACACAT 453
              || || |||||| || | |  |||||| | | | ||  | | ||||  | ||||  | ||
Db       7126 ATGTATACATATGTATATATGTATATAGATACACATGTATACATATGTATATATGTATAT 7185

Qy        454 ATATACACATATATACATATATACACATATATATA-----ATATACACACATATATATAC 508
              | |||||||| ||||||||| || | || ||||||     | ||  | |||||| ||||
Db       7186 AGATACACATGTATACATATGTATATATGTATATAGATACACATGTATACATATGTATAT 7245

Qy        509 ACATATATACACACATATATACACATATATACACATATATATACACATATATACACATAT 568
              |  |||||| | ||| || || |||||| || | || |||||||||  ||| || | || |
Db       7246 ATGTATATAGATACACATGTATACATATGTATATATGTATATACATGTATGTATATATGT 7305

Qy        569 ATACATATATACACATATATATACATATATACACATATATACATATATACACATATATAC 628
              |||  ||||||||  ||| ||||  | ||||   ||||||  |||| || | | || ||
Db       7306 ATATGTATATACATGTATGTATATGTGTATATGTATATATGTATATGTATATACATGTAT 7365

Qy        629 ATATATATACACACATATATACACATACATATACACACACATAGATATACATATATATACAC 688
              ||||||  |||  ||||||||  |  ||||| | |||  || |||   ||||| ||| |
Db       7366 GTATATATACAGCTATATACATGTGTATATATATACAGCTATATACGTATATACATATAT 7425

Qy        689 ATATATATACGTATATATATGTATATATATAT 720
```

```
              || |||||| ||||| |||||| |||||||||
Db       7426 ATGTATATATGTATACATATGTGTATATATAT 7457


RESULT 78
US-10-017-161-2361
; Sequence 2361, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
;  FILE REFERENCE: 084335/0152
;  CURRENT APPLICATION NUMBER: US/10/017,161
;  CURRENT FILING DATE:  2002-12-18
;  PRIOR APPLICATION NUMBER: JP 2001/246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2430
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2361
;   LENGTH: 1486
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: source
;   LOCATION: (1)..(1486)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(1286)
US-10-017-161-2361

  Query Match             5.2%;  Score 258;  DB 15;  Length 1486;
  Best Local Similarity   70.4%;  Pred. No. 2.9e-43;
  Matches  345;  Conservative    0;  Mismatches  145;  Indels    0;  Gaps    0;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | |||||||| || ||| | |||||| ||| |      |  | |||||||||| || | ||
Db        718 TGTATATATACATGTATGTGTATATATACATATGTGTATGTGTATATATACATATGTGTA 777

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              |    |||||||||||| |  |  | |||| | | |||| |  |  |  |||| |||| |
Db        778 TGTGTATATATACATATGTGTATGTGTATATATACATATGTGTATGTGTATATATACATA 837

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              | | |||| | | || | |  | |||| || | || |||| | |||||| |  | ||||
Db        838 TGTGTATATATACATATGTGTATATATGTATATATGTATATATACATATGTGTATATATG 897

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | |  |||||| ||||||||      ||||||   |||||    ||| || |||| |
Db        898 TATATATGTATATATACATATATGTGTGTATATATGTATATATGTGTATTTATACATGTG 957

Qy        467 TACATATATACACATATATATAATATACACACATATATATACACATATATACACACATAT 526
              |  | | || | |||| ||||    || |   | |||||| ||||||||  |    ||||
Db        958 CATGTGTGTATAAATATGTATATATTATATGTGTGTATATATACATATATGTATGTATAT 1017

Qy        527 ATACACATATATACACATATATATACACATATATACACATATATACATATATACACATAT 586
              ||||| ||||| | |||||||||| | | ||| | ||| |||   ||||||| ||| ||||
Db       1018 ATACATATATACATACATATATACATATATACACACATATACGCACATATAGACATATAT 1077

Qy        587 ATATACATATATACACATATATACATATATACACATATATACATATATACACACATATAT 646
              || || ||||| ||||||||||||||||||||    |||||||||  | ||| | ||||||||
Db       1078 ATGTATATATACACACATATATACATATATATGTATATATACACACATATATACATATAT 1137

Qy        647 ACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATAT 706
              |    ||| ||| ||||| | ||| |||||  ||||||||||||||||||||    ||||||
Db       1138 ATGTATATATACACACATATATACATATATGTATATATACACATATATATGTATATATAC 1197

Qy        707 ATGTATATAT 716
              |  |||||||
Db       1198 ACATATATAT 1207
```

RESULT 79
US-10-257-166-11
; Sequence 11, Application US/10257166
; Publication No. US20040023230A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Method and Nucleic Acids for Analysing the Methylation of
; TITLE OF INVENTION: Genes Implicated in Pharmacogenomics
; FILE REFERENCE: 5013.1011
; CURRENT APPLICATION NUMBER: US/10/257,166
; CURRENT FILING DATE: 2002-10-07
; PRIOR APPLICATION NUMBER: PCT/EP01/07470
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-06-29
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 178
; SEQ ID NO 11
; LENGTH: 12069
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (390, 398)
US-10-257-166-11

Query Match 5.2%; Score 257.6; DB 17; Length 12069;
Best Local Similarity 62.9%; Pred. No. 8.9e-43;
Matches 433; Conservative 0; Mismatches 249; Indels 6; Gaps 2;

Qy 226 TTTTATATATATATATATATATATATACACACACACACACATATATATATACACATATAT 285
Db 9239 TTATATATATATAGATATTTATATATATATAGATATACGTGGATATATATATTTATATAT 9298

Qy 286 ATACACATATATACATATATACACATATATACACATATATATACACATACATATGTACAC 345
Db 9299 ATATAGATATTTATATATATATAGATATACGTGGATATAGATATTTATATATATATAGAT 9358

Qy 346 ATATATATACACATATGTATACACATATATACACATATATACACACATATATACACATAT 405
Db 9359 ATTTATATATATATATAGATATTTATATATATATAGATATTTATATATATATAGATATAT 9418

Qy 406 ATACACACATATATACACATATATACACATATATACACATATACACATATATACACATAT 465
Db 9419 GTGGATATATATATTTATATATATAGAGATATTTATATATATATAGATATTTATATATAT 9478

Qy 466 ATACATATATACACATATATATA-----ATATACACACATATATATACACATATATACAC 520
Db 9479 ATAGATATACGTGGATATATATATTTATATATATATAGATATTTATATATATATAGATAT 9538

Qy 521 ACATATATACACATATATACACATATATATACACATATATACACATATATACATATATAC 580
Db 9539 ACGTGGATATATATATTTATACGTAGATATTTATATATATATAGATATACGTGGATATAT 9598

Qy 581 ACATATATATACATATATACACATATATACATATATACACATATATACATATATACACAC 640
Db 9599 ATATTTATACGTAGATATTTATATATATATAGATATACGTGGATATATATATTTATATAT 9658

Qy 641 ATATATACACATACATATACACACACATAGATATACATATATATACACATATATATACGT 700
Db 9659 ATATAGATATTTATATATATATAGATATACGTGGATATATATATTTATATATATATAGAT 9718

Qy 701 ATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGG 760
Db 9719 ATTTATATATATATAGATATACGTGGATATAGATATTTATATATATAGATATTTATAC 9778

Qy 761 GGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATG 820
Db 9779 GTATATAGATATTTATATATATATAGATATTTATATATATATAGATATATTTATATATAT 9838

Qy 821 GCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAA 880
Db 9839 ATAGATATATATGGATATATATATTTATAT-ATATATAATATTTATATATAGATATATAT 9897

Qy 881 GGTAACACTATATATATATATAGTGAAA 908
Db 9898 GGATATAGTATTTATATATATATAGATA 9925

RESULT 80
US-10-292-798-2005
; Sequence 2005, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
; FILE REFERENCE: 084335/166
; CURRENT APPLICATION NUMBER: US/10/292,798
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: 10/017,161
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: JP 2001-246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2070
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2005
; LENGTH: 1486
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; LOCATION: source
; FEATURE:
; LOCATION: (1)..(1486)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (201)..(1286)
US-10-292-798-2005

Query Match 5.2%; Score 256.4; DB 17; Length 1486;
Best Local Similarity 70.2%; Pred. No. 6.2e-43;
Matches 344; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

Qy 227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
Db 718 TGTATATATACATGTATGTGTATATATACATATGTGTATGTGTATATATACATATGTGTA 777

Qy 287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
Db 778 TGTGTATATATACATATGTGTATGTGTATATATACATATGTGTATGTGTATATATACATA 837

Qy 347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
Db 838 TGTGTATATATACATATGTGTATATATGTATATATGTATATATACATATGTGTATATATG 897

Qy 407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
Db 898 TATATATGTATATATACATATATGTGTGTATATATGTATATATGTGTATTTATATATGTG 957

Qy 467 TACATATATACACATATATATAATATACACACATATATATACACATATATACACACATAT 526
Db 958 CATGTGTGTATAAATATGTATATATTATATGTGTGTATATATACATATATGTATGTATAT 1017

Qy 527 ATACACATATATACACATATATATACACATATATACACATATATACATATATACACATAT 586
Db 1018 ATACATATATACATACATATATACATATATACACACATATACGCACATATAGACATATAT 1077

Qy 587 ATATACATATATACACATATATACATATATACACATATATACATATATACACACATATAT 646

```
              || || ||||| ||||||||||||||||||||   ||||||||| | ||| | ||||||||
Db       1078 ATGTATATATACACACATATATACATATATATGTATATATACACACATATATACATATAT 1137

Qy        647 ACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATAT 706
              |   ||| ||| ||||| | ||| |||||  ||||||||||||||||||   ||||||
Db       1138 ATGTATATATACACACATATATACATATATGTATATATACACATATATATGTATATATAC 1197

Qy        707 ATGTATATAT 716
              |  |||||||
Db       1198 ACATATATAT 1207


RESULT 81
US-10-437-963-46068
; Sequence 46068, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 46068
;   LENGTH: 610
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_48971C.1
US-10-437-963-46068

  Query Match             5.1%;  Score 255.8;  DB 18;  Length 610;
  Best Local Similarity   69.6%;  Pred. No. 5.5e-43;
  Matches  364;  Conservative    0;  Mismatches  152;  Indels    7;  Gaps    1;

Qy        197 AAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACA 256
              ||||  || |  | | |   || |||  | || |||  ||  ||   || ||||||   |
Db         39 AAATAACACCGAAAATATTTAATAAGAAAATTAATACGTAGGTACGGATGTATATATGAA 98

Qy        257 CACACACACATATATATATACACATATATATACACATATATACATATATACACATATATA 316
              |   ||  | |||||||   || ||||||    ||| || || ||||   || |
Db         99 AGTATGGACGGAAATATATATGTATGTATATAATAGTATGTATATGTATATCTATGTGAG 158

Qy        317 CACATATATATACACATACATATGTACACATATATATACACATATGTATACACATATATA 376
              |||||||||   ||| |||| || | ||||||||| | |||| |||| | |||||||
Db        159 TGTATATATATATGTATATATATATATATATATATATATATATATATATATATATATATA 218

Qy        377 CACATATATACACACATATATACACATATATACACACATATATACACATATATACACATA 436
              | |||||||| | | |||||||| | |||||||| | |  |||  |  | |||||||| | |||
Db        219 TATATATATATATATATATATATATATATATATATATATGTATGTGTATATATATATATATA 278

Qy        437 TATACACATATACACATATATACACATATATACACATATATACACATATATATAATATACAC 496
              ||||    | |  | ||||||| | ||||||| |||||||| | ||||||||       |
Db        279 TATATGTGTGTGTATATATATATATATATATATATATATATATATATATATAT-------AT 331

Qy        497 ACATATATATACACATATATACACACATATATACACATATATACACATATATATACACAT 556
              | |||||||| ||| | | ||||||| || || | |||||||||| | || || | ||
Db        332 ATATATATATATACACACACACACACACATGTATATATATATACACACACATGTATATAT 391

Qy        557 ATATACACATATATACATATATACACATATATATACATATATACACATATATACATATAT 616
              |||||||||| || | ||| ||| |||||||||| ||||||| ||| || |||||||||
Db        392 ATATACACATGTACATATACATATACATATATATATATATATATATACACATGTACATATAT 451

Qy        617 ACACATATATACATATATACACACATATATACACATACATATACACACACATAGATATAC 676
              |||  || ||| |||||||||||||||||||| ||| | ||||| | | | | | |||
Db        452 ACATGTACATATATATATACACACATATATATACACATATATATATATATACACACATAT 511

Qy        677 ATATATATACACATATATATACGTATATATATGTATATATATA 719
              ||||||||| ||||    |  | | | || |  || || ||||
Db        512 ATATATATATACATGACGACTCTTCTGTACACCTAGATGTATA 554


RESULT 82
US-10-017-161-2363/c
; Sequence 2363, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
;  FILE REFERENCE: 084335/0152
;  CURRENT APPLICATION NUMBER: US/10/017,161
;  CURRENT FILING DATE:  2002-12-18
;  PRIOR APPLICATION NUMBER: JP 2001/246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2430
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2363
;   LENGTH: 1489
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: source
;   LOCATION: (1)..(1489)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(1289)
US-10-017-161-2363

  Query Match             5.1%;  Score 255.4;  DB 15;  Length 1489;
  Best Local Similarity   63.8%;  Pred. No. 1e-42;
  Matches  438;  Conservative    0;  Mismatches  241;  Indels    8;  Gaps    3;

Qy        236 ATATATATATATATATACACACACACACACATATATATATACACATATATATACACATAT 295
              | || | |||||||||||| | | ||  | |||| | |||| | | ||| |||   ||||
Db       1197 ACATGTGTATATATATACATATATACGTATATATGTGTATATATACATACATAGGTATAT 1138

Qy        296 ATACATATATACACATATATACACATATATATACACATACATATGTACACATATATATAC 355
              | | ||||| ||| | ||||| || |||||||  | |||  ||| || |||| ||||||
Db       1137 ACATATATACACATACATATATACGTATATATGTATATATGTATATATACATGTATATAT 1078

Qy        356 ACATATG-TATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              |||||| | || |||||||||||| |||||| ||  |||| || ||||||||||||  |
Db       1077 GTATATGTTGTATACATATATACACGTATATATACGTATATGTATACATATATACACGTA 1018

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              |||||||  |||| || |||||||||||| |||| | |  | || | | || || |||||
Db       1017 TATATACGTATATGTATACATATATACACGTATATATACGTGTATATACATGTATATATA 958

Qy        475 TACACATATATATAATATACACACATATATATACACATATATACACACATATATACACAT 534
              | || |     ||| | | |||||||||| ||   |||||||| |  |||||||| |
Db        957 CGTATATGT-----ATACATATACATATATACACGTGTATATACATATGTATATACACGT 903

Qy        535 ATATACACATATATATACACATATATACACATATATACATATATACACATATATATACAT 594
              |||| |  ||| |||| |||  | || | || ||||| |||||  |||  | |||| ||
Db        902 GTATATATGTATGTATATACACGTGTATATATGTATACGTATATATACACGTGTATATAT 843

Qy        595 ATATACACATATATACATATATACACATATATACATATATACACACATATATACACATAC 654
              |||||  |||||||||  | || | || |||||  ||||||  |||| | |||| | ||
Db        842 GTATACGTATATATACACGTGTATATATGTATACGTATATATACACGTGTATATATGTAT 783

Qy        655 ATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATAT 714
              |  || | | |||    |||| |||| |||||  ||||||| |||| ||||||||||||
Db        782 ACGTATATATACACGTGTATATATATGTATACGTATATATACACGTGTATATATGTATAC 723

Qy        715 ATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAG 774
```

```
Db        722 GTATATATACACGTGTATATATGTATGTATATATACGTGTATATATATGTATACGTATAT 663

Qy        775 AGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGG 834
              | |  |  | |  | |||||     |     | | |     |     | ||| |
Db        662 ATACGTATATATAAATGTAT--ATATGTATATGTGTATATATACGTATATATACATGTAT 605

Qy        835 AAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATA 894
              | |||| || || |  |   | || ||| ||  | |||  ||   |  |    |||||
Db        604 ATATACGTATATATACACATATACATATATGGTACGTATATATACATATGTGTATATATG 545

Qy        895 TATATATAGTGAAAGCAGTGTAAACCT 921
              ||||||||  | |    | || || |
Db        544 TATATATATACATACATATATATACAT 518


RESULT 83
US-10-292-798-2007/c
; Sequence 2007, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
;  FILE REFERENCE: 084335/166
;  CURRENT APPLICATION NUMBER: US/10/292,798
;  CURRENT FILING DATE:  2002-11-13
;  PRIOR APPLICATION NUMBER: 10/017,161
;  PRIOR FILING DATE: 2001-12-18
;  PRIOR APPLICATION NUMBER: JP 2001-246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2070
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2007
;   LENGTH: 1489
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   LOCATION: source
;   FEATURE:
;   LOCATION: (1)..(1489)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(1289)
US-10-292-798-2007

  Query Match             5.1%;  Score 255.4;  DB 17;  Length 1489;
  Best Local Similarity   63.8%;  Pred. No. 1e-42;
  Matches  438;  Conservative    0;  Mismatches  241;  Indels    8;  Gaps    3;

Qy        236 ATATATATATATATATACACACACACACATATATATATACACATATATATACACATAT 295
              | || | |||||||||||| | | ||  | |||| | |||| | | ||| |||   ||||
Db       1197 ACATGTGTATATATATACATATATACGTATATATGTGTATATATACATACATAGGTATAT 1138

Qy        296 ATACATATATACACATATATACACATATATATACACATACATATGTACACATATATATAC 355
              | | ||||| ||| | ||||| || |||||||  | |||  ||| || |||| ||||||
Db       1137 ACATATATACACATACATATATACGTATATATGTATATATGTATATATACATGTATATAT 1078

Qy        356 ACATATG-TATACACATATATACACATATATACACACATATATACACATATATACACACA 414
                ||||| | || |||||||||||| |||||| ||  |||| || ||||||||||||  |
Db       1077 GTATATGTTGTATACATATATACACGTATATATACGTATATGTATACATATATACACGTA 1018

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              |||||||  |||| || ||||||||||||| |||| | |  | || | | || || |||||
Db       1017 TATATACGTATATGTATACATATATACACGTATATATACGTGTATATACATGTATATATA 958

Qy        475 TACACATATATATAATATACACACATATATATACACATATATACACACATATATACACAT 534
                | || |     ||| | | |||||||||| ||   |||||||| |  |||||||||  |
Db        957 CGTATATGT-----ATACATATACATATATACACGTGTATATACATATGTATATACACGT 903

Qy        535 ATATACACATATATATACACATATATACACATATATACATATATACACATATATATACAT 594
              |||||  | |||| | |||| |||  | || | || ||||| |||||| |||  | |||| ||
Db        902 GTATATATGTATGTATATACACGTGTATATATGTATACGTATATATACACGTGTATATAT 843

Qy        595 ATATACACATATATACATATATACACATATATACATATATACACACATATATACACATAC 654
              |||||  |||||||||  | || | || ||||| |||||| |||| | |||| |  ||
Db        842 GTATACGTATATATACACGTGTATATATGTATACGTATATATACACGTGTATATATGTAT 783

Qy        655 ATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATAT 714
              |  || | | |||    |||| |||| |||||  ||||||| |||| ||||||||||||
Db        782 ACGTATATATACACGTGTATATATATGTATACGTATATATACACGTGTATATATGTATAC 723

Qy        715 ATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAG 774
              |||||   |  | ||  ||| |         | |   |      |     || | | | |
Db        722 GTATATATACACGTGTATATATGTATGTATATATACGTGTATATATATGTATACGTATAT 663

Qy        775 AGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGG 834
              | |  |  | |  | |||||     |     | | |     |     | ||| |
Db        662 ATACGTATATATAAATGTAT--ATATGTATATGTGTATATATACGTATATATACATGTAT 605

Qy        835 AAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATA 894
              | |||| || || |  |   | || ||| ||  | |||  ||   |  |    |||||
Db        604 ATATACGTATATATACACATATACATATATGGTACGTATATATACATATGTGTATATATG 545

Qy        895 TATATATAGTGAAAGCAGTGTAAACCT 921
              ||||||||  | |    | || || |
Db        544 TATATATATACATACATATATATACAT 518


RESULT 84
US-10-322-281-123/c
; Sequence 123, Application US/10322281
; Publication No. US20040126762A1
; GENERAL INFORMATION:
;  APPLICANT: David W. Morris
;  APPLICANT:  Marc S. Malandro
;  TITLE OF INVENTION: Novel Compositions and Methods in Cancer
;  FILE REFERENCE: 529452001000
;  CURRENT APPLICATION NUMBER: US/10/322,281
;  CURRENT FILING DATE:  2002-12-17
;  NUMBER OF SEQ ID NOS: 866
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 123
;   LENGTH: 104399
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(104399)
;   OTHER INFORMATION: n = A,T,C or G
US-10-322-281-123

  Query Match             5.1%;  Score 255.4;  DB 18;  Length 104399;
  Best Local Similarity   73.3%;  Pred. No. 6.8e-42;
  Matches  367;  Conservative    0;  Mismatches  131;  Indels    3;  Gaps    3;

Qy        230 ATATATATATATATATATATATACACACACACACATATATATATACACATATATATAC 289
              ||| ||| ||| ||||||| |||| ||| | | ||||| ||| |||| | ||||||||
Db      13721 ATACATACATACATATATACATACGTACATATATACATACATACATACGTACATATATAC 13662

Qy        290 ACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATAT 349
              | |  || | ||||| | ||||| ||||| ||||| ||||  | |  |||| | |||||
Db      13661 ATACGTACATATATACATACATACATACATATATACATACGTACATATATACATACATAC 13602

Qy        350 ATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATAC 409
              |||   |||||| || | || || ||| | | ||| | ||| | ||||| |||||  |||
Db      13601 ATACGTACATATATACATACGTACATATATACATACATACATACATATATACATACGTAC 13542

Qy        410 ACACATATATACACATATATACACATATATACACATATACACATATATACACATATATAC 469
              | | ||| ||||| | |||   ||||||||||| | ||||| | ||||| |||| ||||
Db      13541 ATATATACATACATACATACGTACATATATACATACATACATACATATATACATACATAC 13482

Qy        470 ATATATACACATATATATAATATACACACATATATATACACATATATACACACATATATA 529
              ||||||||| | ||| |||||| | | | | ||  ||| | | ||| | ||| |||||
```

```
Db      13481 ATATATACATACATACATAATACATATATACATTCATATATACATA-ATACATATATACA 13423

Qy        530 CACATATATACACATATATAT-ACACATATATACACATAT-ATACATATATACACATATA 587
              |||||||||||| | ||| || | ||||||||||| | || ||||||| || || |||||
Db      13422 TACATATATACATACATACATAATACATATATACATACATAATACATACATGCATATATA 13363

Qy        588 TATACATATATACACATATATACATATATACACATATATACATATATACACACATATATA 647
              || |||| ||| | | ||| ||| | ||||| | ||| | | ||||| ||| |  || |
Db      13362 CATGCATACATATATACATACACACATACATACATACATACATATATACATACGTACA 13303

Qy        648 CACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATA 707
              | ||||||| | ||| || || ||||| | ||| ||||| ||||| |||| || |||||
Db      13302 TATATACATACATACATACGTACATATATACATACATACATATATACATACATACATATA 13243

Qy        708 TGTATATATATATGCTCCAGA 728
              |  ||| ||| ||  | || |
Db      13242 TACATACATACATAATACATA 13222


RESULT 85
US-10-292-798-607/c
; Sequence 607, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
;  FILE REFERENCE: 084335/166
;  CURRENT APPLICATION NUMBER: US/10/292,798
;  CURRENT FILING DATE:  2002-11-13
;  PRIOR APPLICATION NUMBER: 10/017,161
;  PRIOR FILING DATE: 2001-12-18
;  PRIOR APPLICATION NUMBER: JP 2001-246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2070
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 607
;   LENGTH: 84510
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   LOCATION: source
;   FEATURE:
;   LOCATION: (1)..(84510)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(272)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (29879)..(29999)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (45509)..(45616)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (50871)..(50974)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (58541)..(58638)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (60172)..(60304)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (62427)..(62580)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (71080)..(71140)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (72348)..(72417)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (75647)..(75738)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (77542)..(77668)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (82828)..(82869)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (84170)..(84310)
US-10-292-798-607

  Query Match             5.1%;  Score 255;  DB 17;  Length 84510;
  Best Local Similarity   72.6%;  Pred. No. 7.5e-42;
  Matches  358;  Conservative    0;  Mismatches  130;  Indels    5;  Gaps    2;

Qy        232 ATATATATATATATATATACACACACACACATATATATATACACATATATATACAC 291
              ||||||||| ||||   |||| | | | ||||| || |||||||||| ||  ||||| ||
Db      80292 ATATATATACATATTCTTATATATATATACACATATGTATATATACATATGAATATATAC 80233

Qy        292 ATATATACATATATACACATATATACACATATATATACACATACATATGTACACATATAT 351
              | ||||  |||||||| ||  ||| | | | ||||||| ||  |||| ||||||||||
Db      80232 ACATATGTATATATACATATGAATATATACACATATACATATGAATATATACACATATAC 80173

Qy        352 ATACACATATGTATACACATATATACACATATATACACACATATATACACATATATACAC 411
              |||   |||| || | | ||| |||   |||||||||| |||||    |   ||||| ||
Db      80172 ATATGAATATATACATATATACATATGAATATATACACATATATACGTATGAATATATAC 80113

Qy        412 ACATATATAC----ACATATATACACATATATACACATATACACATATATACACATATAT 467
              ||||||||||      ||||||| ||| ||||| || |||  | ||||||||||||||||
Db      80112 ACATATATACGTATGAATATATATACACATATATACGTATGAATATATATACACATATAT 80053

Qy        468 ACATATATACACATATATATAATATACACACATATATATACACATATATACACACATATA 527
              || |||  | | ||||| |  ||||||  |   ||||||| ||  ||||||  |   |||
Db      80052 ACGTATGAATATATATACA-CATATACGTATGAATATATATACCCATATACGTATGAATA 79994

Qy        528 TACACATATATACACATATATATACACATATATACACATATATACATATATACACATATA 587
              |||||||||||| || ||  |||| | ||| | | | |||  ||| |||||| | | | |
Db      79993 TACACATATATATACGTACGTATATATATACACATATATACGTACGTATATATATACACA 79934

Qy        588 TATACATATATACACATATATACATATATACACATATATACATATATACACACATATATA 647
              |||| |  ||   | |||||||||| ||||| || ||   | |||||||||| |||||
Db      79933 TATATACGTACGTATATATATACACATATATACGTACGAATATATATACACATATATACG 79874

Qy        648 CACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATA 707
              ||  | ||||| ||||| ||| |  |||  ||||||| ||| ||||||||||   |||
Db      79873 TACGAATATATATACACATATATACGTACGAATATATATACACATATATACGTACGAATA 79814

Qy        708 TGTATATATATAT 720
              | |||| | ||||
Db      79813 TATATACACATAT 79801


RESULT 86
US-10-017-161-2375
; Sequence 2375, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
;  FILE REFERENCE: 084335/0152
;  CURRENT APPLICATION NUMBER: US/10/017,161
;  CURRENT FILING DATE:  2002-12-18
;  PRIOR APPLICATION NUMBER: JP 2001/246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2430
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2375
```

```
;   LENGTH: 1114
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: source
;   LOCATION: (1)..(1114)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(914)
US-10-017-161-2375

  Query Match             5.1%;  Score 254.2;  DB 15;  Length 1114;
  Best Local Similarity   64.0%;  Pred. No. 1.6e-42;
  Matches  399;  Conservative    0;  Mismatches  223;  Indels    1;  Gaps    1;

Qy        229 TATATATATATATATATATATATACACACACACACACATATATATATACACATATATATA 288
              ||||| ||||| | ||| |||||||| | | | ||| |||| |||||     || ||||||
Db        212 TATATGTATATTTGTATGTATATACAAATATATACATATATGTATATGTGTATGTATATA 271

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              || ||||||||||||||    || |||| ||| |||||||| ||| | | |  | |||
Db        272 CAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATATA 331

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              | ||| | | || |||   || |||| ||| ||||| ||| |||| |  |  ||||||
Db        332 CAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATATA 391

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              || | ||||| | | || | ||  |||| | | ||||||| |||| |  |  ||||||
Db        392 CAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATATA 451

Qy        469 CATATATACACATATATATAATATACACACATATATATACACATATATACACACATATAT 528
              || ||||| |||||||| |  |||  | | | ||||| ||||||||    |  |||||
Db        452 CAAATATATACATATATGT-GTATGTATATACAAATATATACATATATGTGTATGTATAT 510

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACACATATAT 588
              ||| |||||||| |||| | ||  ||||| | | ||||| | |||| |  |  |||||
Db        511 ACAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATAT 570

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
              | | |||||||| |||| |  || |||| ||| ||||| | |||||     || ||||
Db        571 ACAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATAT 630

Qy        649 ACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATAT 708
              ||| | |||||| | |  |  || || ||| | ||| | | |||||| |||| |||||
Db        631 ACAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATAT 690

Qy        709 GTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAG 768
              | |||||||       | ||  ||       |  |  ||    ||    || | |
Db        691 ACAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATAT 750

Qy        769 GAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAG 828
              ||| | |     | |  || | |    | || | ||     ||      |||
Db        751 ACAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATAT 810

Qy        829 TAGGGGAAATACATAAATGTGTA 851
                  | ||||||| |||||||
Db        811 ACAAATATATACATATATGTGTA 833


RESULT 87
US-10-292-798-2017
; Sequence 2017, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
;  FILE REFERENCE: 084335/166
;  CURRENT APPLICATION NUMBER: US/10/292,798
;  CURRENT FILING DATE:  2002-11-13
;  PRIOR APPLICATION NUMBER: 10/017,161
;  PRIOR FILING DATE: 2001-12-18
;  PRIOR APPLICATION NUMBER: JP 2001-246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2070
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2017
;   LENGTH: 1114
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   LOCATION: source
;   FEATURE:
;   LOCATION: (1)..(1114)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(914)
US-10-292-798-2017

  Query Match             5.1%;  Score 254.2;  DB 17;  Length 1114;
  Best Local Similarity   64.0%;  Pred. No. 1.6e-42;
  Matches  399;  Conservative    0;  Mismatches  223;  Indels    1;  Gaps    1;

Qy        229 TATATATATATATATATATATATACACACACACACACATATATATATACACATATATATA 288
              ||||| ||||| | ||| |||||||| | | | ||| |||| |||||     || ||||||
Db        212 TATATGTATATTTGTATGTATATACAAATATATACATATATGTATATGTGTATGTATATA 271

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              || ||||||||||||||    || |||| ||| |||||||| ||| | | |  | |||
Db        272 CAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATATA 331

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              | ||| | | || |||   || |||| ||| ||||| ||| |||| |  |  ||||||
Db        332 CAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATATA 391

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              || | ||||| | | || | ||  |||| | | ||||||| |||| |  |  ||||||
Db        392 CAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATATA 451

Qy        469 CATATATACACATATATATAATATACACACATATATATACACATATATACACACATATAT 528
              || ||||| |||||||| |  |||  | | | ||||| ||||||||    |  |||||
Db        452 CAAATATATACATATATGT-GTATGTATATACAAATATATACATATATGTGTATGTATAT 510

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACACATATAT 588
              ||| |||||||| |||| | ||  ||||| | | ||||| | |||| |  |  |||||
Db        511 ACAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATAT 570

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
              | | |||||||| |||| |  || |||| ||| ||||| | |||||     || ||||
Db        571 ACAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATAT 630

Qy        649 ACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATAT 708
              ||| | |||||| | |  |  || || ||| | ||| | | |||||| |||| |||||
Db        631 ACAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATAT 690

Qy        709 GTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAG 768
              | |||||||       | ||  ||       |  |  ||    ||    || | |
Db        691 ACAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATAT 750

Qy        769 GAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAG 828
              ||| | |     | |  || | |    | || | ||     ||      |||
Db        751 ACAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATAT 810

Qy        829 TAGGGGAAATACATAAATGTGTA 851
                  | ||||||| |||||||
Db        811 ACAAATATATACATATATGTGTA 833


RESULT 88
US-09-728-552-3/c
; Sequence 3, Application US/09728552
; Publication No. US20030096398A1
; GENERAL INFORMATION:
```

```
;  APPLICANT: Choo, Kong-Hong Andy
;  APPLICANT:  Du Sart, Desiree
;  APPLICANT:  Cancilla, Michael R.
;  TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
;  FILE REFERENCE: Davies Col
;  CURRENT APPLICATION NUMBER: US/09/728,552
;  CURRENT FILING DATE:  2000-12-02
;  PRIOR APPLICATION NUMBER: 09/078,294
;  PRIOR FILING DATE: 1998-05-13
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 80595
;   TYPE: DNA
;   ORGANISM: Nucleotide sequence of HC-contig
US-09-728-552-3

  Query Match             5.1%;  Score 253.6;  DB 10;  Length 80595;
  Best Local Similarity   71.9%;  Pred. No. 1.4e-41;
  Matches  361;  Conservative    0;  Mismatches  134;  Indels    7;  Gaps    2;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | |||| |||| ||| | |||||| ||| | | ||| |  ||| | |||| ||| |||||
Db      16026 TATATACATATTTATGTCTATATACACATATATACATATTTATGTCTATATACACATATA 15967

Qy        287 TACACATATATACATATATACACATATATACACA------TATATATACACATACATATG 340
              |||| || |||   |||||||||||||||||| |      | ||||||||||| |||
Db      15966 TACATATTTATGTCTATATACACATATATACATATTTATGTCTATATACACATATATACA 15907

Qy        341 TACACATATATATACACATATGTATACACATATATACACATATATACACACATATATACA 400
              ||   || | |||| ||| || |||||| || |||    ||||| ||| | ||| |||
Db      15906 TATTTATGTCTATATACACATATATACATATTTATGTCTATATACACATATATACATATT 15847

Qy        401 CATATATACACACATATATACACATATATACACATATATACACATATACACATATATACA 460
              || | || | ||| ||||| |||||| ||    ||||||||||||| |||||| ||
Db      15846 TATGTCTATATACACATATATACATATTTATGTCTATATACACATATATACATATTTATG 15787

Qy        461 CATATATACATATATACACATATATAT-AATATACACACATATATATACACATATATACA 519
                |||||||| ||||| |||||| |||    |||| ||||||||||| | |  ||| |  |
Db      15786 TCTATATACACATATATACATATTTATGTCTATATACACATATATACATATTTATGTCTA 15727

Qy        520 CACATATATACACATATATACACATATATATACACATATATACACATATATACATATATA 579
              | |  | ||| | | ||||  |  | ||||||||||||||||||| || |||   ||||||
Db      15726 TATACACATATATACATATTTATGTCTATATACACATATATACATATTTATGTCTATATA 15667

Qy        580 CACATATATATACATATATACACATATATACATATATACACATATATACATATATACACA 639
              |||||||||| | || |||  | ||||| |||||||||| || |||   ||||||||||
Db      15666 CACATATATACATATTTATGTATATATACACATATATACATATTTATGTCTATATACACA 15607

Qy        640 CATATATACACATACATATACACACACATAGATATACATATATATACACATATATATACG 699
              |||||  |  |  ||  |||| |   |  || ||||   | |||||  ||||| |||||
Db      15606 TATATACATATTTATGTATATATTTATGTATATATGTGTGTATATGTACATACATATATA 15547

Qy        700 TATATATATGTATATATATATG 721
              |||  |||  || |||||||||
Db      15546 CATACATAAATACATATATATG 15525


RESULT 89
US-10-856-888-1/c
; Sequence 1, Application US/10856888
; Publication No. US20040235037A1
; GENERAL INFORMATION:
;  APPLICANT: Blumenfeld, Marta
;  APPLICANT:  Bougueleret, Lydie
;  APPLICANT:  Chumakov, Ilya
;  TITLE OF INVENTION: Polymorphic Markers of Prostate Carcinoma Tumor Antigen-1 (PCTA-1
;  FILE REFERENCE: GEN-T112XC1
;  CURRENT APPLICATION NUMBER: US/10/856,888
;  CURRENT FILING DATE:  2004-05-27
;  PRIOR APPLICATION NUMBER: US/09/326,402
;  PRIOR FILING DATE: 1999-06-04
;  PRIOR APPLICATION NUMBER: 60/088,187
;  PRIOR FILING DATE: 1998-06-05
;  PRIOR APPLICATION NUMBER: 60/102,324
;  PRIOR FILING DATE: 1998-09-28
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 106746
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 1..68647
;   OTHER INFORMATION: 5'regulation  region
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 66647..68647
;   OTHER INFORMATION: promoter
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 97156..106746
;   OTHER INFORMATION: 3'regulation  region
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 68648..68741
;   OTHER INFORMATION: exon0
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 70647..70794
;   OTHER INFORMATION: exon1
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 82208..82296
;   OTHER INFORMATION: exon2
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 83613..83823
;   OTHER INFORMATION: exon3
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 85298..85417
;   OTHER INFORMATION: exon4
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 86389..86445
;   OTHER INFORMATION: exon5
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 87496..87522
;   OTHER INFORMATION: exon6
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 87650..87775
;   OTHER INFORMATION: exon6bis
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 88295..88383
;   OTHER INFORMATION: exon7
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 89484..89649
;   OTHER INFORMATION: exon8
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 92749..97155
;   OTHER INFORMATION: exon9
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 92749..92883
;   OTHER INFORMATION: exon9bis
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 95821..97155
;   OTHER INFORMATION: exon9ter
```

; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 70647..70794
; OTHER INFORMATION: homology with genset EST : A241850
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 68648..68741
; OTHER INFORMATION: homology with genset EST : A241850
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 82208..82229
; OTHER INFORMATION: homology with genset EST : A241850
; FEATURE:
; NAME/KEY: allele
; LOCATION: 278
; OTHER INFORMATION: 99-1601-278 : polymorphic base A or C
; FEATURE:
; NAME/KEY: allele
; LOCATION: 402
; OTHER INFORMATION: 99-1601-402 : polymorphic base A or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 472
; OTHER INFORMATION: 99-1601-472 : polymorphic base A or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 2955
; OTHER INFORMATION: 99-13801-100 : polymorphic base T or C
; FEATURE:
; NAME/KEY: allele
; LOCATION: 12167
; OTHER INFORMATION: 99-13806-166 : polymorphic base G or A
; FEATURE:
; NAME/KEY: allele
; LOCATION: 12536
; OTHER INFORMATION: 99-13799-376 : polymorphic base T or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 17593
; OTHER INFORMATION: 99-13798-297 : polymorphic base T or C
; FEATURE:
; NAME/KEY: allele
; LOCATION: 17606
; OTHER INFORMATION: 99-13798-284 : polymorphic base T or C
; FEATURE:
; NAME/KEY: allele
; LOCATION: 22079
; OTHER INFORMATION: 99-1602-200 : polymorphic base G or C
; FEATURE:
; NAME/KEY: allele
; LOCATION: 28964
; OTHER INFORMATION: 99-13794-186 : polymorphic base T or C
; FEATURE:
; NAME/KEY: allele
; LOCATION: 29003
; OTHER INFORMATION: 99-13794-147 : polymorphic base C or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 31077
; OTHER INFORMATION: 99-13812-384 : polymorphic base T or C
; FEATURE:
; NAME/KEY: allele
; LOCATION: 31766
; OTHER INFORMATION: 99-13805-313 : polymorphic base T or C
; FEATURE:
; NAME/KEY: allele
; LOCATION: 34791
; OTHER INFORMATION: 99-1587-281 : polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 45751
; OTHER INFORMATION: 99-1582-430 : polymorphic base C or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 49847
; OTHER INFORMATION: 99-1585-465 : polymorphic base T or C
; FEATURE:
; NAME/KEY: allele
; LOCATION: 49855
; OTHER INFORMATION: 99-1585-457 : polymorphic base T or C
; FEATURE:
; NAME/KEY: allele
; LOCATION: 49886
; OTHER INFORMATION: 99-1585-426 : polymorphic base G or A
; FEATURE:
; NAME/KEY: allele
; LOCATION: 49900
; OTHER INFORMATION: 99-1585-412 : polymorphic base G or A
; FEATURE:
; NAME/KEY: allele
; LOCATION: 49906
; OTHER INFORMATION: 99-1585-406 : polymorphic base C or A
; FEATURE:
; NAME/KEY: allele
; LOCATION: 49921
; OTHER INFORMATION: 99-1585-391 : polymorphic base C or A
; FEATURE:
; NAME/KEY: allele
; LOCATION: 49939
; OTHER INFORMATION: 99-1585-373 : polymorphic base G or A
; FEATURE:
; NAME/KEY: allele
; LOCATION: 50256
; OTHER INFORMATION: 99-1585-55 : polymorphic base C or A
; FEATURE:
; NAME/KEY: allele
; LOCATION: 54955
; OTHER INFORMATION: 99-1607-373 : polymorphic base T or C
; FEATURE:
; NAME/KEY: allele
; LOCATION: 64239
; OTHER INFORMATION: 99-1577-105 : polymorphic base A or G
; FEATURE:
; NAME/KEY: allele

Query Match 5.1%; Score 253; DB 18; Length 106746;
Best Local Similarity 69.4%; Pred. No. 2.2e-41;
Matches 416; Conservative 0; Mismatches 175; Indels 8; Gaps 5;

```
Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              |||||||||||||||||  |||||||||||||||||| |||||||||||||||  ||||||
Db      38717 AGTTCAAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAAAATT 38658

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   |||||||||| | ||||||||||||||||||||||   |||| | ||||||||||||
Db      38657 AGCTGGGCATGGTGGTGCACGCCTGTAATCCCAGCTACTCAGGAGACTGAGGCAGGAGAA 38598

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCCA 178
              ||||||||||||  || ||||||| ||| || ||| ||| |   | ||| |||||||||
Db      38597 TCACTTGAACCTG--GGAGGTGGAGTTTGTGGTGAGCCAAAACTTTCCCACTGCACTCCA 38540

Qy        179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCAT-TTTATATATAT 237
              |||      |   | |   | || |   | |  | |      | |   ||    | ||||
Db      38539 GCCAGTATATATATATATATATACACACGTACACGTGTACACATATACACACATATGC 38480

Qy        238 ATATATATATATATACACACACACACATATATATATACACATATATATACACATATAT 297
              | |  | || ||| ||||||| |||||    | ||| || ||||||| || |  | || ||
Db      38479 ACACGTGTACATACACACACATACACATGCACATACATGCACATATGTACATGCGTACAT 38420

Qy        298 ACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACAC 357
              ||| | ||  |||||||||||||| || || | |  ||||    |||||| || || | |
Db      38419 ACACACATGTACATATATACACACATGTACATATGTACACCCATACACACATGTACATAT 38360

Qy        358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417
              ||      | ||| | ||  |||||| ||||| |||||| | |  || |||   |||
Db      38359 GTACACCCATACACACATGTACATATGTACACCCATATACATATGTACATATGTACACCC 38300
```

```
Qy        418 ATACACATATATACACATATATACACATATACACATATATACACATATATACATATATAC 477
              |||||||||| |||| || || || |||| |||  || |||   | |   |.| | ||
Db      38299 ATACACATATGTACATATGTACACCCATACACATGTACATATGTACACCCATACACATGT 38240

Qy        478 ACATATATATAATATACACACATATATATACACATATATACACACATATATACACATATA 537
              |||||| |   | || |||| || || |||   ||||  | |||||||| |||| |||||
Db      38239 ACATATGT---ACATCCACATATGTACATATGTATATCCATACACATATGTACATATATA 38183

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATACATAT 596
              |||||||| || | |  || ||| | || | |  |||| ||   |||||||||   ||
Db      38182 CACACATATGTACATATGTACATATATATTTGTGTATATCTATGTATATATATAAGAAT 38124


RESULT 90
US-10-856-888-12/c
; Sequence 12, Application US/10856888
; Publication No. US20040235037A1
; GENERAL INFORMATION:
;  APPLICANT: Blumenfeld, Marta
;  APPLICANT:  Bougueleret, Lydie
;  APPLICANT:  Chumakov, Ilya
;  TITLE OF INVENTION: Polymorphic Markers of Prostate Carcinoma Tumor Antigen-1 (PCTA-1
;  FILE REFERENCE: GEN-T112XC1
;  CURRENT APPLICATION NUMBER: US/10/856,888
;  CURRENT FILING DATE:  2004-05-27
;  PRIOR APPLICATION NUMBER: US/09/326,402
;  PRIOR FILING DATE: 1999-06-04
;  PRIOR APPLICATION NUMBER: 60/088,187
;  PRIOR FILING DATE: 1998-06-05
;  PRIOR APPLICATION NUMBER: 60/102,324
;  PRIOR FILING DATE: 1998-09-28
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
;   LENGTH: 106746
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 1..68647
;   OTHER INFORMATION: 5'regulation  region
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 66647..68647
;   OTHER INFORMATION: promoter
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 97156..106746
;   OTHER INFORMATION: 3'regulation  region
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 68648..68741
;   OTHER INFORMATION: exon0
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 70647..70794
;   OTHER INFORMATION: exon1
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 82208..82296
;   OTHER INFORMATION: exon2
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 83613..83823
;   OTHER INFORMATION: exon3
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 85298..85417
;   OTHER INFORMATION: exon4
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 86389..86445
;   OTHER INFORMATION: exon5
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 87496..87522
;   OTHER INFORMATION: exon6
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 87650..87775
;   OTHER INFORMATION: exon6bis
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 88295..88383
;   OTHER INFORMATION: exon7
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 89484..89649
;   OTHER INFORMATION: exon8
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 92749..97155
;   OTHER INFORMATION: exon9
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 92749..92883
;   OTHER INFORMATION: exon9bis
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 95821..97155
;   OTHER INFORMATION: exon9ter
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 70647..70794
;   OTHER INFORMATION: homology with genset EST : A241850
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 68648..68741
;   OTHER INFORMATION: homology with genset EST : A241850
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 82208..82229
;   OTHER INFORMATION: homology with genset EST : A241850
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 278
;   OTHER INFORMATION: 99-1601-278 : polymorphic base A or C
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 402
;   OTHER INFORMATION: 99-1601-402 : polymorphic base w= A or T
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 472
;   OTHER INFORMATION: 99-1601-472 : polymorphic base A or T
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 2955
;   OTHER INFORMATION: 99-13801-100 : polymorphic base T or C
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 12167
;   OTHER INFORMATION: 99-13806-166 : polymorphic base G or A
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 12536
;   OTHER INFORMATION: 99-13799-376 : polymorphic base T or G
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 17593
;   OTHER INFORMATION: 99-13798-297 : polymorphic base T or C
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 17606
;   OTHER INFORMATION: 99-13798-284 : polymorphic base T or C
```

```
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 22079
;   OTHER INFORMATION: 99-1602-200 : polymorphic base G or C
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 28964
;   OTHER INFORMATION: 99-13794-186 : polymorphic base T or C
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 29003
;   OTHER INFORMATION: 99-13794-147 : polymorphic base C or G
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 31077
;   OTHER INFORMATION: 99-13812-384 : polymorphic base T or C
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 31766
;   OTHER INFORMATION: 99-13805-313 : polymorphic base T or C
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 34791
;   OTHER INFORMATION: 99-1587-281 : polymorphic base A or G
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 45751
;   OTHER INFORMATION: 99-1582-430 : polymorphic base C or T
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 49847
;   OTHER INFORMATION: 99-1585-465 : polymorphic base T or C
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 49855
;   OTHER INFORMATION: 99-1585-457 : polymorphic base T or C
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 49886
;   OTHER INFORMATION: 99-1585-426 : polymorphic base G or A
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 49900
;   OTHER INFORMATION: 99-1585-412 : polymorphic base G or A
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 49906
;   OTHER INFORMATION: 99-1585-406 : polymorphic base C or A
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 49921
;   OTHER INFORMATION: 99-1585-391 : polymorphic base C or A
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 49939
;   OTHER INFORMATION: 99-1585-373 : polymorphic base G or A
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 50256
;   OTHER INFORMATION: 99-1585-55 : polymorphic base C or A
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 54955
;   OTHER INFORMATION: 99-1607-373 : polymorphic base T or C
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 64239
;   OTHER INFORMATION: 99-1577-105 : polymorphic base A or G
;   FEATURE:
;   NAME/KEY: allele

  Query Match             5.1%;  Score 253;  DB 18;  Length 106746;
  Best Local Similarity   69.4%;  Pred. No. 2.2e-41;
  Matches  416;  Conservative    0;  Mismatches  175;  Indels    8;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              |||||||||||||||||  ||||||||||||||||| |||||||||||||||| ||||||
Db      38717 AGTTCAAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAAAATT 38658

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   ||||||||| | ||||||||||||||||||||||   |||| | ||||||||||||
Db      38657 AGCTGGGCATGGTGGTGCACGCCTGTAATCCCAGCTACTCAGGAGACTGAGGCAGGAGAA 38598

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCCA 178
              ||||||||||||  || ||||||| ||| || ||| ||| |   | ||| |||||||||
Db      38597 TCACTTGAACCTG--GGAGGTGGAGTTTGTGGTGAGCCAAAACTTTCCCACTGCACTCCA 38540

Qy        179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCAT-TTTATATATAT 237
              |||       |   | |   | || |   | | | |     | |   ||    | ||||
Db      38539 GCCAGTATATATATATATATATACACACGTACACGTGTACACATATACACACATATGC 38480

Qy        238 ATATATATATATATACACACACACACATATATATATACACATATATATACACATATAT 297
              | |  | || ||| ||||||| |||||   | ||| || ||||||| || |  | || ||
Db      38479 ACACGTGTACATACACACACATACACATGCACATACATGCACATATGTACATGCGTACAT 38420

Qy        298 ACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACAC 357
              ||| | ||  |||||||||||||| || || | |  ||||    ||||| || || | |
Db      38419 ACACACATGTACATATATACACACATGTACATATGTACACCCATACACACATGTACATAT 38360

Qy        358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417
              ||      | ||| | ||  |||||| ||||| |||||| | |  || |||   |||
Db      38359 GTACACCCATACACACATGTACATATGTACACCCATATACATATGTACATATGTACACCC 38300

Qy        418 ATACACATATATACACATATATACACATATACACATATATACACATATATACATATATAC 477
              |||||||||| |||| || || || |||| |||  || |||   | |   | | | ||
Db      38299 ATACACATATGTACATATGTACACCCATACACATGTACATATGTACACCCATACACATGT 38240

Qy        478 ACATATATATAATATACACACATATATATACACATATATACACACATATATACACATATA 537
              |||||| |   | || |||| || || |||   ||||  | ||||||||| |||| |||||
Db      38239 ACATATGT---ACATCCACATATGTACATATGTATATCCATACACATATGTACATATATA 38183

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATACATAT 596
              ||||||||  || | |  || ||| | || | |  |||| ||   |||||||||   ||
Db      38182 CACACATATGTACATATGTACATATATATTTGTGTATATCTATGTATATATATAAGAAT 38124


RESULT 91
US-09-728-552-4/c
; Sequence 4, Application US/09728552
; Publication No. US20030096398A1
; GENERAL INFORMATION:
;  APPLICANT: Choo, Kong-Hong Andy
;  APPLICANT:  Du Sart, Desiree
;  APPLICANT:  Cancilla, Michael R.
;  TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
;  FILE REFERENCE: Davies Col
;  CURRENT APPLICATION NUMBER: US/09/728,552
;  CURRENT FILING DATE:  2000-12-02
;  PRIOR APPLICATION NUMBER: 09/078,294
;  PRIOR FILING DATE: 1998-05-13
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 80246
;   TYPE: DNA
;   ORGANISM: Nucleotide sequence of NC-contig
US-09-728-552-4

  Query Match             5.1%;  Score 252.6;  DB 10;  Length 80246;
  Best Local Similarity   71.9%;  Pred. No. 2.3e-41;
  Matches  360;  Conservative    0;  Mismatches  134;  Indels    7;  Gaps    2;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | || |||| ||| | |||||| | ||| | | ||| |  ||| | |||| ||| | |||
Db      15787 TATACATATTTATGTCTATATACACACATATATACATATTTATGTCTATATACACACATA 15728
```

```
Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              |||| || |||   ||||||||||||||||||| || ||| |   | ||||| || || |||
Db      15727 TACATATTTATGTCTATATACACATATATACATATTTATGTCTATATACACATATATACA 15668

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              ||| |||    ||||  | | | ||| |||   || | || | ||| ||||| ||||||
Db      15667 TATTTATGTCTATATACACATATATACATATTTATGTCTATATACACATATATACATATT 15608

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              |  |   ||||||||||||||||||| || |||    |||||||||||||||||| || ||
Db      15607 T--ATGTCTATATACACATATATACATATTTATGTCTATATACACATATATACATATTTA 15550

Qy        467 TACATATATACACATATATA-----TAATATACACACATATATATACACATATATACACA 521
              |   ||||||||||||||||     | || |  | | | | |||| ||||||  ||
Db      15549 TGTCTATATACACATATATACATATTTATGTCTATATACACATATATACATATTTATGTC 15490

Qy        522 CATATATACACATATATACACATATATATACACATATATACACATATATACATATATACA 581
              |||| ||| ||| |||   || | |||| ||| ||||| |||||| ||  | |||| |
Db      15489 TATATACACATATACATATTTATGTCTATATACACATATATACATATTTATGTCTATATA 15430

Qy        582 CATATATATACATATATACACATATATACATATATACACATATATACATATATACACACA 641
              || ||||||||||||  ||    |||||||| ||||| |||||| ||  |||||| |||||
Db      15429 CACATATATACATATTTATGTCTATATACACATATATACATATTTATGTATATATACACA 15370

Qy        642 TATATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTA 701
              |||||||| ||  ||| |  | | ||| | ||||||||| |||  | |||| |||   ||
Db      15369 TATATACATATTTATGTCTATATACACATATATACATATTTATGTATATATTTATGTATA 15310

Qy        702 TATATATGTATATATATATGC 722
              ||| | ||||||| || || |
Db      15309 TATGTGTGTATATGTACATAC 15289


RESULT 92
US-09-818-657-3/c
; Sequence 3, Application US/09818657
; Publication No. US20030077773A1
; GENERAL INFORMATION:
;  APPLICANT: BRANDON, Rhonda et al.
;  TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
;  TITLE OF INVENTION:  NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
;  TITLE OF INVENTION:  AND USES THEREOF
;  FILE REFERENCE: CL001006-CIP
;  CURRENT APPLICATION NUMBER: US/09/818,657
;  CURRENT FILING DATE:  2001-03-28
;  PRIOR APPLICATION NUMBER: 09/730,002
;  PRIOR FILING DATE: 2000-12-06
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 82938
;   TYPE: DNA
;   ORGANISM: Human
US-09-818-657-3

  Query Match             5.1%;  Score 252.6;  DB 10;  Length 82938;
  Best Local Similarity   70.7%;  Pred. No. 2.3e-41;
  Matches  350;  Conservative    0;  Mismatches  144;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | |||| | |||| |||| |||||||| | | |   | |||  ||||||    ||||||
Db       7454 TATATACACATATGTATACATATATACATATATATGTATATACGTATATAGCTGTATATA 7395

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || ||| || || |||   |  | ||||| | |||| ||||||| ||||||| |||| |
Db       7394 TATACACATGTATATAGCTGTATATATACATACATGTATATACATATACATATATACATA 7335

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              || | ||| |||||  | || | | ||||||| ||||| | |||  |||||||| |||||
Db       7334 TACACATATACATACATGTATATACATATACATATATACATACATGTATATACATATATA 7275

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
                | |   ||| ||    || |||| |||||||||| || || |||| | ||   |||||
Db       7274 CATATGTATACATGTGTATCTATATACATATATACATATGTATACATGTGTATCTATATA 7215

Qy        467 TACATATATACACATATATAT-AATATACACACATATATATACACATATATACACACATA 525
              | |||||  | |  |||| ||   |||  | | | |||||||||| || ||||||    ||
Db       7214 CATATATACATATGTATACATGTGTATCTATATACATATATACATATGTATACATGTGTA 7155

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              | || | | ||||| ||||||  ||||||   | |||   | |||  ||| |||  |||   |||
Db       7154 TCTATATACATATATACATATGTATACATGTGTATCTATATACATATATACATATGTATA 7095

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              || |  || |||| ||||||||||||||| || |||| | ||  |||||||| | ||| |
Db       7094 CATGTGTATCTATATACATATATACATATGTATACATGTGTATCTATATACATATATACA 7035

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              ||   |||||| |  |   | ||| ||||| | |||| || |||| | | || ||||| |
Db       7034 TATGTATACATGTGTATCTATATACATATATACATATGTATACATGTGTGTATGTATACA 6975

Qy        706 TATGTATATATATAT 720
              ||||| | | |||||
Db       6974 TATGTGTGTGTATAT 6960


RESULT 93
US-10-417-375-36/c
; Sequence 36, Application US/10417375
; Publication No. US20040219528A1
; GENERAL INFORMATION:
;  APPLICANT: David W. Morris
;  APPLICANT:  Marc Malandro
;  TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
;  FILE REFERENCE: 529452001600
;  CURRENT APPLICATION NUMBER: US/10/417,375
;  CURRENT FILING DATE:  2003-04-15
;  NUMBER OF SEQ ID NOS: 176
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 36
;   LENGTH: 91697
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-417-375-36

  Query Match             5.1%;  Score 252.2;  DB 18;  Length 91697;
  Best Local Similarity   69.6%;  Pred. No. 3e-41;
  Matches  357;  Conservative    0;  Mismatches  153;  Indels    3;  Gaps    1;

Qy        210 GAAAAAAAAAAAGGCATTTTATATATATATATATATATATATACACACACACACATAT 269
              | | | | |  ||  | || |||| |||||| ||||| ||| | |   | | | ||||
Db      20066 GCATATATATACAGCTATATAGATATGTATATAGATATACATATAGATGTATATATATAT 20007

Qy        270 ATATATACACAT---ATATATACACATATATACATATATACACATATATACACATATATA 326
              |||||||| ||   || || | ||||||| |  |||||  | ||| ||| | | |||||
Db      20006 CTATATACATATGCCATGTAGATACATATACATCTATATGTATATAGATATATAGATATA 19947

Qy        327 TACACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATA 386
              |  | ||| |||| |  | |||||||| || |||| |||| | |||||||||| || ||||
Db      19946 TGTATATAGATATATGTAGATATATATCCATATATTTATATAGATATATACATATGTATA 19887

Qy        387 CACACATATATACACATATATACACACATATATACACATATATACACATATATACACATA 446
              | | |||| || | || |||| ||  |||||   | ||| ||| | | ||||| |||||
Db      19886 TAGATATATTTATATATGTATATACCTATATACGTATATAGATATATACATATATACATA 19827

Qy        447 TACACATATATACACATATATACATATATACACATATATATAATATACACACATATATAT 506
              ||   ||||||| | ||||||||||||||||| || ||||  |||||||  |     |||||
Db      19826 TATGTATATATAGATATATATACATATATACATATGTATACTATATACATAATAGTATAT 19767

Qy        507 ACACATATATACACACATATATACACATATATACACATATATATACACATATATACACAT 566
              ||| || |||| |   ||| ||| |  |||| | |  |||||||  | | || ||   ||
Db      19766 ACATATGTATATATGTATACATAGATGTATACATATGTATATATGTATACATCTATGTAT 19707

Qy        567 ATATACATATATACACATATATATACATATATACACATATATACATATATACACATATAT 626
              | ||| || ||||||  ||| |||||||| ||  | | ||| | ||||| | |  |||||
Db      19706 ACATAGATGTATACATCTATGTATACATAGATGTATACATAGATATATACATATGTATAT 19647
```

```
Qy        627 ACATATATACACACATATATACACATACATATACACACACATAGATATACATATATATAC 686
              | || ||||   |  |||||| |||||  |||| |   | ||||| ||| ||| ||| |
Db      19646 AGATGTATATCAATCTATATATACATATGTATATATGTATATAGAAATATATACATACAT 19587

Qy        687 ACATATATATACGTATATATATGTATATATATA 719
              |  |||||||   |||| |||| || || ||||
Db      19586 ATGTATATATGTATATAGATATATAGATGTATA 19554


RESULT 94
US-10-017-161-2375/c
; Sequence 2375, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
;  FILE REFERENCE: 084335/0152
;  CURRENT APPLICATION NUMBER: US/10/017,161
;  CURRENT FILING DATE:  2002-12-18
;  PRIOR APPLICATION NUMBER: JP 2001/246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2430
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2375
;   LENGTH: 1114
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: source
;   LOCATION: (1)..(1114)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(914)
US-10-017-161-2375

  Query Match             5.1%;  Score 251;  DB 15;  Length 1114;
  Best Local Similarity   63.7%;  Pred. No. 7.2e-42;
  Matches  397;  Conservative    0;  Mismatches  225;  Indels    1;  Gaps    1;

Qy        230 ATATATATATATATATATATACACACACACACATATATATATACACATATATATAC 289
              ||| | |||||| ||||||| |  | | ||| ||||||||||| |||| |  | |||||||
Db        864 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 805

Qy        290 ACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATAT 349
              | | | ||| || |||| |  | |||| ||||| | ||| |  || |||| |  | |||
Db        804 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 745

Qy        350 ATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATAC 409
              ||| ||| |||||||||  |  | |||| ||||| | | |  ||||||    |||||||
Db        744 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 685

Qy        410 ACACATATATACACATATATACACATATATACACATATACACATATATACACATATATAC 469
              | ||| |||||   ||||||    ||||| | ||| ||| |  ||||||    |||||||
Db        684 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 625

Qy        470 ATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATATAT 528
              ||| | | | || ||||||  | || | ||||| | ||| |  ||||||    |||||
Db        624 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 565

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACATATAT 588
              | ||| |||||   ||||||  |  | ||| ||||||||||||  ||||||    |||||
Db        564 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 505

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
              ||||| |||||   ||||||   |||||||| | | ||| || |||| |    |||||||
Db        504 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 445

Qy        649 ACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATAT 708
              | | |||||||   | | ||   |||| | ||| | | | || ||||||  | ||||||
```

```
Db        444 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 385

Qy        709 GTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAG 768
              || | |||||||          |  |||          ||  |   ||     |  | ||| |
Db        384 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 325

Qy        769 GAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAG 828
              | | | |  |  | |   |||||    | |   | || |     |  | | | |||
Db        324 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 265

Qy        829 TAGGGGAAATACATAAATGTGTA 851
                | ||||||| |||| ||
Db        264 ATACACATATACATATATGTATA 242


RESULT 95
US-10-292-798-2017/c
; Sequence 2017, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
;  FILE REFERENCE: 084335/166
;  CURRENT APPLICATION NUMBER: US/10/292,798
;  CURRENT FILING DATE:  2002-11-13
;  PRIOR APPLICATION NUMBER: 10/017,161
;  PRIOR FILING DATE: 2001-12-18
;  PRIOR APPLICATION NUMBER: JP 2001-246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2070
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2017
;   LENGTH: 1114
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   LOCATION: source
;   FEATURE:
;   LOCATION: (1)..(1114)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(914)
US-10-292-798-2017

  Query Match             5.1%;  Score 251;  DB 17;  Length 1114;
  Best Local Similarity   63.7%;  Pred. No. 7.2e-42;
  Matches  397;  Conservative    0;  Mismatches  225;  Indels    1;  Gaps    1;

Qy        230 ATATATATATATATATATATACACACACACACATATATATATACACATATATATAC 289
              ||| | |||||| ||||||| |  | | ||| ||||||||||| |||| |  | |||||||
Db        864 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 805

Qy        290 ACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATAT 349
              | | | ||| || |||| |  | |||| ||||| | ||| |  || |||| |  | |||
Db        804 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 745

Qy        350 ATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATAC 409
              ||| ||| |||||||||  |  | |||| ||||| | | |  ||||||    |||||||
Db        744 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 685

Qy        410 ACACATATATACACATATATACACATATATACACATATACACATATATACACATATATAC 469
              | ||| |||||   ||||||    ||||| | ||| ||| |  ||||||    |||||||
Db        684 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 625

Qy        470 ATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATATAT 528
              ||| | | | || ||||||  | || | ||||| | ||| |  ||||||    |||||
Db        624 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 565

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACACATATAT 588
              | ||| |||||   ||||||  |  | ||| ||||||||||||  ||||||    |||||
```

```
Db        564 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 505

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
              ||||| |||||   ||||||  |||||||| | | ||| || |||| |    |||||||
Db        504 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 445

Qy        649 ACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATAT 708
              | | |||||||   | | ||   |||| | ||| | | | || ||||||  | ||||||
Db        444 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 385

Qy        709 GTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAG 768
               || | |||||||        |  |||        ||  |   ||    |  | ||| |
Db        384 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 325

Qy        769 GAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAG 828
              | | | |  | | |    |||||    | |   | || |     |  | | | |||
Db        324 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 265

Qy        829 TAGGGGAAATACATAAATGTGTA 851
                  | ||||||| |||| ||
Db        264 ATACACATATACATATATGTATA 242


RESULT 96
US-10-052-482-232
; Sequence 232, Application US/10052482
; Publication No. US20040072264A1
; GENERAL INFORMATION:
;  APPLICANT: Engelhard, Eric
;  APPLICANT:  Morris, David
;  TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
;  FILE REFERENCE: A-71087/RMS/DCF
;  CURRENT APPLICATION NUMBER: US/10/052,482
;  CURRENT FILING DATE:  2002-08-15
;  PRIOR APPLICATION NUMBER: US 09/747,377
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: US 09/798,586
;  PRIOR FILING DATE: 2001-03-02
;  NUMBER OF SEQ ID NOS: 241
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 232
;   LENGTH: 96595
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (49183)..(49621)
;   OTHER INFORMATION: "n" at positions 49183 to 49621 can be any base
US-10-052-482-232

  Query Match             5.1%;  Score 251;  DB 17;  Length 96595;
  Best Local Similarity   69.5%;  Pred. No. 5.4e-41;
  Matches  354;  Conservative    0;  Mismatches  152;  Indels    3;  Gaps    1;

Qy        229 TATATATATATATATATATATATACACACACACACATATATATATACACATATATATA 288
              || ||| | |||| | |||||||||| ||| | |||||||| | | |||| ||| ||| |
Db      90982 TACATACACATATGTGTATATATACATACATATACACATATGTGTGTACATATACATACA 91041

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              || | | |||  | | |||| ||| ||||||| | ||| ||| ||   ||  |  |||||
Db      91042 CATACACATATGTGTGTACATATACATACACACACATACACATATGTGTACATGTACATA 91101

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
               ||| ||| | |  |||    | || |||| | ||| ||| ||| ||||    | || ||
Db      91102 CATACACATACACATATGTGTACATGTACATACATACACATACACATATGTGTACATGTA 91161

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              || ||||| | | ||| ||||    | || |||| ||||||||||||| ||| || | ||
Db      91162 CATACATACACATACACATATGTGTACATGTACATATATACACATATACACATATGTGTA 91221

Qy        469 CATATATACACATATATATAATATACACACATATATATACACATATATACACACATATAT 528
              ||| || | | ||| |    |||||||| || | || |    | ||||| |||||||||
Db      91222 CATGTACATATATACAC---ATATACACATATGTGTACATGTACATATATACACATATAC 91278

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACATATAT 588
              ||| || | ||||  || |||||||||||||| ||| || | ||||| || | | | | ||
Db      91279 ACATATGTGTACATGTACATATACACATATACACATATGTGTACATGTACATATACACAT 91338

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
              ||||| ||||    | ||  ||||||| |||  | || |   | ||| |||||||| |
Db      91339 ATACACATATGTGTACATGCACATATACACATGTGTACATGCACATATACACATATGTGT 91398

Qy        649 ACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATAT 708
              ||||||| ||| ||||| ||   | || ||| |||||||||| |||| |  ||||||| ||
Db      91399 ACATACACATATACACATATGTGTGTATATACATATACACACATATGTGTGTATATACAT 91458

Qy        709 GTATATATATATGCTCCAGAGTTCATAAG 737
               |||| ||||||  |       | |||| |
Db      91459 ATATACATATATACATATATATACATATG 91487


RESULT 97
US-10-311-455-1972/c
; Sequence 1972, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determi
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1972
;   LENGTH: 21537
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1972

  Query Match             5.0%;  Score 250;  DB 15;  Length 21537;
  Best Local Similarity   68.1%;  Pred. No. 4.4e-41;
  Matches  391;  Conservative    0;  Mismatches  180;  Indels    3;  Gaps    3;

Qy        181 CCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATA 240
              |||    | || |    || |  |||||     |  | |     | ||||||||||||
Db       6005 CCAAATTAAAATACAATAACGCGATCTCGATTCACTACACTATATATATATATATATATA 5946

Qy        241 TATATATATATACACACACACACATATATATATACACATATATATACACATATATACA 300
              |||||||||||| ||||| |||| |||||||||||| | ||||||||||| | |||||||
Db       5945 TATATATATATATACACATACAC-CATATATATATATATATATATATACATACATATACA 5887

Qy        301 TATATACACATATATACACATATATATACACATACATATGTACACATATATATACACATA 360
               |   |     |   || | | | ||  | | |    |    | |||| | | ||| ||
Db       5886 AAACAATTACTTACTAAATACACAAATCAATAAAATAAAAAATTCATAAACAAACATATC 5827

Qy        361 TGTATACACATATATACACATATATACACACATATATACACATATATACACACATATATA 420
              | | |  |||||||| |  |||||  |      |||| | ||||||| | | |||||||
Db       5826 TCTCTTTACATATATATATCTATATCGATCACACTATATATATATATATATATATATA 5767

Qy        421 CACATATATACACATATATACACATATACACATATATACACATATATACATATATACACA 480
              |||| ||||| | ||||||| | ||| |||||||||| | ||||||| | | | | | |
Db       5766 CACACATATATATATATATATATACACACATATATATATATATATACACACATATA 5707

Qy        481 TATATATAATATACACACATATATATACACATATATACACACATATATACACATATATAC 540
              ||||||| | |||||||||||||||| | ||||||| ||||| ||||| | |||||||
Db       5706 TATATAT-ACATACACACATATATATATATATATATACACACATATATATATATAT 5648
```

```
Qy        541 ACATATATATACACATATATACACATATATACATATATACACATATATATACATATATAC 600
              | ||||||||| | ||||||| | ||||||| ||||||| | ||||| |  ||||||||
Db       5647 ATATATATATATATATATATATATATATATATATATATATATATATATACACGCATATATAT 5588

Qy        601 ACATATATACATATATACACATATATACATATATACACACATATATACACATACATATAC 660
              || ||||||||||||| | |   ||| ||||||| | ||||||||| | ||| |||||
Db       5587 ACGTATATACATATATATAAACTCATATATATATATACATATATATATATATATATAT 5528

Qy        661 ACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATAT 720
              | | | ||| |   ||| |||||| |  | | | ||      || | | ||||  |  ||
Db       5527 ATATATATAAACTCACA-ATATATCCTTAAACAAATTATATAATCTCTATATACCTCAAT 5469

Qy        721 GCTCCAGAGTTCATAAGAGGTAGCAGTTGATTAC 754
              ||   | |  | || |   |  | |  | |||
Db       5468 TTTCTCAAATATAAAACAAAAATAATTATAATAC 5435


RESULT 98
US-09-997-722-196
; Sequence 196, Application US/09997722
; Publication No. US20040072154A1
; GENERAL INFORMATION:
;  APPLICANT: Morris, David
;  APPLICANT:  Engelhard, Eric
;  TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
;  FILE REFERENCE: A-71171/RMS/DCF
;  CURRENT APPLICATION NUMBER: US/09/997,722
;  CURRENT FILING DATE:  2001-11-30
;  PRIOR APPLICATION NUMBER: US 09/747,377
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: US 09/798,586
;  PRIOR FILING DATE: 2001-03-02
;  NUMBER OF SEQ ID NOS: 301
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 196
;   LENGTH: 96596
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-997-722-196

  Query Match             5.0%;  Score 249.6;  DB 11;  Length 96596;
  Best Local Similarity   70.5%;  Pred. No. 1.1e-40;
  Matches  347;  Conservative    0;  Mismatches  144;  Indels    1;  Gaps    1;

Qy        229 TATATATATATATATATATATATACACACACACACACATATATATATACACATATATATA 288
              | | || ||||||| || || || ||   | | |||  | || ||||| |||| || ||
Db      57110 TGTGTACATATATACATGTAGATGCATGTAGATACATGTGTACATATATACATGTAGATG 57169

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              ||  || |||||| ||||||  || |||||  || ||| |   |||||| |  | ||||
Db      57170 CATGTAGATACATGTATACATGTAGATACATGTAGATACATGTATACATGTAGATACATG 57229

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              || |||||  |||  ||| | | || || | | || || | |||| |||||| ||||| |
Db      57230 TAGATACATGTATACATATATACATGTAGATACATGTAGATACATGTATACATATATACA 57289

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              | ||| ||  | | ||||| |||| || | |||| || |||||||||||  || |||
Db      57290 TGTAGATACATGTATACATATATACATGTAGATACATGTATACATATATACATGTAGATA 57349

Qy        469 CATATATACACATATATATAATATACACACATATATATACACATATATACACACATATAT 528
              ||| |||||| ||||| ||  || | |||  || ||| |   ||| ||| | |||| ||
Db      57350 CATGTATACATATATACAT-GTAGATACATGTAGATACATGTATACATATATACATGTAG 57408

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACATATAT 588
              | |||| || | |||| |||| | | ||| ||  | ||| ||  ||||| | ||||| ||
Db      57409 ATACATGTAGATACATGTATACATATATACATGTAGATACATGTATATAGATACATACAT 57468

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
              | | ||||| | ||||| ||||||||||  | |||| |||||||||  ||| ||||||
Db      57469 AGATATATAGATACATAGATACATATATTTATATATGTACATATATCTACATATATATCT 57528

Qy        649 ACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATAT 708
              | ||| | ||| |   ||||| |||| ||||||||| |  || |||||  | |||||||
Db      57529 ATATATACATATATCTACATATATATCTATATATATATATCTACATATATATCTATATAT 57588

Qy        709 GTATATATATAT 720
              ||| | ||||||
Db      57589 GTAGAAATATAT 57600


RESULT 99
US-10-437-963-61122
; Sequence 61122, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 61122
;   LENGTH: 623
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_62586C.1
US-10-437-963-61122

  Query Match             5.0%;  Score 249;  DB 18;  Length 623;
  Best Local Similarity   73.2%;  Pred. No. 1.4e-41;
  Matches  361;  Conservative    0;  Mismatches  125;  Indels    7;  Gaps    3;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | |||||||||||||||||||||||| | | | | | ||||||||||| | |||||||
Db        122 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 181

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | |||||||||| | ||| |||| || | |
Db        182 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 241

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        242 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 301

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | |      | |  |||
Db        302 TATATATATATATATATATATATATATATATATATATATATATATATATATAAGATGAAATGTATG 361

Qy        467 TACATATATACACATATATATAATATACACACAT---ATATATACACATATATACACACA 523
              ||    ||  |||||||||||     || || ||    | |||| ||  | |  | | |
Db        362 GTCAGTGATCAACATATATATATGGCACGCAAATTTAAAATATTCATCTGTTGATAAAAT 421

Qy        524 TATA---TACACATATATACACATATATATACACATATATACACATATATACATATATAC 580
              | ||    |||| || ||| || |||||||||||||| ||    |||| | |    | ||||
Db        422 TTTACACTACAAATGTATGCATATATATATACACATGTAACTACATGTGTGTGTCTATAT 481

Qy        581 ACATATATATACATATATACACATATATACATATATACACATATATACATATATACACAC 640
              | ||||||||| ||||||| | ||||||| ||||||| | ||||||| |||||| |
Db        482 ATATATATATATATATATATATATATATATATATATATATATATATATATATAT-CGTGT 540

Qy        641 ATATATACACATACATATACACACACATAGATATACATATATATACACATATATATACGT 700
              |||||||| | | |    | |  ||| | | ||||| |||||||| | |||||||||  |
Db        541 ATATATACCCTTTCTCGTGCCTACATACATATATATATATATATATATATATATATAT 600

Qy        701 ATATATATGTATA 713
```

```
              |||||||| ||||
Db        601 ATATATATATATA 613


RESULT 100
US-10-437-963-61122/c
; Sequence 61122, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 61122
;   LENGTH: 623
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_62586C.1
US-10-437-963-61122

  Query Match             5.0%;  Score 249;  DB 18;  Length 623;
  Best Local Similarity   72.7%;  Pred. No. 1.4e-41;
  Matches  364;  Conservative    0;  Mismatches  130;  Indels    7;  Gaps    3;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACAC-ATATAT 285
              | ||||||||||||||||||| |||  |  |||    |    |||||||||||| ||||||
Db        589 TATATATATATATATATATATGTATGTAGGCACGAGAAAGGGTATATATACACGATATAT 530

Qy        286 ATACACATATATACATATATACACATATATACACATATATATACACATACATATGTACAC 345
              ||| | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||   ||||
Db        529 ATATATATATATATATATATATATATATATATATATATATATATATATATATAGACACAC 470

Qy        346 ATATATATACACATATGTATACACATATATACACAT-----ATATACACACATATATACA 400
              | || ||   |||| |||||| | ||||  | ||||      | || |    |||  |||
Db        469 ACATGTAGTTACATGTGTATATATATATGCATACATTTGTAGTGTAAAATTTTATCAACA 410

Qy        401 CATATATACACACATATATACACATATATACACATATATACACATATACACATATATACA 460
              ||  |||    |  |      ||||| | || |  |    |   || | || |
Db        409 GATGAATATTTTAAATTTGCGTGCCATATATATATGTTGATCACTGACCATACATTTCAT 350

Qy        461 CATATATACATATATACACATATATATA-ATATACACACATATATATACACATATATACA 519
              | |||||| ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| |
Db        349 CTTATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 290

Qy        520 CACATATATACACATATATACACATATATATACACATATATACACATATATACATATATA 579
              | ||||||| | ||||||| | ||||||||| | ||||||| | ||||||| |||||||
Db        289 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 230

Qy        580 CACATATATATACATATATACACATATATACATATATACACATATATACATATATACACA 639
              | ||||||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | |
Db        229 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 170

Qy        640 CATATATACACATACATATACACACACATAGATATACATATATATACACATATATATACG 699
              ||||||| | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||
Db        169 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 110

Qy        700 TATATATATGTATATATATAT 720
              |||||||| ||||||||||
Db        109 TATATATATATATATATATAT 89


Search completed: April 22, 2005, 07:11:52
```

Job time : 2191.11 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 20, 2005, 22:52:45 ; Search time 2013.07 Seconds
(without alignments)
13936.065 Million cell updates/sec

Title: US-09-939-209A-3_COPY_1_4621
Perfect score: 4621
Sequence: 1 agttcaagaccagcctgagc.....ttgctttttagtcctaaaaa 4621

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5633728 seqs, 3035525691 residues

Total number of hits satisfying chosen parameters: 11267456

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq:*
19: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
20: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:*
21: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
22: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

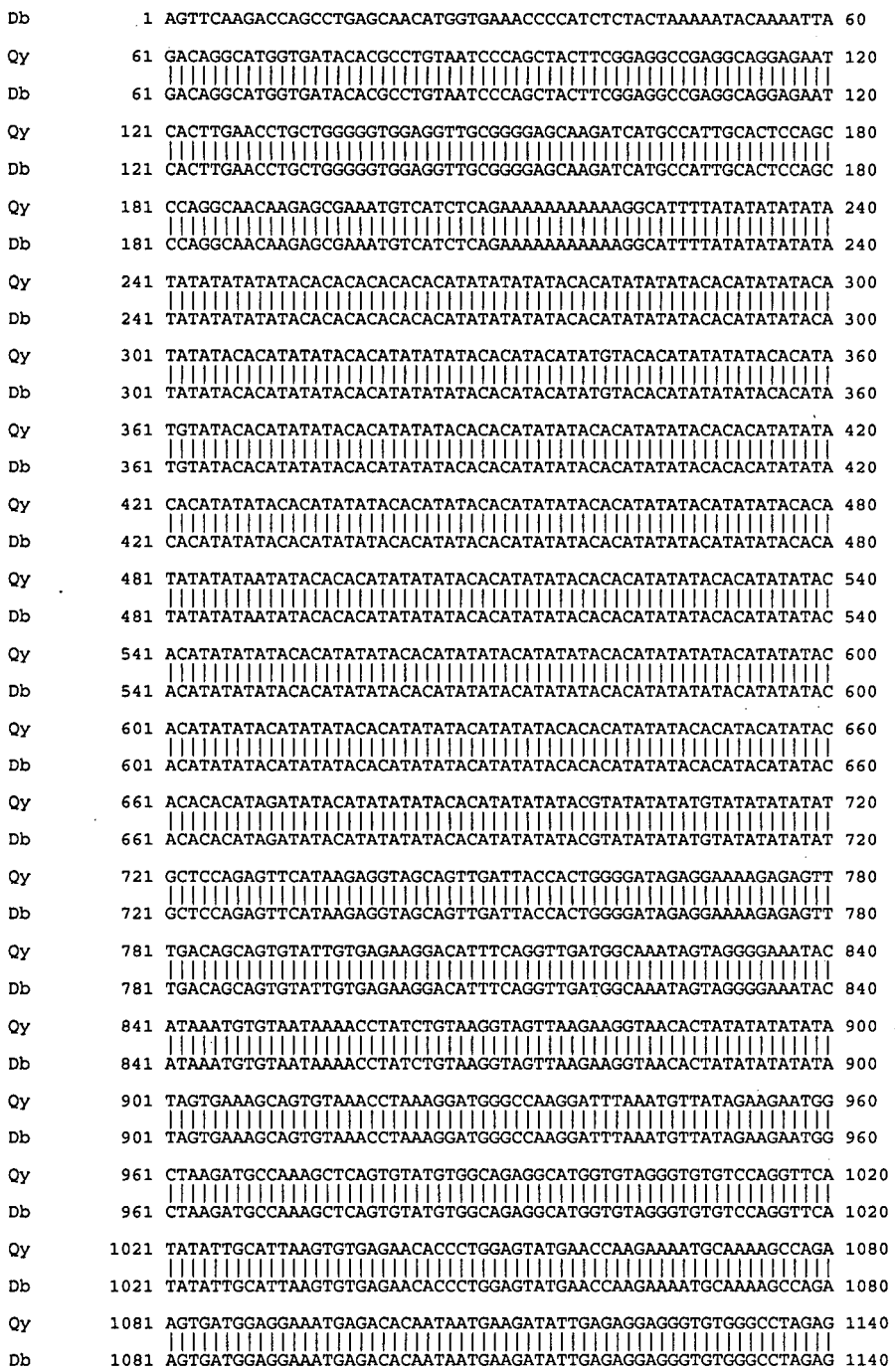
	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	4621	100.0	20300	10	US-09-939-209A-3	Sequence 3, Appli
c	2	356.6	7.7	121600	18	US-10-723-860-1125	Sequence 1125, Ap
	3	345.8	7.5	109730	19	US-10-741-600-17809	Sequence 17809, A
c	4	327.4	7.1	188971	17	US-10-235-192A-27	Sequence 27, Appl
	5	323.6	7.0	163382	18	US-10-367-094-179	Sequence 179, App
c	6	322.2	7.0	26997	17	US-10-221-714A-474	Sequence 474, App
	7	320.6	6.9	3310	13	US-10-027-632-113621	Sequence 113621,
	8	320.6	6.9	3310	13	US-10-027-632-113622	Sequence 113622,
	9	320.6	6.9	3310	17	US-10-027-632-113621	Sequence 113621,
	10	320.6	6.9	3310	17	US-10-027-632-113622	Sequence 113622,
c	11	311.2	6.7	26997	17	US-10-221-714A-473	Sequence 473, App
	12	308	6.7	1990	15	US-10-017-161-2383	Sequence 2383, Ap
	13	308	6.7	1990	17	US-10-292-798-2025	Sequence 2025, Ap
	14	300.6	6.5	1800	18	US-10-437-963-12666	Sequence 12666, A
	15	296	6.4	15418	9	US-09-783-203-1	Sequence 1, Appli
	16	296	6.4	15418	9	US-09-994-427A-1	Sequence 1, Appli
	17	296	6.4	15418	10	US-09-995-419A-1	Sequence 1, Appli
	18	296	6.4	15418	14	US-10-141-220-1	Sequence 1, Appli
	19	296	6.4	15418	14	US-10-023-969-1	Sequence 1, Appli
	20	296	6.4	15418	14	US-10-206-447-1	Sequence 1, Appli
	21	296	6.4	15418	17	US-10-674-836-1	Sequence 1, Appli
	22	296	6.4	15418	18	US-10-811-012-1	Sequence 1, Appli
c	23	294.8	6.4	1486	17	US-10-292-798-2005	Sequence 2005, Ap
	24	293.6	6.4	27067	10	US-09-841-158-5	Sequence 5, Appli
	25	293.6	6.4	27067	18	US-10-801-837-5	Sequence 5, Appli
c	26	293.2	6.3	1486	15	US-10-017-161-2361	Sequence 2361, Ap
c	27	291.6	6.3	1672	15	US-10-017-161-2381	Sequence 2381, Ap
c	28	291.6	6.3	1672	17	US-10-292-798-2023	Sequence 2023, Ap
c	29	273.8	5.9	81924	18	US-10-741-601-5636	Sequence 5636, Ap
	30	267.2	5.8	14603	19	US-10-741-600-17799	Sequence 17799, A
c	31	266.4	5.8	6971	15	US-10-311-455-1210	Sequence 1210, Ap
c	32	265.6	5.7	263852	18	US-10-812-232-6	Sequence 6, Appli
c	33	263.6	5.7	2522	17	US-10-108-260A-543	Sequence 543, App
c	34	263	5.7	227968	18	US-10-723-860-1357	Sequence 1357, Ap
c	35	262.6	5.7	109730	19	US-10-741-600-17809	Sequence 17809, A
c	36	262	5.7	1800	18	US-10-437-963-12666	Sequence 12666, A
	37	262	5.7	1980090	18	US-10-719-993-6815	Sequence 6815, Ap
	38	262	5.7	1980090	19	US-10-741-600-17676	Sequence 17676, A
c	39	261.4	5.7	12069	17	US-10-257-166-11	Sequence 11, Appl
c	40	259.6	5.6	1980090	18	US-10-719-993-6815	Sequence 6815, Ap
c	41	259.6	5.6	1980090	19	US-10-741-600-17676	Sequence 17676, A
	42	258.8	5.6	1489	15	US-10-017-161-2363	Sequence 2363, Ap
	43	258.8	5.6	1489	17	US-10-292-798-2007	Sequence 2007, Ap
c	44	258.6	5.6	163382	18	US-10-367-094-179	Sequence 179, App
	45	258.4	5.6	82938	10	US-09-818-657-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-09-939-209A-3
; Sequence 3, Application US/09939209A
; Publication No. US20030113721A1
; GENERAL INFORMATION:
; APPLICANT: LEVITT, PAT R.
; APPLICANT: MIRNICS, KAROLY
; APPLICANT: KODAVALI, MENKATA CHOWDARI
; APPLICANT: NIMGAONKAR, VISHWAJIT L.
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR FACILITATING THE DIAGNOSIS AND TREATMENT
; TITLE OF INVENTION: SCHIZOPHRENIA
; FILE REFERENCE: 00-539-US
; CURRENT APPLICATION NUMBER: US/09/939,209A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: 60/228,021
; PRIOR FILING DATE: 2000-08-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 20300
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: A genomic sequence containing RGS4 nucleic acid sequence and sequ
; OTHER INFORMATION: ences upstream and downstream to the RGS4 nucleic acid sequence
US-09-939-209A-3

Query Match 100.0%; Score 4621; DB 10; Length 20300;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4621; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAATTA 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||


Db 1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAATTA 60
Qy 61 GACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAAT 120
Db 61 GACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAAT 120
Qy 121 CACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAGC 180
Db 121 CACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAGC 180
Qy 181 CCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATA 240
Db 181 CCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATA 240
Qy 241 TATATATATATACACACACACACACATATATATATACACATATATATACACATATATACA 300
Db 241 TATATATATATACACACACACACATATATATATACACATATATATACACATATATACA 300
Qy 301 TATATACACATATATACACATATATATACACATACATATGTACACATATATATACACATA 360
Db 301 TATATACACATATATACACATATATATACACATACATATGTACACATATATATACACATA 360
Qy 361 TGTATACACATATATACACATATATACACACATATATACACATATATACACACATATATA 420
Db 361 TGTATACACATATATACACATATATACACACATATATACACATATATACACACATATATA 420
Qy 421 CACATATATACACATATATACACATATACACATATATACACATATATACATATATACACA 480
Db 421 CACATATATACACATATATACACATATACACATATATACACATATATACATATATACACA 480
Qy 481 TATATATAATATACACACATATATATACACATATATACACACATATATACACATATATAC 540
Db 481 TATATATAATATACACACATATATATACACATATATACACACATATATACACATATATAC 540
Qy 541 ACATATATATACACATATATACACATATATACATATATACACATATATATACATATATAC 600
Db 541 ACATATATATACACATATATACACATATATACATATATACACATATATATACATATATAC 600
Qy 601 ACATATATACATATATACACATATATACATATATACACACATATATACACATACATATAC 660
Db 601 ACATATATACATATATACACATATATACATATATACACACATATATACACATACATATAC 660
Qy 661 ACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATAT 720
Db 661 ACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATAT 720
Qy 721 GCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTT 780
Db 721 GCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTT 780
Qy 781 TGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATAC 840
Db 781 TGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATAC 840
Qy 841 ATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATATATATA 900
Db 841 ATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATATATATA 900
Qy 901 TAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTATAGAAGAATGG 960
Db 901 TAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTATAGAAGAATGG 960
Qy 961 CTAAGATGCCAAAGCTCAGTGTATGTGGCAGAGGCATGGTGTAGGGTGTGTCCAGGTTCA 1020
Db 961 CTAAGATGCCAAAGCTCAGTGTATGTGGCAGAGGCATGGTGTAGGGTGTGTCCAGGTTCA 1020
Qy 1021 TATATTGCATTAAGTGTGAGAACACCCTGGAGTATGAACCAAGAAAATGCAAAAGCCAGA 1080
Db 1021 TATATTGCATTAAGTGTGAGAACACCCTGGAGTATGAACCAAGAAAATGCAAAAGCCAGA 1080
Qy 1081 AGTGATGGAGGAAATGAGACACAATAATGAAGATATTGAGAGGAGGGTGTGGGCCTAGAG 1140
Db 1081 AGTGATGGAGGAAATGAGACACAATAATGAAGATATTGAGAGGAGGGTGTGGGCCTAGAG 1140
Qy 1141 TGAAGCTTTTCGTGCCAGTACTTCTTTTGAAGGCCCAGTTCTCTTCTCTCTCGGGGGCTC 1200
Db 1141 TGAAGCTTTTCGTGCCAGTACTTCTTTTGAAGGCCCAGTTCTCTTCTCTCTCGGGGGCTC 1200
Qy 1201 CTTCATCTCTCATAGAGTCCACAGCTTTTAAGGGCCAACACTTGAGGTCAGCCTGGCTCT 1260
Db 1201 CTTCATCTCTCATAGAGTCCACAGCTTTTAAGGGCCAACACTTGAGGTCAGCCTGGCTCT 1260
Qy 1261 CTCATTTGAGCTGGATAGAACATTTTAGAGCACCATCTATTCTTCAAGAGGAAGTTTAAA 1320
Db 1261 CTCATTTGAGCTGGATAGAACATTTTAGAGCACCATCTATTCTTCAAGAGGAAGTTTAAA 1320
Qy 1321 AATAAAAGAACCTTGAAGAGGAAAAAATGTAGACATTCAATCTAACCTTTTCATTTTACT 1380
Db 1321 AATAAAAGAACCTTGAAGAGGAAAAAATGTAGACATTCAATCTAACCTTTTCATTTTACT 1380
Qy 1381 AGCCAAAGCTAAATAGAATGCAGGTTACCTGTTTTTCAGCCAGGCACCATCATTTCCTAA 1440
Db 1381 AGCCAAAGCTAAATAGAATGCAGGTTACCTGTTTTTCAGCCAGGCACCATCATTTCCTAA 1440
Qy 1441 TTGTTATAAAATTTATTATTATTGTTGTTATTATTATTATTTGCCATAAGAAGTTTCCCA 1500
Db 1441 TTGTTATAAAATTTATTATTATTGTTGTTATTATTATTATTTGCCATAAGAAGTTTCCCA 1500
Qy 1501 TATCCTTTTAGTATAACAAAAACACAATTCACAAGCATTATAAAACCCATGGTGTCTAAC 1560
Db 1501 TATCCTTTTAGTATAACAAAAACACAATTCACAAGCATTATAAAACCCATGGTGTCTAAC 1560
Qy 1561 TATTAAAAAAATTAAGTGGAACACACTTGTCCCAGCTACTGGGGAGGCTGAGGAGGGAGG 1620
Db 1561 TATTAAAAAAATTAAGTGGAACACACTTGTCCCAGCTACTGGGGAGGCTGAGGAGGGAGG 1620
Qy 1621 ATCACGTGATCCCAGGGGGTCAAGGTTATGGAGAGCTATGATTGTGCCACTGCACTCCAG 1680
Db 1621 ATCACGTGATCCCAGGGGGTCAAGGTTATGGAGAGCTATGATTGTGCCACTGCACTCCAG 1680
Qy 1681 CCTGGGTGACAGGGAAAGACCCTGTCTCTAAAATTTTTTTTAAAAAAACTAAACTGGTTT 1740
Db 1681 CCTGGGTGACAGGGAAAGACCCTGTCTCTAAAATTTTTTTTAAAAAAACTAAACTGGTTT 1740
Qy 1741 TATTACAGAGATTCTGGAGACAGCTACACATAAAAGGGTGGTATGCCTCATATTAGCTAC 1800
Db 1741 TATTACAGAGATTCTGGAGACAGCTACACATAAAAGGGTGGTATGCCTCATATTAGCTAC 1800
Qy 1801 CCAGGGAGGTGGAATGCCAACTTAGGTGGTGTCACCACTATTAAAAATGCCCCAAAGCAA 1860
Db 1801 CCAGGGAGGTGGAATGCCAACTTAGGTGGTGTCACCACTATTAAAAATGCCCCAAAGCAA 1860
Qy 1861 TCAAAACTGAGAACTTCCTGGGAGCTTAGCATTGTGCAAAAGCAGCACAAAACACTTAAA 1920
Db 1861 TCAAAACTGAGAACTTCCTGGGAGCTTAGCATTGTGCAAAAGCAGCACAAAACACTTAAA 1920
Qy 1921 CAATTCACAGTTGTGTTGGAATGGGAAGGCCTGGAAATATAAACCAAAGAGTATATTGTC 1980
Db 1921 CAATTCACAGTTGTGTTGGAATGGGAAGGCCTGGAAATATAAACCAAAGAGTATATTGTC 1980
Qy 1981 TAAATTGATAGAGATTACAATTGCCTGAAAGAAAAAGTTGACTTTTAACTAGAATGTTCA 2040
Db 1981 TAAATTGATAGAGATTACAATTGCCTGAAAGAAAAAGTTGACTTTTAACTAGAATGTTCA 2040
Qy 2041 GAGTAGGTTTACAGAAGAAGCTCTTAAACTGGGCTCCAGTGGATTTGTCAATGCTTTGGA 2100
Db 2041 GAGTAGGTTTACAGAAGAAGCTCTTAAACTGGGCTCCAGTGGATTTGTCAATGCTTTGGA 2100
Qy 2101 AGCTGGTGGGGTGGGAGGGTTGGAGGGGGCATAAAAAGTCATGTTGGTATGCTCTGCTCA 2160
Db 2101 AGCTGGTGGGGTGGGAGGGTTGGAGGGGGCATAAAAAGTCATGTTGGTATGCTCTGCTCA 2160
Qy 2161 AGTCTCCATTCTGTTTCCTTTTCCTCTTTTCAATGTCATGTCCCATTATTTCATTATGGG 2220
Db 2161 AGTCTCCATTCTGTTTCCTTTTCCTCTTTTCAATGTCATGTCCCATTATTTCATTATGGG 2220

```
Qy      2221 CTTCCCTTTATCCAGGATCAATATGCCACCTCTTGGTTGTCTTTTACCTACTTCTCCACC 2280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2221 CTTCCCTTTATCCAGGATCAATATGCCACCTCTTGGTTGTCTTTTACCTACTTCTCCACC 2280

Qy      2281 TCACTATGGAATCGTCCTTGGGTAGCTCCTGTGCTTGGGAACCTGCACGGGCACTTTTCT 2340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2281 TCACTATGGAATCGTCCTTGGGTAGCTCCTGTGCTTGGGAACCTGCACGGGCACTTTTCT 2340

Qy      2341 GATGTCTTGATTCCAGCTTTACTCCTAAAACTTAAATGCTGAGGGGCCAACACCATGGCA 2400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2341 GATGTCTTGATTCCAGCTTTACTCCTAAAACTTAAATGCTGAGGGGCCAACACCATGGCA 2400

Qy      2401 GTGGTAGGGATGGGAATGGGGGTCTTGTAACACACTACATAAACTACACGAAATAAACTA 2460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2401 GTGGTAGGGATGGGAATGGGGGTCTTGTAACACACTACATAAACTACACGAAATAAACTA 2460

Qy      2461 CATGAAACTCAACATGTTTGCAAGACTCAGTTCACATCCATGAGGAGCTCATGCTTCTCC 2520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2461 CATGAAACTCAACATGTTTGCAAGACTCAGTTCACATCCATGAGGAGCTCATGCTTCTCC 2520

Qy      2521 CTCCTGCTCCCCTAGCACACATGATTATCTCTATTTGGAAATGTTTGGCATTTTTGGTGA 2580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2521 CTCCTGCTCCCCTAGCACACATGATTATCTCTATTTGGAAATGTTTGGCATTTTTGGTGA 2580

Qy      2581 AGTGAATGGTTCAATAACTTTCTCCACCATCAGAACAAAAGCTCTTTAAGGTTAGGGATG 2640
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2581 AGTGAATGGTTCAATAACTTTCTCCACCATCAGAACAAAAGCTCTTTAAGGTTAGGGATG 2640

Qy      2641 GGATCATACACACTTCCCTTGTCCAAGTCCCCATCACCCCTTATCTAGACAATTGCTACA 2700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2641 GGATCATACACACTTCCCTTGTCCAAGTCCCCATCACCCCTTATCTAGACAATTGCTACA 2700

Qy      2701 GTTTCCTACACACTCTTCTAACCTCTTGCAGTCTATTTTCATAAAACAGCTAGAGAACTT 2760
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2701 GTTTCCTACACACTCTTCTAACCTCTTGCAGTCTATTTTCATAAAACAGCTAGAGAACTT 2760

Qy      2761 TGAGATGTAAGTCAAAAAATAGAACATGTCGCTCTTTCCCATTGTTTTTGAAATAAAGTT 2820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2761 TGAGATGTAAGTCAAAAAATAGAACATGTCGCTCTTTCCCATTGTTTTTGAAATAAAGTT 2820

Qy      2821 CAACCCCCTTACCAGGGTCAACAAGGCCCTGCAATGATTTGGTCCTGTTAAAAATTCTTT 2880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2821 CAACCCCCTTACCAGGGTCAACAAGGCCCTGCAATGATTTGGTCCTGTTAAAAATTCTTT 2880

Qy      2881 AGCCTTAACTCATGCTGTTCTTCCTTACACTCACTGCATTCTAGCCATTGAGGTTTCTAT 2940
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2881 AGCCTTAACTCATGCTGTTCTTCCTTACACTCACTGCATTCTAGCCATTGAGGTTTCTAT 2940

Qy      2941 GCATCAAACTTTTTTTGGTCCCAGCACTGTGCACATCCTTCTGGGTAGAATGCCCCTTGA 3000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2941 GCATCAAACTTTTTTTGGTCCCAGCACTGTGCACATCCTTCTGGGTAGAATGCCCCTTGA 3000

Qy      3001 TTTGTATAATTAGCACCTCCTTCATCATTTAGGTCTTAGTATAACTACTACCTTCTTAGA 3060
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3001 TTTGTATAATTAGCACCTCCTTCATCATTTAGGTCTTAGTATAACTACTACCTTCTTAGA 3060

Qy      3061 GAAGCTCTGCTTCTTCATCCTATAAAAAAGTAAAATTCCTTACCCTGTTATTTTTTAAGT 3120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3061 GAAGCTCTGCTTCTTCATCCTATAAAAAAGTAAAATTCCTTACCCTGTTATTTTTTAAGT 3120

Qy      3121 CATCCGTGTTTCATTCTGTTAAAGTTCTTATCACAATTTATCATTATTTTATTTACAGTC 3180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3121 CATCCGTGTTTCATTCTGTTAAAGTTCTTATCACAATTTATCATTATTTTATTTACAGTC 3180

Qy      3181 ATGTGCCACATAACAATGTTTCAGTCAGGGATAGAACACAAATGTATCTGGCCCCATAAT 3240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3181 ATGTGCCACATAACAATGTTTCAGTCAGGGATAGAACACAAATGTATCTGGCCCCATAAT 3240

Qy      3241 ATTATAAGCTGAGAAATTTCTATTAACTAGTGATATCGCAGCCATCATAAGTGTAATGCA 3300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3241 ATTATAAGCTGAGAAATTTCTATTAACTAGTGATATCGCAGCCATCATAAGTGTAATGCA 3300

Qy      3301 GGACATTACCTTTTCTATGTTTAGATATGTTAGATACACAAATATATTTCATTGTGTTAT 3360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3301 GGACATTACCTTTTCTATGTTTAGATATGTTAGATACACAAATATATTTCATTGTGTTAT 3360

Qy      3361 AATTTCCTACAGTATTCAGTACAGTAACATGCTGTACAGGTTTGTAACCTAGGAGTAATA 3420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3361 AATTTCCTACAGTATTCAGTACAGTAACATGCTGTACAGGTTTGTAACCTAGGAGTAATA 3420

Qy      3421 GGCTATACCATACAGCTTAGGTGTGTAGTAGGCTATAACCATCTAGGTTTGTGTAAGTAC 3480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3421 GGCTATACCATACAGCTTAGGTGTGTAGTAGGCTATAACCATCTAGGTTTGTGTAAGTAC 3480

Qy      3481 ATTCTATGATATTCCCACAATGATGAAATCACCTAACTACACATTTCTCAGAATGTTTCA 3540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3481 ATTCTATGATATTCCCACAATGATGAAATCACCTAACTACACATTTCTCAGAATGTTTCA 3540

Qy      3541 CTGTTGTGAAGTGACCCATGACTATATTTTCCTATATACTTGATATTTTTGTGCATCTGC 3600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3541 CTGTTGTGAAGTGACCCATGACTATATTTTCCTATATACTTGATATTTTTGTGCATCTGC 3600

Qy      3601 CCATGAGAATGTAGTGTAAGATCAAAGGATGCAAGAATGGGTTCTATCCAGTATAGTACC 3660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3601 CCATGAGAATGTAGTGTAAGATCAAAGGATGCAAGAATGGGTTCTATCCAGTATAGTACC 3660

Qy      3661 CACTACACTGGTGGATGTCAATATGTATTTGTTAGATTAATATCTCAAGAATGAGCACCT 3720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3661 CACTACACTGGTGGATGTCAATATGTATTTGTTAGATTAATATCTCAAGAATGAGCACCT 3720

Qy      3721 TTCTCAGACACATAAAAGATGCTCAATATAAAAGTTTGTTGAACTGAACGTTATTGGCAA 3780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3721 TTCTCAGACACATAAAAGATGCTCAATATAAAAGTTTGTTGAACTGAACGTTATTGGCAA 3780

Qy      3781 ATGTAACATGATCGGATTTAAAGAGGAGCGAAACAGAGGTCTGGCTCAAACACCATACTT 3840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3781 ATGTAACATGATCGGATTTAAAGAGGAGCGAAACAGAGGTCTGGCTCAAACACCATACTT 3840

Qy      3841 CTAGAGTGCATAAGAGGTAGCAGTTGATTACCACTGGCGACAGGAGAAAAAAGAGCTTGA 3900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3841 CTAGAGTGCATAAGAGGTAGCAGTTGATTACCACTGGCGACAGGAGAAAAAAGAGCTTGA 3900

Qy      3901 CCGCAGGGTACTGTGAAGACATTTCAGGTTGATGGCACAGAACAGGGGAAATACATAAAT 3960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3901 CCGCAGGGTACTGTGAAGACATTTCAGGTTGATGGCACAGAACAGGGGAAATACATAAAT 3960

Qy      3961 GTGTGGGAATATTCAGTGGTCTGGGATGACTACATAGTAGAATATAATGAAGAAAAGAGT 4020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3961 GTGTGGGAATATTCAGTGGTCTGGGATGACTACATAGTAGAATATAATGAAGAAAAGAGT 4020

Qy      4021 GGAAGGGAAAGATGAAAAGTTGGAATGGGGATGAATTATGAAAGTACCAGAATGTTATGC 4080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4021 GGAAGGGAAAGATGAAAAGTTGGAATGGGGATGAATTATGAAAGTACCAGAATGTTATGC 4080

Qy      4081 TAAGGAATCTAGATTTTAAAATGTGAGGGCAAATTGAAGTCCTGGGCACGTTACAAAACT 4140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4081 TAAGGAATCTAGATTTTAAAATGTGAGGGCAAATTGAAGTCCTGGGCACGTTACAAAACT 4140

Qy      4141 AGAGGTCATAAAGTTTACCCTAATTTACCAAGATTTCCTAGAGGATCTATAATTGGAATC 4200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4141 AGAGGTCATAAAGTTTACCCTAATTTACCAAGATTTCCTAGAGGATCTATAATTGGAATC 4200

Qy      4201 CAGATCTGCCTCTCTGTAAAGTTCAAGCACTTTCCATGACACCATACTGTTTCTTTCCAC 4260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4201 CAGATCTGCCTCTCTGTAAAGTTCAAGCACTTTCCATGACACCATACTGTTTCTTTCCAC 4260

Qy      4261 CTGCACAATGCAAATGAACTCTTATGAAACTGCTGTTTCTATCCTGGGCTAAATGTTGCA 4320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4261 CTGCACAATGCAAATGAACTCTTATGAAACTGCTGTTTCTATCCTGGGCTAAATGTTGCA 4320

Qy      4321 GAAAAAAGATTTAATCTTTGGGATAAGGCTATTTTGGGTTTTCTCCTACTTCTTGGGAAA 4380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4321 GAAAAAAGATTTAATCTTTGGGATAAGGCTATTTTGGGTTTTCTCCTACTTCTTGGGAAA 4380

Qy      4381 CAAGGTTTTCTTCCCCTGGCTAATTAAGTGTGGTATTGTTCTTCCAGGGAAATCAGTGAT 4440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       4381 CAAGGTTTTCTTCCCCTGGCTAATTAAGTGTGGTATTGTTCTTCCAGGGAAATCAGTGAT 4440

Qy       4441 GCATCACCTGCTGCTATCAAATGTCAGGGTTGGAGTTCCTGATTTATTGCATGTGCCCAC 4500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4441 GCATCACCTGCTGCTATCAAATGTCAGGGTTGGAGTTCCTGATTTATTGCATGTGCCCAC 4500

Qy       4501 AAAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCT 4560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4501 AAAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCT 4560

Qy       4561 GTTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAA 4620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4561 GTTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAA 4620

Qy       4621 A 4621
              |
Db       4621 A 4621
```

RESULT 2
US-10-723-860-1125/c
; Sequence 1125, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1125
; LENGTH: 121600
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-723-860-1125

Query Match 7.7%; Score 356.6; DB 18; Length 121600;
Best Local Similarity 74.7%; Pred. No. 2.2e-62;
Matches 543; Conservative 0; Mismatches 169; Indels 15; Gaps 7;

```
Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAA--- 57
              || || |||||| ||||    ||||  |||||||||| |||||||||||||| |||||
Db      44776 AGATCGAGACCATCCTGGCTAACACAGTGAAACCCCGTCTCTACTAAAAATCCAAAAAAA 44717

Qy         58 --TTAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGG 115
                |||| | ||| |||||     | |||||| ||||||||||  |||||| ||||||||
Db      44716 AGTTAGCCGGGCGTGGTGGCGGGCACCTGTAGTCCCAGCTACTCTGGAGGCTGAGGCAGG 44657

Qy        116 AGAATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCAC 174
              ||||| | ||||| ||  || || |||| |||| | ||| | ||||   |||| |||||
Db      44656 AGAATGGCGTGAACTTG--GGAGGCGGAGCTTGCAGTGAGCCGAGATAGCGCCACTGCAC 44599

Qy        175 TCCAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATA 234
              ||||| |  ||| || ||||||| |   ||||||| |||||||| |      | ||||||
Db      44598 TCCAGTCTGGGCGAC-AGAGCGAGACTCCATCTCAAAAAAAAAATA----TATATATATA 44544

Qy        235 TATATATATATATATATACACACACACACACATATATATATACACATATATATACACATA 294
              || ||||||||| ||||||| | | | | | ||||||| ||| | | |||||||| | |||
Db      44543 TAGATATATATAGATATACAGATATATAGATATATATAGATATATAGATATATATAGATA 44484

Qy        295 TATACATATATACACATATATACACATATATATACACATACATATGTACACATATATATA 354
              |||| |||||||| | |||||||| | |||||||| | | ||| |||| || | ||||||| |
Db      44483 TATAGATATATATAGATATATAGATATATATAGATATATAGATATATATAGATATATAGA 44424

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              | ||||  ||| | | ||||| | | ||||| | | ||||| | | ||||| | | | |
Db      44423 TATATATAGATATATAGATATATATAGATATATAGATATATAGATATATATAGATATATA 44364

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              ||||| | | ||||| | ||||||| | ||||| | ||||||| | ||||||| |||||
Db      44363 GATATATATAGATATATAGATATATATAGATATATAGATATATAGATATATATAGATATA 44304

Qy        475 TACACATATATATAATATACACACATATATATACACATATATACACACATATATACACAT 534
              || |  |||| || ||||| | | ||||| ||| | ||| ||| | | ||||||| | ||
Db      44303 TAGATGTATAGAT-ATATATAGATATATAGATATATATAGATATATAGATATATAGATAT 44245

Qy        535 ATATACACATATATATACACATATATACACATATATACATATATACACATATATATACAT 594
              ||||| | ||||| ||| | ||| ||| | ||||| | ||||| | | ||| ||||| |
Db      44244 ATATAGATATATAGATATATATAGATATATATATAGATATATAGATATATAGATATATAG 44185

Qy        595 ATATACACATATATACATATATACA-CATATATACATATATACACACATATATACACATA 653
              ||||| | ||||||| |||||||  |  ||||||| ||||||| | | ||||| | | |||
Db      44184 ATATATAGATATATAGATATATAGATGATATATAGATATATATAGATATATAGATATATA 44125

Qy        654 CATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATA 713
              ||||| | | | ||||||||| ||| ||||| | ||||||| |  |||| |||| ||||
Db      44124 GATATATAGATATATAGATATATATAGATATATAGATATATAGATATATAGATATATATA 44065

Qy        714 TATATAT 720
              |||||||
Db      44064 GATATAT 44058
```

RESULT 3
US-10-741-600-17809
; Sequence 17809, Application US/10741600
; Publication No. US20050026169A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001499
; CURRENT APPLICATION NUMBER: US/10/741,600
; CURRENT FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 73997
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17809
; LENGTH: 109730
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-741-600-17809

Query Match 7.5%; Score 345.8; DB 19; Length 109730;
Best Local Similarity 73.7%; Pred. No. 3.6e-60;
Matches 537; Conservative 0; Mismatches 177; Indels 15; Gaps 7;

```
Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAATTA 60
              ||||  |||||||||||| |||||||| |||||||| |||| ||||||||||||||||||
Db     107515 AGTTTGAGACCAGCCTGACCAACATGGAGAAACCCCGTCTCCACTAAAAATACAAAATTA 107574

Qy         61 GACAGGCATGGTGATACACGCCTGT-AATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              |  |||||||||   ||  ||||| ||||||||||||||| |||||| ||||||||||||
Db     107575 GCGGGGCATGGTGGCGCATCCCTGTAAATCCCAGCTACTTGGGAGGCTGAGGCAGGAGAA 107634

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCCA 178
              |||||||||  | | || |||||||||||| || ||| | ||||   |||||| |||||||
Db     107635 TCACTTGAA--TCCAGGAGGTGGAGGTTGTGGTGAGCCGAGATGGAGCCATTACACTCCA 107692

Qy        179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAA--------AAGGCATTTTA 230
              |||  |||||||||||| |||   ||||||| ||||| | |         | | || |||
Db     107693 GCCTGGGCAACAAGAGCAAAACTCCATCTCAAAAAAATATATATATATGTATGTATATTA 107752

Qy        231 TATATATATATATATATATACACACACACACACATATATATATACACATATATATACA 290
              |||| ||||| ||||||||||  | | | | | | |||| |||||||| ||| ||||| |
Db     107753 TATAAATATAAATATATATATTTATATAAATATATATATTTATATACATATAAATATAAA 107812

Qy        291 CATATATACATATATACACATATATACACATATATATACACATACATATGTACACATATA 350
              |||||||  |||||||| |||||||    |||| |||| |||||  ||| || | |||||
Db     107813 TATATATATTTATATACATATATATATTTATATTTATATACATATGTATATAAATATATA 107872
```

```
Qy        351 TATACACATATGTATACACATATATACACATATATACACACATATATACACATATATACA 410
              |||  |  ||| || | | ||||||  | || |||| | | |||||||   || |||| |
Db     107873 TATTTATGTATATAAATATATATATTTATATGTATATAAATATATATATTTATGTATATA 107932

Qy        411 CACATATATACACATATATACACATATATACACATATACACATATATACACATATATACA 470
              | |||||||   |||| || | | ||||| | || || |  |||||| | ||||||
Db     107933 AATATATATATTTATATGTATATAAATATATATATCTATATGTATATAAATATATATTTT 107992

Qy        471 TATATACACATATATATAATATACACACATATATATACACATATATACACACATATATAC 530
              |||||  | ||| |||| ||||     | || |||||| | ||| ||    | || ||||
Db     107993 TATATGTATATAAATAT-ATATTTTTATATGTATATAAATATAGATTTTTATATGTATAT 108051

Qy        531 ACATATATACACATATATATACACATATATACACATATATACACATATATACACATATATAT 590
              | |||||||   |||| |||| | |||||||   |||| || ||| ||| | || |||||
Db     108052 AAATATATATTTATATGTATATAAATATATATTTATATGTATATAAATATATATTTATAT 108111

Qy        591 ACATATATACACATAT-ATACATATATACACATATATACATATATACACACATATATACA 649
              |||||  | | |||| | | |||| || | ||| ||| |||||||     ||||||  |
Db     108112 GTATATAAATATATATAACATATATTTATATATAAATATATATATATTTTTATATATGTA 108171

Qy        650 CATACATATACACACACATAGATATACATATATATACACATATATAT-ACGTATATATAT 708
              ||| ||||| | |   |||  |||| | ||||||| | |||||||| || ||||| |||
Db     108172 TATAAATATATATATTTATATGTATATAAATATATATAAATATATATAACATATATTTAT 108231

Qy        709 GTATATATA 717
              |||||||||
Db     108232 GTATATATA 108240
```

RESULT 4
US-10-235-192A-27/c
; Sequence 27, Application US/10235192A
; Publication No. US20040043389A1
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Jeanette
; TITLE OF INVENTION: Methods and Compositions for Identifying
; TITLE OF INVENTION: Risk Factors for Abnormal Lipid Levels and the Diseases
; TITLE OF INVENTION: and Disorders Associated Therewith
; FILE REFERENCE: MMI-011
; CURRENT APPLICATION NUMBER: US/10/235,192A
; CURRENT FILING DATE: 2002-09-04
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 188971
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 145329-145428, 187896-187995, 195894-195993
; OTHER INFORMATION: N = any nucleotide
US-10-235-192A-27

Query Match 7.1%; Score 327.4; DB 17; Length 188971;
Best Local Similarity 67.6%; Pred. No. 2.9e-56;
Matches 601; Conservative 0; Mismatches 261; Indels 27; Gaps 9;

```
Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC--AAAAT 58
              || ||||||||| ||||  |||||||||||||| || ||||||||||||||||  |||||
Db     104585 AGATCAAGACCATCCTGGCCAACATGGTGAAACTCCGTCTCTACTAAAAATACAAAAAAT 104526

Qy         59 TAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGA 118
              |||   ||| ||||| |   | |||||||||||||||||| |||||| || ||||||||
Db     104525 TAGCTGGGCGTGGTGGTGGGCTCCTGTAATCCCAGCTACTTGGGAGGCTGAAGCAGGAGA 104466

Qy        119 ATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTG---CAC 174
              ||| ||||||||  | || || |||||||||| | ||| |||||||  |||||||   |||
Db     104465 ATCGCTTGAACC--CAGGAGGCGGAGGTTGCAGTGAGCCAAGATCGCGCCATTGCACCAC 104408

Qy        175 TCCAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATA 234
              |||||||  ||||| ||||| ||||     ||||| ||||||||||   |    | | |
Db     104407 TCCAGCCTGGGCAAAAAGAGAGAAACTCTGTCTCAAAAAAAAAAAAAAAAAAAAAAAAAA 104348

Qy        235 TATATATATATATATATACACACACACACATATATATATACACATATATATACACATA 294
              ||||||||||||||||| | | | | ||||| | |||||||  ||||||||       |
Db     104347 AATATATATATATATATATATATATACACACACATATATATTCATATATA-------A 104295

Qy        295 TATACATATATACACATATATACACATATATATACACATACATATGTACACATATATATA 354
              | | |||||||  ||||||||   ||  ||||| | || ||||| ||  ||||||||||
Db     104294 TTCATATATATATTCATATATA---ATTCATATATATATTCATATATATTCATATATATA 104238

Qy        355 CACATATGT-ATACACATATATACACATATATACACACATATATACACATATATACACAC 413
              |||||  | ||| || ||||||  ||||||||  || |||||||  |||||||| | | |
Db     104237 TTCATATATAATTCATATATATATTCATATATTCATATATATATTCATATATATATTC 104178

Qy        414 ATATATACACATATATACACATATATACACATATACACATATATACACATATATACATAT 473
              |||||||    ||  | | |||||| || |||||  |||||||| |   ||||| ||||
Db     104177 ATATATA-----ATTCATATATATATTCATATATATTCATATATATATTCATATATATAT 104123

Qy        474 ATACACATATATATAATATACACACATATATATACACATATATACACACATATATACACA 533
              | | |||  |   |||||  |  ||||||||| || |||||||  || |||||| || |
Db     104122 TCATATATAATTCATATATATATTCATATATATTCATATATATATTCATATATATTCATA 104063

Qy        534 TATATACACATATATATACACATATATACACATATATACATATATACACATATATATACA 593
              ||||||  |||||||||  ||||||||  ||||||||  ||  |||| |||| ||| |
Db     104062 TATATATTCATATATATATTCATATATATTCATATATATATTCATATATATATTCATATA 104003

Qy        594 TATATACACATATATACATATATACACATATATACATATATACACACATATATACACATA 653
              ||||| || |||||||   ||||| | ||  |||| ||||||  || | ||||  | | |||
Db     104002 TATATTCATATATATATTCATATATATATTCATATATATATTCATATATATTCATATATA 103943

Qy        654 CATATACACACACATAGATATACA---TATATATACACATATATATACGTATATATATGT 710
              ||  | | | | ||  |||||  |    |||||||  |   |||||||  |  ||||||
Db     103942 TATTCATATATATATTCATATATAATTCATATATATATTCATATATATATTCATATATAA 103883

Qy        711 ATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGA 770
              ||  ||||||  | ||| |  |      |  |||   || ||    | |   |||||
Db     103882 ATTCATATATAATTCATATATATATTCATATAGATATTCATATATAATTCATATAGATAT 103823

Qy        771 AAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTA 830
              | |||  || | |    |  ||   || |  | || |     |     || ||| |
Db     103822 TCATAGATATTCATATATATTCATATAGATATTCATATATTCATATATATTCATATATTC 103763

Qy        831 GGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGA 879
              | | || |||| |  | |    | |   || |  |  |   || | |
Db     103762 ATGTATATTCATATAGATATTCATATATAGATATTCATAGATATTCATA 103714
```

RESULT 5
US-10-367-094-179
; Sequence 179, Application US/10367094
; Publication No. US20040170982A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc Malandro
; TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
; FILE REFERENCE: 529452001500
; CURRENT APPLICATION NUMBER: US/10/367,094
; CURRENT FILING DATE: 2003-02-14
; NUMBER OF SEQ ID NOS: 203
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 179
; LENGTH: 163382
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(163382)
; OTHER INFORMATION: n = A,T,C or G
US-10-367-094-179

Query Match 7.0%; Score 323.6; DB 18; Length 163382;
Best Local Similarity 68.2%; Pred. No. 1.6e-55;
Matches 495; Conservative 0; Mismatches 224; Indels 7; Gaps 3;

```
Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
```

```
Db     144405 AGTTCAAGACCAGCCTGGCCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAAATT 144464

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119

Db     144465 AGCCAGGCGTGGTGGCACGTGCCTGTAATCCCAGCTACCCAGGAGGCTGAGGCAGGAGAA 144524

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179

Db     144525 TTGCTTGAAGC-CAAGAGGCAGAAGTTGCAGTGAGCCGAGATCATGCCATTGCACTCCAG 144583

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239

Db     144584 CCTGGGGGACAAGAACGAGACTTCATGTCAAAAAAAAAAAATACACAGTCTTCCTTGGCT 144643

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299

Db     144644 TAGCTGTATGGGAAATATATATATATATTTATTTATATATATTTATATATATTTATTT 144703

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACAT 359

Db     144704 ATATATATTTATTTATATATATTTATTTATATATATTTATTTATATATTTATTTATATTT 144763

Qy        360 ATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATAT 419

Db     144764 ATTTTTATATATTTATTTATATATATTTGTATATATTTATTTATATATTTATTTATATAT 144823

Qy        420 ACACATATATACACATATATACACATATACACATATATACACATATATACATA-----TA 474

Db     144824 TTATTTATATATTTATATATATTAGTTCATATATATTTATATATTTATATATATTAGTTA 144883

Qy        475 TACACATATATATAATATACACACATATATATACACATATATACACACATATATACACAT 534

Db     144884 TATATATTTATATATTAGTTATATATATTTATATATATTTTTATATATTTATTTATATAT 144943

Qy        535 ATATACACATATATATACACATATATACACATATATACATATATACACATATATATACAT 594

Db     144944 ATGTATATATTAACATATTTATATATATTCATATATTTATATATAAATATATATATTAAT 145003

Qy        595 ATATACACATATATACATATATACACATATATACATATATACACACATATATACACATAC 654

Db     145004 ATATTTATATATATTTATATATTAATATATGTTTATATATTTATACATATTTATGTATAT 145063

Qy        655 ATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATAT 714

Db     145064 TAATATATATATTTATATATATTTATATATGTATTTATATATATTTATATATTTATATTT 145123

Qy        715 ATATAT 720

Db     145124 ATTTAT 145129


RESULT 6
US-10-221-714A-474/c
; Sequence 474, Application US/10221714A
; Publication No. US20040048254A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with
;  TITLE OF INVENTION:  tumor suppressor genes and oncogenes
;  FILE REFERENCE: 5013.1005
;  CURRENT APPLICATION NUMBER: US/10/221,714A
;  CURRENT FILING DATE:  2003-01-21
;  PRIOR APPLICATION NUMBER: PCT/EP01/02955
;  PRIOR FILING DATE: 2001-03-15
;  PRIOR APPLICATION NUMBER: DE 10013847.0
;  PRIOR FILING DATE: 2000-03-15
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 540
; SEQ ID NO 474
;   LENGTH: 26997
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-714A-474

  Query Match             7.0%;  Score 322.2;  DB 17;  Length 26997;
  Best Local Similarity   66.0%;  Pred. No. 1.4e-55;
  Matches  498;  Conservative    0;  Mismatches  253;  Indels    4;  Gaps    2;

Qy        207 TCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATATACACACACACACA 266

Db       6221 TAAATAAAAAAAATATAATATATATATATATATATATATATATACACACATATATATA 6162

Qy        267 TATATATATACACATATATATACACATATATACATATATACACATATATACACATATATA 326

Db       6161 AACATATATACACACACATATACATATATACACACATATATACATATATACACATATATA 6102

Qy        327 TACACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATA 386

Db       6101 TATATATATACATATATATACATATATATATACACATATATATATATACATATATATA 6042

Qy        387 CACACATATATACACATATATACACACATATATACACATATATACACATATATACACATA 446

Db       6041 CATATACATATACACATATATATATATACATATATATACATATACATATACACATATA 5982

Qy        447 TACACATATATACACATATATACATATATACACATATATATA-ATATACACACATATATA 505

Db       5981 TATATATATACATATACATACATATATATACACATATATATATATACATATACATACA 5922

Qy        506 TACACATATATACACACATATATACACATATATACACATATATATACACATATATACACA 565

Db       5921 TATATATACACATATATATATATATACATATATACACGTATATACATATATACACACATA 5862

Qy        566 TATATACATATATACACATATATATACATATATACACATATATACATATATACACA---T 622

Db       5861 TTTATATATACGTATATACATATATATATATATACATATATATATATATATATACATATT 5802

Qy        623 ATATACATATATACACACATATATACACATACATATACACACACATAGATATACATATAT 682

Db       5801 ATATATATATATATATACATATATAAATATACATATATATATATATATACATATATACAT 5742

Qy        683 ATACACATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTA 742

Db       5741 ATATATATATATAAATATATATATATATATACGTATATATATATACACACACACACCATA 5682

Qy        743 GCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAA 802

Db       5681 AAATACTATACACCCATAAAAAAAAAAAAAAATAACATTTACAACAACTTAAATAAAATTA 5622

Qy        803 GGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTAT 862

Db       5621 AAAACCATTATTCTAACCGAAATAACTCAAAAATAAAAAACCAAATATTATATATTCTCA 5562

Qy        863 CTGTAAGGTAGTTAAGAAGGTAACACTATATATATATATAGTGAAAGCAGTGTAAACCTA 922

Db       5561 CTTATAACTAAAAACTAAACTATAAAAACACAAAAACATAAAAATAATACAATAAACTTT 5502

Qy        923 AAGGATGGGCCAAGGATTTAAATGTTATAGAAGAA 957

Db       5501 AAAAACTCTAAAAAAAACATAAAAATAAATAAAAA 5467


RESULT 7
US-10-027-632-113621
; Sequence 113621, Application US/10027632
```

```
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 113621
;   LENGTH: 3310
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-113621

  Query Match             6.9%;  Score 320.6;  DB 13;  Length 3310;
  Best Local Similarity   65.7%;  Pred. No. 1.2e-55;
  Matches  559;  Conservative    0;  Mismatches  284;  Indels    8;  Gaps    6;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACA-AAATT 59
              |||||||| | || |||| ||| ||||||||||||| |||||||| |||||| |||||
Db       1488 AGTTCAAGTCTAGTCTGACCAATATGGTGAAACCCCGTCTCTACTGGAAATACAGAAATT 1547

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              || ||||  ||||  | | |||||||| ||| |||||||  |||||| ||||||||||||
Db       1548 AGCCAGGTGTGGTAGTGCTCGCCTGTAGTCCTAGCTACTCGGGAGGCTGAGGCAGGAGAA 1607

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCCA 178
              || |||||||||  || ||  |||||||| | ||| | ||||| ||||  |||||||||
Db       1608 TCGCTTGAACCTG--GGAGGCAGAGGTTGCAGTGAGCCGAGATCCTGCCCCTGCACTCCA 1665

Qy        179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATA 238
              |||   |  |  ||||||| ||    ||||| ||||||    | | | | || |||||||
Db       1666 GCC--TGGCAACAGAGCGAGATTCTGTCTCAAAAAAAACCCCATGTA-TATACATATATA 1722

Qy        239 TATATATATATATACACACACACACATATATATATACACATATATATACACATATATA 298
              ||||||||  | |  | | |   | |||||||| |||| ||| ||||||||  |||||||
Db       1723 CGTATATATACACATATATATGTATACATATATGTATATACACATATATACGTATATATA 1782

Qy        299 CATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACA 358
              | |||| |||  |||| |  |||| ||| |  || |||   | | | | || ||   |
Db       1783 TGTGTATATACACGTATATATGTATACATATATGTATATACACATATACACATGTATGTA 1842

Qy        359 TATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATA 418
              ||  |||| |  |||||||   ||||||   | | ||||    ||||||||||  |||||
Db       1843 TACATATATATGTATATACGTGTATATATGTATACATATGTGTATATATACACGTATATA 1902

Qy        419 TACACATATATACACATATATACACATATACACATATATACACATATATACATATATACA 478
              |||||||| || || | | || | |||      |||| ||||  ||||||||  || |
Db       1903 TACACATATGTATACGTGTGTATATATACGTGTATATACACACACGTATACATACGTATA 1962

Qy        479 CATATATAT-AATATACACACATATATATACACATATATACACACATATATACACATATA 537
              |||||  |  ||||| ||||| || |||||| || | || |    ||||| ||| | |
Db       1963 TATATACGTGTATATATACACACATGTATACATATGTGTATATGTGTATATATACACACA 2022

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATACATATA 597
              |  | | ||||||    |||||||||||||| || | ||| ||   ||||||  ||| |||
Db       2023 TGTATACATATATGTGTATATATACACATATGTATACATACACGTGTATATACACACATA 2082

Qy        598 TACACATATATACATATATACACATATATACATATATACACACATATATACACATACATA 657
              |  | | ||| || | |||| | | | |||  |||| | | |  | |||||| |  ||
Db       2083 TGTATACATACACGTGTATATATACACATATGTATACATATATGTGTATACATACGCACG 2142

Qy        658 TACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATA 717
              || |||||  ||  |||||||||| | || | |||    | |  || | |||||||| |||
Db       2143 TATACACATGTATGTATACATATTTGTATATATACTCACATATACACATATGTATACATA 2202

Qy        718 TATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGA 777
              |||           | |         ||  |  | |   | | | | |  | | | |
Db       2203 CATGTGTATATACACGCACACATACACATATATGTGCATATACGCACATATATACATATA 2262

Qy        778 GTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAA 837
              |  |||   || |||    | |   | ||       ||    ||  ||       | |
Db       2263 TGTACACATATGTATATATACACACATATATACATATATGTACACATGTACGTACACATA 2322

Qy        838 TACATAAATGT 848
              |||||| ||||
Db       2323 TACATATATGT 2333


RESULT 8
US-10-027-632-113622
; Sequence 113622, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 113622
;   LENGTH: 3310
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-113622

  Query Match             6.9%;  Score 320.6;  DB 13;  Length 3310;
  Best Local Similarity   65.7%;  Pred. No. 1.2e-55;
  Matches  559;  Conservative    0;  Mismatches  284;  Indels    8;  Gaps    6;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACA-AAATT 59
              |||||||| | || |||| ||| ||||||||||||| |||||||| |||||| |||||
Db       1488 AGTTCAAGTCTAGTCTGACCAATATGGTGAAACCCCGTCTCTACTGGAAATACAGAAATT 1547

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              || ||||  ||||  | | |||||||| ||| |||||||  |||||| ||||||||||||
Db       1548 AGCCAGGTGTGGTAGTGCTCGCCTGTAGTCCTAGCTACTCGGGAGGCTGAGGCAGGAGAA 1607

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCCA 178
              || |||||||||  || ||  |||||||| | ||| | ||||| ||||  |||||||||
Db       1608 TCGCTTGAACCTG--GGAGGCAGAGGTTGCAGTGAGCCGAGATCCTGCCCCTGCACTCCA 1665

Qy        179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATA 238
              |||   |  |  ||||||| ||    ||||| ||||||    | | | | || |||||||
```

```
Db       1666 GCC--TGGCAACAGAGCGAGATTCTGTCTCAAAAAAAACCCCATGTA-TATACATATATA 1722

Qy        239 TATATATATATATACACACACACACATATATATATACACATATATATACACATATATA 298
              |||||||| | | | | |   | |||||||| |||| ||| ||||||||  |||||||
Db       1723 CGTATATATACACATATATATGTATACATATATGTATATACACATATATACGTATATATA 1782

Qy        299 CATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACA 358
              | |||| |||  |||| |  |||| ||| |  || |||   | | | | || ||   |
Db       1783 TGTGTATATACACGTATATATGTATACATATATGTATATACACATATACACATGTATGTA 1842

Qy        359 TATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATA 418
              ||  |||| |  |||||||   ||||||   | | ||||    ||||||||||  |||||
Db       1843 TACATATATATGTATATACGTGTATATATGTATACATATGTGTATATATACACGTATATA 1902

Qy        419 TACACATATATACACATATATACACATATACACATATATACACATATATACATATATACA 478
              |||||||||.|| || | | || | |||     ||||| |||||  |||||||  || |
Db       1903 TACACATATGTATACGTGTGTATATATACGTGTATATACACACACGTATACATACGTATA 1962

Qy        479 CATATATAT-AATATACACACATATATATACACATATATACACACATATATACACATATA 537
              |||||  |  ||||| ||||| || |||||| || | || |    ||||| ||| | |
Db       1963 TATATACGTGTATATATACACACATGTATACATATGTGTATATGTGTATATATACACACA 2022

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATACATATA 597
              |  | | ||||||    |||||||||||||| || | ||| ||   |||||| ||| |||
Db       2023 TGTATACATATATGTGTATATATACACATATGTATACATACACGTGTATATACACACATA 2082

Qy        598 TACACATATATACATATATACACATATATACATATATACACACATATATACACATACATA 657
              |  | | ||| || | |||| | | | |||  |||| | | |  | |||||| |  ||
Db       2083 TGTATACATACACGTGTATATATACACATATGTATACATATATGTGTATACATACGCACG 2142

Qy        658 TACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATA 717
              || |||||  ||  |||||||||| | || | ||| |  | |  || | |||||||| |||
Db       2143 TATACACATGTATGTATACATATTTGTATATATACTCACATATACACATATGTATACATA 2202

Qy        718 TATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGA 777
              |||           | |         ||  |  | |  |  | | | |  | | | |
Db       2203 CATGTGTATATACACGCACACATACACATATATGTGCATATACGCACATATATACATATA 2262

Qy        778 GTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAA 837
              |  |||   || |||   | |  | ||      ||    ||  ||       | |
Db       2263 TGTACACATATGTATATATACACACATATATACATATATGTACACATGTACGTACACATA 2322

Qy        838 TACATAAATGT 848
              |||||| ||||
Db       2323 TACATATATGT 2333


RESULT 9
US-10-027-632-113621
; Sequence 113621, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720

;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 113621
;   LENGTH: 3310
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-113621

  Query Match             6.9%;  Score 320.6;  DB 17;  Length 3310;
  Best Local Similarity   65.7%;  Pred. No. 1.2e-55;
  Matches  559;  Conservative    0;  Mismatches  284;  Indels    8;  Gaps    6;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACA-AAATT 59
              |||||||| | || |||| ||| |||||||||||||| ||||||||  |||||||  |||||
Db       1488 AGTTCAAGTCTAGTCTGACCAATATGGTGAAACCCCGTCTCTACTGGAAATACAGAAATT 1547

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              || ||||  ||||  | | |||||||| ||| |||||||  |||||| ||||||||||||
Db       1548 AGCCAGGTGTGGTAGTGCTCGCCTGTAGTCCTAGCTACTCGGGAGGCTGAGGCAGGAGAA 1607

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCCA 178
              || |||||||||  || ||  |||||||| | ||| | ||||| ||||  |||||||||
Db       1608 TCGCTTGAACCTG--GGAGGCAGAGGTTGCAGTGAGCCGAGATCCTGCCCCTGCACTCCA 1665

Qy        179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATA 238
              |||  | |  ||||||| ||     ||||| ||||||    | | | | || |||||||
Db       1666 GCC--TGGCAACAGAGCGAGATTCTGTCTCAAAAAAAACCCCATGTA-TATACATATATA 1722

Qy        239 TATATATATATATACACACACACACATATATATATACACATATATATACACATATATA 298
              ||||||||| | | | |   | |||||||| |||| ||| |||||||  |||||||
Db       1723 CGTATATATACACATATATATGTATACATATATGTATATACACATATATACGTATATATA 1782

Qy        299 CATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACA 358
              | |||| |||  |||| |  |||| ||| |  || |||   | | | | || ||   |
Db       1783 TGTGTATATACACGTATATATGTATACATATATGTATATACACATATACACATGTATGTA 1842

Qy        359 TATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATA 418
              ||  |||| |  |||||||   ||||||   | | ||||    ||||||||||  |||||
Db       1843 TACATATATATGTATATACGTGTATATATGTATACATATGTGTATATATACACGTATATA 1902

Qy        419 TACACATATATACACATATATACACATATACACATATATACACATATATACATATATACA 478
              |||||||| || || | | || | |||     ||||| |||||  |||||||  || |
Db       1903 TACACATATGTATACGTGTGTATATATACGTGTATATACACACACGTATACATACGTATA 1962

Qy        479 CATATATAT-AATATACACACATATATATACACATATATACACACATATATACACATATA 537
              |||||  |  ||||| ||||| || |||||| || | || |    ||||| ||| | |
Db       1963 TATATACGTGTATATATACACACATGTATACATATGTGTATATGTGTATATATACACACA 2022

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATACATATA 597
              |  | | ||||||    |||||||||||||| || | ||| ||   |||||| ||| |||
Db       2023 TGTATACATATATGTGTATATATACACATATGTATACATACACGTGTATATACACACATA 2082

Qy        598 TACACATATATACATATATACACATATATACATATATACACACATATATACACATACATA 657
              |  | | ||| || | |||| | | | |||  |||| | | |  | |||||| |  ||
Db       2083 TGTATACATACACGTGTATATATACACATATGTATACATATATGTGTATACATACGCACG 2142

Qy        658 TACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATA 717
              || |||||  ||  |||||||||| | || | ||| |  | |  || | |||||||| |||
Db       2143 TATACACATGTATGTATACATATTTGTATATATACTCACATATACACATATGTATACATA 2202

Qy        718 TATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGA 777
              |||           | |         ||  |  | |  |  | | | |  | | | |
Db       2203 CATGTGTATATACACGCACACATACACATATATGTGCATATACGCACATATATACATATA 2262

Qy        778 GTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAA 837
              |  |||   || |||   | |  | ||      ||    ||  ||       | |
Db       2263 TGTACACATATGTATATATACACACATATATACATATATGTACACATGTACGTACACATA 2322

Qy        838 TACATAAATGT 848
              |||||| ||||
Db       2323 TACATATATGT 2333


RESULT 10
```

```
US-10-027-632-113622
; Sequence 113622, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 113622
;   LENGTH: 3310
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-113622

  Query Match             6.9%;  Score 320.6;  DB 17;  Length 3310;
  Best Local Similarity   65.7%;  Pred. No. 1.2e-55;
  Matches  559;  Conservative    0;  Mismatches  284;  Indels    8;  Gaps    6;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACA-AAATT 59
              |||||||| | || |||| ||| ||||||||||||| ||||||||  ||||||| |||||
Db       1488 AGTTCAAGTCTAGTCTGACCAATATGGTGAAACCCCGTCTCTACTGGAAATACAGAAATT 1547

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              || ||||  ||||  | | |||||||| ||| |||||||  |||||| ||||||||||||
Db       1548 AGCCAGGTGTGGTAGTGCTCGCCTGTAGTCCTAGCTACTCGGGAGGCTGAGGCAGGAGAA 1607

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCCA 178
              || |||||||||| || ||  |||||||| | ||| | ||||| ||||  |||||||||
Db       1608 TCGCTTGAACCTG--GGAGGCAGAGGTTGCAGTGAGCCGAGATCCTGCCCCTGCACTCCA 1665

Qy        179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATA 238
              |||   |  |  ||||||| ||    ||||| ||||||    | | | | || |||||||
Db       1666 GCC--TGGCAACAGAGCGAGATTCTGTCTCAAAAAAAACCCCATGTA-TATACATATATA 1722

Qy        239 TATATATATATATACACACACACACATATATATATACACATATATATACACATATATA 298
              |||||||| | | | | |   | |||||||| |||| ||| |||||||||  |||||||
Db       1723 CGTATATATACACATATATATGTATACATATATGTATATACACATATATACGTATATATA 1782

Qy        299 CATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACA 358
              | |||| |||  |||| |  |||| ||| |  || |||   | | | | || ||   |
Db       1783 TGTGTATATACACGTATATATGTATACATATATGTATATACACATATACACATGTATGTA 1842

Qy        359 TATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATA 418
              ||  |||| |  |||||||  ||||||  | | ||||   |||||||||  |||||
Db       1843 TACATATATATGTATATACGTGTATATATGTATACATATGTGTATATATACACGTATATA 1902

Qy        419 TACACATATATACACATATATACACATATACACATATATACACATATATACATATATACA 478
              |||||||| || || | | || | |||     |||| ||||  |||||||  || |
Db       1903 TACACATATGTATACGTGTGTATATATACGTGTATATACACACACGTATACATACGTATA 1962

Qy        479 CATATATAT-AATATACACACATATATATACACATATATACACACATATATACACATATA 537
              |||||  |  |||| |||| || |||||| || | || |    |||||| ||| | |
Db       1963 TATATACGTGTATATATACACACATGTATACATATGTGTATATGTGTATATATACACACA 2022

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATACATATA 597
              |  | | ||||||    |||||||||||||| || | ||| ||   |||||| ||| |||
Db       2023 TGTATACATATATGTGTATATATACACATATGTATACATACACGTGTATATACACACATA 2082

Qy        598 TACACATATATACATATATACACATATATACATATATACACACATATATACACATACATA 657
              |  | | ||| || | |||| | | | |||  |||| | | |  | |||||| |  ||
Db       2083 TGTATACATACACGTGTATATATACACATATGTATACATATATGTGTATACATACGCACG 2142

Qy        658 TACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATA 717
              || |||||  ||  ||||||||| | || | |||   | |  || | ||||||||| |||
Db       2143 TATACACATGTATGTATACATATTTGTATATATACTCACATATACACATATGTATACATA 2202

Qy        718 TATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGA 777
              |||           | |         || |  | | |  | | | | |  | | | |
Db       2203 CATGTGTATATACACGCACACATACACATATATGTGCATATACGCACATATATACATATA 2262

Qy        778 GTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAA 837
              |  |||   || |||   | |  | ||      ||   ||  ||      | |
Db       2263 TGTACACATATGTATATATACACACATATATACATATATGTACACATGTACGTACACATA 2322

Qy        838 TACATAAATGT 848
              |||||| ||||
Db       2323 TACATATATGT 2333


RESULT 11
US-10-221-714A-473/c
; Sequence 473, Application US/10221714A
; Publication No. US20040048254A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with
;  TITLE OF INVENTION:  tumor suppressor genes and oncogenes
;  FILE REFERENCE: 5013.1005
;  CURRENT APPLICATION NUMBER: US/10/221,714A
;  CURRENT FILING DATE:  2003-01-21
;  PRIOR APPLICATION NUMBER: PCT/EP01/02955
;  PRIOR FILING DATE: 2001-03-15
;  PRIOR APPLICATION NUMBER: DE 10013847.0
;  PRIOR FILING DATE: 2000-03-15
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 540
; SEQ ID NO 473
;   LENGTH: 26997
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-714A-473

  Query Match             6.7%;  Score 311.2;  DB 17;  Length 26997;
  Best Local Similarity   76.9%;  Pred. No. 2.7e-53;
  Matches  409;  Conservative    0;  Mismatches  113;  Indels   10;  Gaps    2;

Qy        203 CATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACA 262
              | | ||| ||| | | ||        | ||||||||||||||||| | | || || | | |
Db      21339 CTTTTCATAAATACATAATATTCCATAATATATATATATATATACACACGCATATATA 21280

Qy        263 CACATATATATACACATATATATACACATATATACATATATACACATATATACACATA 322
              ||||||||   | ||| ||| ||||| | |||||||||||||| ||| ||||||| |  ||
Db      21279 CACATATACCCACACATATACATATATATATATACATATACACATATATATATCTA 21220

Qy        323 TATATACACATACATATGT-----ACACATATATATACACATATGTATACACATATATAC 377
              |||||||| | |||  |      | | ||| | ||| |||||| |||| ||| |||||||
Db      21219 CATATACACACATATACATATATAATATATACACATATACATATATATATACACATATAC 21160
```

```
Qy        378 ACATATATACACACATATATACACATATATACACACATATATACACATATATACACATAT 437
              ||||||| | || || ||||| |||||||||| | ||||||||||||| |||| ||| |||
Db      21159 ACATATACATACGCACATATAAACATATATATATACATATATACACGTATACACATATAC 21100

Qy        438 ATACACATATACACATATATACACATATATACATATATACACATATATATA-----ATAT 492
              ||| ||| ||| | |||||||||| ||||| ||||||| | | ||| |||||     |||
Db      21099 ATATACACATATATATATATACATATATACACATATACATATATACATATATATATAC 21040

Qy        493 ACACACATATATATACACATATATACACACATATATACACATATATACACATATATATAC 552
              | | | | | ||||||| ||||| | | | ||||||| ||||||||||||||||| |||
Db      21039 ATATATACACATATACATATATACATATATATATATACATATATACACATATACATAT 20980

Qy        553 ACATATATACACATATATACATATATACACATATATATACATATATACACATATATACAT 612
              | | ||||| | |||||||||||||||||||||||| ||| ||| ||| | ||||||| |
Db      20979 ATACATATATATATATATACATATATACACATATACATATATACATATATATATATAC 20920

Qy        613 ATATACACATATATACATATATACACACATATATACACATACATATACACACACATAGAT 672
              ||||| ||| |||||||||||||| | ||||||| | ||| ||||| | | | ||| ||
Db      20919 ATATATACACATATACATATATACATATATATATATATATATATATATATATATACAT 20860

Qy        673 ATACATATATATACACATATATATACGTATATATATGTATATATATATGCTC 724
              ||| |||||||||| |||||| ||  |||||||| |||||||||| | |
Db      20859 ATATATATATATACATATATATCTATATATATATATATATATATATATACAC 20808
```

RESULT 12
US-10-017-161-2383
; Sequence 2383, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 084335/0152
; CURRENT APPLICATION NUMBER: US/10/017,161
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: JP 2001/246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2430
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2383
; LENGTH: 1990
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: source
; LOCATION: (1)..(1990)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (201)..(1790)
US-10-017-161-2383

Query Match 6.7%; Score 308; DB 15; Length 1990;
Best Local Similarity 80.2%; Pred. No. 3.9e-53;
Matches 401; Conservative 0; Mismatches 90; Indels 9; Gaps 3;

```
Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | || | | |||| |||||||| | | | | ||||||| || ||||||||||||||||||
Db        826 TATACACACATATGTATATATACACATATATACACACATATGTATATATACACATATATA 885

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              |||||||| || ||||| ||| ||||| |||||||| |||| | | |||||| ||||||
Db        886 CACACATATGTATATATACACATATATACACACATATGTATATATACACATATATACACA 945

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||| || | |||   ||| | | |  || |  |||||| ||||||||||||||||||| |
Db        946 CATATGTATATATACACATATATACACGTATATGTATATATACACATATATACACATACA 1005

Qy        407 TACACA----CATATATACACATATATACACATATATACACATATACACATATATACACA 462
              ||||||    ||||| |||||||||||||||| ||||||||||| ||||||||||
Db       1006 TACACATATGTATATACACACATATATACACATACATACACATATACGTATATATACACA 1065

Qy        463 TATATACATATATACACATATATATAATATACACACATATATATACACATATATACACAC 522
              |||| ||||| |||||||||| | ||||| |||||||||| ||| ||   || | ||
Db       1066 TATACACATACATACACATATACGT-ATATATACACATATATACACATATGCGTATATAC 1124

Qy        523 ATATATACACATATATACACATATATATACACATATATACACATATATACATATATACAC 582
              |||||||||||||||||||    ||||||  |||||||||||||||||||| ||||||
Db       1125 ATATATACACATATATACAC----ATATACGTATATATACACATATATACACATATACGT 1180

Qy        583 ATATATATACATATATACACATATATACATATATACACATATATACATATATACACACAT 642
              ||||||| |||||||||||||||||||| ||||| |||||| |||||||||||| ||
Db       1181 ATATATACACATATATACACATATATACACATATATGTATATACACATATATACACATAT 1240

Qy        643 ATATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTAT 702
              |||| | | ||||||| ||||| ||| |||| ||| ||||| |||  ||| ||  |||
Db       1241 GTATATATACACATATATACACATATATGTATATATACATATATACACGTATGTATATAT 1300

Qy        703 ATATATGTATATATATATGC 722
              | | || ||| ||||||| |
Db       1301 ACACATATATGTATATATAC 1320

RESULT 13
US-10-292-798-2025
; Sequence 2025, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
; FILE REFERENCE: 084335/166
; CURRENT APPLICATION NUMBER: US/10/292,798
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: 10/017,161
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: JP 2001-246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2070
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2025
; LENGTH: 1990
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; LOCATION: source
; FEATURE:
; LOCATION: (1)..(1990)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (201)..(1790)
US-10-292-798-2025

Query Match 6.7%; Score 308; DB 17; Length 1990;
Best Local Similarity 80.2%; Pred. No. 3.9e-53;
Matches 401; Conservative 0; Mismatches 90; Indels 9; Gaps 3;

```
Qy   227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | || | | |||| |||||||| | | | | ||||||| || ||||||||||||||||||
Db   826 TATACACACATATGTATATATACACATATATACACACATATGTATATATACACATATATA 885

Qy   287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              |||||||| || ||||| ||| ||||| |||||||| |||| | | |||||| ||||||
Db   886 CACACATATGTATATATACACATATATACACACATATGTATATATACACATATATACACA 945

Qy   347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||| || | ||| ||| | | | || | |||||| ||||||||||||||||||| |
Db   946 CATATGTATATATACACATATATACACGTATATGTATATATACACATATATACACATACA 1005

Qy   407 TACACA----CATATATACACATATATACACATATATACACATATACACATATATACACA 462
              |||||| ||||| |||||||||||||||| ||||||||||| ||||||||||
```
```

```
Db       1006 TACACATATGTATATATACACACATATATACACATACATACACATATACGTATATATATACACA 1065

Qy        463 TATATACATATATATACACATATATATAATATACACACATATATATACACATATATACACAC 522
              |||| ||||| |||||||||||  | ||||| |||||||||| ||| ||   || | ||
Db       1066 TATACACATACATACACATATACGT-ATATATACACATATATACACATATGCGTATATAC 1124

Qy        523 ATATATACACATATATACACATATATATACACATATATACACATATATACATATATACAC 582
              ||||||||||||||||||||    ||||||  ||||||||||||||||||| ||||||
Db       1125 ATATATACACATATATACAC----ATATACGTATATATACACATATATACACATATACGT 1180

Qy        583 ATATATATACATATATACACATATATACATATATACACATATATACATATATACACACAT 642
              ||||||| |||||||||||||||||||| |||||   ||||| |||||||||||| ||
Db       1181 ATATATACACATATATACACATATATACACATATATGTATATACACATATATACACATAT 1240

Qy        643 ATATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTAT 702
              |||| | | ||||||| ||||| |||  |||| ||| ||||| |||  ||| ||  |||
Db       1241 GTATATATACACATATATACACATATATGTATATATACATATATACACGTATGTATATAT 1300

Qy        703 ATATATGTATATATATATGC 722
              | | || ||| ||||||| |
Db       1301 ACACATATATGTATATATAC 1320


RESULT 14
US-10-437-963-12666
; Sequence 12666, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 12666
;   LENGTH: 1800
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_18773C.1
US-10-437-963-12666

  Query Match             6.5%;  Score 300.6;  DB 18;  Length 1800;
  Best Local Similarity   78.2%;  Pred. No. 1.2e-51;
  Matches  387;  Conservative    0;  Mismatches  104;  Indels    4;  Gaps    2;

Qy        229 TATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 288
              |||||||||||| |||||||||| | | | ||| | ||| |||||||| ||||| |||
Db       1298 TATATATATATACATATATATATATATATATACATATATACATATATACATATATACATA 1357

Qy        289 CACATATATACATATATACACATATATAC---ACATATATATACACATACATATGTACAC 345
              ||||||| | ||||| ||||| ||||||    ||||||| ||| | |||||||| ||
Db       1358 TACATATACATATATACACACACATATACATAACATATACATATATATACATATATATGT 1417

Qy        346 ATATATATACACATATGTATACACATATATACACATATATACACACATATATACACATAT 405
              ||||||| | | |||| || | | ||| ||| | || |||| | |||||||| ||||||
Db       1418 ATATATACATATATATATACATATATACATATATGTATATATACATATATATACATAT 1477

Qy        406 ATACACACATATATACACATATATACACATATATACACATATACACATATATACACATAT 465
              ||| ||  |||||||| ||||||||  |||||||| ||||| |||||  || |. | |
Db       1478 ATATACGTATATATACATATATATACGTATATATACATATATATACATACGTATATACGT 1537

Qy        466 ATACATATATACACATATATATAATATACACACATATATATACACATATATACACACATA 525
              |||||||  || | ||| |||| |||||||| || |||||||||| ||||||||| || ||
Db       1538 ATACATACGTATATATACATAT-ATATACATACGTATATATACATATATATACATACGTA 1596

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| ||||||||| |||||||||| | ||||| | | ||||| | ||||| |
Db       1597 TATATACATATATATACATATATATATATATACATATATATACATATATACATTTTCG 1656

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| |||||||    ||||||| |||| || | | |||||
Db       1657 TATATATATATATATATATATATATATATATGTATATATATATATGTATATATATATA 1716

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | |   ||| ||||| |||| |||| | |||||| ||  ||||||
Db       1717 TATATATATATATATATATGTATATATATATATATGTATATATATATATGTATATATATA 1776

Qy        706 TATGTATATATATAT 720
              ||||||||||||||||
Db       1777 TATGTATATATATAT 1791


RESULT 15
US-09-783-203-1
; Sequence 1, Application US/09783203
; Patent No. US20020098582A1
; GENERAL INFORMATION:
;  APPLICANT: Geron Corporation
;  APPLICANT:  Gold, Joseph
;  APPLICANT:  Lebkowski, Jane
;  TITLE OF INVENTION: Tpacked stem cells
;  FILE REFERENCE: 096/003
;  CURRENT APPLICATION NUMBER: US/09/783,203
;  CURRENT FILING DATE:  2001-02-13
;  PRIOR APPLICATION NUMBER: 60/253,443
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/253,357
;  PRIOR FILING DATE: 2000-11-27
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
;   LENGTH: 15418
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-783-203-1

  Query Match             6.4%;  Score 296;  DB 9;  Length 15418;
  Best Local Similarity   69.2%;  Pred. No. 2.9e-50;
  Matches  479;  Conservative    0;  Mismatches  200;  Indels   13;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||| |||||||||||  ||| || | |||||||||||| |||||||| || ||||||
Db        646 AGTTCGAGACCAGCCTGGCCAATATAGCGAAACCCCATCTGTACTAAAAACACAAAAATT 705

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   ||  |||||| ||| |||||||||||||| ||||  ||||||  ||||||||||
Db        706 AGCTGGGTGTGGTGACACATGCCTGTAATCCCAGGTACTCAGGAGGCTAAGGCAGGAGAA 765

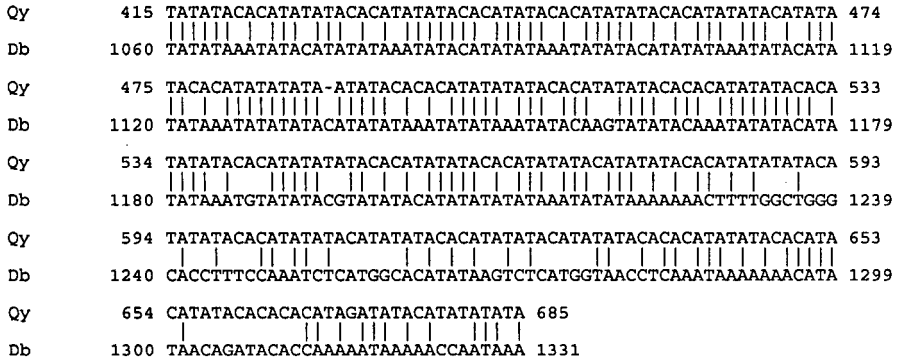
Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              |  |||||||  |  ||  | || |||     |  ||||||   ||| ||||||||||
Db        766 TTGCTTGAACTGGGAGGCAGAGGTGGT-----GAGCCAAGATTGCACCAGTGCACTCCAG 820

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  ||  || |||| ||||   ||||||| ||| ||| |||   | |  ||||| |||
Db        821 CCTTGGTGAC-AGAGTGAAACTCCATCTCAAAAACAAACAAACAAAATACATATACATAA 879

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              |||||||  | | | | | ||| | ||| ||||||| ||| ||||||| |  |||||||
Db        880 ATATATATGCACATATATATACATATATAAATATATATACACATATATAAATCTATATAC 939

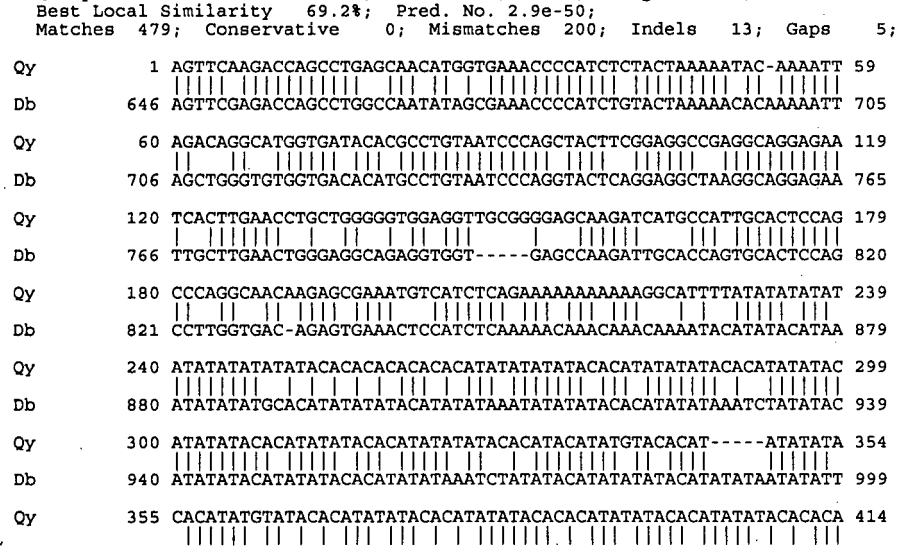
Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACAT-----ATATATA 354
              ||||||||| ||||| ||| ||||| ||  | |||||||| || ||||     |||||||
Db        940 ATATATACATATATACACATATATAAATCTATATACATATATATACATATATAATATATT 999

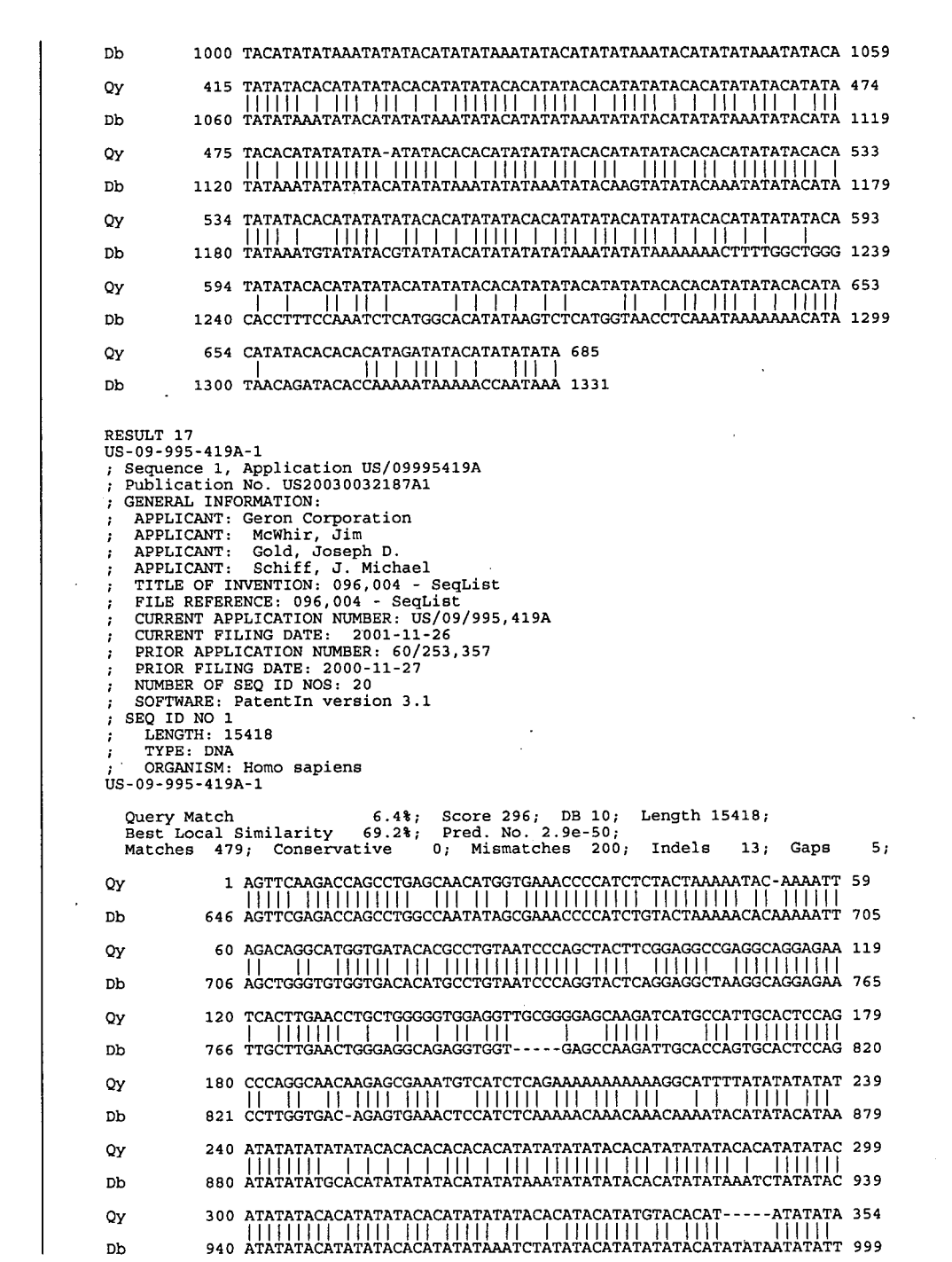
Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              |||||| || | | ||| ||| | | ||||||| | ||| ||||| ||||| | | |||
Db       1000 TACATATATAAATATATACATATATAAATATACATATATAAATACATATATAAATATACA 1059
```



RESULT 16
US-09-994-427A-1
; Sequence 1, Application US/09994427A
; Patent No. US20020128221A1
; GENERAL INFORMATION:
; APPLICANT: Geron Corporation
; APPLICANT: Schiff, J. Michael
; TITLE OF INVENTION: GLYCOSYLTRANSFERASE VECTORS FOR TREATING CANCER
; FILE REFERENCE: 083,002
; CURRENT APPLICATION NUMBER: US/09/994,427A
; CURRENT FILING DATE: 2002-02-26
; PRIOR APPLICATION NUMBER: 60/253,395
; PRIOR FILING DATE: 2000-11-27
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 15418
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-994-427A-1

Query Match 6.4%; Score 296; DB 9; Length 15418;
Best Local Similarity 69.2%; Pred. No. 2.9e-50;
Matches 479; Conservative 0; Mismatches 200; Indels 13; Gaps 5;





RESULT 17
US-09-995-419A-1
; Sequence 1, Application US/09995419A
; Publication No. US20030032187A1
; GENERAL INFORMATION:
; APPLICANT: Geron Corporation
; APPLICANT: McWhir, Jim
; APPLICANT: Gold, Joseph D.
; APPLICANT: Schiff, J. Michael
; TITLE OF INVENTION: 096,004 - SeqList
; FILE REFERENCE: 096,004 - SeqList
; CURRENT APPLICATION NUMBER: US/09/995,419A
; CURRENT FILING DATE: 2001-11-26
; PRIOR APPLICATION NUMBER: 60/253,357
; PRIOR FILING DATE: 2000-11-27
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 15418
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-995-419A-1

Query Match 6.4%; Score 296; DB 10; Length 15418;
Best Local Similarity 69.2%; Pred. No. 2.9e-50;
Matches 479; Conservative 0; Mismatches 200; Indels 13; Gaps 5;

```
Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              |||||| || | | ||| ||| | | ||||||| | ||| ||||| ||||| | | |||
Db       1000 TACATATATAAATATATACATATATAAATATACATATATAAATACATATATAAATATACA 1059

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              |||||| | ||| ||| | | ||||||| ||||| | ||||| | | ||| ||| | |||
Db       1060 TATATAAATATACATATATAAATATACATATATAAATATATACATATATAAATATACATA 1119

Qy        475 TACACATATATATA-ATATACACACATATATATACACATATATACACACATATATACACA 533
              || | ||||||||| ||||| | | ||||| ||| |||  |||| ||| |||||||| |
Db       1120 TATAAATATATATACATATATAAATATATAAATATACAAGTATATACAAATATATACATA 1179

Qy        534 TATATACACATATATATACACATATATACACATATATACATATATACACATATATATACA 593
              |||| |   |||||  || | | ||||| | ||| ||| ||| | | || | |   |
Db       1180 TATAAATGTATATACGTATATACATATATATATAAATATATAAAAAAACTTTTGGCTGGG 1239

Qy        594 TATATACACATATATACATATATACACATATATACATATATACACACATATATACACATA 653
              | |   || || |      | | | |  |      || | || ||| | | |||||
Db       1240 CACCTTTCCAAATCTCATGGCACATATAAGTCTCATGGTAACCTCAAATAAAAAAACATA 1299

Qy        654 CATATACACACACATAGATATACATATATATA 685
              |     || | ||| | |   ||| |
Db       1300 TAACAGATACACCAAAAATAAAAACCAATAAA 1331
```

RESULT 18
US-10-141-220-1
; Sequence 1, Application US/10141220
; Publication No. US20030040111A1
; GENERAL INFORMATION:
; APPLICANT: Geron Corporation
; APPLICANT: Gold, Joseph
; APPLICANT: Lebkowski, Jane
; TITLE OF INVENTION: Tpacked stem cells
; FILE REFERENCE: 096/003
; CURRENT APPLICATION NUMBER: US/10/141,220
; CURRENT FILING DATE: 2002-05-07
; PRIOR APPLICATION NUMBER: US/09/783,203
; PRIOR FILING DATE: 2001-02-13
; PRIOR APPLICATION NUMBER: 60/253,443
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/253,357
; PRIOR FILING DATE: 2000-11-27
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 15418
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-141-220-1

Query Match 6.4%; Score 296; DB 14; Length 15418;
Best Local Similarity 69.2%; Pred. No. 2.9e-50;
Matches 479; Conservative 0; Mismatches 200; Indels 13; Gaps 5;

```
Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||| |||||||||||  ||| || | ||||||||||| |||||||| || ||||||
Db        646 AGTTCGAGACCAGCCTGGCCAATATAGCGAAACCCCATCTGTACTAAAAACACAAAAATT 705

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   ||  |||||| ||| |||||||||||||| |||| ||||||  ||||||||||
Db        706 AGCTGGGTGTGGTGACACATGCCTGTAATCCCAGGTACTCAGGAGGCTAAGGCAGGAGAA 765

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              | |||||||  |  ||  | || |||     |   ||||||    ||| ||||||||||
Db        766 TTGCTTGAACTGGGAGGCAGAGGTGGT-----GAGCCAAGATTGCACCAGTGCACTCCAG 820

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  ||  || |||| ||||   ||||||| ||| ||| |||   | |  ||||| |||
Db        821 CCTTGGTGAC-AGAGTGAAACTCCATCTCAAAAACAAACAAACAAAATACATATACATAA 879

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              |||||||| | | | | ||| | ||| ||||||| ||| ||||||| |  |||||||
Db        880 ATATATATGCACATATATATACATATATAAATATATATACACATATATAAATCTATATAC 939

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACAT-----ATATATA 354
              |||||||| ||||| ||| ||||| ||  | |||||||| || ||||     ||||||
Db        940 ATATATACATATATACACATATATAAATCTATATACATATATATACATATATAATATATT 999

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              |||||| || | | ||| ||| | | ||||||| | ||| ||||| ||||| | | |||
Db       1000 TACATATATAAATATATACATATATAAATATACATATATAAATACATATATAAATATACA 1059

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              |||||| | ||| ||| | | ||||||| ||||| | ||||| | | ||| ||| | |||
Db       1060 TATATAAATATACATATATAAATATACATATATAAATATATACATATATAAATATACATA 1119

Qy        475 TACACATATATATA-ATATACACACATATATATACACATATATACACACATATATACACA 533
              || | ||||||||| ||||| | | ||||| ||| |||  |||| ||| |||||||| |
Db       1120 TATAAATATATATACATATATAAATATATAAATATACAAGTATATACAAATATATACATA 1179

Qy        534 TATATACACATATATATACACATATATACACATATATACATATATACACATATATATACA 593
              |||| |   |||||  || | | ||||| | ||| ||| ||| | | || | |   |
Db       1180 TATAAATGTATATACGTATATACATATATATATAAATATATAAAAAAACTTTTGGCTGGG 1239

Qy        594 TATATACACATATATACATATATACACATATATACATATATACACACATATATACACATA 653
              | |   || || |      | | | |  |      || | || ||| | | |||||
Db       1240 CACCTTTCCAAATCTCATGGCACATATAAGTCTCATGGTAACCTCAAATAAAAAAACATA 1299

Qy        654 CATATACACACACATAGATATACATATATATA 685
              |     || | ||| | |   ||| |
Db       1300 TAACAGATACACCAAAAATAAAAACCAATAAA 1331
```

RESULT 19
US-10-023-969-1
; Sequence 1, Application US/10023969
; Publication No. US20030095989A1
; GENERAL INFORMATION:
; APPLICANT: Geron Corporation
; APPLICANT: Irving, John
; APPLICANT: Lebkowski, Jane
; TITLE OF INVENTION: Chimeric Cytolytic Viruses for Cancer Treatment
; FILE REFERENCE: 084,002
; CURRENT APPLICATION NUMBER: US/10/023,969
; CURRENT FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: 60/256,418
; PRIOR FILING DATE: 2000-12-18
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 15418
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-023-969-1

Query Match 6.4%; Score 296; DB 14; Length 15418;
Best Local Similarity 69.2%; Pred. No. 2.9e-50;
Matches 479; Conservative 0; Mismatches 200; Indels 13; Gaps 5;

```
Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||| |||||||||||  ||| || | ||||||||||| |||||||| || ||||||
Db        646 AGTTCGAGACCAGCCTGGCCAATATAGCGAAACCCCATCTGTACTAAAAACACAAAAATT 705

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   ||  |||||| ||| |||||||||||||| |||| ||||||  ||||||||||
Db        706 AGCTGGGTGTGGTGACACATGCCTGTAATCCCAGGTACTCAGGAGGCTAAGGCAGGAGAA 765

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              | |||||||  |  ||  | || |||     |   ||||||    ||| ||||||||||
Db        766 TTGCTTGAACTGGGAGGCAGAGGTGGT-----GAGCCAAGATTGCACCAGTGCACTCCAG 820

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  ||  || |||| ||||   ||||||| ||| ||| |||   | |  ||||| |||
Db        821 CCTTGGTGAC-AGAGTGAAACTCCATCTCAAAAACAAACAAACAAAATACATATACATAA 879
```

```
Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              ||||||||  | | | | | ||| | ||| ||||||| ||| ||||||| |  |||||||
Db        880 ATATATATGCACATATATATACATATATAAATATATATACACATATATAAATCTATATAC 939

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACAT-----ATATATA 354
              ||||||||| ||||| ||| ||||| ||  | |||||||| || ||||     ||||||
Db        940 ATATATACATATATACACATATATAAATCTATATACATATATATACATATATAATATATT 999

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              |||||| || | | ||| ||| | | ||||||| | ||| ||||| ||||| | | |||
Db       1000 TACATATATAAATATATACATATATAAATATACATATATAAATACATATATAAATATACA 1059

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              |||||| | ||| ||| | | ||||||| ||||| | ||||| | | ||| ||| | |||
Db       1060 TATATAAATATATACATATATAAATATACATATATAAATATATACATATATAAATATACATA 1119

Qy        475 TACACATATATATA-ATATACACACATATATATACACATATATACACACATATATACACA 533
              || | |||||||| ||||| | | ||||| ||| |||  |||| ||| |||||||| |
Db       1120 TATAAATATATATACATATATAAATATATAAATATACAAGTATATACAAATATATACATA 1179

Qy        534 TATATACACATATATATACACATATATACACATATATACATATATACACATATATATACA 593
              |||| |   |||||  || | | ||||| | ||| ||| ||| | | || | |   |
Db       1180 TATAAATGTATATACGTATATACATATATATATAAATATATAAAAAAACTTTTGGCTGGG 1239

Qy        594 TATATACACATATATACATATATACACATATATACATATATACACACATATATACACATA 653
              |  |   || || |      | | | |  | |      ||  | || ||| | | |||||
Db       1240 CACCTTTCCAAATCTCATGGCACATATAAGTCTCATGGTAACCTCAAATAAAAAAACATA 1299

Qy        654 CATATACACACACATAGATATACATATATATA 685
              |       || | ||| | |   ||| |
Db       1300 TAACAGATACACCAAAAATAAAAACCAATAAA 1331
```

RESULT 20
US-10-206-447-1
; Sequence 1, Application US/10206447
; Publication No. US20030099616A1
; GENERAL INFORMATION:
; APPLICANT: Geron Corporation
; APPLICANT: Irving, John
; APPLICANT: Karpf, David
; APPLICANT: Schiff, Michael
; TITLE OF INVENTION: DUAL SPECIFICITY TUMOR KILLING VECTORS DRIVEN BY THE TELOMERASE P
; FILE REFERENCE: 085/002
; CURRENT APPLICATION NUMBER: US/10/206,447
; CURRENT FILING DATE: 2002-07-25
; PRIOR APPLICATION NUMBER: 60/308,029
; PRIOR FILING DATE: 2001-07-25
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 15418
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-206-447-1

Query Match 6.4%; Score 296; DB 14; Length 15418;
Best Local Similarity 69.2%; Pred. No. 2.9e-50;
Matches 479; Conservative 0; Mismatches 200; Indels 13; Gaps 5;

```
Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||| |||||||||||  ||| || | |||||||||||| ||||||||| || ||||||
Db        646 AGTTCGAGACCAGCCTGGCCAATATAGCGAAACCCCATCTGTACTAAAAACACAAAAATT 705

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   ||  |||||| ||| |||||||||||||| ||||  ||||||  ||||||||||
Db        706 AGCTGGGTGTGGTGACACATGCCTGTAATCCCAGGTACTCAGGAGGCTAAGGCAGGAGAA 765

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              |  |||||||  |  ||  | || |||     |  ||||||   ||| ||||||||||
Db        766 TTGCTTGAACTGGGAGGCAGAGGTGGT-----GAGCCAAGATTGCACCAGTGCACTCCAG 820
```

```
Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  ||  || |||| ||||   ||||||| ||| ||| |||   | |  ||||| |||
Db        821 CCTTGGTGAC-AGAGTGAAACTCCATCTCAAAAACAAACAAACAAAATACATATACATAA 879

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              |||||||  | | | | | ||| | ||| ||||||| ||| ||||||| |  |||||||
Db        880 ATATATATGCACATATATATACATATATAAATATATATACACATATATAAATCTATATAC 939

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACAT-----ATATATA 354
              |||||||| ||||| ||| ||||| ||  | |||||||| || ||||     ||||||
Db        940 ATATATACATATATACACATATATAAATCTATATACATATATATACATATATAATATATT 999

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              |||||| || | | ||| ||| | | ||||||| | ||| ||||| ||||| | | |||
Db       1000 TACATATATAAATATATACATATATAAATATACATATATAAATACATATATAAATATACA 1059

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              |||||| | ||| ||| | | ||||||| ||||| | ||||| | | ||| ||| | |||
Db       1060 TATATAAATATACATATATAAATATACATATATAAATATATACATATATAAATATACATA 1119

Qy        475 TACACATATATATA-ATATACACACATATATATACACATATATACACACATATATACACA 533
              || | |||||||||| ||||| | | ||||| ||| |||  |||| ||| |||||||| |
Db       1120 TATAAATATATATACATATATAAATATATAAATATACAAGTATATACAAATATATACATA 1179

Qy        534 TATATACACATATATATACACATATATACACATATATACATATATACACATATATATACA 593
              |||| |   |||||  || | | ||||| | ||| ||| ||| | | || | |   |
Db       1180 TATAAATGTATATACGTATATACATATATATATAAATATATAAAAAAACTTTTGGCTGGG 1239

Qy        594 TATATACACATATATACATATATACACATATATACATATATACACACATATATACACATA 653
              |  |   || || |      | | | |  | |      ||  | || ||| | | |||||
Db       1240 CACCTTTCCAAATCTCATGGCACATATAAGTCTCATGGTAACCTCAAATAAAAAAACATA 1299

Qy        654 CATATACACACACATAGATATACATATATATA 685
              |       || | ||| | |   ||| |
Db       1300 TAACAGATACACCAAAAATAAAAACCAATAAA 1331
```

RESULT 21
US-10-674-836-1
; Sequence 1, Application US/10674836
; Publication No. US20040072787A1
; GENERAL INFORMATION:
; APPLICANT: Morin, Gregg B.
; APPLICANT: Lichtsteiner, Serge
; APPLICANT: Vasserot, Alain
; APPLICANT: Adams, Robert R.
; APPLICANT: Geron Corporation
; TITLE OF INVENTION: Telomerase Reverse Transcriptase Transcriptional
; TITLE OF INVENTION: Regulatory Sequences and Methods of Using
; FILE REFERENCE: 019/246P
; CURRENT APPLICATION NUMBER: US/10/674,836
; CURRENT FILING DATE: 2003-09-29
; PRIOR APPLICATION NUMBER: US/09/244,438
; PRIOR FILING DATE: 1999-02-04
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 15418
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Human TERT promoter
US-10-674-836-1

Query Match 6.4%; Score 296; DB 17; Length 15418;
Best Local Similarity 69.2%; Pred. No. 2.9e-50;
Matches 479; Conservative 0; Mismatches 200; Indels 13; Gaps 5;

```
Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||| |||||||||||  ||| || | |||||||||||| ||||||||| || ||||||
Db        646 AGTTCGAGACCAGCCTGGCCAATATAGCGAAACCCCATCTGTACTAAAAACACAAAAATT 705

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
```

```
                ||   ||  |||||| ||| ||||||||||||||| ||||  ||||||  ||||||||||
Db        706 AGCTGGGTGTGGTGACACATGCCTGTAATCCCAGGTACTCAGGAGGCTAAGGCAGGAGAA 765

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
                |  |||||||  |  ||  | || |||     |   ||||||    ||| ||||||||||
Db        766 TTGCTTGAACTGGGAGGCAGAGGTGGT-----GAGCCAAGATTGCACCAGTGCACTCCAG 820

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
                ||  ||  || |||| ||||   ||||||| ||| ||| |||   | |  ||||| |||
Db        821 CCTTGGTGAC-AGAGTGAAACTCCATCTCAAAAACAAACAAACAAAATACATATACATAA 879

Qy        240 ATATATATATATACACACACACACACATATATATATACACATATATATACACATATATAC 299
                ||||||||  | | | | | ||| | ||| |||||||| ||| |||||||  |  |||||||
Db        880 ATATATATGCACATATATATACATATATAAATATATATACACATATATAAATCTATATAC 939

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACAT-----ATATATA 354
                |||||||||  ||||| ||| |||||  ||   | ||||||||  || ||||      ||||||
Db        940 ATATATACATATATACACATATATAAATCTATATACATATATATACATATATAATATATT 999

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
                ||||||  || |  | ||| ||| |  | ||||||| |  ||| |||||  |||||  |  | |||
Db       1000 TACATATATAAATATATACATATATAAATATACATATATAAATACATATATAAATATACA 1059

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
                ||||||  | |||  ||| |  | |||||||  |||||  |  |||||  | |  |||  ||| |||
Db       1060 TATATAAATATACATATATAAATATACATATATAAATATATACATATATAAATATACATA 1119

Qy        475 TACACATATATATA-ATATACACACATATATATACACATATATACACACATATATACACA 533
                ||  | |||||||||  |||||  |  |  |||||  |||  |||   ||||  |||  |||||||||  |
Db       1120 TATAAATATATATACATATATAAATATATAAATATACAAGTATATACAAATATATACATA 1179

Qy        534 TATATACACATATATATACACATATATACACATATATACATATATACACATATATATACA 593
                ||||  |.  |||||  ||  |  |  |||||  |  |||  |||  |||  |  |  ||  |  |    |
Db       1180 TATAAATGTATATACGTATATACATATATATATAAATATATAAAAAAACTTTTGGCTGGG 1239

Qy        594 TATATACACATATATACATATATACACATATATACATATATACACACATATATACACATA 653
                |  |    ||  ||  |       |  |  |  |   |  ||   |  ||  |||  |  |  ||||||
Db       1240 CACCTTTCCAAATCTCATGGCACATATAAGTCTCATGGTAACCTCAAATAAAAAAACATA 1299

Qy        654 CATATACACACACATAGATATACATATATATA 685
                |           ||  |  |||  |  |    |||  |
Db       1300 TAACAGATACACCAAAAATAAAAACCAATAAA 1331


RESULT 22
US-10-811-012-1
; Sequence 1, Application US/10811012
; Publication No. US20040152189A1
; GENERAL INFORMATION:
;  APPLICANT: Geron Corporation
;  APPLICANT:  McWhir, Jim
;  APPLICANT:  Gold, Joseph D.
;  APPLICANT:  Schiff, J. Michael
;  TITLE OF INVENTION: Selective Antibody Targeting of Undifferentiated Stem Cells
;  FILE REFERENCE: 096,006D- SeqList
;  CURRENT APPLICATION NUMBER: US/10/811,012
;  CURRENT FILING DATE:  2004-03-26
;  PRIOR APPLICATION NUMBER: 09/995,419
;  PRIOR FILING DATE: 2001-11-26
;  PRIOR APPLICATION NUMBER: 60/253,357
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/253,443
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/253,395
;  PRIOR FILING DATE: 2000-11-27
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 15418
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-811-012-1
```

```
  Query Match             6.4%;  Score 296;  DB 18;  Length 15418;
  Best Local Similarity   69.2%;  Pred. No. 2.9e-50;
  Matches  479;  Conservative    0;  Mismatches  200;  Indels   13;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
                |||||  ||||||||||||  ||| || | ||||||||||||  |||||||||  || ||||||
Db        646 AGTTCGAGACCAGCCTGGCCAATATAGCGAAACCCCATCTGTACTAAAAACACAAAAATT 705

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
                ||   ||  ||||||  ||| ||||||||||||||  ||||  ||||||  ||||||||||
Db        706 AGCTGGGTGTGGTGACACATGCCTGTAATCCCAGGTACTCAGGAGGCTAAGGCAGGAGAA 765

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
                |  |||||||  |  ||  | || |||     |   ||||||    ||| ||||||||||
Db        766 TTGCTTGAACTGGGAGGCAGAGGTGGT-----GAGCCAAGATTGCACCAGTGCACTCCAG 820

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
                ||  ||  || |||| ||||   ||||||| ||| ||| |||   | |  ||||| |||
Db        821 CCTTGGTGAC-AGAGTGAAACTCCATCTCAAAAACAAACAAACAAAATACATATACATAA 879

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
                ||||||||  | | | | | ||| | ||| |||||||| ||| |||||||  |  |||||||
Db        880 ATATATATGCACATATATATACATATATAAATATATATACACATATATAAATCTATATAC 939

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACAT-----ATATATA 354
                |||||||||  ||||| ||| |||||  ||   | ||||||||  || ||||      ||||||
Db        940 ATATATACATATATACACATATATAAATCTATATACATATATATACATATATAATATATT 999

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
                ||||||  || |  | ||| ||| |  | ||||||| |  ||| |||||  |||||  |  | |||
Db       1000 TACATATATAAATATATACATATATAAATATACATATATAAATACATATATAAATATACA 1059

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
                ||||||  | |||  ||| |  | |||||||  |||||  |  |||||  | |  |||  ||| |||
Db       1060 TATATAAATATACATATATAAATATACATATATAAATATATACATATATAAATATACATA 1119

Qy        475 TACACATATATATA-ATATACACACATATATATACACATATATACACACATATATACACA 533
                ||  | |||||||||  |||||  |  |  |||||  |||  |||   ||||  |||  |||||||||  |
Db       1120 TATAAATATATATACATATATAAATATATAAATATACAAGTATATACAAATATATACATA 1179

Qy        534 TATATACACATATATATACACATATATACACATATATACATATATACACATATATATACA 593
                ||||  |   |||||  ||  |  |  |||||  |  |||  |||  |||  |  |  ||  |  |    |
Db       1180 TATAAATGTATATACGTATATACATATATATATAAATATATAAAAAAACTTTTGGCTGGG 1239

Qy        594 TATATACACATATATACATATATACACATATATACATATATACACACATATATACACATA 653
                |  |    ||  ||  |       |  |  |  |   |  ||   |  ||  |||  |  |  ||||||
Db       1240 CACCTTTCCAAATCTCATGGCACATATAAGTCTCATGGTAACCTCAAATAAAAAAACATA 1299

Qy        654 CATATACACACACATAGATATACATATATATA 685
                |           ||  |  |||  |  |    |||  |
Db       1300 TAACAGATACACCAAAAATAAAAACCAATAAA 1331


RESULT 23
US-10-292-798-2005/c
; Sequence 2005, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
;  FILE REFERENCE: 084335/166
;  CURRENT APPLICATION NUMBER: US/10/292,798
;  CURRENT FILING DATE:  2002-11-13
;  PRIOR APPLICATION NUMBER: 10/017,161
;  PRIOR FILING DATE: 2001-12-18
;  PRIOR APPLICATION NUMBER: JP 2001-246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2070
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2005
```

```
;   LENGTH: 1486
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   LOCATION: source
;   FEATURE:
;   LOCATION: (1)..(1486)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(1286)
US-10-292-798-2005

  Query Match             6.4%;  Score 294.8;  DB 17;  Length 1486;
  Best Local Similarity   74.9%;  Pred. No. 1.8e-50;
  Matches  397;  Conservative    0;  Mismatches  127;  Indels    6;  Gaps    2;

Qy        197 AAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACA 256
              | || |   | || | || | | |      || || | | ||  | ||||||| |||||||
Db        999 ATATATACACACATATAATATATACATATTTATACACACATGCACATATATAAATACACA 940

Qy        257 CACACACACATATATATATACACATATATATACACATATATACATATATACACATATATA 316
              | | ||| ||||| | | | | || |||||| ||||||||||||||||||||||||| ||
Db        939 TATATACATATATACACACATATATGTATATATACATATATACATATATACACATATGTA 880

Qy        317 CACATATATATACACATACATATGTACACATATATATACACATATGTATACACATATATA 376
              | ||| ||||| ||||| |||   | |  ||||||||||||||||||| | | | |||
Db        879 TATATACATATATACATATATACACATATGTATATATACACATATGTATATATACACATA 820

Qy        377 CACATATATACACACATATATACACATATATACACACATATATACACATATATACACATA 436
              ||||||| || | | | | |||||||||| || | | | | |||||||||| || | |||
Db        819 CACATATGTATATATACACATACACATATGTATATATACACATACACATATGTATATATA 760

Qy        437 TATACACATATACACATATATACACATATATACATATATACACATATATATA-----ATA 491
              | | ||| |||   |||||||||||| ||  |||||||||||||||| ||      | |
Db        759 CACATACACATATGTATATATACACATACATGTATATATACACATATATGTATATATACA 700

Qy        492 TACACACATATATATACACATATATACACACATATATACACATATATACACATATATATA 551
              | |||||| |||| ||| | | ||| |   |||||||||||| | | | ||| | |||
Db        699 CATACACATGTATACACATACACATATATGTATATATACACATACACATATATACACATA 640

Qy        552 CACATATATACACATATATACATATATACACATATATATACATATATACACATATATACA 611
              ||| ||||||||||||||||| ||||||||||||||   ||||||||||||||| |||
Db        639 TACACATATACACATATATACACATATACACATATATAAGTATATATACACATATACACA 580

Qy        612 TATATACACATATATACATATATACACACATATATACACATACATAT-ACACACACATAG 670
              ||||  | ||||| |||||||||||| |||  || | ||||| || | ||||| |||
Db        579 TATACGTATATATACACATATATACACATATACGTATATATACACATAATACACATATAC 520

Qy        671 ATATACATATATATACACATATATATACGTATATATATGTATATATATAT 720
              |||| ||| | ||| | | | ||||||||||||||| ||||||| | ||
Db        519 GTATAAATACACATATATACACATATACGTATATATACGTATATACACAT 470


RESULT 24
US-09-841-158-5
; Sequence 5, Application US/09841158
; Publication No. US20030166046A1
; GENERAL INFORMATION:
;  APPLICANT: BEASLEY, Ellen M.
;  TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
;  TITLE OF INVENTION:  NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND
;  TITLE OF INVENTION:  USES THEREOF
;  FILE REFERENCE: CL001229
;  CURRENT APPLICATION NUMBER: US/09/841,158
;  CURRENT FILING DATE:  2001-04-25
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
;   LENGTH: 27067
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(27067)
;   OTHER INFORMATION: n = A,T,C or G
US-09-841-158-5

  Query Match             6.4%;  Score 293.6;  DB 10;  Length 27067;
  Best Local Similarity   75.8%;  Pred. No. 1.1e-49;
  Matches  391;  Conservative    0;  Mismatches  119;  Indels    6;  Gaps    2;

Qy        230 ATATATATATATATATATATATACACACACACACATATAT-ATATACACATATATATA 288
              || ||||  |  | |  |  ||| |      | ||| |||||| | ||| | ||||| | |
Db      26509 ATGTATAACTTAAAAATTTCATATATGTGTATACATATATATGAAATATATATATACAGA 26568

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              |||| ||||| || |||||| |||||||||||||||||  | | | ||||  |||||||
Db      26569 CACACATATATATGTATACATATATATACACATATATATGTATACATATATACACACATA 26628

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              ||| || |||||| |||||||| ||||||||||||||||||||||| ||| | | ||| | |
Db      26629 TATGTATACATATATATACACACATATACACATATATACACACACATACATATATACACA 26688

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              |||| ||||| |||||||||||||| ||  |||||||| |  |||| | | ||| | |
Db      26689 CACATATATACACACATATATACACACATGCACACATATATATGTATACATATATACACA 26748

Qy        469 CATATATACACATATATATA-----ATATACACACATATATATACACATATATACACACA 523
              ||| |||||  ||||||| |      ||||||||||||||||||||||| ||||||||||
Db      26749 CATGTATACGTATATATACACACATATATACACACATATATATACACACATATACACACA 26808

Qy        524 TATATACACATATATACACATATATATACACATATATACACATATATACATATATACACA 583
              || | |||||||||||||| |||||||||| ||||| ||| | ||| |||| || |||
Db      26809 TACACACACATATATACACACATATATACACACATATATACACACATATATATGTATACA 26868

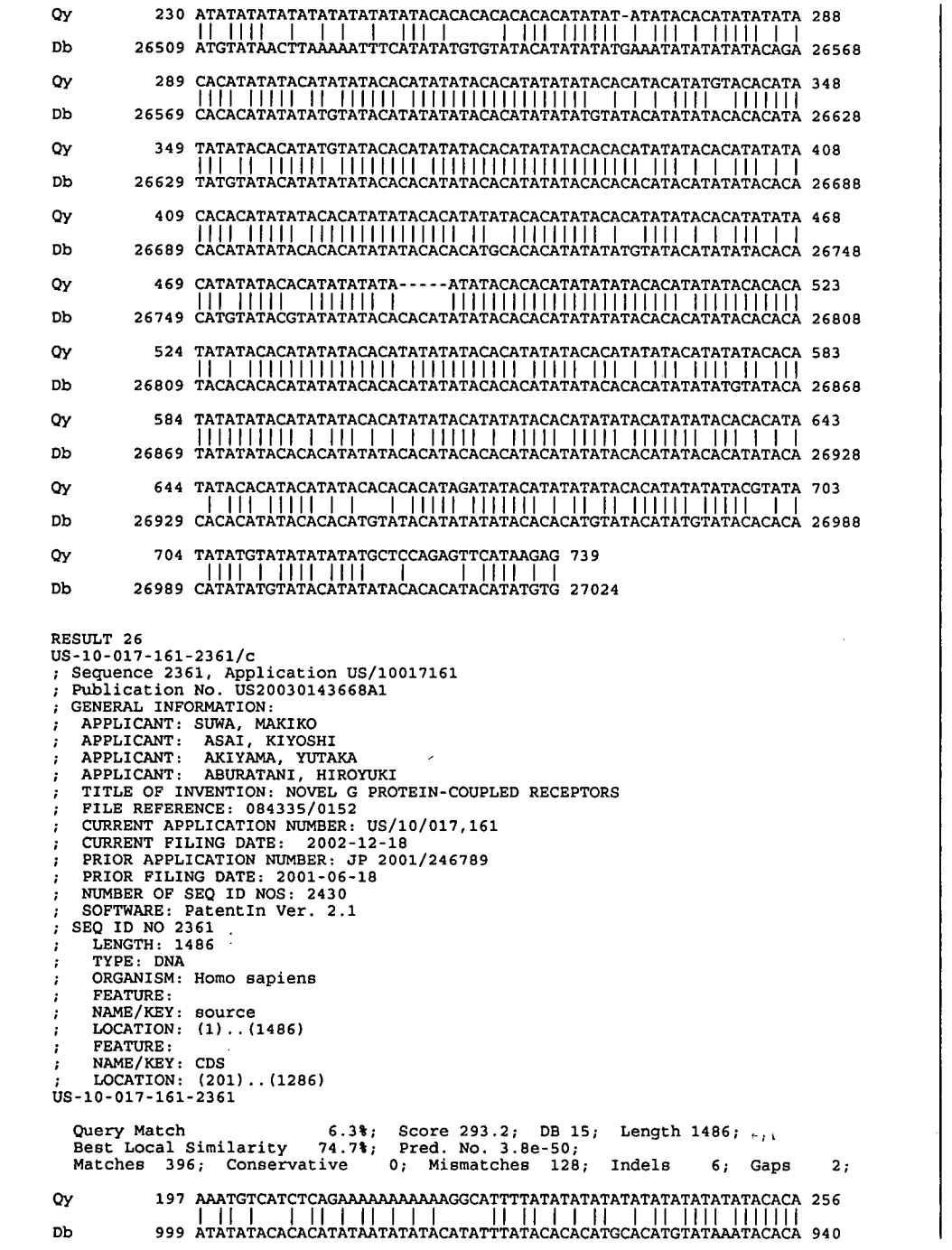
Qy        584 TATATATACATATATACACATATATACATATATACACATATATACATATATACACACATA 643
              |||||||||| | ||| | | | ||||| | ||||| ||||| ||||||| ||| | | |
Db      26869 TATATATACACACATATATACACATACACACATACATATATACACATATACACATATACA 26928

Qy        644 TATACACATACATATATACACACACATAGATATACATATATATACACATATATATACGTATA 703
              | ||| ||||| | |   | ||||| ||||||| | || || |||||| |||||  | |
Db      26929 CACACATATACACACATGTATACATATATATACACACATGTATACATATGTATACACACA 26988

Qy        704 TATATGTATATATATATGCTCCAGAGTTCATAAGAG 739
              |||| | |||| ||||    |     | |||| | |
Db      26989 CATATATGTATACATATATACACACATACATATGTG 27024

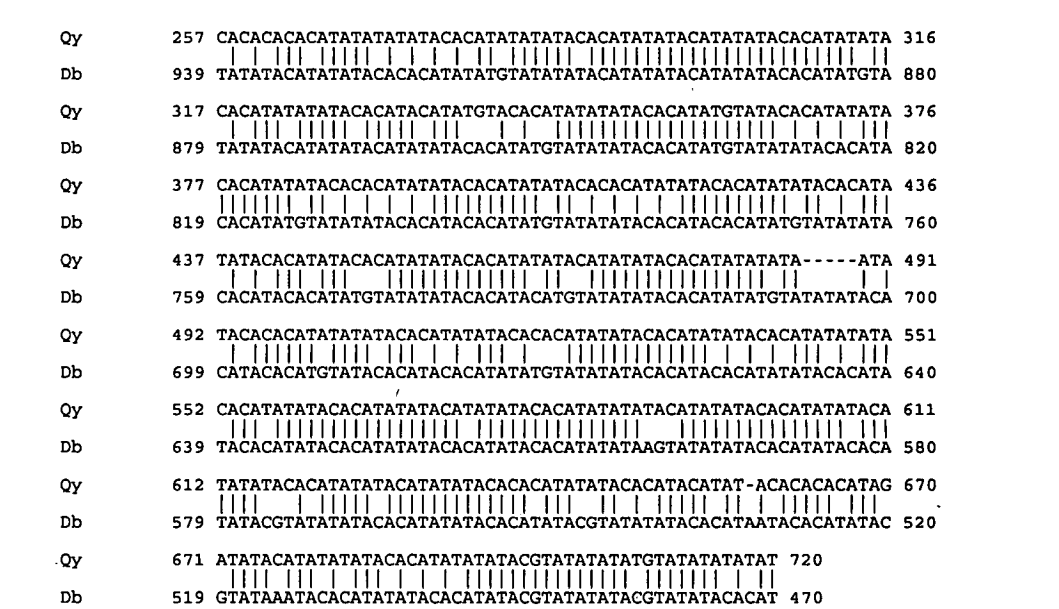

RESULT 25
US-10-801-837-5
; Sequence 5, Application US/10801837
; Publication No. US20040152167A1
; GENERAL INFORMATION:
;  APPLICANT: BEASLEY, Ellen M. et al.
;  TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
;  TITLE OF INVENTION:  NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND
;  TITLE OF INVENTION:  USES THEREOF
;  FILE REFERENCE: CL001229-DIV
;  CURRENT APPLICATION NUMBER: US/10/801,837
;  CURRENT FILING DATE:  2004-03-17
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
;   LENGTH: 27067
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(27067)
;   OTHER INFORMATION: n = A,T,C or G
US-10-801-837-5

  Query Match             6.4%;  Score 293.6;  DB 18;  Length 27067;
  Best Local Similarity   75.8%;  Pred. No. 1.1e-49;
  Matches  391;  Conservative    0;  Mismatches  119;  Indels    6;  Gaps    2;
```



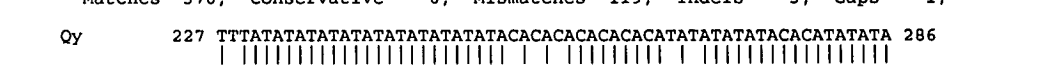
RESULT 26
US-10-017-161-2361/c
; Sequence 2361, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 084335/0152
; CURRENT APPLICATION NUMBER: US/10/017,161
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: JP 2001/246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2430
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2361
; LENGTH: 1486
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: source
; LOCATION: (1)..(1486)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (201)..(1286)
US-10-017-161-2361

Query Match 6.3%; Score 293.2; DB 15; Length 1486;
Best Local Similarity 74.7%; Pred. No. 3.8e-50;
Matches 396; Conservative 0; Mismatches 128; Indels 6; Gaps 2;



RESULT 27
US-10-017-161-2381/c
; Sequence 2381, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 084335/0152
; CURRENT APPLICATION NUMBER: US/10/017,161
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: JP 2001/246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2430
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2381
; LENGTH: 1672
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: source
; LOCATION: (1)..(1672)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (201)..(1472)
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (208)..(307)
; OTHER INFORMATION: a, t, c, g, unknown or other
US-10-017-161-2381

Query Match 6.3%; Score 291.6; DB 15; Length 1672;
Best Local Similarity 75.5%; Pred. No. 8.7e-50;
Matches 376; Conservative 0; Mismatches 119; Indels 3; Gaps 1;



```
Db       1390 TATATATATATATATATATATATATATATACACACACACATATATACACATATATA 1331

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              |||||| | | | |||| ||| |||| |||||||||||||||||| ||| | | |
Db       1330 TACACACACATATATACACATATATACACACATATATATACACATATATACACATATA 1271

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||| | | | |||| | | | |||| |||||||||||||||||| |||| || ||||
Db       1270 TATACACATATATATACACATATATATGTACACATATATACACACATGTATATACGTATA 1211

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              |    |||||||||||||| | ||| |  ||| |  | ||||||||||||||  | |||||
Db       1210 TGTGTACATATATACACACACATATATGTATGTGTATATATACACATATATGTATATATA 1151

Qy        467 TACATATATACACATATATATA---ATATACACACATATATATACACATATATACACACA 523
              |||| | | | |||| ||||||    |||| | ||| | ||| | | || |||| |
Db       1150 TACACACACAGACATGTATATATGTGTATATATACACACATAGATATATGTATATATGTG 1091

Qy        524 TATATACACATATATACACATATATATACACATATATACACATATATACATATATACACA 583
              |||||||||| | ||| |  | |||||||||| |||||    |||||||| | | | | |
Db       1090 TATATACACACACATATATGTGTATATACACACATATATGTGTATATACACACACATATA 1031

Qy        584 TATATATACATATATACACATATATACATATATACACATATATACATATATACACACATA 643
              | | |||| | | | ||| |||| |  ||||| ||||| |||||  | |||| ||||| |
Db       1030 TGTGTATATACACACACATATATGTGTATATACACACACATATATGTGTATATACACACA 971

Qy        644 TATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATA 703
              ||| |  |  |||||||||||||  |||||  | |||||||||||||| ||  |||
Db        970 CATATATGTGTATATACACACACATGTATATATGTGTATATACACATATATGTATATATG 911

Qy        704 TATATGTATATATATATG 721
              | ||| || | |||||||
Db        910 TGTATATACACATATATG 893


RESULT 28
US-10-292-798-2023/c
; Sequence 2023, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
;  FILE REFERENCE: 084335/166
;  CURRENT APPLICATION NUMBER: US/10/292,798
;  CURRENT FILING DATE:  2002-11-13
;  PRIOR APPLICATION NUMBER: 10/017,161
;  PRIOR FILING DATE: 2001-12-18
;  PRIOR APPLICATION NUMBER: JP 2001-246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2070
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2023
;   LENGTH: 1672
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   LOCATION: source
;   FEATURE:
;   LOCATION: (1)..(1672)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(1472)
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (208)..(307)
;   OTHER INFORMATION: a, t, c, g, unknown or other
US-10-292-798-2023

  Query Match             6.3%;  Score 291.6;  DB 17;  Length 1672;
  Best Local Similarity   75.5%;  Pred. No. 8.7e-50;
  Matches  376;  Conservative    0;  Mismatches  119;  Indels    3;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACATATATATATACACATATATA 286
              | |||||||||||||||||||||||| | | ||||||||| | ||||||||||||||||
Db       1390 TATATATATATATATATATATATATATATACACACACACACATATATACACATATATA 1331

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              |||||| | | | |||| ||| |||| |||||||||||||||||| |||    | | |
Db       1330 TACACACACATATATACACATATATACACACATATATATACACATATATACACATATA 1271

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||| | | | |||| | | | |||| |||||||||||||||||| |||| || ||||
Db       1270 TATACACATATATATACACATATATATGTACACATATATACACACATGTATATACGTATA 1211

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              |    |||||||||||||| | ||| |  ||| |  | ||||||||||||||  | |||||
Db       1210 TGTGTACATATATACACACACATATATGTATGTGTATATATACACATATATGTATATATA 1151

Qy        467 TACATATATACACATATATATA---ATATACACACATATATATACACATATATACACACA 523
              |||| | | | |||| ||||||    |||| | ||| | ||| | | || |||| |
Db       1150 TACACACACAGACATGTATATATGTGTATATATACACACATAGATATATGTATATATGTG 1091

Qy        524 TATATACACATATATACACATATATATACACATATATACACATATATACATATATACACA 583
              |||||||||| | ||| |  | |||||||||| |||||    |||||||| | | | | |
Db       1090 TATATACACACACATATATGTGTATATACACACATATATGTGTATATACACACACATATA 1031

Qy        584 TATATATACATATATACACATATATACATATATACACATATATACATATATACACACATA 643
              | | |||| | | | ||| |||| |  ||||| ||||| |||||  | |||| ||||| |
Db       1030 TGTGTATATACACACACATATATGTGTATATACACACACATATATGTGTATATACACACA 971

Qy        644 TATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATA 703
              ||| |  |  |||||||||||||  |||||  | |||||||||||||| ||  |||
Db        970 CATATATGTGTATATACACACACATGTATATATGTGTATATACACATATATGTATATATG 911

Qy        704 TATATGTATATATATATG 721
              | ||| || | |||||||
Db        910 TGTATATACACATATATG 893


RESULT 29
US-10-741-601-5636/c
; Sequence 5636, Application US/10741601
; Publication No. US20040166519A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  STENOSIS, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001500
;  CURRENT APPLICATION NUMBER: US/10/741,601
;  CURRENT FILING DATE:  2003-12-22
;  NUMBER OF SEQ ID NOS: 26415
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5636
;   LENGTH: 81924
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(81924)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-2
US-10-741-601-5636

  Query Match             5.9%;  Score 273.8;  DB 18;  Length 81924;
  Best Local Similarity   65.9%;  Pred. No. 2.3e-45;
  Matches  458;  Conservative    0;  Mismatches  232;  Indels    5;  Gaps    4;

Qy         42 TACTAAAAATACAAAATTAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCG 101
              ||  ||||| | ||| ||||   |||||||||  |||| ||||||||||||||| || |
Db      38267 TAAAAAAAAAAAAAAGTTAGTGGGGCATGGTGGCACACACCTGTAATCCCAGCTATTTGG 38208

Qy        102 GAGGCCGAGGCAGGAGAATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGC-AAGA 160
              ||||| |||||||||||||  ||||||||  | || ||||||||||||| | ||||  | |
Db      38207 GAGGCTGAGGCAGGAGAATTGCTTGAACC--CAGGAGGTGGAGGTTGCAGTGAGCTGAAA 38150
```

```
Qy       161 TCATGCCATTGCACTCCAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAA 220
             |  ||| ||||||||||||||    |  || |||| || |     ||||| || | | | |
Db     38149 TTGTGCTATTGCACTCCAGCCTGAGTGAC-AGAGTGAGACTCTGTCTCAAAATATATATA 38091

Qy       221 AGGCATTTTATATATATATATATATATATACACACACACACATATATATATACACA 280
             | |||| | | | | |||||| | ||||||||| | || | | | || | |
Db     38090 TACACACACACATATGTGTGTGTGTATATACATACACACATATATGTGTGTGTATATA 38031

Qy       281 TATATATACACATATATACATATATACACATATATACACATATATATACACATACATATG 340
             ||| ||| ||| ||| ||| | |||    |||||||  |||| | | | ||   |||
Db     38030 CATACATATACACATACACACACATATGTGTATATACATGTATACACATATATGTGTATA 37971

Qy       341 TACACATATATATACACATATGTATACACATATATACACATATATACACACATATATACA 400
             ||||| | | | || |  | | |||||||||||| |     |||||     | ||| | | |
Db     37970 TACACGTGTGTGTATATGTGTATATACACATATGTGTGTGTATATGTGTATATACACATA 37911

Qy       401 CATATATACACACATATATACACATATATACACATATATACACATATACACATATATACA 460
             | | || |    |||||||||||||| |     |||||    ||||||||||||| |
Db     37910 TGTGTGTATATGTGTATATACACATATGTGTGTGTATATGTGTATATACACATATGTGTG 37851

Qy       461 CATATATACATATATACACATATATATAATATACACACATATATATACACATATATACAC 520
             |||||   |||||||||||||| | |  | || |    ||||||||  ||||||||||
Db     37850 TGTATATGTGTATATACACATATGTGTG-TGTATATGTGTATATATATGTATATATACAC 37792

Qy       521 ACATATATACACATATATACACATATATATACACATATATACACATATATACATATATAC 580
             | | | | ||| |||| |     | ||| || | | | | | |  | | ||  |||||||
Db     37791 ATACACACACATATATGTTTGTGTGTATGTATATACACACATATGTGTGTATGTATATAC 37732

Qy       581 ACATATATATACATATATACACATATATACATATATACACATATATACATATATACACAC 640
             ||| |||| |   |||| ||| |||| ||| ||||| ||| | |  |||||||| ||
Db     37731 ACACATATGTGTGTATACACATATATGCACACATATATACACGTGTGTATATATACCCAT 37672

Qy       641 ATATATACACATACATATACACACACATAGATATACATATATATACACATATATATACGT 700
             |||| || | ||   |||| | |||  |   ||||||||  | |  | |||| ||  ||
Db     37671 ATATGTATATATGTGTATATATACATGTGTGTATACATACGTGTGTATATATACATGTGT 37612

Qy       701 ATATATATGTATATATATATGCTCCAGAGTTCATA 735
             |||| ||   | | ||||| |    | ||  | |
Db     37611 GTATACATACGTGTGTATATACATGTGTGTATACA 37577


RESULT 30
US-10-741-600-17799
; Sequence 17799, Application US/10741600
; Publication No. US20050026169A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001499
;  CURRENT APPLICATION NUMBER: US/10/741,600
;  CURRENT FILING DATE:  2003-12-22
;  NUMBER OF SEQ ID NOS: 73997
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17799
;   LENGTH: 14603
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-741-600-17799

  Query Match             5.8%;  Score 267.2;  DB 19;  Length 14603;
  Best Local Similarity   71.5%;  Pred. No. 2.4e-44;
  Matches  449;  Conservative    0;  Mismatches  168;  Indels   11;  Gaps    7;

Qy         1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC--AAAAT 58
             || ||||||||| ||||  |||||||||||||| || |||||||||||||||  |||||
Db     10438 AGATCAAGACCATCCTGGCCAACATGGTGAAACTCCGTCTCTACTAAAAATACAAAAAAT 10497

Qy        59 TAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGA 118
             |||   ||| ||||| |   | ||||||||||||||||||| |||||| || ||||||||
Db     10498 TAGCTGGGCGTGGTGGTGGGCTCCTGTAATCCCAGCTACTTGGGAGGCTGAAGCAGGAGA 10557

Qy       119 ATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTG---CAC 174
             ||| ||||||||  | || || ||||||||| | ||| |||||||  |||||||   |||
Db     10558 ATCGCTTGAACC--CAGGAGGCGGAGGTTGCAGTGAGCCAAGATCGCGCCATTGCACCAC 10615

Qy       175 TCCAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAGGCATTTTATATA 234
             |||||||  ||||| ||||| ||||      |||||| ||||||||||    |     | | |
Db     10616 TCCAGCCTGGGCAAAAAGAGAGAAACTCTGTCTCAAAAAAAAAAAAAAAAAAAAAAAAA 10675

Qy       235 TATATATATATATATATACACACACACACATATATATATACACATATATATACACATA 294
             ||||||||||||||||| | |||||||||||||||||||  | | ||| ||  | | |||
Db     10676 AATATATATATATATATATACACACACATATATATTCATATATA-ATTCATATATA 10734

Qy       295 TATACATATATACACATATATACACATATATATACACATACATATGTACACATATATATA 354
             ||| ||||||||           |   ||||||| |   | |||  ||  |||||||| |
Db     10735 TATTCATATATAATTCATATATAATTCATATATATATTCATATAGATATTCATATATA-A 10793

Qy       355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
               |||||  |||  |||| |||  |||||||  || ||| |||  ||||||| || | |
Db     10794 TTCATATAGATATTCATAGATATTCATATATATTCATATAGATATTCATATATTCATATA 10853

Qy       415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
             |||  | | ||  ||  | ||  ||| | |||| || ||||| | |   ||| | ||  |
Db     10854 TATTCATATATTCATGTATATTCATATAGATATTCATATATAGATATTCATAGATATTCA 10913

Qy       475 TACACATATATATAATATACACACATATATATACACATATATACACACATATATACACAT 534
             || | | |||| | ||||| | |   |||||||  | ||  ||| | | ||  ||  | |
Db     10914 TATATAGATAT-TCATATAGATATTCATATATAGATATTCATATAGATATTCATGTATAG 10972

Qy       535 ATATACACATATATATACACATATATACACATATATACATATATACACATATATATACAT 594
             |||| || |||||||| || ||||| | |   ||||| |||||| || || |||||  ||
Db     10973 ATATTCATATATATATTCATATATAGATATTCATATAGATATATTCATATCTATATGAAT 11032

Qy       595 ATATACACATATATACATATATACACAT 622
             ||||| |   |||||  ||||||   ||
Db     11033 ATATATATGAATATATCTATATATGAAT 11060


RESULT 31
US-10-311-455-1210/c
; Sequence 1210, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1210
;   LENGTH: 6971
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: 2741
;   OTHER INFORMATION: n is a or g or c or t
US-10-311-455-1210

  Query Match             5.8%;  Score 266.4;  DB 15;  Length 6971;
  Best Local Similarity   71.1%;  Pred. No. 2.6e-44;
  Matches  367;  Conservative    0;  Mismatches  146;  Indels    3;  Gaps    1;
```

```
Qy        212 AAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACATATAT 271
              | | | | |    |  | ||||||||  |||||||||| |||||| | | | | ||||||
Db       6472 ATACATATATACCTATATATATATACCTATATATATAAATACACATATATATATATAT 6413

Qy        272 ATATACACATATATATACACATATATACATATATACACATATATACACATATATATACAC 331
              |||   | |||||||| |||  |||| ||||||||||  |||| |||||||||   |
Db       6412 ATACCAATATATATATATACACGTATATATATATACACACTTATATACATATATACCTAT 6353

Qy        332 ATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACACAC 391
              ||| |||  |||||||||||||| | |||  |||| | ||| |||||||| |||| | |
Db       6352 ATATATAAATACACATATATATATATACCTATATATATAAATACACATTTATATATAT 6293

Qy        392 ATATATACACATATATACACACATATATACACATATATACACATATATACACATATACAC 451
              | ||| |||  |||||| | | |    || | | ||| |||  |||||| | ||||||
Db       6292 AAATACACATTTATATATATATACGCCTATATAAATACACATTTATATATATATATACCT 6233

Qy        452 ATATATACACATATATACATATATACACATATATATAATATACACACATATATATACACA 511
              ||||||| || |||||| ||| ||||  |||||||||| | |||   ||||||||| | |
Db       6232 ATATATATACCTATATATATAAATACCTATATATATAAAACACATTTATATATATATATA 6173

Qy        512 TATATACACACATATATACACATATATACACATATATATACACATATATACACATATATA 571
                |||| | | ||| | |   ||||||| || | ||||||  | |  |||| | | | |||
Db       6172 CCTATATATAAATACACATTTATATATATACCTCTATATATATACCTATATATAAAAATA 6113

Qy        572 CATATATACACATATATATACATATATACACATATATACATATATACACATATATACATA 631
              || || || | |||    || ||||| | | | | ||  |||||||| | |||||| |||
Db       6112 CACATTTATATATACGCCTATATATAAAAATACACATTTATATATACGCCTATATATATA 6053

Qy        632 TATACACACATATATACACATACATATACACACACATAGATATACATATATAT---ACAC 688
                |||||||  |||||| | ||||   || | | | | | |  || ||||||||   ||||
Db       6052 AATACACATTTATATATATATACGCCTATATATATAAATACCTATATATATATAAAACAC 5993

Qy        689 ATATATATACGTATATATATGTATATATATATGCTC 724
              || |||||| ||||||||  |||||||| || | |
Db       5992 ATTTATATATATATATATACCTATATATAAATACAC 5957


RESULT 32
US-10-812-232-6/c
; Sequence 6, Application US/10812232
; Publication No. US20040265961A1
; GENERAL INFORMATION:
;  APPLICANT: Liu, Wei
;  APPLICANT:  Wu, Leeying
;  APPLICANT:  Channavajhala, Padma L,
;  APPLICANT:  Lin, Lih-Ling
;  APPLICANT:  Zhang, Yuhua
;  TITLE OF INVENTION: Novel Proteins Homologous to Kinase Suppressor of Ras
;  FILE REFERENCE: 01997.026700
;  CURRENT APPLICATION NUMBER: US/10/812,232
;  CURRENT FILING DATE:  2004-03-29
;  PRIOR APPLICATION NUMBER: US 60/457,928
;  PRIOR FILING DATE: 2003-03-28
;  PRIOR APPLICATION NUMBER: US 60/491,283
;  PRIOR FILING DATE: 2003-07-31
;  NUMBER OF SEQ ID NOS: 254
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6
;   LENGTH: 263852
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3814)..(4319)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (6552)..(6571)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (17206)..(17225)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (19680)..(19699)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (24320)..(24339)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (26572)..(26591)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (32986)..(33005)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (56103)..(56724)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (59672)..(59691)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (61525)..(61544)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (71035)..(71054)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (79771)..(79807)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (91343)..(91801)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (141749)..(141768)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (144739)..(144758)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (164171)..(164340)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (173689)..(173708)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (175051)..(175070)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (192924)..(192993)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (202010)..(202029)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (247837)..(247856)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
```

```
;   NAME/KEY: misc_feature
;   LOCATION: (251852)..(252421)
;   OTHER INFORMATION: n is a, c, g, or t
US-10-812-232-6

  Query Match             5.7%;  Score 265.6;  DB 18;  Length 263852;
  Best Local Similarity   73.7%;  Pred. No. 1.9e-43;
  Matches  367;  Conservative    0;  Mismatches  124;  Indels    7;  Gaps    2;

Qy        229 TATATATATATATATATATATATATACACACACACACACATATATATATACACATATATATA 288
              || |||||||||||  ||| ||| | ||| ||| ||| || |  | ||||||| ||  |||
Db      94187 TACATATATATATACATACATACATACATACATATGTCCACATACACACATGCATA 94128

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              | ||||||| | ||||| | ||||||| ||| | | | |  ||| ||||  |||| | |
Db      94127 TATATATATATACATATATATATATATATACACACACATTCATGCATAAATACAAACA 94068

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACAT------ATATACACA 402
              |||| |||||||  |||| ||| ||| ||| | |||| |||||||      ||| |||||
Db      94067 TATACACACATACATATAAACACATACACAGAGATATGCACACATACATACATACACACA 94008

Qy        403 TATATACACACATATATACACATATATACACATATATACACATATACACATATATACACA 462
              || ||| | ||||  |||||||| |||||| | | ||||||  | || |||||| |||
Db      94007 TACATATATACATGCATACACATAAATACACACAGACACACATGCATACGTATATATACA 93948

Qy        463 TATATACATATATACACATATATATAATATACACACATATATATACACATATATACACAC 522
              |||||  ||| ||| ||||||| |  | ||||| | | | |||||||||| ||||||||||
Db      93947 TATATGTATACATATACATATACA-CACATACATAAACACATATACACAGATATACACAC 93889

Qy        523 ATATATACACATATATACACATATATATACACATATATACACATATATACATATATACAC 582
              ||| | | ||| | | | |||||| |||||| ||| | | ||||||  | | | |||  |
Db      93888 ATACACAAACACACAGAGACATATGTATACATATACACATACATATGCAAACACATAAGC 93829

Qy        583 ATATATATACATATATACACATATATACATATATACACATATATACATATATACACACAT 642
              | ||| |||||||||||| | ||||||| ||| | ||||| ||||||| ||||| | |
Db      93828 ACATACATACATATATACATACATATACACATACATACATACATACATAAATACATAGAG 93769

Qy        643 ATATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTAT 702
              | | |||   ||||||| |||||   | | ||||| ||| ||||||| | | | |||| |
Db      93768 ACACACATGCACATATATACACATGCATACATACACATAAATACACACAGACACACGTGT 93709

Qy        703 ATATATGTATATATATAT 720
              |||||   | ||| | ||
Db      93708 ATATACACACATACACAT 93691


RESULT 33
US-10-108-260A-543/c
; Sequence 543, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
;  APPLICANT: HELIX RESEARCH INSTITUTE
;  TITLE OF INVENTION: No. US20040005560A1el full length cDNA
;  FILE REFERENCE: H1-A0106
;  CURRENT APPLICATION NUMBER: US/10/108,260A
;  CURRENT FILING DATE:  2002-03-27
;  NUMBER OF SEQ ID NOS: 5458
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 543
;   LENGTH: 2522
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-108-260A-543

  Query Match             5.7%;  Score 263.6;  DB 17;  Length 2522;
  Best Local Similarity   68.9%;  Pred. No. 6.2e-44;
  Matches  376;  Conservative    0;  Mismatches  169;  Indels    1;  Gaps    1;

Qy        236 ATATATATATATATATATACACACACACACATATATATATACACATATATATACACATAT 295
              ||  |  | | |||||| | | ||||||| |  |||||||    ||||||||||||  |
Db       2217 ATTCAACTGTGTATATATATATACACACATACGTATATATGTGTATATATATACACACGT 2158

Qy        296 ATACATATATACACATATATACACATATATATACACATACATATGTACACATATATATAC 355
              ||| ||| | | | || |||||| | | |   | | ||||| || ||   ||| ||| ||
Db       2157 ATATATACACATATATGTATACATACACACGCATATATACACATATATGTATACATACAC 2098

Qy        356 ACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACAT 415
              || ||| ||||||||||||||  | | ||| |||| ||||||||||||||||||  | ||||
Db       2097 ACGTATATATACACATATATGTATACATACACACGCATATATACACATATATGTATACAT 2038

Qy        416 ATATACACATATATACACATATATACACATATACACATATATACACATATATACATAT 475
              | | ||  |||||||||||||||||| || | |||||||   ||| | | | ||| ||  ||
Db       2037 ACACACGTATATATACACATATATGCATACATACACACGCATATATACACATATATGCAT 1978

Qy        476 ACACATATATATAATATACACACATATATATACACATATATACACACATATATACACATA 535
              ||| | |||  | ||||| |||||||||| || ||||| ||||  | | ||| ||| |||
Db       1977 ACATACATACGT-ATATATACACATATATGTATACATACATACGTATACATACACATATA 1919

Qy        536 TATACACATATATATACACATATATACACATATATACATATATACACATATATATACATA 595
              | || ||||| ||   || ||||| ||| |||| || | ||| | || ||||||||||| |
Db       1918 TGTATACATACATGCGCATATATACACATATATGTATACATACATACGTATATATACACA 1859

Qy        596 TATACACATATATACATATATACACATATATACATATATACACACATATATACACATACA 655
              ||||   ||| |||||||  || ||||| | | || |||||| |||||  || | |||||
Db       1858 TATATGTATACATACATACGTATACATACACATATGTATACATACATACGTATATATACA 1799

Qy        656 TATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATA 715
              ||| |  | | | || | |||||||||||||||||| ||   ||| ||  |||||||
Db       1798 CATATATGTATACATACATGCGTATATATACACATATATGTATACATACATGCGTATATA 1739

Qy        716 TATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGA 775
              || |    | | ||||     | || |   ||   || | ||| |   | | |
Db       1738 TACACATATATGCATACATACATATGAACACATATATATGTGTGTGTATATATACATATA 1679

Qy        776 GAGTTT 781
              |||
Db       1678 AGATTT 1673


RESULT 34
US-10-723-860-1357/c
; Sequence 1357, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
;  APPLICANT: Aziz, Natasha
;  APPLICANT:  Ginsburg, Wendy M.
;  APPLICANT:  Zlotnik, Albert
;  TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
;  TITLE OF INVENTION:  Methods for Screening for Soft Tissue Sarcoma Modulators
;  FILE REFERENCE: 05882.0193.NPUS01
;  CURRENT APPLICATION NUMBER: US/10/723,860
;  CURRENT FILING DATE:  2003-11-26
;  PRIOR APPLICATION NUMBER: 60/429,739
;  PRIOR FILING DATE: 2002-11-26
;  NUMBER OF SEQ ID NOS: 8393
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1357
;   LENGTH: 227968
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-723-860-1357

  Query Match             5.7%;  Score 263;  DB 18;  Length 227968;
  Best Local Similarity   71.6%;  Pred. No. 6e-43;
  Matches  373;  Conservative    0;  Mismatches  145;  Indels    3;  Gaps    2;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||  |||||||||||  ||||||||||||||||||||||||||||||||||| ||||||
Db       5103 AGTTTGAGACCAGCCTGGCCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAAATT 5044

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              || |||||||||||  ||| |||||||||||||||||| | |||||| || | |||||||
Db       5043 AGCCAGGCATGGTGGCACATGCCTGTAATCCCAGCTACCTGGGAGGCTGAAGTAGGAGAA 4984

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              || || ||||| |  || ||  |||| ||| | |||| |        | | |   ||
```

```
Db       4983 TCTCTGGAACCAG--GGAGGCAGAGGCTGCAGTGAGCCAAGATCATTCCTGGGTGTCAGA 4926

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239

Db       4925 GAGAGACTCCATCCCTGCCCGTCCCCCGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4866

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299

Db       4865 AAAAAAATATATATATATATATATATATATATATATATATACACACACACATAT 4806

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACAT 359

Db       4805 ATATACATATATATACACACATATATATACATATATATACATATATATACATATATATAT 4746

Qy        360 ATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATAT 419

Db       4745 ACGTATATATATACATATATATATACACATATATATACACATATATATACATATATATAT 4686

Qy        420 ACACATATATACACATATATACACATATACACATATATACACATATATACATATATACAC 479

Db       4685 ACACATATATATACACATATATATACATATATATATACACATATATATACATATATAT 4626

Qy        480 ATATATATAATATACACACATATATATACACATATATACAC 520

Db       4625 ATGTATATTTTATTGAGATCACATTAAATGTGTACATTCAC 4585


RESULT 35
US-10-741-600-17809/c
; Sequence 17809, Application US/10741600
; Publication No. US20050026169A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001499
;  CURRENT APPLICATION NUMBER: US/10/741,600
;  CURRENT FILING DATE:  2003-12-22
;  NUMBER OF SEQ ID NOS: 73997
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17809
;   LENGTH: 109730
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-741-600-17809

  Query Match             5.7%;  Score 262.6;  DB 19;  Length 109730;
  Best Local Similarity   72.0%;  Pred. No. 5.3e-43;
  Matches  384;  Conservative    0;  Mismatches  144;  Indels    5;  Gaps    3;

Qy        204 ATCTCAGAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACAC 263

Db     108244 ATGTTATATATACATAAATATATGTTATATATATTTATATATATTTATATACATATAAAT 108185

Qy        264 ACATATATATATACACATATATATACACATATATACATATATACACATATATA-CACATA 322

Db     108184 ATATATATTTATATACATATATAAAAATATATATATATTTATATATAAATATATGTTATA 108125

Qy        323 TATATACACATACATATGTACACATATATATACACATATGTATACACATATATACACATA 382

Db     108124 TATATTTATATACATATAAATATATATTTATATACATATAAATATATATTTATATACATA 108065

Qy        383 TATACACACATATATACACATATATACACACATATATACACATATATACACATATATACA 442

Db     108064 TAAATATATATTTATATACATATAAAAATCTATATTTATATACATATAAAAATATATATT 108005

Qy        443 CATATACACATATATACACATATATACATATATACACATATATATA---ATATACACACA 499

Db     108004 TATATACATATAAAAATATATATTTATATACATATAGATATATATATTTATATACATATA 107945

Qy        500 TATATATACACATATATACACACATATATACACATATATACACATATATATACACATATA 559

Db     107944 AATATATATATTTATATACATAAATATATATATTTATATACATATAAATATATATATTTA 107885
```

```
Qy        560 TACACATATATACATATATACACATATATATACATATATACACATATATACATATATACA 619

Db     107884 TATACATAAATATATATATTTATATACATATGTATATAAATATAAATATATATATGTATA 107825

Qy        620 CATATATACATATATACACACATATATACACATACATATACACACACATAGATATACATA 679

Db     107824 TAAATATATATATTTATATTTATATGTATATAAATATATATATTTATATAAATATATATA 107765

Qy        680 TATATACACATAT-ATATACGTATATATATGTATATATATATGCTCCAGAGTT 731

Db     107764 TTTATATTTATATAATATACATACATATATATATATTTTTTTGAGATGGAGTT 107712


RESULT 36
US-10-437-963-12666/c
; Sequence 12666, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 12666
;   LENGTH: 1800
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_18773C.1
US-10-437-963-12666

  Query Match             5.7%;  Score 262;  DB 18;  Length 1800;
  Best Local Similarity   72.0%;  Pred. No. 1.1e-43;
  Matches  370;  Conservative    0;  Mismatches  140;  Indels    4;  Gaps    2;

Qy        229 TATATATATATATATATATATATACACACACACACATATATATATACACATATATATA 288

Db       1796 TATATATATATATATACATATATATATATACATATATATATATACATATATATATATA 1737

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348

Db       1736 CATATATATATATATATATATATATATATACATATATATATATACATATATATATATA 1677

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408

Db       1676 TATATATATATATATATATACGAAAATGTATATATGTATATATATGTATATATATATATA 1617

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468

Db       1616 TATGTATATATATGTATATATACGTATGTATATATATGTATATATACGTATGTATATATA 1557

Qy        469 CATATATACACATATATATAATATACACACATATATATACACATATATACACACATATAT 528

Db       1556 TGTATATATACGTATGTATACG-TATATACGTATGTATATATATGTATATATACGTATAT 1498

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACATATAT 588

Db       1497 ATATGTATATATACGTATATATATGTATATATATGTATATATACATATATATGTATATAT 1438

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATA---TACACACATATA 645

Db       1437 GTATATATATATGTATATATACATATATATGTATATATATGTATATGTTATGTATATGTG 1378
```

```
Qy        646 TACACATACATATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              |    ||| || || |   | ||  |||||  |||||||  | ||||  ||||  ||||||
Db       1377 TGTGTATATATGTATATGTATATGTATATATGTATATATGTATATATGTATATATATATA 1318

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAG 739
              ||| ||| |||||||      |  |  ||||| |
Db       1317 TATATATGTATATATATATATGTATATATAAGCG 1284


RESULT 37
US-10-719-993-6815
; Sequence 6815, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001496
;  CURRENT APPLICATION NUMBER: US/10/719,993
;  CURRENT FILING DATE:  2003-11-24
;  NUMBER OF SEQ ID NOS: 55342
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6815
;   LENGTH: 1980090
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1980090)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-719-993-6815

  Query Match             5.7%;  Score 262;  DB 18;  Length 1980090;
  Best Local Similarity   70.7%;  Pred. No. 2.5e-42;
  Matches  420;  Conservative    0;  Mismatches  165;  Indels    9;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC--AAAAT 58
              ||||| ||||||||||  |||||||| ||||||||||||||||||||||||  |||||
Db    1507113 AGTTCGAGACCAGCCTGGTCAACATGGTAAAACCCCATCTCTACTAAAAATACAAAAAAT 1507172

Qy         59 TAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGA 118
              ||| | ||  |||||      |||||||||||||||||||  |||||| ||| || ||||
Db    1507173 TAGCCGGGTGTGGTGGCGAGTGCCTGTAATCCCAGCTACTCTGGAGGCTGAGACATGAGA 1507232

Qy        119 ATCACTTGAACCTGCTGGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCC 177
              ||| ||||||||||  || || |||||||| | ||| | |||||  |||| | ||||||
Db    1507233 ATCGCTTGAACCTG--GGAGGCGGAGGTTGCAGTGAGCCGAGATCGCGCCACTACACTCC 1507290

Qy        178 AGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATAT 237
              ||||  ||||||||||||||||     ||||| ||||||||||   |    | | | ||
Db    1507291 AGCCTGGGCAACAAGAGCGAAACTCTGTCTCAAAAAAAAAAAAA---AAAAAAAAAAAAT 1507347

Qy        238 ATATATATATATATACACACACACACATATATATATACACATATATATACACATATAT 297
              |||||||||||||| | | | |   || ||||||||    |||||||  | |||| |
Db    1507348 ATATATATATATATATATATATATGTATGTATATATATGTGTATATATGTATATATGT 1507407

Qy        298 ACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACAC 357
              ||||||||   |||||||    || |||||   ||| || |||| |  | ||||||
Db    1507408 GCATATATATGTATATATATGTGTAGATATATGTATATATGTGTATATGTGTATATATGT 1507467

Qy        358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417
              |||| | || | || | || | || |||| |    | |||| |  |||||| |    |||
Db    1507468 ATATATGTATATATGTGTATATATGTATATATGTGTGTATATATGTATATATATGTGTAT 1507527

Qy        418 ATACACATATATACACATATATACACATATACACATATATACACATATATACATATATAC 477
              |||   |||||||    ||||||   ||||     ||||||   |||||||  | | |
Db    1507528 ATATGTATATATATGTGTATATATGTATATGTGTGTATATATGTATATATATGTGTGTGT 1507587

Qy        478 ACATATATATAATATACACACATATATATACACATATATACACACATATATACACATATA 537
              | |||| ||| ||||      ||||||||  | ||||||  | |  |||| |     ||||
Db    1507588 ATATATGTAT-ATATGTGTGTATATATATGTATATATATGTATATATATGTGTGTGTATA 1507646

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATA 591
              || |  |||||||   |||||||    | |||| ||||||  | |||||||||
Db    1507647 TATATGTATATATATGTATATATATGTGTGTATATATATATGTATATATATATA 1507700


RESULT 38
US-10-741-600-17676
; Sequence 17676, Application US/10741600
; Publication No. US20050026169A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001499
;  CURRENT APPLICATION NUMBER: US/10/741,600
;  CURRENT FILING DATE:  2003-12-22
;  NUMBER OF SEQ ID NOS: 73997
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17676
;   LENGTH: 1980090
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1980090)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-741-600-17676

  Query Match             5.7%;  Score 262;  DB 19;  Length 1980090;
  Best Local Similarity   70.7%;  Pred. No. 2.5e-42;
  Matches  420;  Conservative    0;  Mismatches  165;  Indels    9;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC--AAAAT 58
              ||||| ||||||||||  |||||||| ||||||||||||||||||||||||  |||||
Db    1507113 AGTTCGAGACCAGCCTGGTCAACATGGTAAAACCCCATCTCTACTAAAAATACAAAAAAT 1507172

Qy         59 TAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGA 118
              ||| | ||  |||||      |||||||||||||||||||  |||||| ||| || ||||
Db    1507173 TAGCCGGGTGTGGTGGCGAGTGCCTGTAATCCCAGCTACTCTGGAGGCTGAGACATGAGA 1507232

Qy        119 ATCACTTGAACCTGCTGGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCC 177
              ||| ||||||||||  || || |||||||| | ||| | |||||  |||| | ||||||
Db    1507233 ATCGCTTGAACCTG--GGAGGCGGAGGTTGCAGTGAGCCGAGATCGCGCCACTACACTCC 1507290

Qy        178 AGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATAT 237
              ||||  ||||||||||||||||     ||||| ||||||||||   |    | | | ||
Db    1507291 AGCCTGGGCAACAAGAGCGAAACTCTGTCTCAAAAAAAAAAAAA---AAAAAAAAAAAAT 1507347

Qy        238 ATATATATATATATACACACACACACATATATATATACACATATATATACACATATAT 297
              |||||||||||||| | | | |   || ||||||||    |||||||  | |||| |
Db    1507348 ATATATATATATATATATATATATGTATGTATATATATGTGTATATATGTATATATGT 1507407

Qy        298 ACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACAC 357
              ||||||||   |||||||    || |||||   ||| || |||| |  | ||||||
Db    1507408 GCATATATATGTATATATATGTGTAGATATATGTATATATGTGTATATGTGTATATATGT 1507467

Qy        358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417
              |||| | || | || | || | || |||| |    | |||| |  |||||| |    |||
Db    1507468 ATATATGTATATATGTGTATATATGTATATATGTGTGTATATATGTATATATATGTGTAT 1507527

Qy        418 ATACACATATATACACATATATACACATATACACATATATACACATATATACATATATAC 477
              |||   |||||||    ||||||   ||||     ||||||   |||||||  | | |
Db    1507528 ATATGTATATATATGTGTATATATGTATATGTGTGTATATATGTATATATATGTGTGTGT 1507587

Qy        478 ACATATATATAATATACACACATATATATACACATATATACACACATATATACACATATA 537
              | |||| ||| ||||      ||||||||  | ||||||  | |  |||| |     ||||
Db    1507588 ATATATGTAT-ATATGTGTGTATATATATGTATATATATGTATATATATGTGTGTGTATA 1507646

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATA 591
              || |  |||||||   |||||||    | |||| ||||||  | |||||||||
Db    1507647 TATATGTATATATATGTATATATATGTGTGTATATATATATGTATATATATATA 1507700


RESULT 39
```

```
US-10-257-166-11/c
; Sequence 11, Application US/10257166
; Publication No. US20040023230A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Method and Nucleic Acids for Analysing the Methylation of
;  TITLE OF INVENTION:  Genes Implicated in Pharmacogenomics
;  FILE REFERENCE: 5013.1011
;  CURRENT APPLICATION NUMBER: US/10/257,166
;  CURRENT FILING DATE:  2002-10-07
;  PRIOR APPLICATION NUMBER: PCT/EP01/07470
;    DE 10032529.7
;    DE 10043826.1
;  PRIOR FILING DATE: 2001-06-29
;    2000-06-30
;    2000-09-01
;  NUMBER OF SEQ ID NOS: 178
; SEQ ID NO 11
;   LENGTH: 12069
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (390, 398)
US-10-257-166-11

  Query Match             5.7%;  Score 261.4;  DB 17;  Length 12069;
  Best Local Similarity   70.5%;  Pred. No. 3.5e-43;
  Matches  363;  Conservative    0;  Mismatches  151;  Indels    1;  Gaps    1;

Qy        204 ATCTCAGAAAAAAAAAAAGGCATTTTATATATATATATATATATACACACACACAC 263
              || ||   | | | | ||     | |||||||||||| |||| |||||| | | | | |
Db       9846 ATATCTATATATATATAAATATATCTATATATATATAAATATCTATATATATATAAATAT 9787

Qy        264 ACATATATATATACACATATATATACACATATATACATATATACACATATATACACATAT 323
              |||||  |||| | || |||||| | ||| |||  ||||| ||| |||||  | ||||
Db       9786 CTATATACGTATAAATATCTATATATATATAAATATCTATATCCACGTATATCTATATAT 9727

Qy        324 ATATACACATACATATGTACACATATATATACACATATGTATACACATATATACACATAT 383
              ||||| | ||  |||| || | | ||||||| |   | |||||   ||||||| | | ||
Db       9726 ATATAAATATCTATATATATATAAATATATATATCCACGTATATCTATATATATATAAAT 9667

Qy        384 ATACACACATATATACACATATATACACACATATATACACATATATACACATATATACAC 443
              ||   | | ||||||| | |||||||  ||| |||||  | ||||||| | |||| |||
Db       9666 ATCTATATATATATAAATATATATATCCACGTATATCTATATATATATAAATATCTACGT 9607

Qy        444 ATATACACATATATACACATATATACATATATACACATATATATAATATACACACATATA 503
              ||| | | |||||| ||| |||||  ||||||| | | |||| | |  || | | |||||
Db       9606 ATAAATATATATATCCACGTATATCTATATATATATAAATATCT-ACGTATAAATATATA 9548

Qy        504 TATACACATATATACACACATATATACACATATATACACATATATATACACATATATACA 563
              ||| ||| |||||  | | ||||||| | || |||| | ||||| ||| | ||||  ||
Db       9547 TATCCACGTATATCTATATATATATAAATATCTATATATATATAAATATATATATCCACG 9488

Qy        564 CATATATACATATATACACATATATATACATATATACACATATATACATATATACACATA 623
              |||| || ||||||| | |||| |||| ||||||| | || | || ||||||| | |||
Db       9487 TATATCTATATATATATAAATATCTATATATATATAAATATCTCTATATATATAAATATA 9428

Qy        624 TATACATATATACACACATATATACACATACATATACACACACATAGATATACATATATA 683
              ||||   | |||.|   ||||||| | | | || || | | | ||| ||||  |||||||
Db       9427 TATATCCACATATATCTATATATATATAAATATCTATATATATATAAATATCTATATATA 9368

Qy        684 TACACATATATATACGTATATATATGTATATATAT 718
              || | | |||| ||  ||||||||  ||| |||||
Db       9367 TATATAAATATCTATATATATATAAATATCTATAT 9333


RESULT 40
US-10-719-993-6815/c
; Sequence 6815, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001496
;  CURRENT APPLICATION NUMBER: US/10/719,993
;  CURRENT FILING DATE:  2003-11-24
;  NUMBER OF SEQ ID NOS: 55342
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6815
;   LENGTH: 1980090
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1980090)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-2
US-10-719-993-6815

  Query Match             5.6%;  Score 259.6;  DB 18;  Length 1980090;
  Best Local Similarity   84.6%;  Pred. No. 7.9e-42;
  Matches  303;  Conservative    0;  Mismatches   54;  Indels    1;  Gaps    1;

Qy        270 ATATATACACATATATATACACATATATACATATATACACATATATACACATATATATAC 329
              ||||||| | ||| ||||| | ||| | ||||||||| ||||||||| ||||||||||||
Db    1507701 ATATATATATATACATATATATATACACACATATATATACATATATATACATATATATAC 1507642

Qy        330 ACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACAC 389
              ||| |||||| || |||||||||||| |||| || |||||||||||| ||||| ||||
Db    1507641 ACACACATATATATACATATATATACACATATATATACACACATATATACATATATACACAC 1507582

Qy        390 ACATATATACACATATATACACACATATATACACATATATACACATATATACACATATAC 449
              ||||||||| |||||||||||||||||||| | | ||| | | | ||||| | | ||| ||
Db    1507581 ACATATATATACATATATACACACATATACATATATACACATATATATACATATATACAC 1507522

Qy        450 ACATATATACACATATATACATATATACACATATATATA-ATATACACACATATATATAC 508
              | ||||||||| ||||| ||| ||||| |||||||| | ||||| ||| ||||| |||
Db    1507521 ATATATATACATATATACACACATATATACATATATACACATATATACATATATACATAT 1507462

Qy        509 ACATATATACACACATATATACACATATATACACATATATATACACATATATACACATAT 568
              | | | ||| |||||||||||| |||| |||||||||||||||| |||||| |||||||
Db    1507461 ATACACATATACACATATATACATATATCTACACATATATATACATATATATGCACATAT 1507402

Qy        569 ATACATATATACACATATATATACATATATACACATATATACATATATACACATATAT 626
              |||||||||||||||||||||||||||||| ||| | ||||||| ||||||| | |||| |
Db    1507401 ATACATATATACACATATATATACATACATATATATATATATATATATATATATTT 1507344


RESULT 41
US-10-741-600-17676/c
; Sequence 17676, Application US/10741600
; Publication No. US20050026169A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001499
;  CURRENT APPLICATION NUMBER: US/10/741,600
;  CURRENT FILING DATE:  2003-12-22
;  NUMBER OF SEQ ID NOS: 73997
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17676
;   LENGTH: 1980090
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1980090)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-2
US-10-741-600-17676
```

```
  Query Match             5.6%;  Score 259.6;  DB 19;  Length 1980090;
  Best Local Similarity   84.6%;  Pred. No. 7.9e-42;
  Matches  303;  Conservative    0;  Mismatches   54;  Indels    1;  Gaps    1;

Qy        270 ATATATACACATATATATACACATATATACATATATACACATATATACACATATATATAC 329
              ||||||| | ||| ||||| | ||| | ||||||||| ||||||||| ||||||||||||
Db    1507701 ATATATATATATACATATATATATACACACATATATATACATATATATACATATATATAC 1507642

Qy        330 ACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACAC 389
              ||| |||||| || |||||||||||||| |||| || |||||||||||||| ||||| ||||
Db    1507641 ACACACATATATATACATATATATACATATATATACACACATATATACATATATACACAC 1507582

Qy        390 ACATATATACACATATATACACACATATATACACATATATACACATATATACACATATAC 449
              |||||||| |||||||||||||||||||| | | ||| | | | ||||| | | ||| ||
Db    1507581 ACATATATATACATATATACACACATATACATATATACACATATATATACATATATACAC 1507522

Qy        450 ACATATATACACATATATACATATATACACATATATATA-ATATACACACATATATATAC 508
              | |||||||||| ||||| ||| ||||| ||||||||| | ||||| ||| ||||| |||
Db    1507521 ATATATATACATATATACACACATATATACATATATACACATATATACATATATACATAT 1507462

Qy        509 ACATATATACACACATATATACACATATATACACATATATATACACATATATACACATAT 568
              | | | ||| |||||||||||||| |||| |||||||||||||||| |||||| |||||||
Db    1507461 ATACACATATACACATATATACATATATCTACACATATATATACATATATATGCACATAT 1507402

Qy        569 ATACATATATACACATATATATACATATATACACATATATACATATATACACATATAT 626
              |||||||||||||||||||||||||||| ||| | ||||||| ||||||| | |||| |
Db    1507401 ATACATATATACACATATATATACATACATATATATATATATATATATATATATTT 1507344


RESULT 42
US-10-017-161-2363
; Sequence 2363, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
;  FILE REFERENCE: 084335/0152
;  CURRENT APPLICATION NUMBER: US/10/017,161
;  CURRENT FILING DATE:  2002-12-18
;  PRIOR APPLICATION NUMBER: JP 2001/246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2430
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2363
;   LENGTH: 1489
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: source
;   LOCATION: (1)..(1489)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(1289)
US-10-017-161-2363

  Query Match             5.6%;  Score 258.8;  DB 15;  Length 1489;
  Best Local Similarity   72.9%;  Pred. No. 4.8e-43;
  Matches  360;  Conservative    0;  Mismatches  132;  Indels    2;  Gaps    2;

Qy        229 TATATATATATATATATATATATACACACACACACATATATATATACACATATATATA 288
              | |||||||||| ||| |||||||| ||| | | |||||||| |||||| || ||   |||
Db        519 TGTATATATATGTATGTATATATATACATATATACACATATGTATATATACGTACC-ATA 577

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              |  |||||   |||||| || |||||| |||| |||||| || || ||||  ||| |||
Db        578 TATGTATATGTGTATATATACGTATATATACATGTATATATACGTATATATACACATATA 637

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              ||||| |||| | |||| || |||||| || |||| | | | ||| |   | ||||| |
Db        638 CATATATACATTTATATATACGTATATATACGTATACATATATATACACGTATATATACA 697


RESULT 43
US-10-292-798-2007
; Sequence 2007, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
;  FILE REFERENCE: 084335/166
;  CURRENT APPLICATION NUMBER: US/10/292,798
;  CURRENT FILING DATE:  2002-11-13
;  PRIOR APPLICATION NUMBER: 10/017,161
;  PRIOR FILING DATE: 2001-12-18
;  PRIOR APPLICATION NUMBER: JP 2001-246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2070
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2007
;   LENGTH: 1489
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   LOCATION: source
;   FEATURE:
;   LOCATION: (1)..(1489)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(1289)
US-10-292-798-2007

  Query Match             5.6%;  Score 258.8;  DB 17;  Length 1489;
  Best Local Similarity   72.9%;  Pred. No. 4.8e-43;
  Matches  360;  Conservative    0;  Mismatches  132;  Indels    2;  Gaps    2;

Qy        229 TATATATATATATATATATATATACACACACACACATATATATATACACATATATATA 288
              | |||||||||| ||| |||||||| ||| | | |||||||| |||||| || ||   |||
Db        519 TGTATATATATGTATGTATATATATACATATATACACATATGTATATATACGTACC-ATA 577

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              |  |||||   |||||| || |||||| |||| |||||| || || ||||  ||| |||
Db        578 TATGTATATGTGTATATATACGTATATATACATGTATATATACGTATATATACACATATA 637

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              ||||| |||| | |||| || |||||| || |||| | | | ||| |   | ||||| |
Db        638 CATATATACATTTATATATACGTATATATACGTATACATATATATACACGTATATATACA 697
```

```
Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              ||| |||||| ||    |||||| || |||| | | ||| ||    |||||| || |||| |
Db        698 TACATATATACACGTGTATATATACGTATACATATATACACGTGTATATATACGTATACA 757

Qy        469 CATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATATA 527
              |||||||||| | |||||| |  || ||| ||||| |     |||||||  | | |||
Db        758 TATATATACACGTGTATATATACGTATACATATATACACGTGTATATATACGTATACATA 817

Qy        528 TACACATATATACACATATATATACACATATATACACATATATACATATATACACATATA 587
              || |||  | || | |||  |||||| ||||| ||   ||||||  |  |||||| |||||
Db        818 TATACACGTGTATATATACGTATACATATATACACGTGTATATATACGTATACATATATA 877

Qy        588 TATACATATATACACATATATACATATATACACATATATACATATATACACACATATATA 647
              |    |||||||| | ||||| | |   |  ||| |||  |||||||||   ||||||
Db        878 CACGTGTATATACATACATATATACACGTGTATATACATATGTATATACACGTGTATATA 937

Qy        648 CACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATA 707
              |||  |||||| | || || |||||||| |||||||| |||||||||| ||||||
Db        938 TGTATATGTATACATATACGTATATATACATGTATATACACGTATATATACGTGTATATA 997

Qy        708 TGTATATATATATG 721
              |||||| ||||| |
Db        998 TGTATACATATACG 1011


RESULT 44
US-10-367-094-179/c
; Sequence 179, Application US/10367094
; Publication No. US20040170982A1
; GENERAL INFORMATION:
;  APPLICANT: David W. Morris
;  APPLICANT:  Marc Malandro
;  TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
;  FILE REFERENCE: 529452001500
;  CURRENT APPLICATION NUMBER: US/10/367,094
;  CURRENT FILING DATE:  2003-02-14
;  NUMBER OF SEQ ID NOS: 203
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 179
;   LENGTH: 163382
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(163382)
;   OTHER INFORMATION: n = A,T,C or G
US-10-367-094-179

  Query Match             5.6%;  Score 258.6;  DB 18;  Length 163382;
  Best Local Similarity   70.0%;  Pred. No. 4.2e-42;
  Matches  362;  Conservative    0;  Mismatches  154;  Indels    1;  Gaps    1;

Qy        204 ATCTCAGAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACAC 263
              || | |  | | | |||       |||| ||||||| |||| |||| | ||| | | |
Db     145185 ATATAAATATATATAAATAAATAAATATAAATATATAAATATGTATAAATACATATATAA 145126

Qy        264 ACATATATATATACACATATATATACACATATATACATATATACACATATATACACATAT 323
              | | ||||| ||| | | |||||||| | ||| ||| ||| ||| | ||| ||| | ||||
Db     145125 ATAAATATAAATATATAAATATATATAAATACATATATAAATATATATAAATATATATAT 145066

Qy        324 ATATACACATACATATGTACACATATATATACACATATGTATACACATATATACACATAT 383
                ||| ||||| ||||||| | ||||||| ||| ||||  ||| | | ||||| | | ||
Db     145065 TAATATACATAAATATGTATAAATATATAAACATATATTAATATATAAATATATATAAAT 145006

Qy        384 ATACACACATATATACACATATATACACACATATATACACATATATACACATATATACAC 443
              |||   | |||||||   ||||||| | | ||  ||| | ||| |||     ||||| ||
Db     145005 ATATTAATATATATATTTATATATAAATATATGAATATATATAAATATGTTAATATATAC 144946

Qy        444 ATATACACATATATACACATATATACATATATACACATATATATAATATACACACATATA 503
              ||||| | | ||| | | ||| | | ||||||| | |||||||    ||| | | |||||
Db     144945 ATATATATAAATAAATATATAAAAATATATATAAATATATATAACTAATATATAAATATA 144886

Qy        504 TATACACATATATACACACATATATACACATATATACACATATATATACACATATATACA 563
              |||| ||  ||||| | | |||||||  | ||||||  ||  |||||||  | |||||||
Db     144885 TATA-ACTAATATATATAAATATATAAATATATATGAACTAATATATATAAATATATAAA 144827

Qy        564 CATATATACATATATACACATATATATACATATATACACATATATACATATATACACATA 623
              | ||||| | ||| | | ||| ||| | ||||| | | |||||||  ||| ||| | | |
Db     144826 TAAATATATAAATAAATATATAAATAAATATATACAAATATATATAAATAAATATATAAA 144767

Qy        624 TATACATATATACACACATATATACACATACATATACACACACATAGATATACATATATA 683
              ||| ||||| | | | ||||| | | ||| ||||| | | | ||| ||| | | |||||
Db     144766 AATAAATATAAATAAATATATAAATAAATATATATAAATAAATATATATAAATAAATATA 144707

Qy        684 TACACATATATATACGTATATATATGTATATATATAT 720
              || | ||| |||||  || |||||| |||| ||| ||
Db     144706 TATAAATAAATATATATAAATATATATATAAATAAAT 144670


RESULT 45
US-09-818-657-3
; Sequence 3, Application US/09818657
; Publication No. US20030077773A1
; GENERAL INFORMATION:
;  APPLICANT: BRANDON, Rhonda et al.
;  TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
;  TITLE OF INVENTION:  NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
;  TITLE OF INVENTION:  AND USES THEREOF
;  FILE REFERENCE: CL001006-CIP
;  CURRENT APPLICATION NUMBER: US/09/818,657
;  CURRENT FILING DATE:  2001-03-28
;  PRIOR APPLICATION NUMBER: 09/730,002
;  PRIOR FILING DATE: 2000-12-06
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 82938
;   TYPE: DNA
;   ORGANISM: Human
US-09-818-657-3

  Query Match             5.6%;  Score 258.4;  DB 10;  Length 82938;
  Best Local Similarity   70.5%;  Pred. No. 3.4e-42;
  Matches  361;  Conservative    0;  Mismatches  146;  Indels    5;  Gaps    1;

Qy        214 AAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACATATATAT 273
              | | | | |    | |||| | | | |||| |||| |||||||||    | |||||| |||
Db       6946 ATACATATGTATATATATACACACACATATGTATACATACACACATGTATACATATGTAT 7005

Qy        274 ATACACATATATATACACATATATACATATATACACATATATACACATATATATACACAT 333
              |||   ||||| |||||||| ||||||||| || | || |||| | ||| | ||  | |
Db       7006 ATATGTATATAGATACACATGTATACATATGTATATATGTATATAGATACACATGTATAC 7065

Qy        334 ACATATGTACACATATATATACACATATGTATACACATATATACACATATATACACACAT 393
              | || | || |  |||||| | ||| |||||||||| || |||| |  |||||| | |||
Db       7066 ATATGTATATATGTATATAGATACACATGTATACATATGTATATATGTATATAGATACAC 7125

Qy        394 ATATACACATATATACACACATATATACACATATATACACATATATACACATATACACAT 453
              || || |||||| || | |  |||||| | | | ||  | | ||||  | ||||  | ||
Db       7126 ATGTATACATATGTATATATGTATATAGATACACATGTATACATATGTATATATGTATAT 7185

Qy        454 ATATACACATATATACATATATACACATATATATA-----ATATACACACATATATATAC 508
              | |||||||| ||||||||| || | || ||||||     | ||  | |||||| ||||
Db       7186 AGATACACATGTATACATATGTATATATGTATATAGATACACATGTATACATATGTATAT 7245

Qy        509 ACATATATACACACATATATACACATATATACACATATATATACACATATATACACATAT 568
              |  |||||| | ||| || || |||||| || | || ||||||||  ||| || | || |
Db       7246 ATGTATATAGATACACATGTATACATATGTATATATGTATATACATGTATGTATATATGT 7305

Qy        569 ATACATATATACACATATATATACATATATACACATATATACATATATACACATATATAC 628
              |||  ||||||||  ||| ||||  | ||||   ||||||  |||| || | | || ||
Db       7306 ATATGTATATACATGTATGTATATGTGTATATGTATATATGTATATGTATATACATGTAT 7365

Qy        629 ATATATACACACATATATACACATACATATACACACACATAGATATACATATATATACAC 688
              |||||| |||  ||||||||  |  ||||| | |||  || |||   ||||| ||| |
```

```
Db       7366 GTATATATACAGCTATATACATGTGTATATATATACAGCTATATACGTATATACATATAT 7425

Qy        689 ATATATATACGTATATATATGTATATATATAT 720
              || |||||| |||||| |||||| |||||||||
Db       7426 ATGTATATATGTATACATATGTGTATATATAT 7457


RESULT 46
US-10-017-161-2361
; Sequence 2361, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
;  FILE REFERENCE: 084335/0152
;  CURRENT APPLICATION NUMBER: US/10/017,161
;  CURRENT FILING DATE:  2002-12-18
;  PRIOR APPLICATION NUMBER: JP 2001/246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2430
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2361
;   LENGTH: 1486
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: source
;   LOCATION: (1)..(1486)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(1286)
US-10-017-161-2361

  Query Match             5.6%;  Score 258;  DB 15;  Length 1486;
  Best Local Similarity   70.4%;  Pred. No. 7e-43;
  Matches  345;  Conservative    0;  Mismatches  145;  Indels    0;  Gaps    0;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||| || ||| | ||||||| ||| |      |  | |||||||||| || | ||
Db        718 TGTATATATACATGTATGTGTATATATACATATGTGTATGTGTATATATACATATGTGTA 777

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              |    ||||||||||||| |   | |||| | | |||| |  |   |   |||| |||| | |
Db        778 TGTGTATATATACATATGTGTATGTGTATATATACATATGTGTATGTGTATATATACATA 837

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              | | |||| | | || | |  | |||| || | || |||| | |||||| |  | ||||
Db        838 TGTGTATATATACATATGTGTATATATGTATATATGTATATATACATATGTGTATATATG 897

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | |  |||||| ||||||||      ||||||   |||||    ||| || |||| |
Db        898 TATATATGTATATATACATATATGTGTGTATATATGTATATATGTGTATTTATACATGTG 957

Qy        467 TACATATATACACATATATATAATATACACACATATATATACACATATATACACACATAT 526
              |  | | || | |||| ||||    || |     | |||||| ||||||||  |    ||||
Db        958 CATGTGTGTATAAATATGTATATATTATATGTGTGTATATATACATATATGTATGTATAT 1017

Qy        527 ATACACATATATACACATATATATACACATATATACACATATATACATATATACACATAT 586
              ||||| ||||| | |||||||||| | | ||| | ||| |||   |||||||| ||| ||||
Db       1018 ATACATATATACATACATATATACATATATACACACATATACGCACATATAGACATATAT 1077

Qy        587 ATATACATATATACACATATATACATATATACACATATATACATATATACACACATATAT 646
              || || ||||| ||||||||||||||||||||    |||||||||  | ||| | |||||||||
Db       1078 ATGTATATATACACACATATATACATATATGTATATATACACACATATATACATATAT 1137

Qy        647 ACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATAT 706
              |    ||| ||| ||||| | ||| |||||  ||||||||||||||||||||    ||||||
Db       1138 ATGTATATATACACACATATATACATATATGTATATATACACATATATATGTATATATAC 1197

Qy        707 ATGTATATAT 716
```

```
              |  ||||||||
Db       1198 ACATATATAT 1207


RESULT 47
US-10-257-166-11
; Sequence 11, Application US/10257166
; Publication No. US20040023230A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Method and Nucleic Acids for Analysing the Methylation of
;  TITLE OF INVENTION:  Genes Implicated in Pharmacogenomics
;  FILE REFERENCE: 5013.1011
;  CURRENT APPLICATION NUMBER: US/10/257,166
;  CURRENT FILING DATE:  2002-10-07
;  PRIOR APPLICATION NUMBER: PCT/EP01/07470
;    DE 10032529.7
;    DE 10043826.1
;  PRIOR FILING DATE: 2001-06-29
;    2000-06-30
;    2000-09-01
;  NUMBER OF SEQ ID NOS: 178
; SEQ ID NO 11
;   LENGTH: 12069
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (390, 398)
US-10-257-166-11

  Query Match             5.6%;  Score 257.6;  DB 17;  Length 12069;
  Best Local Similarity   62.9%;  Pred. No. 2.1e-42;
  Matches  433;  Conservative    0;  Mismatches  249;  Indels    6;  Gaps    2;

Qy        226 TTTTATATATATATATATATATATATACACACACACACACATATATATATACACATATAT 285
              || |||||||||| |||| |||||||| | | | | ||    |||||||||   ||||||
Db       9239 TTATATATATATAGATATTTATATATATATAGATATACGTGGATATATATATTTATATAT 9298

Qy        286 ATACACATATATACATATATACACATATATACACATATATATACACATACATATGTACAC 345
              ||| | |||| || |||||||| | |||||      |||||  |||   ||| |||| || |
Db       9299 ATATAGATATTTATATATATATAGATATACGTGGATATAGATATTTATATATATATAGAT 9358

Qy        346 ATATATATACACATATGTATACACATATATACACATATATACACACATATATACACATAT 405
              || |||||| | ||||  |||    |||||||| | | ||||  | | |||||||| | ||||
Db       9359 ATTTATATATATATATAGATATTTATATATATATAGATATTTATATATATATAGATATAT 9418

Qy        406 ATACACACATATATACACATATATACACATATATACACATATACACATATATACACATAT 465
              |  | | ||||||  | |||||||  | |||| || | |||||  | |||| || | ||||
Db       9419 GTGGATATATATATTTATATATATAGAGATATTTATATATATATAGATATTTATATATAT 9478

Qy        466 ATACATATATACACATATATATA-----ATATACACACATATATATACACATATATACAC 520
              ||| |||||      |||||||||      |||||  | |  |||| |||| | |||||  | |
Db       9479 ATAGATATACGTGGATATATATATTTATATATATATAGATATTTATATATATATAGATAT 9538

Qy        521 ACATATATACACATATATACACATATATATACACATATATACACATATATACATATATAC 580
              || |   ||| | |||| || || || ||||   | |||||||  | |||||      |||||
Db       9539 ACGTGGATATATATATTTATACGTAGATATTTATATATATATAGATATACGTGGATATAT 9598

Qy        581 ACATATATATACATATATACACATATATACATATATACACATATATACATATATACACAC 640
              | || ||||   | ||||  | ||||||| | ||||||    |||||  |||| || | |
Db       9599 ATATTTATACGTAGATATTTATATATATATAGATATACGTGGATATATATATTTATATAT 9658

Qy        641 ATATATACACATACATATACACACACATAGATATACATATATATACACATATATATACGT 700
              |||||  | |  || ||||| | | | |||   |  | ||||||||  | |||||||||  |
Db       9659 ATATAGATATTTATATATATATAGATATACGTGGATATATATATTTATATATATATAGAT 9718

Qy        701 ATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGG 760
              || ||||| |||||| |||| |       |  |||          |  |   ||    |
```

```
Db       9719 ATTTATATATATATAGATATACGTGGATATAGATATTTATATATATATAGATATTTATAC 9778

Qy        761 GGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATG 820

Db       9779 GTATATAGATATTTATATATATATAGATATTTATATATATATAGATATATTTATATATAT 9838

Qy        821 GCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAA 880

Db       9839 ATAGATATATATGGATATATATATTTATAT-ATATATAATATTTATATATAGATATATAT 9897

Qy        881 GGTAACACTATATATATATATAGTGAAA 908

Db       9898 GGATATAGTATTTATATATATATAGATA 9925


RESULT 48
US-10-292-798-2005
; Sequence 2005, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
;  FILE REFERENCE: 084335/166
;  CURRENT APPLICATION NUMBER: US/10/292,798
;  CURRENT FILING DATE:  2002-11-13
;  PRIOR APPLICATION NUMBER: 10/017,161
;  PRIOR FILING DATE: 2001-12-18
;  PRIOR APPLICATION NUMBER: JP 2001-246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2070
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2005
;   LENGTH: 1486
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   LOCATION: source
;   FEATURE:
;   LOCATION: (1)..(1486)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(1286)
US-10-292-798-2005

  Query Match             5.5%;  Score 256.4;  DB 17;  Length 1486;
  Best Local Similarity   70.2%;  Pred. No. 1.5e-42;
  Matches  344;  Conservative    0;  Mismatches  146;  Indels    0;  Gaps    0;

Qy        227 TTTATATATATATATATATATATATACACACACACACATATATATATACACATATATA 286

Db        718 TGTATATATACATGTATGTGTATATATACATATGTGTATGTGTATATATACATATGTGTA 777

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346

Db        778 TGTGTATATATACATATGTGTATGTGTATATATACATATGTGTATGTGTATATATACATA 837

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406

Db        838 TGTGTATATATACATATGTGTATATATGTATATATGTATATATACATATGTGTATATATG 897

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466

Db        898 TATATATGTATATATACATATATGTGTGTATATATGTATATATGTGTATTTATATATGTG 957

Qy        467 TACATATATACACATATATATAATATACACACATATATATACACATATATACACACATAT 526

Db        958 CATGTGTGTATAAATATGTATATATTATATGTGTGTATATATACATATATGTATGTATAT 1017

Qy        527 ATACACATATATACACATATATATACACATATATACACATATATACATATATACACATAT 586

Db       1018 ATACATATATACATACATATATACATATATACACACATATACGCACATATAGACATATAT 1077

Qy        587 ATATACATATATACACATATATACATATATACACATATATACATATATACACACATATAT 646

Db       1078 ATGTATATATACACACATATATACATATATATGTATATATACACACATATATACATATAT 1137

Qy        647 ACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATAT 706

Db       1138 ATGTATATATACACACATATATACATATATGTATATATACACATATATATGTATATATAC 1197

Qy        707 ATGTATATAT 716

Db       1198 ACATATATAT 1207


RESULT 49
US-10-437-963-46068
; Sequence 46068, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 46068
;   LENGTH: 610
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_48971C.1
US-10-437-963-46068

  Query Match             5.5%;  Score 255.8;  DB 18;  Length 610;
  Best Local Similarity   69.6%;  Pred. No. 1.4e-42;
  Matches  364;  Conservative    0;  Mismatches  152;  Indels    7;  Gaps    1;

Qy        197 AAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACA 256

Db         39 AAATAACACCGAAAATATTTAATAAGAAAATTAATACGTAGGTACGGATGTATATATGAA 98

Qy        257 CACACACACATATATATATACACATATATATACACATATATACATATATACACATATATA 316

Db         99 AGTATGGACGGAAATATATATGTATGTATATAATAGTATGTATATGTATATCTATGTGAG 158

Qy        317 CACATATATATACACATACATATGTACACATATATATACACATATGTATACACATATATA 376

Db        159 TGTATATATATATGTATATATATATATATATATATATATATATATATATATATATATATA 218

Qy        377 CACATATATACACACATATATACACATATATACACACATATATACACATATATACACATA 436

Db        219 TATATATATATATATATATATATATATATATATATATATGTATGTGTATATATATATATA 278

Qy        437 TATACACATATACACATATATACACATATATACATATATACACATATATATAATATACAC 496

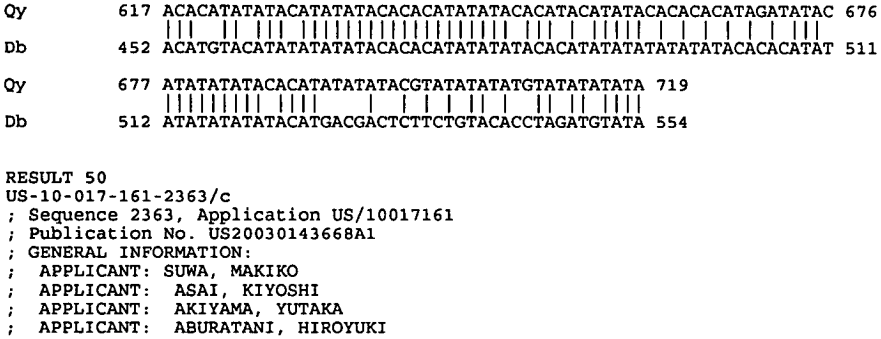
Db        279 TATATGTGTGTGTATATATATATATATATATATATATATATATATATATATAT-------AT 331

Qy        497 ACATATATATACACATATATACACACATATATACACATATATACACATATATATACACAT 556

Db        332 ATATATATATATACACACACACACACATGTATATATATATACACACACATGTATATAT 391

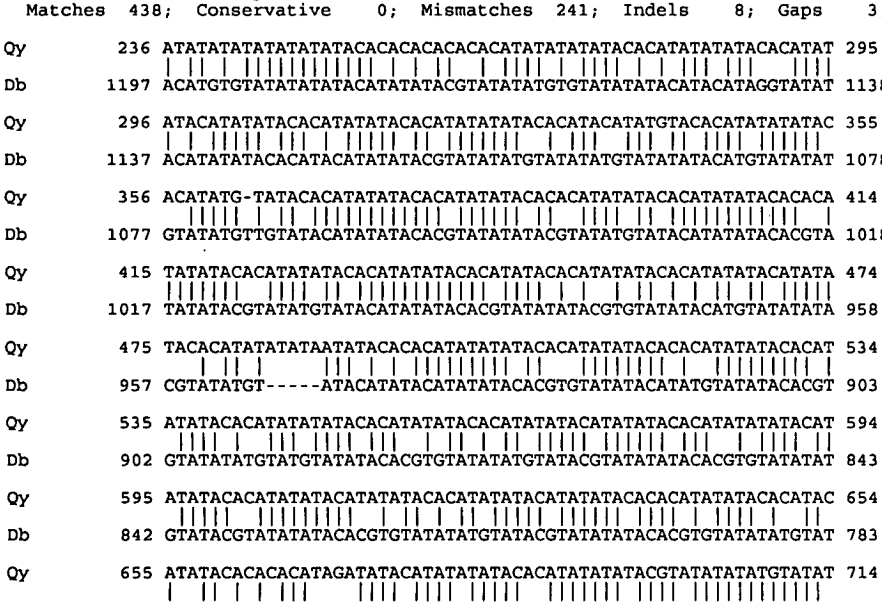
Qy        557 ATATACACATATATACATATATACACATATATATACATATATACACATATATACATATAT 616

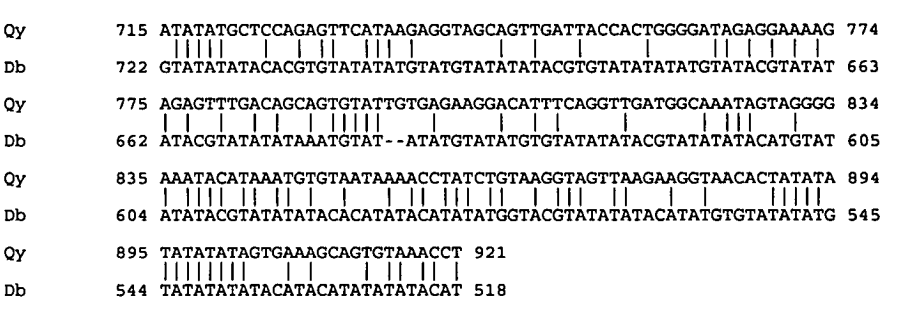
Db        392 ATATACACATGTACATATACATATACATATATATATATATATATACACATGTACATATAT 451
```



RESULT 50
US-10-017-161-2363/c
; Sequence 2363, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 084335/0152
; CURRENT APPLICATION NUMBER: US/10/017,161
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: JP 2001/246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2430
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2363
; LENGTH: 1489
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: source
; LOCATION: (1)..(1489)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (201)..(1289)
US-10-017-161-2363

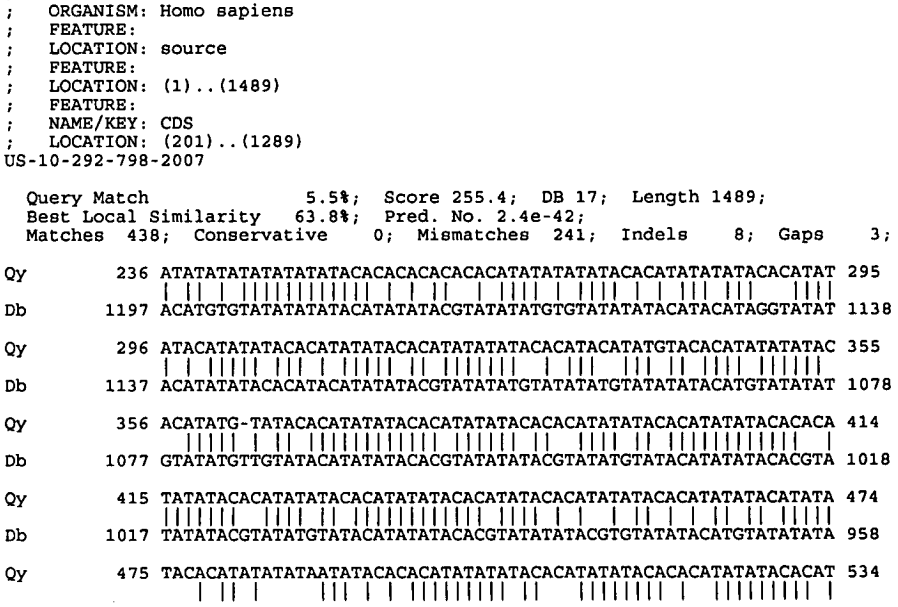
Query Match 5.5%; Score 255.4; DB 15; Length 1489;
Best Local Similarity 63.8%; Pred. No. 2.4e-42;
Matches 438; Conservative 0; Mismatches 241; Indels 8; Gaps 3;





RESULT 51
US-10-292-798-2007/c
; Sequence 2007, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
; FILE REFERENCE: 084335/166
; CURRENT APPLICATION NUMBER: US/10/292,798
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: 10/017,161
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: JP 2001-246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2070
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2007
; LENGTH: 1489
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; LOCATION: source
; FEATURE:
; LOCATION: (1)..(1489)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (201)..(1289)
US-10-292-798-2007

Query Match 5.5%; Score 255.4; DB 17; Length 1489;
Best Local Similarity 63.8%; Pred. No. 2.4e-42;
Matches 438; Conservative 0; Mismatches 241; Indels 8; Gaps 3;



```
Db        957 CGTATATGT-----ATACATATACATATATACACGTGTATATACATATGTATATACACGT 903

Qy        535 ATATACACATATATATACACATATATACACATATATACATATATACACATATATATACAT 594
              |||| |  ||| |||| |||  | || | || ||||| |||||| |||  | |||| ||
Db        902 GTATATATGTATGTATATACACGTGTATATATGTATACGTATATATACACGTGTATATAT 843

Qy        595 ATATACACATATATACATATATACACATATATACATATATACACACATATATACACATAC 654
              |||||  |||||||||  | || | || ||||| |||||| |||| | |||| |  ||
Db        842 GTATACGTATATATACACGTGTATATATGTATACGTATATATACACGTGTATATATGTAT 783

Qy        655 ATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATAT 714
              |  || | | |||    |||| |||| |||||  |||||| |||| ||||||||||||
Db        782 ACGTATATATACACGTGTATATATATGTATACGTATATATACACGTGTATATATGTATAC 723

Qy        715 ATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAG 774
              |||||   | | ||  ||| |     |  |  |     |    || | | | |
Db        722 GTATATATACACGTGTATATATGTATGTATATATACGTGTATATATATGTATACGTATAT 663

Qy        775 AGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGG 834
              | |  | | | |||||    |     | | |     |     | |||   |
Db        662 ATACGTATATATAAATGTAT--ATATGTATATGTGTATATATACGTATATATACATGTAT 605

Qy        835 AAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATA 894
              | |||| || || |  |   | || ||| ||  | |||  ||   | |     |||||
Db        604 ATATACGTATATATACACATATACATATATGGTACGTATATATACATATGTGTATATATG 545

Qy        895 TATATATAGTGAAAGCAGTGTAAACCT 921
              ||||||||  | |    | || || |
Db        544 TATATATATACATACATATATATACAT 518


RESULT 52
US-10-322-281-123/c
; Sequence 123, Application US/10322281
; Publication No. US20040126762A1
; GENERAL INFORMATION:
;  APPLICANT: David W. Morris
;  APPLICANT:  Marc S. Malandro
;  TITLE OF INVENTION: Novel Compositions and Methods in Cancer
;  FILE REFERENCE: 529452001000
;  CURRENT APPLICATION NUMBER: US/10/322,281
;  CURRENT FILING DATE:  2002-12-17
;  NUMBER OF SEQ ID NOS: 866
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 123
;   LENGTH: 104399
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(104399)
;   OTHER INFORMATION: n = A,T,C or G
US-10-322-281-123

  Query Match           5.5%;  Score 255.4;  DB 18;  Length 104399;
  Best Local Similarity   73.3%;  Pred. No. 1.6e-41;
  Matches  367;  Conservative    0;  Mismatches  131;  Indels    3;  Gaps    3;

Qy        230 ATATATATATATATATATATATACACACACACACATATATATATACACATATATATAC 289
              ||| ||| ||| ||||||| |||| ||| | | ||||| ||| ||||  | ||||||||
Db      13721 ATACATACATACATATATACATACGTACATATATACATACATACATACGTACATATATAC 13662

Qy        290 ACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATAT 349
              | |  || | ||||| | ||||| ||||| ||||| ||||  | | ||||  | |||||
Db      13661 ATACGTACATATATACATACATACATACATATATACATACGTACATATATACATACATAC 13602

Qy        350 ATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATAC 409
              |||   |||||| || | || || ||| | | ||| | ||| | ||||| |||||  |||
Db      13601 ATACGTACATATATACATACGTACATATATACATACATACATACATATATACATACGTAC 13542

Qy        410 ACACATATATACACATATATACACATATATACACATATACACATATATACACATATATAC 469
              | | ||| ||||| | |||   |||||||||||| | ||||| | ||||| ||||| ||||
Db      13541 ATATATACATACATACATACGTACATATATACATACATACATACATATATACATACATAC 13482

Qy        470 ATATATACACATATATATAATATACACACATATATATACACATATATACACACATATATA 529
              |||||||||  | ||| ||||||  | | | | ||  ||| | | ||| | ||| ||||| |
Db      13481 ATATATACATACATACATAATACATATATACATTCATATATACATA-ATACATATATACA 13423

Qy        530 CACATATATACACATATATAT-ACACATATATACACATAT-ATACATATATACACATATA 587
              |||||||||||| | ||| || | ||||||||||| | || |||||| || || |||||
Db      13422 TACATATATACATACATACATAATACATATATACATACATAATACATACATGCATATATA 13363

Qy        588 TATACATATATACACATATATACATATATACACATATATACATATATACACACATATATA 647
              || |||| ||| | | ||| ||| | ||||| | ||| | | ||||| ||| |  || |
Db      13362 CATGCATACATATATACATACACACACATACATACATACATACATATATACATACGTACA 13303

Qy        648 CACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATA 707
              | ||||||| | ||| || || ||||| | ||| ||||| ||||| |||| || |||||
Db      13302 TATATACATACATACATACGTACATATATACATACATACATATATACATACATACATATA 13243

Qy        708 TGTATATATATATGCTCCAGA 728
              |  ||| ||| ||  | || |
Db      13242 TACATACATACATAATACATA 13222


RESULT 53
US-10-292-798-607/c
; Sequence 607, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
;  FILE REFERENCE: 084335/166
;  CURRENT APPLICATION NUMBER: US/10/292,798
;  CURRENT FILING DATE:  2002-11-13
;  PRIOR APPLICATION NUMBER: 10/017,161
;  PRIOR FILING DATE: 2001-12-18
;  PRIOR APPLICATION NUMBER: JP 2001-246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2070
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 607
;   LENGTH: 84510
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   LOCATION: source
;   FEATURE:
;   LOCATION: (1)..(84510)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(272)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (29879)..(29999)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (45509)..(45616)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (50871)..(50974)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (58541)..(58638)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (60172)..(60304)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (62427)..(62580)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (71080)..(71140)
```

```
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (72348)..(72417)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (75647)..(75738)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (77542)..(77668)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (82828)..(82869)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (84170)..(84310)
US-10-292-798-607

  Query Match             5.5%;  Score 255;  DB 17;  Length 84510;
  Best Local Similarity   72.6%;  Pred. No. 1.7e-41;
  Matches  358;  Conservative    0;  Mismatches  130;  Indels    5;  Gaps    2;

Qy        232 ATATATATATATATATATATACACACACACACATATATATATACACATATATATACAC 291
              |||||||| |||| |||| | | | ||||| || |||||||||| ||  ||||| ||
Db      80292 ATATATATACATATTCTTATATATATATACACATATGTATATATACATATGAATATATAC 80233

Qy        292 ATATATACATATATACACATATATACACATATATATACACATACATATGTACACATATAT 351
              | ||||  |||||||| ||  ||| | | | ||||||| ||  |||| ||||||||||
Db      80232 ACATATGTATATATACATATGAATATATACACATATACATATGAATATATACACATATAC 80173

Qy        352 ATACACATATGTATACACATATATACACATATATACACACATATATACACATATATACAC 411
              |||  |||| || | | ||| |||  |||||||||| |||||  |  |||||  ||
Db      80172 ATATGAATATATACATATATACATATGAATATATACACATATATACGTATGAATATATAC 80113

Qy        412 ACATATATAC----ACATATATACACATATATACACATATACACATATATACACATATAT 467
              |||||||||      ||||||| ||| ||||| || |||  | ||||||||||||||||
Db      80112 ACATATATACGTATGAATATATATACACATATATACGTATGAATATATATACACATATAT 80053

Qy        468 ACATATATACACATATATATAATATACACACATATATATACACATATATACACACATATA 527
              || |||  | | ||||| |  ||||||  |  ||||||| ||  ||||||  |  |||
Db      80052 ACGTATGAATATATATACA-CATATACGTATGAATATATATACCCATATACGTATGAATA 79994

Qy        528 TACACATATATACACATATATATACACATATATACACATATATACATATATACACATATA 587
              |||||||||||| || ||  |||| | ||| | | | |||  ||| |||||| | | | |
Db      79993 TACACATATATATACGTACGTATATATATACACATATATACGTACGTATATATATACACA 79934

Qy        588 TATACATATATACACATATATACATATATACACATATATACATATATACACACATATATA 647
              |||| |  ||   | |||||||| ||||| || ||   | ||||||||||| |||||
Db      79933 TATATACGTACGTATATATATACACATATATACGTACGAATATATATACACATATATACG 79874

Qy        648 CACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATA 707
              ||  | ||||| ||||| ||| |  |||  ||||||| ||| ||||||||||   |||
Db      79873 TACGAATATATATACACATATATACGTACGAATATATATACACATATATACGTACGAATA 79814

Qy        708 TGTATATATATAT 720
              | |||| | ||||
Db      79813 TATATACACATAT 79801


RESULT 54
US-10-017-161-2375
; Sequence 2375, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
;  FILE REFERENCE: 084335/0152
;  CURRENT APPLICATION NUMBER: US/10/017,161
;  CURRENT FILING DATE:  2002-12-18
;  PRIOR APPLICATION NUMBER: JP 2001/246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2430
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2375
;   LENGTH: 1114
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: source
;   LOCATION: (1)..(1114)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(914)
US-10-017-161-2375

  Query Match             5.5%;  Score 254.2;  DB 15;  Length 1114;
  Best Local Similarity   64.0%;  Pred. No. 3.8e-42;
  Matches  399;  Conservative    0;  Mismatches  223;  Indels    1;  Gaps    1;

Qy        229 TATATATATATATATATATATACACACACACACATATATATATACACATATATATA 288
              ||||| ||||| | ||| |||||||| | | | ||| |||| |||||    || ||||||
Db        212 TATATGTATATTTGTATGTATATACAAATATATACATATATGTATATGTGTATGTATATA 271

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              || ||||||||||||||    || |||| ||| |||||||| |||  | | |  | |||
Db        272 CAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATATA 331

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              | ||| | | || |||     || |||| ||| ||||| ||| |||| |   | ||||||
Db        332 CAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATATA 391

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              || | ||||| | | || | ||   ||||| | | ||||||| |||| |   |  ||||||
Db        392 CAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATATA 451

Qy        469 CATATATACACATATATATAATATACACACATATATATACACATATATACACACATATAT 528
              || ||||| |||||||| |  |||  | | | | ||||| ||||||||     | |||||
Db        452 CAAATATATACATATATGT-GTATGTATATACAAATATATACATATATGTGTATGTATAT 510

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACATATAT 588
              ||| |||||||| |||| | ||   ||||| | | ||||| | |||| |   |  |||||
Db        511 ACAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATAT 570

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
              | | |||||||||| |||| |  || |||| ||| ||||| | |||||     || ||||
Db        571 ACAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATAT 630

Qy        649 ACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATAT 708
              ||| | |||||||  | |   |  || || ||| | ||| | | ||||||  |||| |||||
Db        631 ACAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATAT 690

Qy        709 GTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAG 768
              | |||||||        | ||   ||      |  | |   ||    ||   || | |
Db        691 ACAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATAT 750

Qy        769 GAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAG 828
              ||| | |    | |  || | |    | ||  | ||      ||      |||
Db        751 ACAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATAT 810

Qy        829 TAGGGGAAATACATAAATGTGTA 851
                  | ||||||| |||||||
Db        811 ACAAATATATACATATATGTGTA 833


RESULT 55
US-10-292-798-2017
; Sequence 2017, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
```

```
;  FILE REFERENCE: 084335/166
;  CURRENT APPLICATION NUMBER: US/10/292,798
;  CURRENT FILING DATE:  2002-11-13
;  PRIOR APPLICATION NUMBER: 10/017,161
;  PRIOR FILING DATE: 2001-12-18
;  PRIOR APPLICATION NUMBER: JP 2001-246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2070
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2017
;   LENGTH: 1114
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   LOCATION: source
;   FEATURE:
;   LOCATION: (1)..(1114)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(914)
US-10-292-798-2017

  Query Match             5.5%;  Score 254.2;  DB 17;  Length 1114;
  Best Local Similarity   64.0%;  Pred. No. 3.8e-42;
  Matches  399;  Conservative    0;  Mismatches  223;  Indels    1;  Gaps    1;

Qy        229 TATATATATATATATATATATATACACACACACACATATATATATACACATATATATA 288
              ||||| ||||| | ||| |||||||| | | | ||| |||| |||||    || ||||||
Db        212 TATATGTATATTTGTATGTATATACAAATATATACATATATGTATATGTGTATGTATATA 271

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              || |||||||||||||    || |||| ||| |||||||| ||| | | |  | |||
Db        272 CAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATATA 331

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              | ||| | | || |||    || |||| ||| ||||| ||| |||| |  |  ||||||
Db        332 CAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATATA 391

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              || | ||||| | | || | ||   ||||| | | ||||||| |||| |  |  ||||||
Db        392 CAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATATA 451

Qy        469 CATATATACACATATATATAATATACACACATATATATACACATATATACACACATATAT 528
              || ||||| |||||||| |  |||  | | | | ||||| ||||||||    |  |||||
Db        452 CAAATATATACATATATGT-GTATGTATATACAAATATATACATATATGTGTATGTATAT 510

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACATATAT 588
              ||| |||||||| |||| | ||   ||||| | | ||||| | |||| |  |  |||||
Db        511 ACAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATAT 570

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
              | | |||||||||| |||| |  || |||| ||| ||||| | |||||    || ||||
Db        571 ACAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATAT 630

Qy        649 ACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATAT 708
              ||| | ||||||| | |  |  || || ||| | ||| | | ||||||  |||| |||||
Db        631 ACAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATAT 690

Qy        709 GTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAG 768
              | |||||||       | ||   ||      |  |  |   ||    ||    || |
Db        691 ACAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATAT 750

Qy        769 GAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAG 828
              |||  | |    | |   || | |    | ||  | ||          ||      |||
Db        751 ACAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATAT 810

Qy        829 TAGGGGAAATACATAAATGTGTA 851
                    | ||||||| |||||||
Db        811 ACAAATATATACATATATGTGTA 833


RESULT 56
US-09-728-552-3/c
; Sequence 3, Application US/09728552
; Publication No. US20030096398A1
; GENERAL INFORMATION:
;  APPLICANT: Choo, Kong-Hong Andy
;  APPLICANT:  Du Sart, Desiree
;  APPLICANT:  Cancilla, Michael R.
;  TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
;  FILE REFERENCE: Davies Col
;  CURRENT APPLICATION NUMBER: US/09/728,552
;  CURRENT FILING DATE:  2000-12-02
;  PRIOR APPLICATION NUMBER: 09/078,294
;  PRIOR FILING DATE: 1998-05-13
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 80595
;   TYPE: DNA
;   ORGANISM: Nucleotide sequence of HC-contig
US-09-728-552-3

  Query Match             5.5%;  Score 253.6;  DB 10;  Length 80595;
  Best Local Similarity   71.9%;  Pred. No. 3.3e-41;
  Matches  361;  Conservative    0;  Mismatches  134;  Indels    7;  Gaps    2;

Qy        227 TTTATATATATATATATATATATATACACACACACACATATATATATACACATATATA 286
              | |||| |||| ||| | |||||| ||| | | ||| |  ||| | |||| ||| |||||
Db      16026 TATATACATATTTATGTCTATATACACATATATACATATTTATGTCTATATACACATATA 15967

Qy        287 TACACATATATACATATATACACATATATACACA------TATATATACACATACATATG 340
              |||| || |||   |||||||||||||||||| |       | ||||||||||| |||
Db      15966 TACATATTTATGTCTATATACACATATATACATATTTATGTCTATATACACATATATACA 15907

Qy        341 TACACATATATATACACATATGTATACACATATATACACATATATACACACATATATACA 400
              ||  '|| | |||| ||| || |||||| || |||    |||||| ||| | ||| |||
Db      15906 TATTTATGTCTATATACACATATATACATATTTATGTCTATATACACATATATACATATT 15847

Qy        401 CATATATACACACATATATACACATATATACACATATATACACATATACACATATATACA 460
              || | || | ||| ||||| |||||| ||    ||||||||||||| |||||| ||
Db      15846 TATGTCTATATACACATATATACATATTTATGTCTATATACACATATATACATATTTATG 15787

Qy        461 CATATATACATATATACACATATATAT-AATATACACACATATATATACACATATATACA 519
              |||||||| ||||| |||||| |||    |||| ||||||||||| | |  ||| | |
Db      15786 TCTATATACACATATATACATATTTATGTCTATATACACATATATACATATTTATGTCTA 15727

Qy        520 CACATATATACACATATATACACATATATATACACATATATACACATATATACATATATA 579
              | | | ||| | | ||||  |  | ||||||||||||||||||| || |||   ||||||
Db      15726 TATACACATATATACATATTTATGTCTATATACACATATATACATATTTATGTCTATATA 15667

Qy        580 CACATATATATACATATATACACATATATACATATATACACATATATACATATATACACA 639
              |||||||||| | || |||  | ||||| ||||||||||| || |||   ||||||||||
Db      15666 CACATATATACATATTTATGTATATATACACATATATACATATTTATGTCTATATACACA 15607

Qy        640 CATATATACACATACATATACACACACATAGATATACATATATATACACATATATATACG 699
              ||||| | |  ||  |||| |   |  || ||||   | |||||  ||||| |||||
Db      15606 TATATACATATTTATGTATATATTTATGTATATATGTGTGTATATGTACATACATATATA 15547

Qy        700 TATATATATGTATATATATATG 721
              ||| |||  || |||||||||
Db      15546 CATACATAAATACATATATATG 15525


RESULT 57
US-10-856-888-1/c
; Sequence 1, Application US/10856888
; Publication No. US20040235037A1
; GENERAL INFORMATION:
;  APPLICANT: Blumenfeld, Marta
;  APPLICANT:  Bougueleret, Lydie
;  APPLICANT:  Chumakov, Ilya
;  TITLE OF INVENTION: Polymorphic Markers of Prostate Carcinoma Tumor Antigen-1 (PCTA-1
;  FILE REFERENCE: GEN-T112XC1
;  CURRENT APPLICATION NUMBER: US/10/856,888
;  CURRENT FILING DATE:  2004-05-27
```

; PRIOR APPLICATION NUMBER: US/09/326,402
; PRIOR FILING DATE: 1999-06-04
; PRIOR APPLICATION NUMBER: 60/088,187
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/102,324
; PRIOR FILING DATE: 1998-09-28
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 106746
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..68647
; OTHER INFORMATION: 5'regulation region
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 66647..68647
; OTHER INFORMATION: promoter
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 97156..106746
; OTHER INFORMATION: 3'regulation region
; FEATURE:
; NAME/KEY: exon
; LOCATION: 68648..68741
; OTHER INFORMATION: exon0
; FEATURE:
; NAME/KEY: exon
; LOCATION: 70647..70794
; OTHER INFORMATION: exon1
; FEATURE:
; NAME/KEY: exon
; LOCATION: 82208..82296
; OTHER INFORMATION: exon2
; FEATURE:
; NAME/KEY: exon
; LOCATION: 83613..83823
; OTHER INFORMATION: exon3
; FEATURE:
; NAME/KEY: exon
; LOCATION: 85298..85417
; OTHER INFORMATION: exon4
; FEATURE:
; NAME/KEY: exon
; LOCATION: 86389..86445
; OTHER INFORMATION: exon5
; FEATURE:
; NAME/KEY: exon
; LOCATION: 87496..87522
; OTHER INFORMATION: exon6
; FEATURE:
; NAME/KEY: exon
; LOCATION: 87650..87775
; OTHER INFORMATION: exon6bis
; FEATURE:
; NAME/KEY: exon
; LOCATION: 88295..88383
; OTHER INFORMATION: exon7
; FEATURE:
; NAME/KEY: exon
; LOCATION: 89484..89649
; OTHER INFORMATION: exon8
; FEATURE:
; NAME/KEY: exon
; LOCATION: 92749..97155
; OTHER INFORMATION: exon9
; FEATURE:
; NAME/KEY: exon
; LOCATION: 92749..92883
; OTHER INFORMATION: exon9bis
; FEATURE:
; NAME/KEY: exon
; LOCATION: 95821..97155
; OTHER INFORMATION: exon9ter
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 70647..70794
; OTHER INFORMATION: homology with genset EST : A241850
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 68648..68741
; OTHER INFORMATION: homology with genset EST : A241850
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 82208..82229
; OTHER INFORMATION: homology with genset EST : A241850
; FEATURE:
; NAME/KEY: allele
; LOCATION: 278
; OTHER INFORMATION: 99-1601-278 : polymorphic base A or C
; FEATURE:
; NAME/KEY: allele
; LOCATION: 402
; OTHER INFORMATION: 99-1601-402 : polymorphic base A or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 472
; OTHER INFORMATION: 99-1601-472 : polymorphic base A or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 2955
; OTHER INFORMATION: 99-13801-100 : polymorphic base T or C
; FEATURE:
; NAME/KEY: allele
; LOCATION: 12167
; OTHER INFORMATION: 99-13806-166 : polymorphic base G or A
; FEATURE:
; NAME/KEY: allele
; LOCATION: 12536
; OTHER INFORMATION: 99-13799-376 : polymorphic base T or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 17593
; OTHER INFORMATION: 99-13798-297 : polymorphic base T or C
; FEATURE:
; NAME/KEY: allele
; LOCATION: 17606
; OTHER INFORMATION: 99-13798-284 : polymorphic base T or C
; FEATURE:
; NAME/KEY: allele
; LOCATION: 22079
; OTHER INFORMATION: 99-1602-200 : polymorphic base G or C
; FEATURE:
; NAME/KEY: allele
; LOCATION: 28964
; OTHER INFORMATION: 99-13794-186 : polymorphic base T or C
; FEATURE:
; NAME/KEY: allele
; LOCATION: 29003
; OTHER INFORMATION: 99-13794-147 : polymorphic base C or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 31077
; OTHER INFORMATION: 99-13812-384 : polymorphic base T or C
; FEATURE:
; NAME/KEY: allele
; LOCATION: 31766
; OTHER INFORMATION: 99-13805-313 : polymorphic base T or C
; FEATURE:
; NAME/KEY: allele
; LOCATION: 34791
; OTHER INFORMATION: 99-1587-281 : polymorphic base A or G
; FEATURE:
; NAME/KEY: allele

```
;   LOCATION: 45751
;   OTHER INFORMATION: 99-1582-430 : polymorphic base C or T
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 49847
;   OTHER INFORMATION: 99-1585-465 : polymorphic base T or C
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 49855
;   OTHER INFORMATION: 99-1585-457 : polymorphic base T or C
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 49886
;   OTHER INFORMATION: 99-1585-426 : polymorphic base G or A
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 49900
;   OTHER INFORMATION: 99-1585-412 : polymorphic base G or A
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 49906
;   OTHER INFORMATION: 99-1585-406 : polymorphic base C or A
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 49921
;   OTHER INFORMATION: 99-1585-391 : polymorphic base C or A
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 49939
;   OTHER INFORMATION: 99-1585-373 : polymorphic base G or A
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 50256
;   OTHER INFORMATION: 99-1585-55 : polymorphic base C or A
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 54955
;   OTHER INFORMATION: 99-1607-373 : polymorphic base T or C
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 64239
;   OTHER INFORMATION: 99-1577-105 : polymorphic base A or G
;   FEATURE:
;   NAME/KEY: allele

  Query Match             5.5%;  Score 253;  DB 18;  Length 106746;
  Best Local Similarity   69.4%;  Pred. No. 5e-41;
  Matches  416;  Conservative    0;  Mismatches  175;  Indels    8;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              |||||||||||||||||  |||||||||||||||||| |||||||||||||||  ||||||
Db      38717 AGTTCAAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAAAATT 38658

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||  |||||||||  | ||||||||||||||||||||||  |||| | ||||||||||||
Db      38657 AGCTGGGCATGGTGGTGCACGCCTGTAATCCCAGCTACTCAGGAGACTGAGGCAGGAGAA 38598

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCCA 178
              ||||||||||||  || |||||||| ||| || ||| ||| |    | ||| |||||||||
Db      38597 TCACTTGAACCTG--GGAGGTGGAGTTTGTGGTGAGCCAAAACTTTCCCACTGCACTCCA 38540

Qy        179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCAT-TTTATATATAT 237
              |||      |   | |   | || |   | | | |      | |    ||     | ||||
Db      38539 GCCAGTATATATATATATATATATACACACGTACACGTGTACACATATACACACATATGC 38480

Qy        238 ATATATATATATATACACACACACACACATATATATATACACATATATATACACATATAT 297
              | |  | || ||| ||||||| |||||   | ||| || ||||||| || |  | || ||
Db      38479 ACACGTGTACATACACACACATACACATGCACATACATGCACATATGTACATGCGTACAT 38420

Qy        298 ACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACAC 357
              ||| | ||  ||||||||||||| || || | |  ||||    ||||| || || | |
Db      38419 ACACACATGTACATATATACACACATGTACATATGTACACCCATACACACATGTACATAT 38360
```

```
Qy        358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417
              ||      | ||| | ||  |||||| ||||| |||||| | |  || |||   |||
Db      38359 GTACACCCATACACACATGTACATATGTACACCCATATACATATGTACATATGTACACCC 38300

Qy        418 ATACACATATATACACATATATACACATATACACATATATACACATATATACATATATAC 477
              |||||||||| |||| || || || |||| |||  || |||   | |   | | | ||
Db      38299 ATACACATATGTACATATGTACACCCATACACATGTACATATGTACACCCATACACATGT 38240

Qy        478 ACATATATATAATATACACACATATATATACACATATATACACACATATATACACATATA 537
              ||||||  |   | || |||| || || |||   ||||  | |||||||| |||| |||||
Db      38239 ACATATGT---ACATCCACATATGTACATATGTATATCCATACACATATGTACATATATA 38183

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATACATAT 596
              ||||||||| || | |  || ||| | || | |  |||| ||   |||||||||   ||
Db      38182 CACACATATGTACATATGTACATATATATTTGTGTATATCTATGTATATATATAAGAAT 38124
```

```
RESULT 58
US-10-856-888-12/c
; Sequence 12, Application US/10856888
; Publication No. US20040235037A1
; GENERAL INFORMATION:
;  APPLICANT: Blumenfeld, Marta
;  APPLICANT:  Bougueleret, Lydie
;  APPLICANT:  Chumakov, Ilya
;  TITLE OF INVENTION: Polymorphic Markers of Prostate Carcinoma Tumor Antigen-1 (PCTA-1)
;  FILE REFERENCE: GEN-T112XC1
;  CURRENT APPLICATION NUMBER: US/10/856,888
;  CURRENT FILING DATE:  2004-05-27
;  PRIOR APPLICATION NUMBER: US/09/326,402
;  PRIOR FILING DATE: 1999-06-04
;  PRIOR APPLICATION NUMBER: 60/088,187
;  PRIOR FILING DATE: 1998-06-05
;  PRIOR APPLICATION NUMBER: 60/102,324
;  PRIOR FILING DATE: 1998-09-28
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
;   LENGTH: 106746
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 1..68647
;   OTHER INFORMATION: 5'regulation  region
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 66647..68647
;   OTHER INFORMATION: promoter
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 97156..106746
;   OTHER INFORMATION: 3'regulation  region
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 68648..68741
;   OTHER INFORMATION: exon0
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 70647..70794
;   OTHER INFORMATION: exon1
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 82208..82296
;   OTHER INFORMATION: exon2
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 83613..83823
;   OTHER INFORMATION: exon3
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 85298..85417
;   OTHER INFORMATION: exon4
```

```
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 86389..86445
;   OTHER INFORMATION: exon5
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 87496..87522
;   OTHER INFORMATION: exon6
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 87650..87775
;   OTHER INFORMATION: exon6bis
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 88295..88383
;   OTHER INFORMATION: exon7
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 89484..89649
;   OTHER INFORMATION: exon8
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 92749..97155
;   OTHER INFORMATION: exon9
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 92749..92883
;   OTHER INFORMATION: exon9bis
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 95821..97155
;   OTHER INFORMATION: exon9ter
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 70647..70794
;   OTHER INFORMATION: homology with genset EST : A241850
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 68648..68741
;   OTHER INFORMATION: homology with genset EST : A241850
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 82208..82229
;   OTHER INFORMATION: homology with genset EST : A241850
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 278
;   OTHER INFORMATION: 99-1601-278 : polymorphic base A or C
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 402
;   OTHER INFORMATION: 99-1601-402 : polymorphic base w= A or T
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 472
;   OTHER INFORMATION: 99-1601-472 : polymorphic base A or T
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 2955
;   OTHER INFORMATION: 99-13801-100 : polymorphic base T or C
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 12167
;   OTHER INFORMATION: 99-13806-166 : polymorphic base G or A
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 12536
;   OTHER INFORMATION: 99-13799-376 : polymorphic base T or G
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 17593
;   OTHER INFORMATION: 99-13798-297 : polymorphic base T or C
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 17606
;   OTHER INFORMATION: 99-13798-284 : polymorphic base T or C
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 22079
;   OTHER INFORMATION: 99-1602-200 : polymorphic base G or C
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 28964
;   OTHER INFORMATION: 99-13794-186 : polymorphic base T or C
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 29003
;   OTHER INFORMATION: 99-13794-147 : polymorphic base C or G
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 31077
;   OTHER INFORMATION: 99-13812-384 : polymorphic base T or C
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 31766
;   OTHER INFORMATION: 99-13805-313 : polymorphic base T or C
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 34791
;   OTHER INFORMATION: 99-1587-281 : polymorphic base A or G
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 45751
;   OTHER INFORMATION: 99-1582-430 : polymorphic base C or T
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 49847
;   OTHER INFORMATION: 99-1585-465 : polymorphic base T or C
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 49855
;   OTHER INFORMATION: 99-1585-457 : polymorphic base T or C
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 49886
;   OTHER INFORMATION: 99-1585-426 : polymorphic base G or A
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 49900
;   OTHER INFORMATION: 99-1585-412 : polymorphic base G or A
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 49906
;   OTHER INFORMATION: 99-1585-406 : polymorphic base C or A
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 49921
;   OTHER INFORMATION: 99-1585-391 : polymorphic base C or A
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 49939
;   OTHER INFORMATION: 99-1585-373 : polymorphic base G or A
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 50256
;   OTHER INFORMATION: 99-1585-55 : polymorphic base C or A
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 54955
;   OTHER INFORMATION: 99-1607-373 : polymorphic base T or C
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 64239
;   OTHER INFORMATION: 99-1577-105 : polymorphic base A or G
;   FEATURE:
;   NAME/KEY: allele
```

```
  Query Match             5.5%;  Score 253;  DB 18;  Length 106746;
  Best Local Similarity   69.4%;  Pred. No. 5e-41;
  Matches  416;  Conservative    0;  Mismatches  175;  Indels    8;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              |||||||||||||||||  ||||||||||||||||| ||||||||||||||| ||||||
Db      38717 AGTTCAAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAAAATT 38658

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   |||||||| | ||||||||||||||||||||||  |||| | ||||||||||||
Db      38657 AGCTGGGCATGGTGGTGCACGCCTGTAATCCCAGCTACTCAGGAGACTGAGGCAGGAGAA 38598

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCCA 178
              ||||||||||||  || |||||||  ||| || ||| ||| |    | ||| |||||||||
Db      38597 TCACTTGAACCTG--GGAGGTGGAGTTTGTGGTGAGCCAAAACTTTCCCACTGCACTCCA 38540

Qy        179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCAT-TTTATATATAT 237
              |||    |   | |   | || |   | |  | |     | |   ||    | ||||
Db      38539 GCCAGTATATATATATATATATATACACACGTACACGTGTACACATATACACACATATGC 38480

Qy        238 ATATATATATATATACACACACACACATATATATATACACATATATATACACATATAT 297
              | |  | || ||| ||||||| |||||    | ||| || ||||||| || |  | || ||
Db      38479 ACACGTGTACATACACACACATACACATGCACATACATGCACATATGTACATGCGTACAT 38420

Qy        298 ACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACAC 357
              ||| | ||  |||||||||||||| || || | |  ||||    |||||| || || | |
Db      38419 ACACACATGTACATATATACACACATGTACATATGTACACCCATACACACATGTACATAT 38360

Qy        358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417
              ||      | ||| | ||  |||||| ||||| |||||| | |  || |||   |||
Db      38359 GTACACCCATACACACATGTACATATGTACACCCATATACATATGTACATATGTACACCC 38300

Qy        418 ATACACATATATACACATATATACACATATACACATATATACACATATATACATATATAC 477
              |||||||||| |||| || || || |||| |||  || |||   | |   | | | ||
Db      38299 ATACACATATGTACATATGTACACCCATACACATGTACATATGTACACCCATACACATGT 38240

Qy        478 ACATATATATAATATACACACATATATATACACATATATACACACATATATACACATATA 537
              ||||||  |   | || |||| || || |||   ||||  | ||||||| |||| |||||
Db      38239 ACATATGT---ACATCCACATATGTACATATGTATATCCATACACATATGTACATATATA 38183

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATACATAT 596
              ||||||||  || | |  || ||| | || | |  |||| ||   |||||||||   ||
Db      38182 CACACATATGTACATATGTACATATATATTTGTGTATATCTATGTATATATATAAGAAT 38124


RESULT 59
US-09-728-552-4/c
; Sequence 4, Application US/09728552
; Publication No. US20030096398A1
; GENERAL INFORMATION:
;  APPLICANT: Choo, Kong-Hong Andy
;  APPLICANT:  Du Sart, Desiree
;  APPLICANT:  Cancilla, Michael R.
;  TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
;  FILE REFERENCE: Davies Col
;  CURRENT APPLICATION NUMBER: US/09/728,552
;  CURRENT FILING DATE:  2000-12-02
;  PRIOR APPLICATION NUMBER: 09/078,294
;  PRIOR FILING DATE: 1998-05-13
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 80246
;   TYPE: DNA
;   ORGANISM: Nucleotide sequence of NC-contig
US-09-728-552-4

  Query Match             5.5%;  Score 252.6;  DB 10;  Length 80246;
  Best Local Similarity   71.9%;  Pred. No. 5.3e-41;
  Matches  360;  Conservative    0;  Mismatches  134;  Indels    7;  Gaps    2;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | || |||| ||| | |||||| | ||| | | ||| |  ||| | |||| ||| | |||
Db      15787 TATACATATTTATGTCTATATACACACATATATACATATTTATGTCTATATACACACATA 15728

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              |||| || |||   |||||||||||||||||| || ||| |  | ||||| || || |||
Db      15727 TACATATTTATGTCTATATACACATATATACATATTTATGTCTATATACACATATATACA 15668

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              ||| |||    ||||  | | | ||| |||   || | || | ||| |||||| ||||||
Db      15667 TATTTATGTCTATATACACATATATACATATTTATGTCTATATACACATATATACATATT 15608

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              |  |   |||||||||||||||||| || |||    ||||||||||||||| || ||
Db      15607 T--ATGTCTATATACACATATATACATATTTATGTCTATATACACATATATACATATTTA 15550

Qy        467 TACATATATACACATATATA-----TAATATACACACATATATATACACATATATACACA 521
              |   ||||||||||||||||     | || |  | | | | ||||| ||||||  ||
Db      15549 TGTCTATATACACATATATACATATTTATGTCTATATACACATATATACATATTTATGTC 15490

Qy        522 CATATATACACATATATACACATATATATACACATATATACACATATATACATATATACA 581
              ||||| ||| ||| |||   || | |||| ||| ||||| |||||| ||  | |||| |
Db      15489 TATATACACATATACATATTTATGTCTATATACACATATATACATATTTATGTCTATATA 15430

Qy        582 CATATATATACATATATACACATATATACATATATACACATATATACATATATACACACA 641
              || ||||||||||| ||    |||||||| ||||| |||||| ||  |||||| |||||
Db      15429 CACATATATACATATTTATGTCTATATACACATATATACATATTTATGTATATATACACA 15370

Qy        642 TATATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTA 701
              |||||||| ||  || |  | | ||| | |||||||||| |||  | |||| |||  ||
Db      15369 TATATACATATTTATGTCTATATACACATATATACATATTTATGTATATATTTATGTATA 15310

Qy        702 TATATATGTATATATATATGC 722
              ||| | |||||||  || || |
Db      15309 TATGTGTGTATATGTACATAC 15289


RESULT 60
US-09-818-657-3/c
; Sequence 3, Application US/09818657
; Publication No. US20030077773A1
; GENERAL INFORMATION:
;  APPLICANT: BRANDON, Rhonda et al.
;  TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
;  TITLE OF INVENTION:  NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
;  TITLE OF INVENTION:  AND USES THEREOF
;  FILE REFERENCE: CL001006-CIP
;  CURRENT APPLICATION NUMBER: US/09/818,657
;  CURRENT FILING DATE:  2001-03-28
;  PRIOR APPLICATION NUMBER: 09/730,002
;  PRIOR FILING DATE: 2000-12-06
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 82938
;   TYPE: DNA
;   ORGANISM: Human
US-09-818-657-3

  Query Match             5.5%;  Score 252.6;  DB 10;  Length 82938;
  Best Local Similarity   70.7%;  Pred. No. 5.4e-41;
  Matches  350;  Conservative    0;  Mismatches  144;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | |||| | |||| |||| ||||||||| | | |    | |||  ||||||    ||||||
Db       7454 TATATACACATATGTATACATATATACATATATATGTATATACGTATATAGCTGTATATA 7395

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || ||| || || |||   |  | ||||| | |||| ||||||| ||||||| |||| |
Db       7394 TATACACATGTATATAGCTGTATATATACATACATGTATATACATATACATATATACATA 7335

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              || | |||  |||||   | || | | ||||||| ||||| | |||   |||||||| |||||
Db       7334 TACACATATACATACATGTATATACATATACATATATACATACATGTATATACATATATA 7275
```

```
Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              | |   ||| ||    || |||| |||||||||| || || |||| | ||   |||||
Db       7274 CATATGTATACATGTGTATCTATATACATATATACATATGTATACATGTGTATCTATATA 7215

Qy        467 TACATATATACACATATATAT-AATATACACACATATATATACACATATATACACACATA 525
              | ||||| | |  |||| ||   |||  | | | |||||||||| || ||||||    ||
Db       7214 CATATATACATATGTATACATGTGTATCTATATACATATATACATATGTATACATGTGTA 7155

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              | ||  | | ||||| |||||| ||||||  | |||  | ||| ||| ||| |||   |||
Db       7154 TCTATATACATATATACATATGTATACATGTGTATCTATATACATATATACATATGTATA 7095

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              ||  |  || |||| |||||||||||||| || |||| | ||  |||||||| | ||| |
Db       7094 CATGTGTATCTATATACATATATACATATGTATACATGTGTATCTATATACATATATACA 7035

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              ||   |||||| |  |   | ||| ||||| | |||| || |||| | | || ||||| |
Db       7034 TATGTATACATGTGTATCTATATACATATATACATATGTATACATGTGTGTATGTATACA 6975

Qy        706 TATGTATATATATAT 720
              ||||| | | |||||
Db       6974 TATGTGTGTGTATAT 6960
```

RESULT 61
US-10-417-375-36/c
; Sequence 36, Application US/10417375
; Publication No. US20040219528A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc Malandro
; TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
; FILE REFERENCE: 529452001600
; CURRENT APPLICATION NUMBER: US/10/417,375
; CURRENT FILING DATE: 2003-04-15
; NUMBER OF SEQ ID NOS: 176
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 36
; LENGTH: 91697
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-417-375-36

Query Match 5.5%; Score 252.2; DB 18; Length 91697;
Best Local Similarity 69.6%; Pred. No. 6.8e-41;
Matches 357; Conservative 0; Mismatches 153; Indels 3; Gaps 1;

```
Qy        210 GAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACACATAT 269
              | | | | | |  ||  | || |||| |||||| ||||| ||| | |   | | | ||||
Db      20066 GCATATATATACAGCTATATAGATATGTATATAGATATACATATAGATGTATATATATAT 20007

Qy        270 ATATATACACAT---ATATATACACATATATACATATATACACATATATACACATATATA 326
              |||||||| ||   || || | |||||||| |  |||||  | ||| ||| | | |||||
Db      20006 CTATATACATATGCCATGTAGATACATATACATCTATATGTATATAGATATATAGATATA 19947

Qy        327 TACACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATA 386
              |  | ||| |||| |  | |||||||| || |||| |||| | ||||||||| || ||||
Db      19946 TGTATATAGATATATGTAGATATATATCCATATATTTATATAGATATATACATATGTATA 19887

Qy        387 CACACATATATACACATATATACACACATATATACACATATATACACATATATACACATA 446
              | | |||| || | || |||| ||  |||||   | ||| ||| | | ||||| |||||
Db      19886 TAGATATATTTATATATGTATATACCTATATACGTATATAGATATATACATATATACATA 19827

Qy        447 TACACATATATACACATATATACATATATACACATATATATAATATACACACATATATAT 506
              ||   ||||||| | ||||||||||||||||| || ||||  |||||||  |    |||||
Db      19826 TATGTATATATAGATATATATACATATATACATATGTATACTATATACATAATAGTATAT 19767

Qy        507 ACACATATATACACACATATATACACATATATACACATATATATACACATATATACACAT 566
              ||| || |||| |   ||| ||| |  |||| | |  |||||||  | | || ||   ||
Db      19766 ACATATGTATATATGTATACATAGATGTATACATATGTATATATGTATACATCTATGTAT 19707

```
Qy        567 ATATACATATATACACATATATATACATATATACACATATATACATATATACACATATAT 626
              | ||| || ||||||  ||| |||||||| ||  | | ||| | ||||| | |  |||||
Db      19706 ACATAGATGTATACATCTATGTATACATAGATGTATACATAGATATATACATATGTATAT 19647

Qy        627 ACATATATACACACATATATACACATACATATACACACACATAGATATACATATATATAC 686
              | || ||||   |  |||||| ||||  ||||  |    | ||||| ||| ||| ||| |
Db      19646 AGATGTATATCAATCTATATATACATATGTATATATGTATATAGAAATATATACATACAT 19587

Qy        687 ACATATATATACGTATATATATGTATATATATA 719
              |  |||||||   |||| |||| || || ||||
Db      19586 ATGTATATATGTATATAGATATATAGATGTATA 19554
```

RESULT 62
US-10-017-161-2375/c
; Sequence 2375, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 084335/0152
; CURRENT APPLICATION NUMBER: US/10/017,161
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: JP 2001/246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2430
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2375
; LENGTH: 1114
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: source
; LOCATION: (1)..(1114)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (201)..(914)
US-10-017-161-2375

Query Match 5.4%; Score 251; DB 15; Length 1114;
Best Local Similarity 63.7%; Pred. No. 1.7e-41;
Matches 397; Conservative 0; Mismatches 225; Indels 1; Gaps 1;

```
Qy        230 ATATATATATATATATATATATACACACACACACATATATATATACACATATATATAC 289
              ||| | |||||| |||||||  |  | | ||| ||||||||||| |||| |  | |||||||
Db        864 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 805

Qy        290 ACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATAT 349
              | | | ||| || |||| |  | |||| ||||| | ||| |  || |||| |  | |||
Db        804 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 745

Qy        350 ATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATAC 409
              ||| ||| |||||||||  |  | |||| ||||| | | |  |||||    |||||||
Db        744 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 685

Qy        410 ACACATATATACACATATATACACATATATACACATATACACATATATACACATATATAC 469
              | ||| |||||   ||||||    ||||| | ||| ||| |  ||||||    |||||||
Db        684 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 625

Qy        470 ATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATATAT 528
              ||| | | | || |||||  | || | ||||| | ||| |  ||||||     |||||
Db        624 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 565

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACATATAT 588
              | ||| |||||  ||||||  |  | ||| ||||||||||||  ||||||    |||||
Db        564 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 505

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
              ||||| |||||  ||||||  ||||||||  | | ||| || |||| |    |||||||
Db        504 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 445
```
```

```
Qy        649 ACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATAT 708
              | | |||||||    | | ||   |||| | ||| | | | || ||||||  | ||||||
Db        444 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 385

Qy        709 GTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAG 768
              || | |||||||        |  |||        ||  |   ||    |  | ||| |
Db        384 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 325

Qy        769 GAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAG 828
              | | | |  | | |    |||||    | |   | || |      |  | | |||
Db        324 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 265

Qy        829 TAGGGGAAATACATAAATGTGTA 851
                  | ||||||| |||| ||
Db        264 ATACACATATACATATATGTATA 242


RESULT 63
US-10-292-798-2017/c
; Sequence 2017, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
;  FILE REFERENCE: 084335/166
;  CURRENT APPLICATION NUMBER: US/10/292,798
;  CURRENT FILING DATE:  2002-11-13
;  PRIOR APPLICATION NUMBER: 10/017,161
;  PRIOR FILING DATE: 2001-12-18
;  PRIOR APPLICATION NUMBER: JP 2001-246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2070
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2017
;   LENGTH: 1114
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   LOCATION: source
;   FEATURE:
;   LOCATION: (1)..(1114)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(914)
US-10-292-798-2017

  Query Match             5.4%;  Score 251;  DB 17;  Length 1114;
  Best Local Similarity   63.7%;  Pred. No. 1.7e-41;
  Matches  397;  Conservative    0;  Mismatches  225;  Indels    1;  Gaps    1;

Qy        230 ATATATATATATATATATATATACACACACACACACATATATATATACACATATATATAC 289
              ||| | |||||| |||||||  |  | | ||| ||||||||| |||| |  | |||||||
Db        864 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 805

Qy        290 ACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATAT 349
              | | | ||| || |||| |  | |||| ||||| | ||| |  || |||| |  | |||
Db        804 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 745

Qy        350 ATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATAC 409
              ||| ||| |||||||||  |  | |||| |||| | | | |  ||||||    |||||||
Db        744 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 685

Qy        410 ACACATATATACACATATATACACATATATACACATATACACATATATACACATATATAC 469
              | ||| |||||   ||||||    ||||| | ||| ||| |  ||||||    |||||||
Db        684 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 625

Qy        470 ATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATATAT 528
              ||| | | | || ||||||  | || | ||||| | ||| |  ||||||     |||||
Db        624 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 565
```

```
Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACATATAT 588
              | ||| |||||   ||||||  |  | ||| ||||||||||||  ||||||    |||||
Db        564 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 505

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
              ||||| |||||   ||||||   |||||||| | | ||| || |||| |    |||||||
Db        504 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 445

Qy        649 ACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATAT 708
              | | |||||||    | | ||   |||| | ||| | | | || ||||||  | ||||||
Db        444 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 385

Qy        709 GTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAG 768
              || | |||||||        |  |||        ||  |   ||    |  | ||| |
Db        384 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 325

Qy        769 GAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAG 828
              | | | |  | | |    |||||    | |   | || |      |  | | |||
Db        324 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 265

Qy        829 TAGGGGAAATACATAAATGTGTA 851
                  | ||||||| |||| ||
Db        264 ATACACATATACATATATGTATA 242


RESULT 64
US-10-052-482-232
; Sequence 232, Application US/10052482
; Publication No. US20040072264A1
; GENERAL INFORMATION:
;  APPLICANT: Engelhard, Eric
;  APPLICANT:  Morris, David
;  TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
;  FILE REFERENCE: A-71087/RMS/DCF
;  CURRENT APPLICATION NUMBER: US/10/052,482
;  CURRENT FILING DATE:  2002-08-15
;  PRIOR APPLICATION NUMBER: US 09/747,377
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: US 09/798,586
;  PRIOR FILING DATE: 2001-03-02
;  NUMBER OF SEQ ID NOS: 241
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 232
;   LENGTH: 96595
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (49183)..(49621)
;   OTHER INFORMATION: "n" at positions 49183 to 49621 can be any base
US-10-052-482-232

  Query Match             5.4%;  Score 251;  DB 17;  Length 96595;
  Best Local Similarity   69.5%;  Pred. No. 1.2e-40;
  Matches  354;  Conservative    0;  Mismatches  152;  Indels    3;  Gaps    1;

Qy        229 TATATATATATATATATATATACACACACACACATATATATATACACATATATATA 288
              || ||| | |||| | |||||||||| ||| | |||||||| | | |||| ||| ||| |
Db      90982 TACATACACATATGTGTATATATACATACATATACACATATGTGTGTACATATACATACA 91041

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              || | | |||  | | |||| ||| ||||||| | ||| ||| ||   ||  |  |||||
Db      91042 CATACACATATGTGTGTACATATACATACACACACATACACATATGTGTACATGTACATA 91101

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              ||| ||| | |  |||    | || |||| | ||| ||| ||| ||||      | || ||
Db      91102 CATACACATACACATATGTGTACATGTACATACATACACATACATGTGTACATGTA 91161

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              || ||||| | | ||| ||||    | || |||| |||||||||||||| ||| || | ||
Db      91162 CATACATACACATACACATATGTGTACATGTACATATATACACATATACACATATGTGTA 91221
```

```
Qy        469 CATATATACACATATATATAATATACACACATATATATACACATATATACACACATATAT 528
              ||| || | | ||| |    |||||||| || | || |    | ||||| |||||||||
Db      91222 CATGTACATATATACAC---ATATACACATATGTGTACATGTACATATATACACATATAC 91278

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACATATAT 588
              ||| || | ||||  || ||||||||||||| ||| || | |||||| || | | | | ||
Db      91279 ACATATGTGTACATGTACATATACACATATACACATATGTGTACATGTACATATACACAT 91338

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
              ||||| ||||    | ||  ||||||| ||| | || |    | ||| |||||||| |
Db      91339 ATACACATATGTGTACATGCACATATACACATGTGTACATGCACATATACACATATGTGT 91398

Qy        649 ACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATAT 708
              ||||||| ||| ||||| ||    | || ||| |||||||| |||| |  ||||||| ||
Db      91399 ACATACACATATACACATATGTGTGTATATACATATACACACATATGTGTGTATATACAT 91458

Qy        709 GTATATATATATGCTCCAGAGTTCATAAG 737
              |||| ||||||  |      | |||| |
Db      91459 ATATACATATATACATATATATACATATG 91487


RESULT 65
US-10-311-455-1972/c
; Sequence 1972, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1972
;   LENGTH: 21537
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1972

  Query Match             5.4%;  Score 250;  DB 15;  Length 21537;
  Best Local Similarity   68.1%;  Pred. No. 1e-40;
  Matches  391;  Conservative    0;  Mismatches  180;  Indels    3;  Gaps    3;

Qy        181 CCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATA 240
              |||.   | || |    || |  |||||     |  | |       | ||||||||||||
Db       6005 CCAAATTAAAATACAATAACGCGATCTCGATTCACTACACTATATATATATATATATATA 5946

Qy        241 TATATATATATACACACACACACACATATATATATACACATATATATACACATATATACA 300
              |||||||||||| ||||| |||| |||||||||||| | |||||||||| | |||||||
Db       5945 TATATATATATATACACATACAC-CATATATATATATATATATATATACATACATATACA 5887

Qy        301 TATATACACATATATACACATATATATACACATACATATGTACACATATATATACACATA 360
              |   |   |   || | | | | ||  | | |    |   | |||| | | ||| ||
Db       5886 AAACAATTACTTACTAAATACACAAATCAATAAAATAAAAAATTCATAAACAAACATATC 5827

Qy        361 TGTATACACATATATACACATATATACACACATATATACACATATATACACACATATATA 420
              | | |  |||||||||| |  |||||  |       |||| | ||||||| | | |||||||
Db       5826 TCTCTTTACATATATATATCTATATCGATCACACTATATATATATATATATATATATA 5767

Qy        421 CACATATATACACATATATACACATATACACATATATACACATATATACATATATACACA 480
              |||| ||||| | ||||||| | ||| |||||||||| | ||||||| | | | | | |
Db       5766 CACACATATATATATATATATATACACACATATATATATATATATACACACATATA 5707

Qy        481 TATATATAATATACACACATATATATACACATATATACACACATATATACACATATATAC 540
              ||||||| | |||||||||||||||||| | ||||||| ||||| ||||| | |||||||
Db       5706 TATATAT-ACATACACACATATATATATATATATATACACACATATATATATATATAT 5648

Qy        541 ACATATATATACACATATATACACATATATACATATATACACATATATATACATATATAC 600
              | ||||||||| | ||||||| | ||||||| ||||||| | ||||| |  ||||||||
Db       5647 ATATATATATATATATATATATATATATATATATATATATATATATACACGCATATATAT 5588

Qy        601 ACATATATACATATATACACATATATACATATATACACACATATATACACATACATATAC 660
              || ||||||||||||||| | |   ||| |||||||| | |||||||||| | ||| |||||
Db       5587 ACGTATATACATATATATAAACTCATATATATATATATACATATATATATATATATATAT 5528

Qy        661 ACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATAT 720
              | | | ||| |   ||| |||||| |  | | | ||       || | | ||||  |  ||
Db       5527 ATATATATAAACTCACA-ATATATCCTTAAACAAATTATATAATCTCTATATACCTCAAT 5469

Qy        721 GCTCCAGAGTTCATAAGAGGTAGCAGTTGATTAC 754
              ||   | |  | || |   |  | |  | |||
Db       5468 TTTCTCAAATATAAAACAAAAATAATTATAATAC 5435


RESULT 66
US-09-997-722-196
; Sequence 196, Application US/09997722
; Publication No. US20040072154A1
; GENERAL INFORMATION:
;  APPLICANT: Morris, David
;  APPLICANT:  Engelhard, Eric
;  TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
;  FILE REFERENCE: A-71171/RMS/DCF
;  CURRENT APPLICATION NUMBER: US/09/997,722
;  CURRENT FILING DATE:  2001-11-30
;  PRIOR APPLICATION NUMBER: US 09/747,377
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: US 09/798,586
;  PRIOR FILING DATE: 2001-03-02
;  NUMBER OF SEQ ID NOS: 301
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 196
;   LENGTH: 96596
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-997-722-196

  Query Match             5.4%;  Score 249.6;  DB 11;  Length 96596;
  Best Local Similarity   70.5%;  Pred. No. 2.4e-40;
  Matches  347;  Conservative    0;  Mismatches  144;  Indels    1;  Gaps    1;

Qy        229 TATATATATATATATATATATATACACACACACACACATATATATATACACATATATATA 288
              | | || ||||||| || || || ||   | | |||  | || ||||| |||| || ||
Db      57110 TGTGTACATATATACATGTAGATGCATGTAGATACATGTGTACATATATACATGTAGATG 57169

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              ||  || |||||| ||||||  || |||||  || |||  |  ||||||  |  | ||||
Db      57170 CATGTAGATACATGTATACATGTAGATACATGTAGATACATGTATACATGTAGATACATG 57229

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              || |||||  |||  ||| | | || || | | || || | |||| |||||| |||||
Db      57230 TAGATACATGTATACATATATACATGTAGATACATGTAGATACATGTATACATATATACA 57289

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              | ||| ||  | | ||||| |||| || | |||| || ||||||||||  || |||
Db      57290 TGTAGATACATGTATACATATATACATGTAGATACATGTATACATATATACATGTAGATA 57349

Qy        469 CATATATACACATATATATAATATACACACATATATATACACATATATACACACATATAT 528
              ||| |||||| ||||| ||  || | |||  || |||  |   ||| ||| | |||| ||
Db      57350 CATGTATACATATATACAT-GTAGATACATGTAGATACATGTATACATATATACATGTAG 57408

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACATATAT 588
              | |||| || | |||| |||| | | ||| ||  | ||| ||  ||||| | ||||| ||
Db      57409 ATACATGTAGATACATGTATACATATATACATGTAGATACATGTATATAGATACATACAT 57468

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
```

```
Db      57469 AGATATATAGATACATAGATACATATATTTATATATGTACATATATCTACATATATATCT 57528

Qy        649 ACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATAT 708
Db      57529 ATATATACATATATCTACATATATATCTATATATATATATCTACATATATATCTATATAT 57588

Qy        709 GTATATATATAT 720
Db      57589 GTAGAAATATAT 57600
```

RESULT 67
US-10-437-963-61122
; Sequence 61122, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 61122
; LENGTH: 623
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_62586C.1
US-10-437-963-61122

Query Match 5.4%; Score 249; DB 18; Length 623;
Best Local Similarity 73.2%; Pred. No. 3.4e-41;
Matches 361; Conservative 0; Mismatches 125; Indels 7; Gaps 3;

```
Qy        227 TTTATATATATATATATATATATATACACACACACACATATATATATACAC-ATATAT 285
Db        122 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 181

Qy        286 ATACACATATATACATATATACACATATATACACATATATATACACATACATATGTACAC 345
Db        529 ATATATATATATATATATATATATATATATATATATATATATATATATATAGACACAC 470

Qy        346 ATATATATACACATATGTATACACATATATACACAT-----ATATACACACATATATACA 400
Db        469 ACATGTAGTTACATGTGTATATATATATGCATACATTTGTAGTGTAAAATTTTATCAACA 410

Qy        401 CATATATACACACATATATACACATATATACACATATATACACATATACACATATATACA 460
Db        409 GATGAATATTTTAAATTTGCGTGCCATATATATATGTTGATCACTGACCATACATTTCAT 350

Qy        461 CATATATACATATATACACATATATATA-ATATACACACATATATATACACATATATACA 519
Db        349 CTTATATATATATATATATATATATATATATATATATATATATATATATATATATATA 290

Qy        520 CACATATATACACATATATACACATATATATACACATATATACACATATATACATATA 579
Db        289 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 230

Qy        580 CACATATATATACATATATACACATATATACATATATACACATATATACATATATACACA 639
Db        229 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 170

Qy        640 CATATATACACATACATATACACACACATAGATATACATATATATACACATATATATACG 699
Db        169 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 110

Qy        700 TATATATATGTATATATATAT 720
Db        109 TATATATATATATATATATAT 89
```

```
Db        541 ATATATACCCTTTCTCGTGCCTACATACATATATATATATATATATATATATATAT 600

Qy        701 ATATATATGTATA 713
Db        601 ATATATATATATA 613
```

RESULT 68
US-10-437-963-61122/c
; Sequence 61122, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 61122
; LENGTH: 623
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_62586C.1
US-10-437-963-61122

Query Match 5.4%; Score 249; DB 18; Length 623;
Best Local Similarity 72.7%; Pred. No. 3.4e-41;
Matches 364; Conservative 0; Mismatches 130; Indels 7; Gaps 3;

```
Qy        227 TTTATATATATATATATATATATATACACACACACACATATATATATACAC-ATATAT 285
Db        589 TATATATATATATATATATATATGTATGTAGGCACGAGAAAGGGTATATATACACGATATAT 530

Qy        286 ATACACATATATACATATATACACATATATACACATATATATACACATACATATGTACAC 345
Db        529 ATATATATATATATATATATATATATATATATATATATATATATATATATAGACACAC 470

Qy        346 ATATATATACACATATGTATACACATATATACACAT-----ATATACACACATATATACA 400
Db        469 ACATGTAGTTACATGTGTATATATATATGCATACATTTGTAGTGTAAAATTTTATCAACA 410

Qy        401 CATATATACACACATATATACACATATATACACATATATACACATATACACATATATACA 460
Db        409 GATGAATATTTTAAATTTGCGTGCCATATATATATGTTGATCACTGACCATACATTTCAT 350

Qy        461 CATATATACATATATACACATATATATA-ATATACACACATATATATACACATATATACA 519
Db        349 CTTATATATATATATATATATATATATATATATATATATATATATATATATATATATA 290

Qy        520 CACATATATACACATATATACACATATATATACACATATATACACATATATACATATA 579
Db        289 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 230

Qy        580 CACATATATATACATATATACACATATATACATATATACACATATATACATATATACACA 639
Db        229 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 170

Qy        640 CATATATACACATACATATACACACACATAGATATACATATATATACACATATATATACG 699
Db        169 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 110

Qy        700 TATATATATGTATATATATAT 720
Db        109 TATATATATATATATATATAT 89
```

```
RESULT 69
US-10-741-600-17651/c
; Sequence 17651, Application US/10741600
; Publication No. US20050026169A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001499
;  CURRENT APPLICATION NUMBER: US/10/741,600
;  CURRENT FILING DATE:  2003-12-22
;  NUMBER OF SEQ ID NOS: 73997
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17651
;   LENGTH: 141121
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-741-600-17651

  Query Match             5.4%;  Score 248.8;  DB 19;  Length 141121;
  Best Local Similarity   67.1%;  Pred. No. 4.2e-40;
  Matches  365;  Conservative    2;  Mismatches  174;  Indels    3;  Gaps    1;

Qy        181 CCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATA 240
              |||    ||  || |  | | | ||     |||||  |  |      |  |||||||| |
Db      77519 CCAATGCACCCGACCTCAGTATAATTCTTAAAAAATATTATATATATACATATATATGTT 77460

Qy        241 TATATATATATACACACACACACATATATATATACACATATATATACACATATATACA 300
              ||||| ||||||||||||| |||   |||||| |:|| | | ||  |:  | ||||| |||
Db      77459 TATATGTATATACACACATACATGTATATATGTRTATATACATGCAYGTATATATACACA 77400

Qy        301 TATATACACATATATACACATATATATACACATACATATGTACACATATATATACACATA 360
              |||| |||  ||| | |  ||||| | | ||||  |||| |||||||||| | |   ||
Db      77399 TATACACATGTATGTGCGTATATACACATACATGTATATATACACATATACACATGTATG 77340

Qy        361 TGTATACACATATATACACATATATACACACATATATACACATATATACACACATATATA 420
              ||  || | ||| ||| ||| || ||    | | |||| ||||||||||||||  | | |
Db      77339 TGCGTATATATACATATACACATGTATGTGCGTGTATATACATATATACACATGTGTGTG 77280

Qy        421 CACATATATACACATATATACACATATACACATATATACACATATATACATATATACACA 480
              |  |||||||| ||||| |||  | |   | | |||| ||||||||||||||||||||
Db      77279 CGTGTATATACATATATACACATGTGTGTGCGTGTATATACATATATACATATATACACA 77220

Qy        481 TATATAT---AATATACACACATATATATACACATATATACACACATATATACACATATA 537
              |||| |    ||||| ||| ||||| ||| ||| | ||      |||||||| ||||
Db      77219 GATATGTGTGTATATATACATATATACATATACACACATGTGTGTATATATACATATATA 77160

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATACATATA 597
              | | | | || |    ||||||||||||||||||| | |||||  ||  | || |||||
Db      77159 CATATACACATGTGTGTATATATACACATATATACACACATACATGTACGTGTATATATA 77100

Qy        598 TACACATATATACATATATACACATATATACATATATACACACATATATACACATACATA 657
              | | ||| | |||  | |   | ||||| | ||||| |||||| | | || | |||||||
Db      77099 CATATATACACACACGTGTGTATATATACATATATACACACACGTGTGTATATATACATA 77040

Qy        658 TACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATA 717
              || ||||| ||| | | || | | | |  | | ||||||||  |||||||   |  | |
Db      77039 TATACACATATATACACACGTGTGTGTGTATACATATATACACATATATACACACGTGTG 76980

Qy        718 TATG 721
              | ||
Db      76979 TGTG 76976


RESULT 70
US-10-017-161-2365/c
; Sequence 2365, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
;  FILE REFERENCE: 084335/0152
;  CURRENT APPLICATION NUMBER: US/10/017,161
;  CURRENT FILING DATE:  2002-12-18
;  PRIOR APPLICATION NUMBER: JP 2001/246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2430
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2365
;  LENGTH: 1342
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
;  NAME/KEY: source
;  LOCATION: (1)..(1342)
;  FEATURE:
;  NAME/KEY: CDS
;  LOCATION: (201)..(1142)
;  FEATURE:
;  NAME/KEY: modified_base
;  LOCATION: (64)
;  OTHER INFORMATION: a, t, c, g, unknown or other
;  FEATURE:
;  NAME/KEY: modified_base
;  LOCATION: (66)
;  OTHER INFORMATION: a, t, c, g, unknown or other
;  FEATURE:
;  NAME/KEY: modified_base
;  LOCATION: (74)
;  OTHER INFORMATION: a, t, c, g, unknown or other
;  FEATURE:
;  NAME/KEY: modified_base
;  LOCATION: (413)..(434)
;  OTHER INFORMATION: a, t, c, g, unknown or other
;  FEATURE:
;  NAME/KEY: modified_base
;  LOCATION: (956)..(1055)
;  OTHER INFORMATION: a, t, c, g, unknown or other
US-10-017-161-2365

  Query Match             5.4%;  Score 248.4;  DB 15;  Length 1342;
  Best Local Similarity   69.9%;  Pred. No. 6.4e-41;
  Matches  348;  Conservative    0;  Mismatches  143;  Indels    7;  Gaps    2;

Qy        229 TATATATATATATATATATATACACACACACACATATATATATACACATATATATA 288
              || | | | |||| | |||||||||| | ||||||  | | | ||||| |  | ||||||
Db        891 TACACACACATATGTGTATATATACATATACACACCTACACACATATATATGTGTATATA 832

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              ||||||| ||| | ||| |  | |||| | | ||||| | |||||  | | |||||||
Db        831 TACATATACACACACATATATGTGTATATATACATATACATACATATGTGTATACACATG 772

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              |||||   || || | || || ||||| |||||||| | ||| |||| || |||||||||
Db        771 TATATGTGCACATATGTATACGTATATGCACATATACATACATATATGTATACATATATA 712

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              |||| |||| | ||  ||| || || ||| | |||||||  |||| | ||  |||| |||
Db        711 CACATATATGTGCATGTATGTATACGTATGTGCACATATGTACATGTGTATGCATAAATA 652

Qy        469 CATATATACACATATATATA-ATATACACACATATATATACA------CATATATACACA 521
              |||||||  |||| | |||| ||| | ||||||||| ||||       | ||| | |||
Db        651 CATATATGTACATGTGTATACATAAATACACATATATGTACATGCGTGTACATAAATACA 592

Qy        522 CATATATACACATATATACACATATATATACACATATATACACATATATACATATATACA 581
              |||||||  ||||   || ||||| ||| ||| |||| ||||  | ||||||||||||||
Db        591 CATATATGTACATGCGTATACATAAATACACATATATGTACATGTGTATACATATATACA 532

Qy        582 CATATATATACATATATACACATATATACATATATACACATATATACATATATACACACA 641
              ||||||| |||||   || |||||||||| |||| |||| | |  | ||| | |||||
Db        531 CATATATGTACATGCGTATACATATATACACATATGTACATGTGTGTACATAAATACACA 472
```

```
Qy        642 TATATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTA 701
              ||| |||| ||   || | | | ||| | || |||||
Db        471 TATGTACATATGTGTACATAAATACACATATGTACATNNNNNNNNNNNNNNNNNNNNNNNG 412

Qy        702 TATATATGTATATATATA 719
              || || | ||||||||||
Db        411 TACATGTATATATATATA 394


RESULT 71
US-10-292-798-2009/c
; Sequence 2009, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
;  FILE REFERENCE: 084335/166
;  CURRENT APPLICATION NUMBER: US/10/292,798
;  CURRENT FILING DATE:  2002-11-13
;  PRIOR APPLICATION NUMBER: 10/017,161
;  PRIOR FILING DATE: 2001-12-18
;  PRIOR APPLICATION NUMBER: JP 2001-246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2070
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2009
;   LENGTH: 1342
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   LOCATION: source
;   FEATURE:
;   LOCATION: (1)..(1342)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(1142)
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (64)..(64)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (66)..(66)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (74)..(74)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (413)..(434)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (956)..(1055)
;   OTHER INFORMATION: a, t, c, g, unknown or other
US-10-292-798-2009

  Query Match             5.4%;  Score 248.4;  DB 17;  Length 1342;
  Best Local Similarity   69.9%;  Pred. No. 6.4e-41;
  Matches  348;  Conservative    0;  Mismatches  143;  Indels    7;  Gaps    2;

Qy        229 TATATATATATATATATATATATACACACACACACATATATATATACACATATATATA 288
              || | | | |||| | |||||||||| | ||||||  | | | ||||| |  | ||||||
Db        891 TACACACACATATGTGTATATATACATATACACACCTACACACATATATATGTGTATATA 832

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              ||||||| ||| | ||| |  | |||| | | ||||| | |||||  | | |||||||
Db        831 TACATATACACACACATATATGTGTATATATACATATACATACATATGTGTATACACATG 772

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              |||||   || || | || || ||||| |||||||| | ||| |||| || |||||||||
Db        771 TATATGTGCACATATGTATACGTATATGCACATATACATACATATATGTATACATATATA 712

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              |||| |||| | ||  ||| || || ||| | |||||||  |||| | ||  |||| |||
Db        711 CACATATATGTGCATGTATGTATACGTATGTGCACATATGTACATGTGTATGCATAAATA 652

Qy        469 CATATATACACATATATATA-ATATACACACATATATATACA------CATATATACACA 521
              |||||||  |||| | |||| ||| | ||||||||||| ||||       | ||| | |||
Db        651 CATATATGTACATGTGTATACATAAATACACATATATGTACATGCGTGTACATAAATACA 592

Qy        522 CATATATACACATATATACACATATATATACACATATATACACATATATACATATATACA 581
              |||||||  ||||   || ||||| ||| ||| |||| ||||  | ||||||||||||||
Db        591 CATATATGTACATGCGTATACATAAATACACATATATGTACATGTGTATACATATATACA 532

Qy        582 CATATATATACATATATACACATATATACATATATACACATATATACATATATACACACA 641
              ||||||| |||||   || ||||||||||| |||| |||| | |  | ||| | |||||
Db        531 CATATATGTACATGCGTATACATATATACACATATGTACATGTGTGTACATAAATACACA 472

Qy        642 TATATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTA 701
              ||| |||| ||   || | | | ||| | || |||||
Db        471 TATGTACATATGTGTACATAAATACACATATGTACATNNNNNNNNNNNNNNNNNNNNNNG 412

Qy        702 TATATATGTATATATATA 719
              || || | ||||||||||
Db        411 TACATGTATATATATATA 394


RESULT 72
US-10-017-161-2127/c
; Sequence 2127, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
;  FILE REFERENCE: 084335/0152
;  CURRENT APPLICATION NUMBER: US/10/017,161
;  CURRENT FILING DATE:  2002-12-18
;  PRIOR APPLICATION NUMBER: JP 2001/246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2430
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2127
;   LENGTH: 1363
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: source
;   LOCATION: (1)..(1363)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(1163)
US-10-017-161-2127

  Query Match             5.4%;  Score 248.2;  DB 15;  Length 1363;
  Best Local Similarity   71.7%;  Pred. No. 7.1e-41;
  Matches  367;  Conservative    0;  Mismatches  139;  Indels    6;  Gaps    3;

Qy        227 TTTATATATATATATATATATATACACACACACACATATATATATACACATATATA 286
              | || | ||||||| |||||| || | | | || |||| |   |||||||   | |||||
Db       1136 TGTACACATATATACATATATGTACATATATACGCACATACGCATATATATGTACATATA 1077

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              |||||| |||  |||||||  |||||||||||||||  ||| |  |||||| ||||||
Db       1076 TACACACATACGCATATATGTACATATATACACATATGCATATATGTACATATATACACA 1017

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||  ||| |  ||  |||| ||| |||| || |||| |||| | ||| | || ||||||
```

```
Db       1016 TATGCATATATGTACATATATACACATATGCATATATGTACATATATACACACGCATATA 957

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              |  ||| ||||| ||||||||    || |||||     |||||||| ||| | |  | ||
Db        956 TGTACATATATACACACATAT----ACGTATATGTGTGTATACACACATACATATGTGTA 901

Qy        467 TACATATATACACATATATATAATATACACACATATATATACACATATATACACACATAT 526
              |||| | ||||  |||||  | || |||||| ||| ||||| || |||||| ||||||||
Db        900 TACACACATACGTATATACGT-ATGTACACATATACATATATACGTATATATACACATAT 842

Qy        527 ATACACATATATACACATATATATACACATATATACACATATATACATATATACACATAT 586
              | ||  ||||| |||||||| ||| |  || ||| ||  | ||| | |||||    ||
Db        841 ACACCTATATACACACATATACATATATGTACATATACGCACATATACATATATGTGTAC 782

Qy        587 ATATACATATATACACATATATACATATATACACATATATACA-TATATACACACATATA 645
              ||||| | | ||||| ||||  |||||||||||||| || |||| ||||||||||||| ||
Db        781 ATATATACACATACATATATGCACATATATACACACATGTACACTATATACACACATGTA 722

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              ||| || || ||| | ||||| |   | |||  || | ||||  | | ||| ||| | |
Db        721 CACATATGCACATATATACACACATGCACACACGTACACACACGCACACATATGTACACA 662

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAG 737
              ||||| | || | | | |    |  |||| |
Db        661 CATGTACACATGTGTACACATATGCCCATATG 630


RESULT 73
US-10-292-798-1773/c
; Sequence 1773, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
;  FILE REFERENCE: 084335/166
;  CURRENT APPLICATION NUMBER: US/10/292,798
;  CURRENT FILING DATE:  2002-11-13
;  PRIOR APPLICATION NUMBER: 10/017,161
;  PRIOR FILING DATE: 2001-12-18
;  PRIOR APPLICATION NUMBER: JP 2001-246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2070
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1773
;   LENGTH: 1363
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   LOCATION: source
;   FEATURE:
;   LOCATION: (1)..(1363)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(1163)
US-10-292-798-1773

  Query Match             5.4%;  Score 248.2;  DB 17;  Length 1363;
  Best Local Similarity   71.7%;  Pred. No. 7.1e-41;
  Matches  367;  Conservative    0;  Mismatches  139;  Indels    6;  Gaps    3;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | || | ||||||| |||||| || | | | || |||| |   |||||||   | |||||
Db       1136 TGTACACATATATACATATATGTACATATATACGCACATACGCATATATATGTACATATA 1077

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              |||||| |||  |||||||  ||||||||||||||||  |||  |  |||||||  ||||||
Db       1076 TACACACATACGCATATATGTACATATATACACATATGCATATATGTACATATATACACA 1017

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||  |||  |  ||  |||| ||| |||| || |||| |||| | ||| | || ||||||
```

```
Db       1016 TATGCATATATGTACATATATACACATATGCATATATGTACATATATACACACGCATATA 957

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              |  ||| ||||| ||||||||    || |||||     |||||||| ||| | |  | ||
Db        956 TGTACATATATACACACATAT----ACGTATATGTGTGTATACACACATACATATGTGTA 901

Qy        467 TACATATATACACATATATATAATATACACACATATATATACACATATATACACACATAT 526
              |||| | ||||  |||||  | || |||||| ||| ||||| || |||||| ||||||||
Db        900 TACACACATACGTATATACGT-ATGTACACATATACATATATACGTATATATACACACATAT 842

Qy        527 ATACACATATATACACATATATATACACATATATACACATATATACATATATACACATAT 586
              | ||  ||||| |||||||| ||| |  || ||| ||  | ||| | |||||    ||
Db        841 ACACCTATATACACACATATACATATATGTACATATACGCACATATACATATATGTGTAC 782

Qy        587 ATATACATATATACACATATATACATATATACACATATATACA-TATATACACACATATA 645
              ||||| | | ||||| ||||  |||||||||||||| || |||| ||||||||||||| ||
Db        781 ATATATACACATACATATATGCACATATATACACACATGTACACTATATACACACATGTA 722

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              ||| || || ||| | ||||| |   | |||  || | ||||  | | ||| ||| | |
Db        721 CACATATGCACATATATACACACATGCACACACGTACACACACGCACACATATGTACACA 662

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAG 737
              ||||| | || | | | |    |  |||| |
Db        661 CATGTACACATGTGTACACATATGCCCATATG 630


RESULT 74
US-10-437-963-57702
; Sequence 57702, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 57702
;   LENGTH: 614
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_59488C.1
US-10-437-963-57702

  Query Match             5.4%;  Score 247.8;  DB 18;  Length 614;
  Best Local Similarity   71.2%;  Pred. No. 6.1e-41;
  Matches  341;  Conservative    0;  Mismatches  137;  Indels    1;  Gaps    1;

Qy        226 TTTTATATATATATATATATATATATACACACACACACACATATATATATACACATATAT 285
              ||  | |||||||||||||||||||||| |·| | | | | ||||||||||| | ||||||
Db        136 TTGAACATATATATATATATATATATATATATATATATATATATATATATATATATATAT 195

Qy        286 ATACACATATATACATATATACACATATATACACATATATATACACATACATATGTACAC 345
              ||| | |||||||   |  ||  || ||     |      || | |  |||     | || |
Db        196 ATATATATATATAATTTGATTTACGTACTCCCGGGAGTACAGAAGCATTTTCCTATATAT 255

Qy        346 ATATATATACACATATGTATACACATATATACACATATATACACACATATATACACATAT 405
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | ||||
Db        256 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 315

Qy        406 ATACACACATATATACACATATATACACATATATACACATATACACATATATACACATAT 465
              ||| | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | ||||
Db        316 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 375
```

```
Qy        466 ATACATATATACACATATATATAATATACACACATATATATACACATATATACACACATA 525
              ||| ||||||| | |||||||| ||||| | | ||||||||| | ||||||| | | |||
Db        376 ATATATATATATATATATATAT-ATATATATATATATATATATATATATATATATATATA 434

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db        435 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 494

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| |||||||| |    |     |  |           ||
Db        495 TATATATATATATATATATATATATATATATACGGAGGGAGTATTTGACTGTTCGTCTTA 554

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATAT 704
              | || | || | |    | | ||  | |   |||| ||||   ||| |||||  ||||||
Db        555 TTCAAAAACTTTTTGTGAAAAATGTAAAACTATATGTATATGTATACATATAAGTATAT 613


RESULT 75
US-10-723-860-1125
; Sequence 1125, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
;  APPLICANT: Aziz, Natasha
;  APPLICANT:  Ginsburg, Wendy M.
;  APPLICANT:  Zlotnik, Albert
;  TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
;  TITLE OF INVENTION:  Methods for Screening for Soft Tissue Sarcoma Modulators
;  FILE REFERENCE: 05882.0193.NPUS01
;  CURRENT APPLICATION NUMBER: US/10/723,860
;  CURRENT FILING DATE:  2003-11-26
;  PRIOR APPLICATION NUMBER: 60/429,739
;  PRIOR FILING DATE: 2002-11-26
;  NUMBER OF SEQ ID NOS: 8393
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1125
;   LENGTH: 121600
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-723-860-1125

  Query Match             5.4%;  Score 247.4;  DB 18;  Length 121600;
  Best Local Similarity   68.6%;  Pred. No. 7.6e-40;
  Matches  341;  Conservative    0;  Mismatches  156;  Indels    0;  Gaps    0;

Qy        224 CATTTTATATATATATATATATATATATACACACACACACATATATATATACACATAT 283
              ||   |||||||||| ||||||| |||||| |    | | |   ||||||  |||| | || |
Db      43894 CAGCCTATATATATCTATATATCTATATATATCTATATATCTATATATCTATATATATCT 43953

Qy        284 ATATACACATATATACATATATACACATATATACACATATATATACACATACATATGTAC 343
              |||||   |||||||  ||||||   ||||||  | || | |||| |  || |||| ||
Db      43954 ATATATCTATATATATCTATATATCTATATATCTATATCTCTATATATCTATATATCTAT 44013

Qy        344 ACATATATATACACATATGTATACACATATATACACATATATACACACATATATACACAT 403
              |  | ||||||   || | ||||    |||||    ||||||  | | || |||| | ||
Db      44014 ATCTCTATATATCTATCTATATATCTCTATATCTCTATATATCTATATATCTATATATAT 44073

Qy        404 ATATACACACATATATACACATATATACACATATATACACATATACACATATATACACAT 463
              ||||  |   ||||||  | |||| || | |||| || | || || | || |||| |  |
Db      44074 CTATATATCTATATATCTATATATCTATATATATCTATATATCTATATATCTATATATCT 44133

Qy        464 ATATACATATATACACATATATATAATATACACACATATATATACACATATATACACACA 523
              |||||  ||||||  |  ||||||| || || | |  |||||||  | |||| || | |
Db      44134 ATATATCTATATATATCTATATATCATCTATATATCTATATATCTATATATCTATATATC 44193

Qy        524 TATATACACATATATACACATATATATACACATATATACACATATATACATATATACACA 583
              ||||||   ||||||  | |||||||| | ||||||  | |||| || ||||||  | |
Db      44194 TATATATCTATATATCTATATATATATCTATATATATCTATATATCTATATATATCTATA 44253

Qy        584 TATATATACATATATACACATATATACATATATACACATATATACATATATACACACATA 643
              ||| |||| || |||| | || |||| || |||| | || ||||||| |||| |   |||
Db      44254 TATCTATATATCTATATATATCTATATATCTATATATATCTATACATCTATATATCTATA 44313

Qy        644 TATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATA 703
              ||||   ||| || || | |   ||| |||| |||||| || | |||| ||||  | ||
Db      44314 TATATCTATATATCTATATATCTATATATATCTATATATCTATATATATCTATATATCTA 44373

Qy        704 TATATGTATATATATAT 720
              ||||| | ||||||| |
Db      44374 TATATATCTATATATCT 44390


RESULT 76
US-10-221-714A-313/c
; Sequence 313, Application US/10221714A
; Publication No. US20040048254A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with
;  TITLE OF INVENTION:  tumor suppressor genes and oncogenes
;  FILE REFERENCE: 5013.1005
;  CURRENT APPLICATION NUMBER: US/10/221,714A
;  CURRENT FILING DATE:  2003-01-21
;  PRIOR APPLICATION NUMBER: PCT/EP01/02955
;  PRIOR FILING DATE: 2001-03-15
;  PRIOR APPLICATION NUMBER: DE 10013847.0
;  PRIOR FILING DATE: 2000-03-15
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 540
; SEQ ID NO 313
;   LENGTH: 6286
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-714A-313

  Query Match             5.3%;  Score 245;  DB 17;  Length 6286;
  Best Local Similarity   57.3%;  Pred. No. 6.4e-40;
  Matches  443;  Conservative    0;  Mismatches  330;  Indels    0;  Gaps    0;

Qy        174 CTCCAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATAT 233
              | ||  |||| |      |   | || |  | | | | |||  || |   |  | |||
Db       4589 CACCCACCCATCCCCATTCATAAACATTTACTATAAAATAAATCAATAAATAAATAATAA 4530

Qy        234 ATATATATATATATATATACACACACACACACATATATATATACACATATATATACACAT 293
              ||  ||| ||| ||| ||| | | | | | ||||| ||| ||| | | ||| ||| | |
Db       4529 ATTAATAAATAAATAAATAAATAAATAAATACATAAATAAATAAATAAATAAATAAATAA 4470

Qy        294 ATATACATATATACACATATATACACATATATATACACATACATATGTACACATATATAT 353
              ||| | | ||| | | ||| ||||| | ||| ||| | | | | ||  | | ||| |||
Db       4469 ATAAATAAATAAATAAATAAATACATAAATAAATAAATAAATAAATAAATAAATAAATAA 4410

Qy        354 ACACATATGTATACACATATATACACATATATACACACATATATACACATATATACACAC 413
              | | |||  || | ||||| ||| | | ||| | | | | ||| | | ||| ||| | |
Db       4409 ATAAATAAATAAATACATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 4350

Qy        414 ATATATACACATATATACACATATATACACATATACACATATATACACATATATACATAT 473
              ||| ||||| | ||| | | ||| ||| | | ||| | | ||| | | ||| ||| |||
Db       4349 ATAAATACATAAATAAATAAATAAATAAATAAATAAATAAATACATAAATAAATAAATAA 4290

Qy        474 ATACACATATATATAATATACACACATATATATACACATATATACACACATATATACACA 533
              ||| | | ||| ||||    |  ||| ||| ||| | | ||| | | | | ||| | | |
Db       4289 ATAAATAAATAAATAAATAAACAACAAATAAATAAATAAATAAATAAATAAATAAATAAA 4230

Qy        534 TATATACACATATATATACACATATATACACATATATACATATATACACATATATATACA 593
              || ||| | | ||| ||| | | ||| | | ||| ||| ||| ||| | | ||| ||| |
```

```
Db       4229 TAAATAAATACATAAATAAATAAATAAATAAATACATAAATAAATAAATAAATAAATAAA 4170

Qy        594 TATATACACATATATACATATATACACATATATACATATATACACACATATATACACATA 653
              || ||| | | ||| | | ||| | | ||| ||| ||| ||| | | ||| ||| | | |
Db       4169 TAAATAAATAAATACATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAA 4110

Qy        654 CATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATA 713
              | ||| | ||| | | | ||| ||| ||| | | ||| ||| |   ||| ||| || | |
Db       4109 TAAATAAATACATAAATAAATAAATAAATAAATAAATAAATAAATAAATACATAAATAAA 4050

Qy        714 TATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAA 773
              || |||       | |  ||||       |       | | |    | | |   |||
Db       4049 TAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATACATAAATAAATAAA 3990

Qy        774 GAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGG 833
              | |      | |  | |  ||    | |   | |  | |           |||||
Db       3989 TAAATAAATAAATAAATAAATAAATAAATAAATACATAAATAAATAAATAAATACATAAA 3930

Qy        834 GAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATAT 893
              |||||  ||||||   |||  |||   ||   |||   | | || ||   |  | || ||
Db       3929 TAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAACAACAAATAAAT 3870

Qy        894 ATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATG 946
              | ||| |||   |||  | |  | |  ||||  |      ||  |  |||| |
Db       3869 AAATAAATAAATAAATAAATAAATAAATAAATAAATACATAAATAAATAAACG 3817
```

RESULT 77
US-10-437-963-66535/c
; Sequence 66535, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 66535
; LENGTH: 575
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_67477C.1
US-10-437-963-66535

Query Match 5.3%; Score 243.6; DB 18; Length 575;
Best Local Similarity 69.9%; Pred. No. 4.3e-40;
Matches 362; Conservative 0; Mismatches 144; Indels 12; Gaps 2;

```
Qy        215 AAAAAAAGGCATTTTATATATATATATATATATATACACACACACACACATATATATA 274
              |||  |  ||  | |||||||||||||||||||||||| | | | | | | |||||||||
Db        550 AAAGGATTGCTATATATATATATATATATATATATATATATATATATATATATATATA 491

Qy        275 TACACATATATATACACATATATACATATATACACATATATACACATATATATACACATA 334
              || | |||||||| | ||||||| ||||||| | ||||||| | | |    | |  || |
Db        490 TATATATATATATATATATATATATATATATATATATATATATATACAGGCAAATGCAAA 431

Qy        335 CATATGTACACATATATATACACATATGTATACACATATATACACATATATACACACATA 394
              ||    ||||     |   | |      |     | |     | |||  | || |||
Db        430 CAATATTACAAGATCAACAAGCCTACAAGACGATATTCCAGTTGGAAATAGTCCCAAATA 371

Qy        395 TATACACATATATACACACATATATACACATATATACACATATATACACATATACACATA 454
              |||   || ||     | | ||| | |  |||||| |   |      |||| |  ||
Db        370 AATATGGATGTAGGTGGAGAAATAAAAAGCTATATAAAGCAAGTGGAGCATAAAGGCAGC 311
```

```
Qy        455 TATAC----------ACATATATACATATATACACATATATATA-ATATACACACATAT 502
              |||||          | ||||||| ||||||| | ||||||||| ||||| | | ||||
Db        310 TATACAATGTAGTAGTAGATATATATATATATATATATATATATATATATATATATAT 251

Qy        503 ATATACACATATATACACACATATATACACATATATACACATATATATACACATATATAC 562
              ||||| | ||||||| | | ||||||| | ||||||| | ||||||||| | |||||||
Db        250 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 191

Qy        563 ACATATATACATATATACACATATATATACATATATACACATATATACATATATACACAT 622
              | ||||||| ||||||| | ||||||||| ||||||| | ||||||| ||||||| | ||
Db        190 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 131

Qy        623 ATATACATATATACACACATATATACACATACATATACACACACATAGATATACATATAT 682
              ||||| ||||||| | | ||||||| | ||| |||||||||||||| | ||||| ||||||
Db        130 ATATATATATATATATATATATATATATATATATATACACACACACATATATATATAT 71

Qy        683 ATACACATATATATACGTATATATATGTATATATATAT 720
              ||| | |||||||||  |||||||||| ||||||||||
Db         70 ATATATATATATATATATATATATATATATATATATAT 33
```

RESULT 78
US-10-017-161-2385/c
; Sequence 2385, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 084335/0152
; CURRENT APPLICATION NUMBER: US/10/017,161
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: JP 2001/246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2430
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2385
; LENGTH: 1309
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: source
; LOCATION: (1)..(1309)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (201)..(1109)
US-10-017-161-2385

Query Match 5.3%; Score 243.2; DB 15; Length 1309;
Best Local Similarity 70.7%; Pred. No. 7.5e-40;
Matches 355; Conservative 0; Mismatches 138; Indels 9; Gaps 2;

```
Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | || |||| || || ||| ||||||| || |     || |||| ||||||| ||||  ||
Db        553 TATACATATGTACATCTATGTATATACGCATATGTGCATATATGTATATACGCATACGTA 494

Qy        287 TACACATATATACATATATACACATATATACACATATATATACAC------ATACATATG 340
              ||    |||||||  ||||| |||||| | || |||| |||| ||       |||||||||
Db        493 TATGTGTATATACGCATATATACATATGTGCATATATGTATATACGCATATATACATATG 434

Qy        341 TACACATATATATACACATATGTATACACATATATACACATATATACACACATATATACA 400
              | || |||| |||| ||  || |||||| || |   | | || |||| || ||||||||||
Db        433 TGCATATATGTATATACGCATATATACATATGTGCATATATGTATATACGCATATATACA 374

Qy        401 CATATATACACACATATATACACATATATACACATATATACACATATACACATATATACA 460
              || |  | | |  |||||||| |||||||||| | || | || || | || ||||  | |
Db        373 TATGTGCATATATGTATATACGCATATATACATACATGTGCATATGTGCATATATTCATA 314

Qy        461 CATATATACATATATACACATATATATAATATACACACATATATATACACATATATACAC 520
              || ||  || ||||   | ||||   || |||| ||| |||| |||| |||| ||||
```

```
Db        313 AGTACATCTATGTATATGTACATAT---ATGTACATACACATATGTACATATATGTACAT 257

Qy        521 ACATATATACACATATATACACATATATATACACATATATACACATATATACATATATAC 580
              ||| ||||  |||||||  ||||| | ||| ||||||||  |||||||  | | |||||
Db        256 ACACATATGTACATATATGTACATACACATATACATATATGTACATATACATACATATAT 197

Qy        581 ACATATATATACATATATACACATATATACATATATACACATATATACATATATACACAC 640
              ||| ||| || ||||  ||| |||  || ||||| | | || |||| ||| | | | ||
Db        196 GCATGTATGTATATATGCACATATACCTATATATACATATATGTATATATACACATATAC 137

Qy        641 ATATATACACATACATATACACACACATAGATATACATATATATACACATATATATACGT 700
              ||||||  |||||  | || | ||||| | ||  | |||| ||||| |||||||  |  |
Db        136 ATATATGTACATATGTGTATATACACACATATGCATATATGTATACGCATATATGCATAT 77

Qy        701 ATATATATGTATATATATATGC 722
              || |||||  | | ||||||||
Db         76 ATGTATATACACACATATATGC 55


RESULT 79
US-10-292-798-2027/c
; Sequence 2027, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
;  FILE REFERENCE: 084335/166
;  CURRENT APPLICATION NUMBER: US/10/292,798
;  CURRENT FILING DATE:  2002-11-13
;  PRIOR APPLICATION NUMBER: 10/017,161
;  PRIOR FILING DATE: 2001-12-18
;  PRIOR APPLICATION NUMBER: JP 2001-246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2070
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2027
;   LENGTH: 1309
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   LOCATION: source
;   FEATURE:
;   LOCATION: (1)..(1309)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(1109)
US-10-292-798-2027

  Query Match             5.3%;  Score 243.2;  DB 17;  Length 1309;
  Best Local Similarity   70.7%;  Pred. No. 7.5e-40;
  Matches  355;  Conservative    0;  Mismatches  138;  Indels    9;  Gaps    2;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | || |||| || || ||| ||||||| || |    || |||| ||||||| ||||  ||
Db        553 TATACATATGTACATCTATGTATATACGCATATGTGCATATATGTATATACGCATACGTA 494

Qy        287 TACACATATATACATATATACACATATATACACATATATATACAC------ATACATATG 340
              ||   |||||||  ||||| ||||||  | || |||| |||| ||      |||||||||
Db        493 TATGTGTATATACGCATATATACATATGTGCATATATGTATATACGCATATATACATATG 434

Qy        341 TACACATATATATACACATATGTATACACATATATACACATATATACACACATATATACA 400
              | || |||| |||| ||   || |||||| || |  | | || |||| || ||||||||||
Db        433 TGCATATATGTATATACGCATATATACATATGTGCATATATGTATATACGCATATATACA 374

Qy        401 CATATATACACACATATATACACATATATACACATATATACACATATACACATATATACA 460
              ||  |  | | |  ||||||| ||||||||||  | || | || || | || ||||  | |
Db        373 TATGTGCATATATGTATATACGCATATATACATACATGTGCATATGTGCATATATTCATA 314

Qy        461 CATATATACATATATACACATATATATAATATACACACATATATATACACATATATACAC 520
                || ||  || ||||   | ||||   || |||| ||| |||| |||| |||| ||||
```

```
Db        313 AGTACATCTATGTATATGTACATAT---ATGTACATACACATATGTACATATATGTACAT 257

Qy        521 ACATATATACACATATATACACATATATATACACATATATACACATATATACATATATAC 580
              ||| ||||  |||||||  ||||| | ||| ||||||||  |||||||  | | |||||
Db        256 ACACATATGTACATATATGTACATACACATATACATATATGTACATATACATACATATAT 197

Qy        581 ACATATATATACATATATACACATATATACATATATACACATATATACATATATACACAC 640
              ||| ||| || ||||  ||| |||  || ||||| | | || |||| ||| | | | ||
Db        196 GCATGTATGTATATATGCACATATACCTATATATACATATATGTATATATACACATATAC 137

Qy        641 ATATATACACATACATATACACACACATAGATATACATATATATACACATATATATACGT 700
              ||||||  |||||  | || | ||||| | ||  | |||| ||||| |||||||  |  |
Db        136 ATATATGTACATATGTGTATATACACACATATGCATATATGTATACGCATATATGCATAT 77

Qy        701 ATATATATGTATATATATATGC 722
              || |||||  | | ||||||||
Db         76 ATGTATATACACACATATATGC 55


RESULT 80
US-10-437-963-18717/c
; Sequence 18717, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 18717
;   LENGTH: 566
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_24249C.1
US-10-437-963-18717

  Query Match             5.2%;  Score 242.6;  DB 18;  Length 566;
  Best Local Similarity   68.5%;  Pred. No. 6.9e-40;
  Matches  354;  Conservative    0;  Mismatches  154;  Indels    9;  Gaps    1;

Qy        228 TTATATATATATATATATATATATACACACACACACACATATATATATACACATATATAT 287
              |||| | | | ||| ||| ||  ||| |   ||  |  ||||||| |||||| | | |
Db        555 TTATGTGTGTGTATGTATGTACGTACGTATGTACGTATGTATATATGTACACACACACAC 496

Qy        288 ACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACAT 347
              ||||||||||| ||||||| | ||||||| | |||||||||| |  || |||| || | ||
Db        495 ACACATATATATATATATATATATATATATATATATATATAAAAGTATATATATATATAT 436

Qy        348 ATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATAT 407
              ||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | ||||||
Db        435 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 376

Qy        408 ACACACATATATACACATATATACACATATATACACATATACACATATATACACATATAT 467
              | | | ||         ||   |  |||||  |  ||          |   |  |||
Db        375 ATATATATTATGTGTTATGAGTTGCCATATGAATGCAGGGCATGGCGAAACTAGCTATGC 316

Qy        468 ACATATATACAC---------ATATATATAATATACACACATATATATACACATATATAC 518
              |  |  | |||         || | | |       | | ||||||||  | |||||||
Db        315 TCGCAACTCCACAGTTCAGGTATGTGTGTGTGTTATATATATATATATATATATATAT 256

Qy        519 ACACATATATACACATATATACACATATATATACACATATATACACATATATACATATAT 578
              | | |||| || | ||||||||||||||||||| | | ||||||| |||||||| ||||||
Db        255 ATATATATGTATATATATATACACATATATACATATATATATACATATATATATAT 196
```

```
Qy        579 ACACATATATATACATATATACACATATATACATATATACACATATATACATATATACAC 638
              ||| ||||||||||||||||| | ||| ||| ||||||| | | | ||| ||||||| |
Db        195 ACATATATATATACATATATATATATATACATATATATATATATACACATATATATATAT 136

Qy        639 ACATATATACACATACATATACACACACATAGATATACATATATATACACATATATATAC 698
              | | | ||| | ||| ||||| | ||||||| ||||| ||| ||||| | |||||||||
Db        135 ATACACATATATATATATATATATACACATATATATATATACATATATATATATATAT 76

Qy        699 GTATATATATGTATATATATATGCTCCAGAGTTCATA 735
              |||||||||| ||| | ||  |  |||    || |||
Db         75 GTATATATATATATGTGTAATTCGTCCGATATTAATA 39


RESULT 81
US-10-437-963-57702/c
; Sequence 57702, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 57702
;   LENGTH: 614
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_59488C.1
US-10-437-963-57702

  Query Match             5.2%;  Score 242.2;  DB 18;  Length 614;
  Best Local Similarity   70.7%;  Pred. No. 8.7e-40;
  Matches  336;  Conservative    0;  Mismatches  138;  Indels    1;  Gaps    1;

Qy        208 CAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATATACACACACACACACAT 267
              |||  ||| |      |  | |||||||||||||||||||||||| | | | | | | ||
Db        545 CAGTCAAATACTCCCTCCGTATATATATATATATATATATATATATATATATATATATATAT 486

Qy        268 ATATATATACACATATATATACACATATATACATATATACACATATATACACATATATAT 327
              |||||||||  | |||||||||  | |||||||  |||||||  | |||||||  | ||||||||
Db        485 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 426

Qy        328 ACACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATAC 387
              | | ||| |||| || | |||||||||  | ||||  |||| | |||||||  | |||||||
Db        425 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 366

Qy        388 ACACATATATACACATATATACACACATATATACACATATATACACATATATACACATAT 447
              | | |||||||  | |||||||  | |  |||||||  | |||||||  | |||||||  | ||||
Db        365 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 306

Qy        448 ACACATATATACACATATATACATATATACACATATATATA-ATATACACACATATATAT 506
              | | |||||||  | |||||||  |||||||  | |||||||||  |||||  | |  |||||
Db        305 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAGGA 246

Qy        507 ACACATATATACACACATATATACACATATATACACATATATATACACATATATACACAT 566
              | |    | |  || |          || || ||  |  ||||||||  | |||||||  | ||
Db        245 AAATGCTTCTGTACTCCCGGGAGTACGTAAATCAAATTATATATATATATATATATATAT 186

Qy        567 ATATACATATATACACATATATATACATATATACACATATATACATATATACACATATAT 626
              |||||  |||||||  | |||||||||  |||||||  |||||||  |  | |||  |
Db        185 ATATATATATATATATATATATATATATATATATATATATATATATGTTCAAACAAACTCCA 126

Qy        627 ACATATATACACACATATATACACATACATATACACACACATAGATATACATATA 681
              |  ||     | ||  || ||||| |||||   | |   | ||  | |||||
Db        125 CAGTGAATTTCATCTTAACTAGACATATATATAACAATAATAAAATTAAAATATA 71


RESULT 82
US-10-067-514-1
; Sequence 1, Application US/10067514
; Publication No. US20030054531A1
; GENERAL INFORMATION:
;  APPLICANT: Gretarsdottir, Solveig
;  APPLICANT:  Jonsdottir, Sif
;  APPLICANT:  Reynisdottir, Sigridur Th.
;  TITLE OF INVENTION: HUMAN STROKE GENE
;  FILE REFERENCE: 2345.2010-003
;  CURRENT APPLICATION NUMBER: US/10/067,514
;  CURRENT FILING DATE:  2002-02-04
;  PRIOR APPLICATION NUMBER: US 09/811/352
;  PRIOR FILING DATE: 2001-03-19
;  NUMBER OF SEQ ID NOS: 84
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 1691139
;   TYPE: DNA
;   ORGANISM: Human
US-10-067-514-1

  Query Match             5.2%;  Score 242;  DB 14;  Length 1691139;
  Best Local Similarity   70.0%;  Pred. No. 3.2e-38;
  Matches  368;  Conservative    1;  Mismatches  151;  Indels    6;  Gaps    3;

Qy        196 GAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACAC 255
              ||  ||| |    | | | |||| | |    | |||||||| ||||||||  |||||
Db     867636 GAGTTGTGAATATATATATAAAATATGTGTGTGTATATATACATATATATGCATATATGT 867695

Qy        256 ACACACACACATATATATATACACATATATATACACATATATACATATATACACATATAT 315
              ||  | | |||||||| |||| | |  |||||  || ||||||||||||||||||||||||
Db     867696 ACCTATATACATATAT-TATATATACCTATATGAACCTATATACATATATACACATATAT 867754

Qy        316 ACACATATATATACACATACATATGTACACATATAT----ATACACATATGTATACACAT 371
              || |||||| ||| ||||| || :  || |||||     ||| ||| || | |  ||
Db     867755 GAACCTATATACACATATACACATAYGTACCTATATACACATATACACATATGTGTACCC 867814

Qy        372 ATATACACATATATACACACATATATACACATATATACACACATATATACACATATATAC 431
              |||||||||||||  ||| |  ||  || | | | ||| ||||||||  |  ||  ||||||
Db     867815 ATATACACATATACACATATGTACCTATATACACATATACACATATGTGTACCCATATAC 867874

Qy        432 ACATATATACACATATACACATATATACACATATATACATATATACACATATATATAATA 491
              |||||||  ||| || |  ||  ||||||||||||| ||||||  |  ||  |||||  | |
Db     867875 ACATATACACATATGTGTACCCATATACACATATACACATATGTGTACCCATATA-CACA 867933

Qy        492 TACACACATATATATACACATATATACACACATATATACACATATATACACATATATATA 551
              || |||||||| | ||| |||||| ||| | | | |||    ||   | | | | |||||
Db     867934 TATACACATATGTGTACCCATATACACATATACACATATGTGTACCCATATACACATATA 867993

Qy        552 CACATATATACACATATATACATATATACACATATATATACATATATACACATATATACA 611
              |||||||  |  ||  |||||||  |||||||||||| | |||  ||||||||||||| |||
Db     867994 CACATATGTGTACCCATATACACATATACACATATGTGTACCCATATACACATATACACA 868053

Qy        612 TATATACACATATATACATATATACACACATATATACACATACATATACACACACATAGA 671
              | | ||| |||||| |||||||  |||     ||   | | | ||||||||||    ||
Db     868054 TGTGTACCCATATACACATATACACATGTGTACCCATATACACATATACACATGTGTACC 868113

Qy        672 TATACATATATATACACATATATATACGTATATATATGTATATATA 717
              ||| | | |||||  |||||  | |||  |||||    ||| | ||
Db     868114 CATATACACATATATGCATATGTGTACCCATATACGCATATGTGTA 868159


RESULT 83
US-10-419-723-1
; Sequence 1, Application US/10419723
; Publication No. US20040014099A1
; GENERAL INFORMATION:
```

```
;  APPLICANT: Gretarsdottir, Solveig
;  APPLICANT:  Jonsdottir, Sif
;  APPLICANT:  Reynisdottir, Sigridur Th.
;  APPLICANT:  Thorleifsson, Gudmar
;  APPLICANT:  Gulcher, Jeffrey
;  TITLE OF INVENTION: SUSCEPTIBILITY GENE FOR HUMAN STROKE;
;  TITLE OF INVENTION:  METHODS OF TREATMENT
;  FILE REFERENCE: 2345.2010-005
;  CURRENT APPLICATION NUMBER: US/10/419,723
;  CURRENT FILING DATE:  2003-04-18
;  PRIOR APPLICATION NUMBER: 10/255,120
;  PRIOR FILING DATE: 2002-09-25
;  PRIOR APPLICATION NUMBER: 10/067,514
;  PRIOR FILING DATE: 2002-02-04
;  PRIOR APPLICATION NUMBER: 09/811,352
;  PRIOR FILING DATE: 2001-03-19
;  NUMBER OF SEQ ID NOS: 87
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 1691139
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-419-723-1

  Query Match             5.2%;  Score 242;  DB 17;  Length 1691139;
  Best Local Similarity   70.0%;  Pred. No. 3.2e-38;
  Matches  368;  Conservative    1;  Mismatches  151;  Indels    6;  Gaps    3;

Qy        196 GAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACAC 255
              ||  ||| |    | | | |||| | |    | |||||||| ||||||||  |||||
Db     867636 GAGTTGTGAATATATATATAAAATATGTGTGTGTATATATACATATATATGCATATATGT 867695

Qy        256 ACACACACACATATATATATACACATATATATACACATATATACATATATACACATATAT 315
              ||  | | |||||||| |||| | |  |||||  || ||||||||||||||||||||||
Db     867696 ACCTATATACATATAT-TATATATACCTATATGAACCTATATACATATATACACATATAT 867754

Qy        316 ACACATATATATACACATACATATGTACACATATAT----ATACACATATGTATACACAT 371
              || |||||| ||| ||||| || :  || |||||    ||| ||| || | |  ||
Db     867755 GAACCTATATACACATATACACATAYGTACCTATATACACATATACACATATGTGTACCC 867814

Qy        372 ATATACACATATATACACACATATATACACATATATACACACATATATACACATATATAC 431
              |||||||||||||| ||| |  ||  || | | | ||| |||||||| |  ||  ||||||
Db     867815 ATATACACATATACACATATGTACCTATATACACATATACACATATGTGTACCCATATAC 867874

Qy        432 ACATATATACACATATACACATATATACACATATATACATATATACACATATATATAATA 491
              ||||||| ||| || |  ||  |||||||||||||| |||||| |  ||  |||||  | |
Db     867875 ACATATACACATATGTGTACCCATATACACATATACACATATGTGTACCCATATA-CACA 867933

Qy        492 TACACACATATATATACACATATATACACACATATATACACATATATACACATATATATA 551
              || |||||||| | ||| |||||| ||| | | | |||    ||   | | | | |||||
Db     867934 TATACACATATGTGTACCCATATACACATATACACATATGTGTACCCATATACACATATA 867993

Qy        552 CACATATATACACATATATACATATATACACATATATATACATATATACACATATATACA 611
              |||||||  |   ||  ||||||| |||||||||||| | |||  ||||||||||||| |||
Db     867994 CACATATGTGTACCCATATACACATATACACATATGTGTACCCATATACACATATACACA 868053

Qy        612 TATATACACATATATACATATATACACACATATATACACATACATATACACACACATAGA 671
              | | ||| ||||||| ||||||| |||    ||    | | | ||||||||||    ||
Db     868054 TGTGTACCCATATACACATATACACATGTGTACCCATATACACATATACACATGTGTACC 868113

Qy        672 TATACATATATATACACATATATATACGTATATATATGTATATATA 717
              ||| | | |||||  ||||| | |||  |||||    ||| | ||
Db     868114 CATATACACATATATGCATATGTGTACCCATATACGCATATGTGTA 868159


RESULT 84
US-10-017-161-2211/c
; Sequence 2211, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
;  FILE REFERENCE: 084335/0152
;  CURRENT APPLICATION NUMBER: US/10/017,161
;  CURRENT FILING DATE:  2002-12-18
;  PRIOR APPLICATION NUMBER: JP 2001/246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2430
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2211
;   LENGTH: 1441
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: source
;   LOCATION: (1)..(1441)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(1241)
US-10-017-161-2211

  Query Match             5.2%;  Score 241.2;  DB 15;  Length 1441;
  Best Local Similarity   69.5%;  Pred. No. 2e-39;
  Matches  357;  Conservative    0;  Mismatches  153;  Indels    4;  Gaps    2;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | || |||||||||| ||| | |||| |  | | | |||||||||| |||||||||| |||
Db        702 TATACATATATATAGATACACATATGCGTATATATACACATATATGTATACACATACATA 643

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              | |  ||||| ||||||||  | | | ||||| || | |||||| |||  |||  |||
Db        642 TGCGTATATACACATATATGTATACACATACATATGTGTATACATATATGTATACACATG 583

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||| |  | |||| |||  |||||||||   ||||| ||| |    |||| |||||||
Db        582 CATATGTGTATATATTTATGAACATATATATGTATATACACATATGCGTATATACATATA 523

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              |  | ||| ||||  | |||| ||   | | ||| |  |||||||||| |||    ||||
Db        522 TGTATACACATATGTATATATTTATGTACACATATATGTATACACATACATATGTGTATA 463

Qy        467 TACATATATACACATATATATAATATACACACATATATATACACATATATACACACATAT 526
              ||  ||| |||||||||||    |||||||| | |||| |  | |||| |||||  || |
Db        462 TATTTATGTACACATATAT---GTATACACATACATATGTGTATATATTTACACATATGT 406

Qy        527 ATACACATATATACACATATATATACACATATATACACATATATACATATATACACATAT 586
              ||||||||| ||| |  | ||||||  | | | ||| || |||||| ||| | |  | |
Db        405 GTACACATACATATATGTGTATATATTTACACACACATATGTATACACATACATATGTGT 346

Qy        587 ATATACATATATACACATATATACATATATACACATATATACATA-TATACACACATATA 645
              |||||  |||  |||||||| || | | ||||  || | ||  ||| |  |||| || ||
Db        345 ATATATTTATGCACACATATGTATACACATACGTATGTGTATATATTGCACACATATGTA 286

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              ||||||||||||    | | ||  ||  ||| |||| || ||| ||| ||| || ||||
Db        285 TACACATACATATGTGTATATATTTATGCACACATATGTATACACATACATATGTGTATA 226

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAG 739
              | | |||||| | |||        | |||| | |
Db        225 TGTTTATATACACATGTATACACATGCATATGTG 192


RESULT 85
US-10-292-798-1857/c
; Sequence 1857, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
;  FILE REFERENCE: 084335/166
```

```
;  CURRENT APPLICATION NUMBER: US/10/292,798
;  CURRENT FILING DATE:  2002-11-13
;  PRIOR APPLICATION NUMBER: 10/017,161
;  PRIOR FILING DATE: 2001-12-18
;  PRIOR APPLICATION NUMBER: JP 2001-246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2070
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1857
;   LENGTH: 1441
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   LOCATION: source
;   FEATURE:
;   LOCATION: (1)..(1441)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(1241)
US-10-292-798-1857

  Query Match             5.2%;  Score 241.2;  DB 17;  Length 1441;
  Best Local Similarity   69.5%;  Pred. No. 2e-39;
  Matches  357;  Conservative    0;  Mismatches  153;  Indels    4;  Gaps    2;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATATAC 286
              | || ||||||||| ||| | |||| |  | | | |||||||||| |||||||||| |||
Db        702 TATACATATATATAGATACACATATGCGTATATATACACATATATGTATACACATACATA 643

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              | |  ||||| ||||||||  | | | ||||| || | |||||| ||| |||  |||
Db        642 TGCGTATATACACATATATGTATACACATACATATGTGTATACATATATGTATACACATG 583

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              ||||  |  | ||||  |||  |||||||||   ||||| ||| |    |||| |||||||
Db        582 CATATGTGTATATATTTATGAACATATATATGTATATACACATATGCGTATATACATATA 523

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              |  | |||  ||||  | |||| ||   | | ||| |  |||||||||| |||    ||||
Db        522 TGTATACACATATGTATATATTTATGTACACATATATGTATACACATACATATGTGTATA 463

Qy        467 TACATATATACACATATATATAATATACACACATATATATACACATATATACACACATAT 526
              ||  ||| |||||||||||    |||||||| | |||| |  | |||| |||||| || |
Db        462 TATTTATGTACACATATAT---GTATACACATACATATGTGTATATATTTACACATATGT 406

Qy        527 ATACACATATATACACATATATATACACATATATACACATATATACATATATACACATAT 586
              ||||||||  ||| |  | ||||||   | | | ||| || |||||| ||| | |  | |
Db        405 GTACACATACATATATGTGTATATATTTACACACACATATGTATACACATACATATGTGT 346

Qy        587 ATATACATATATACACATATATACATATATACACATATATACATA-TATACACACATATA 645
              |||||  |||  |||||||| || | | ||||  || | ||  ||| |  ||||| || ||
Db        345 ATATATTTATGCACACATATGTATACACATACGTATGTGTATATATTGCACACATATGTA 286

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              ||||||||||||     | | ||   ||  ||| |||| || ||| ||| ||| || ||||
Db        285 TACACATACATATGTGTATATATTTATGCACACATATGTATACACATACATATGTGTATA 226

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAG 739
              | | |||||| | |||       | |||| | |
Db        225 TGTTTATATACACATGTATACACATGCATATGTG 192


RESULT 86
US-10-017-161-2387
; Sequence 2387, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
;  FILE REFERENCE: 084335/0152
;  CURRENT APPLICATION NUMBER: US/10/017,161
;  CURRENT FILING DATE:  2002-12-18
;  PRIOR APPLICATION NUMBER: JP 2001/246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2430
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2387
;   LENGTH: 1276
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: source
;   LOCATION: (1)..(1276)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(1076)
US-10-017-161-2387

  Query Match             5.2%;  Score 240.8;  DB 15;  Length 1276;
  Best Local Similarity   69.9%;  Pred. No. 2.3e-39;
  Matches  341;  Conservative    0;  Mismatches  142;  Indels    5;  Gaps    1;

Qy        230 ATATATATATATATATATATATACACACACACACACATATATATATACACATATATATAC 289
              | ||||| |||| || |||||||||| |   ||| | |||  |||||||| || || ||
Db        793 ACATATACATATGTACATATATACATATGTACATATATACGTATATACATATGTACATCT 852

Qy        290 ACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATAT 349
              |  |||||||  ||||   |||||||  | |||   ||||  |||  | | || ||  ||
Db        853 ATGTATATACGCATATGTGCATATATGTATATACGCATACGTATATGTGTATATACGCAT 912

Qy        350 ATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATAC 409
              |||||||  ||  || |||  |  |||||||  |||||||||| |   |  |||  |  |||||||
Db        913 ATATACATATGTGCATATATGTATATACGCATATATACATATGTGCATATATGTATATAC 972

Qy        410 ACACATATATACACATATATACACATATATACACATATACACATATATACACATATATAC 469
              || ||||| | |   |  ||| |  ||||||| |||||| |||||| | || |||| ||
Db        973 GCATATATACATATGTGCATATATGTATATACGCATATATACATATGTGCATATATGTAT 1032

Qy        470 ATATATACACATATATATAATATACACACATATATATACACATATATACACACATATATA 529
              |||   | | |||       ||| |    |||||  |   | | ||  |||    |||| |||
Db       1033 ATACGCATATATA-----CATACATGTGCATATGTGCATATATTCATAAGTACATCTATG 1087

Qy        530 CACATATATACACATATATATACACATATATACACATATATACATATATACACATATATA 589
              | || || | | || || |||||||||| |||| |||| |||||  | | |   || |||
Db       1088 TATATGTACATATATGTACATACACATATGTACATATATGTACATACACATATGTACATA 1147

Qy        590 TACATATATACACATATATACATATATACACATATATACATATATACACACATATATACA 649
              ||  || |||||||||||| | |||| |||| ||| ||||||||||  ||   || |||| |
Db       1148 TATGTACATACACATATACATATATGTACATATACATACATATATGCATGTATGTATATA 1207

Qy        650 CATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATATG 709
              |||||||  | | ||| |||||  |||||||||||||||| |||  | || |||||
Db       1208 TGCACATATACCTATATATACATATATGTATATATACACATATACATATATGTACATATG 1267

Qy        710 TATATATA 717
              | ||||||
Db       1268 TGTATATA 1275


RESULT 87
US-10-292-798-2029
; Sequence 2029, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
;  FILE REFERENCE: 084335/166
;  CURRENT APPLICATION NUMBER: US/10/292,798
;  CURRENT FILING DATE:  2002-11-13
;  PRIOR APPLICATION NUMBER: 10/017,161
```

```
;  PRIOR FILING DATE: 2001-12-18
;  PRIOR APPLICATION NUMBER: JP 2001-246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2070
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2029
;    LENGTH: 1276
;    TYPE: DNA
;    ORGANISM: Homo sapiens
;    FEATURE:
;    LOCATION: source
;    FEATURE:
;    LOCATION: (1)..(1276)
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (201)..(1076)
US-10-292-798-2029

  Query Match             5.2%;  Score 240.8;  DB 17;  Length 1276;
  Best Local Similarity   69.9%;  Pred. No. 2.3e-39;
  Matches  341;  Conservative    0;  Mismatches  142;  Indels    5;  Gaps    1;

Qy        230 ATATATATATATATATATATATACACACACACACATATATATATACACATATATATAC 289
              | ||||| |||| || |||||||||| |    ||| | |||  |||||||| || || ||
Db        793 ACATATACATATGTACATATATACATATGTACATATATACGTATATACATATGTACATCT 852

Qy        290 ACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATAT 349
              |  |||||||  ||||   |||||||  | |||   ||||  |||  | | || ||  ||
Db        853 ATGTATATACGCATATGTGCATATATGTATATACGCATACGTATATGTGTATATACGCAT 912

Qy        350 ATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATAC 409
              ||||||| || || ||| |  |||||||  |||||||||| |   |  ||| |  |||||||
Db        913 ATATACATATGTGCATATATGTATATACGCATATATACATATGTGCATATATGTATATAC 972

Qy        410 ACACATATATACACATATATACACATATATACACATATACACATATATACACATATATAC 469
              || ||||| | |  |  ||| |  |||||||  |||||| ||||||  | || |||| ||
Db        973 GCATATATACATATGTGCATATATGTATATACGCATATATACATATGTGCATATATGTAT 1032

Qy        470 ATATATACACATATATATAATATACACACATATATATACACATATATACACACATATATA 529
              |||   | | |||       ||| |    |||||  |  | | ||  |||   |||| |||
Db       1033 ATACGCATATATA-----CATACATGTGCATATGTGCATATATTCATAAGTACATCTATG 1087

Qy        530 CACATATATACACATATATATACACATATATACACATATATACATATATACACATATATA 589
              | || || | | || || |||||||||| |||| |||| |||||| | | |  || |||
Db       1088 TATATGTACATATATGTACATACACATATGTACATATATGTACATACACATATGTACATA 1147

Qy        590 TACATATATACACATATATACATATATACACATATATACATATATACACACATATATACA 649
              ||  || ||||||||||| | |||| |||| ||| |||||||||| ||   || |||| |
Db       1148 TATGTACATACACATATACATATATGTACATATACATACATATATGCATGTATGTATATA 1207

Qy        650 CATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATATG 709
                |||||||  | | ||| |||||  ||||||||||||||| |||  | || |||||
Db       1208 TGCACATATACCTATATATACATATATGTATATATACACATATACATATATGTACATATG 1267

Qy        710 TATATATA 717
              | ||||||
Db       1268 TGTATATA 1275


RESULT 88
US-10-437-963-76607
; Sequence 76607, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 76607
;    LENGTH: 534
;    TYPE: DNA
;    ORGANISM: Oryza sativa
;    FEATURE:
;    OTHER INFORMATION: Clone ID: PAT_MRT4530_76588C.1
US-10-437-963-76607

  Query Match             5.2%;  Score 240.6;  DB 18;  Length 534;
  Best Local Similarity   76.4%;  Pred. No. 1.7e-39;
  Matches  308;  Conservative    0;  Mismatches   94;  Indels    1;  Gaps    1;

Qy        293 TATATACATATATACACATATATACACATATATATACACATACATATGTACACATATATA 352
              ||| |    ||    | | ||||| ||||| ||||| | ||| |||| || | |||||||
Db         17 TATTTCACAATTATAAGAAATATATACATACATATATATATATATATATATATATATA 76

Qy        353 TACACATATGTATACACATATATACACATATATACACACATATATACACATATATACACA 412
              || | |||| |||| |||||||| | ||||||||| ||||| ||||| |||||||||  |
Db         77 TATATATATATATACATATATATATATATATACATACATACATACATATATATACATA 136

Qy        413 CATATATACACATATATACACATATATACACATATACACATATATACACATATATACATA 472
              |||| ||||| | ||||| | ||||||||| | ||||| ||||||| | ||||||| |||
Db        137 CATACATACATACATATATATATATATATACATACATACATATATATATATATATATATA 196

Qy        473 TATACACATATATATA-ATATACACACATATATATACACATATATACACACATATATACA 531
              |||| | ||||||||| ||||| | | ||||||||| | ||||||| ||| | ||||| |
Db        197 TATATATATATATATATATATATATATATATATATATATATATATACATACATATATA 256

Qy        532 CATATATACACATATATATACACATATATACACATATATACATATATACACATATATATA 591
              |||||||||||||||||||||||| ||||| ||| ||||||||||||||| | |||| ||||
Db        257 CATATATACACATATATATACATATATACACATATATATACATATATATATATGTATA 316

Qy        592 CATATATACACATATATACATATATACACATATATACATATATACACACATATATACACA 651
              |||||||||||||||||| | ||||| | ||||||| | ||| ||||||| | | | |
Db        317 CATATATACACATATATATACATATATATATATATATACATACACACACACACATATC 376

Qy        652 TACATATACACACACATAGATATACATATATATACACATATAT 694
                  |||   |   || | | || | |  |   |    ||| ||
Db        377 ATGATAGTAAGTTACTTGGCTAGAGAATTTGTTTTGAATACAT 419


RESULT 89
US-10-741-601-5636
; Sequence 5636, Application US/10741601
; Publication No. US20040166519A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  STENOSIS, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001500
;  CURRENT APPLICATION NUMBER: US/10/741,601
;  CURRENT FILING DATE:  2003-12-22
;  NUMBER OF SEQ ID NOS: 26415
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5636
;    LENGTH: 81924
;    TYPE: DNA
;    ORGANISM: Homo sapiens
;    FEATURE:
;    NAME/KEY: misc_feature
;    LOCATION: (1)...(81924)
;    OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-2
US-10-741-601-5636

  Query Match             5.2%;  Score 240.6;  DB 18;  Length 81924;
  Best Local Similarity   69.3%;  Pred. No. 1.6e-38;
  Matches  343;  Conservative    0;  Mismatches  149;  Indels    3;  Gaps    1;
```

```
Qy        233 TATATATATATATATATATACACACACACACACATATATATATACACATATATATACACA 292
              ||| |||| | |  ||| || |||||||| | |||| | | |  ||||| | ||  |
Db      37545 TATGTATACACACGTATGTATACACACACGTATGTATACACACATGTATATACACACGTA 37604

Qy        293 TATATACATATATACACATATATACACATATATATACACATACATATGTACACATATATA 352
              | |||||| | ||  | ||||| |||| ||| |||||||| |  ||| || ||| |||||
Db      37605 TGTATACACACATGTATATATACACACGTATGTATACACACATGTATATATACACATATA 37664

Qy        353 TACACATATGTATACACATATATACACATATATACACACATATATACACATATATACACA 412
              |||| |||||  || | ||| | ||  ||||||      |||||    |||| |||||
Db      37665 TACATATATGGGTATATATACACACGTGTATATATGTGTGCATATATGTGTATACACACA 37724

Qy        413 CATATATACACATATATACACATATATACACATATACACATATATACACATATATACATA 472
              || | |  | ||| |||||||  ||||     |||| ||||| | ||| | |||||  |
Db      37725 TATGTGTGTATATACATACACACATATGTGTGTATATACATACACACAAACATATATGTG 37784

Qy        473 TATACACATATATATA---ATATACACACATATATATACACATATATACACACATATATA 529
              | |  |  | ||||||   |||||  | ||| |||||  |||||  ||||    | ||| | |
Db      37785 TGTGTATGTGTATATATACATATATATACACATATACACACACATATGTGTATATACACA 37844

Qy        530 CACATATATACACATATATATACACATATATACACATATATACATATATACACATATATA 589
              | | | | ||| || | |||| ||| |||||||||  | |||  | |||| ||| |||||
Db      37845 TATACACACACATATGTGTATATACACATATACACACACATATGTGTATATACACATATA 37904

Qy        590 TACATATATACACATATATACATATATACACATATATACATATATACACACATATATACA 649
              |||  ||||    |||||  |||||||  |||||  ||||   ||||||||||  ||| | |||
Db      37905 CACACATATGTGTATATACACATATACACACACATATGTGTATATACACATATACACACA 37964

Qy        650 CATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATATG 709
              | |  |||||||||  | ||   |||||||  |||||||||||||  | |  ||||  | |||
Db      37965 CGTGTATATACACATATATGTGTATACATGTATATACACATATGTGTGTGTATGTGTATA 38024

Qy        710 TATATATATATGCTC 724
              | |||  |||||  | |
Db      38025 TGTATGTATATACAC 38039


RESULT 90
US-10-437-963-46068/c
; Sequence 46068, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 46068
;   LENGTH: 610
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_48971C.1
US-10-437-963-46068

  Query Match             5.2%;  Score 240.4;  DB 18;  Length 610;
  Best Local Similarity   69.2%;  Pred. No. 2e-39;
  Matches  342;  Conservative    0;  Mismatches  151;  Indels    1;  Gaps    1;

Qy        226 TTTTATATATATATATATATATATATACACACACACACACATATATATATACACATATAT 285
              ||   |||| || ||  | || | |          ||   | ||||||||||||    | |||
Db        559 TTGAGTATACATCTAGGTGTACAGAAGAGTCGTCATGTATATATATATATATGTGTGTAT 500

Qy        286 ATACACATATATACATATATACACATATATACACATATATATACACATACATATGTACAC 345
              ||| | ||||||   ||||||  |  | | || | ||||||  ||||  || |||||||||
Db        499 ATATATATATATGTGTATATATATGTGTGTATATATATATGTACATGTATATATGTACAT 440

Qy        346 ATATATATACACATATGTATACACATATATACACATATATACACACATATATACACATAT 405
              | ||||||  | ||||  ||||    ||||  ||  |  ||  ||     | |||||||  |||| |
Db        439 GTGTATATATATATATATATATGTATATGTATATGTACATGTGTATATATATATACATGT 380

Qy        406 ATACACACATATATACACATATATACACATATATACACATATACACATATATACACATAT 465
              |    | |||||||  ||||  |  |      |  |  |  | |||||  | |||||||  | ||||
Db        379 GTGTGTATATATATATACATGTGTGTGTGTGTGTGTATATATATATATATATATATATAT 320

Qy        466 ATACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACAT 524
              |||  |||||||  | |||||||||  ||||||||||  |||||||  |  |||||||  | |||
Db        319 ATATATATATATATATATATATATATATACACACACATATATATATATATATATACAC 260

Qy        525 ATATACACATATATACACATATATATACACATATATACACATATATACATATATACACAT 584
              |||  |  | |||||||  | |||||||||  | |||||||  | |||||||  |||||||  | ||
Db        259 ATACATATATATATATATATATATATATATATATATATATATATATATATATATATAT 200

Qy        585 ATATATACATATATACACATATATACATATATACACATATATACATATATACACACATAT 644
              |||||||  |||||||  | |||||||  | |||||  | |||  |  ||| |||  | | ||||||
Db        199 ATATATATATATATATATATATATATATACATATATATATACACTCACATAGATATACATAT 140

Qy        645 ATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATAT 704
              | | ||     |||||||  |||  |||  ||  | | |  |||     ||||||||  |    ||
Db        139 ACATACTATTATATACATACATATATATTTCCGTCCATACTTTCATATATACATCCGTAC 80

Qy        705 ATATGTATATATAT 718
              || ||||  || |
Db         79 CTACGTATTAATTT 66


RESULT 91
US-10-311-455-1842/c
; Sequence 1842, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1842
;   LENGTH: 13573
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1842

  Query Match             5.2%;  Score 240.4;  DB 15;  Length 13573;
  Best Local Similarity   69.2%;  Pred. No. 8e-39;
  Matches  537;  Conservative    0;  Mismatches  206;  Indels   33;  Gaps   14;

Qy          4 TCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAA--TACAAAATTAG 61
              | || |||| ||| | ||| ||    ||| ||| |||||||| |||||     |||||||
Db       6406 TTAAAACCAACCTAAACAATATAACAAAATCCCGTCTCTACAAAAAAATACAAAAATTAA 6347

Qy         62 ACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAATC 121
              ||  | |  | | |||| ||| |||||||| ||||||    |  |  |    |  | ||||
Db       6346 CCAAACTTAATAACACACACCTATAATCCCAACTACTTAAAAAACTAAAATAAAAAAATC 6287

Qy        122 ACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAGCC 181
```

```
                ||||  ||||              |  |       | ||  ||| | ||| | ||||||| ||
Db       6286 ACTTAAACCCAAAAAATAAAAAC-TACAATAAACCATAATCGTACCACTACACTCCAACC 6228

Qy        182 CAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATAT---A 238
                |  ||| || |   ||   | |    | |||||| | |       | |||||||||    |
Db       6227 TAAACAATAAAACAAAACCCTAACAAAAAAAAAAATATATATATATATATATATATAA 6168

Qy        239 TATATATATATATACACACACACACATATATATAT-ACACATATATATACACATATAT 297
                || ||||||||||||   | | |||   |||||||||| | ||||||||||| || |||||
Db       6167 TACATATATATATACGTATATACATTTATATATATATAATACATATATATATACGTATAT 6108

Qy        298 ACATATATACACATAT-ATACACATATATATACACATACATATGTACACATATATATACA 356
                |||| |||| | |||| ||||| ||||||||||  ||| | || || | ||||||| | |
Db       6107 ACATTTATATATATATAATACATATATATATACGTATATACATTTATATATATATA-ATA 6049

Qy        357 CATATGTATACACATATATACACATATAT-------ACACACATATATACACATATATAC 409
                ||||| |||| || ||||||||  |||||       | ||| ||||||| || ||||||||
Db       6048 CATATATATATACGTATATACATTTATATATATATAATACATATATATATACGTATATAC 5989

Qy        410 ACACATAT-------ATACACATATATACACATATATACACATATACACATATATACACA 462
                |    ||||       ||||| ||||||| || ||||||||  |||| | ||||| | |||
Db       5988 ATTTATATATATATAATACATATATATATACGTATATACATTTATATATATATA-ATACA 5930

Qy        463 TATATACATATATACACATATATATAATATACACACATATATATACACATATATACACAC 522
                |||||| |||||||||  |||||||    | ||| |||||||||| ||||| |   |
Db       5929 TATATATATATATACATTTATATATATATAATACATATATATATACATATATACATTTAT 5870

Qy        523 ATATAT---ACACATATATACACATATATATACACATATATACACATATATACATATATA 579
                ||||||    | |||||||||| | | |||||||||  ||||||  | |||  |||||||||||
Db       5869 ATATATATAATACATATATATATACATATATACATTTATATATATATA-ATACATATATA 5811

Qy        580 CACATATATATACATATATACACAT---ATATACATATATACACATATATACATATATAC 636
                | | ||||||||||| |||| | ||    ||| | ||||||| ||||||||||||  ||||
Db       5810 TATACATATATACATTTATATATATATAATACATATATATATACATATATACATTTATAT 5751

Qy        637 ACACATATATACACATACATATACACACACATAGATATACATATAT-ATACACATATATA 695
                | | ||| ||||| ||| ||||||| | | |  |||| |||||| ||||| |||||||
Db       5750 ATATATA-ATACATATATATATACATATATACATTTATATATATATAATACATATATATA 5692

Qy        696 TACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGAT 751
                ||| |||||| || ||||||||||   | || |   |   | |  ||    || ||
Db       5691 TACATATATACATTTATATATATATAATACATATATATATACATATATACATTTAT 5636


RESULT 92
US-10-719-993-6943
; Sequence 6943, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001496
;  CURRENT APPLICATION NUMBER: US/10/719,993
;  CURRENT FILING DATE:  2003-11-24
;  NUMBER OF SEQ ID NOS: 55342
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6943
;   LENGTH: 145806
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-719-993-6943

  Query Match             5.2%;  Score 239;  DB 18;  Length 145806;
  Best Local Similarity   64.2%;  Pred. No. 4.4e-38;
  Matches  437;  Conservative    3;  Mismatches  228;  Indels   13;  Gaps    5;

Qy          7 AGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATTAGACAG 65
                ||  || ||||   |||||||||||  ||||||||||||||||||| ||||||||
Db      68913 AGCTCATCCTGGCTAACATGGTGAAAAGCCATCTCTACTAAAAATACAAAAATTAGCTGA 68972

Qy         66 GCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAATCACTT 125
               ||||||||  ||  ||||||| || ||||||||  |||||| |||||||||||||| |||

Db      68973 GCATGGTGGCACGTGCCTGTAGTCTCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCTT 69032

Qy        126 GAACCTGCTGGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCCAGCCCAG 184
               |||||||  || ||||||||||| | | ||| |||||||||||||| ||||||||||||  |
Db      69033 GAACCTG--GGAGGTGGAGGTTACAGTGAGCCAAGATCATGCCACTGCACTCCAGCCTGG 69090

Qy        185 GCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATA 244
               || |||        | | |   || | ||||||||  |       | ||||||||||: | |
Db      69091 GCGACAGAGCAAGACTCTTGCCTAAAAAAAAAAATA------TATATATATATAWGTGTG 69144

Qy        245 TATATATACACACACACACATATATATATACACATATATATACACATATATACATATA 304
               | | | |                | |||||||   | |  |||| | |||||  |||||||
Db      69145 TGTGTGTGTGTGTGTGTGTACATATATACGTATACGTATACATACATACATACATACG 69204

Qy        305 TACACATATATACACATATATATACACATACATATGTACACATATATATACACATATGTA 364
               || ||||| ||||| || |||| | |||||| |||   || | ||| |||||  ||| | |
Db      69205 TATACATACATACATATGTATACATACATACGTATACGCATACATACATACATATACGCA 69264

Qy        365 TACACATATATACACATATATACACACATATATACACATATATACACACATATATACACA 424
               |||| | |||||| |||| |||||||| |||   || ||  ||| | | | |  || |  ||
Db      69265 TACATACATATACGCATACATACACATATACGTATACGCATACATATATACGTACATGCA 69324

Qy        425 TATATACACATATATACACATATACACATATATACACATATATACATATATACACATATA 484
               || ||||| ||||| ||| | ||  |  |||||||| |:|||  |||||||||| |||||
Db      69325 TACATACATATATACACATACATGTATGTATATACATAYATACGCATATATACATATATA 69384

Qy        485 TATAATATACACACATATATATACACATATATACACACATATATACACATATATACACAT 544
                  |||| |  | |||||||  | |||   : |   || |||| |||| |||    ||
Db      69385 ---CGTATATATGCGTATATATGTATATACGCRTATGTATGTATATACATGTATGTGTAT 69441

Qy        545 ATATATACACATATATACACATATATACATATATACACATATATATACATATATACACAT 604
               ||||| ||| |||||| |  | |  |  |  |  |  |  |  |  |  | |  ||
Db      69442 ATATACACATATATATGAAGTTTTTATCTGACTTTATTAGCCACTAGAGCTTGTTTGCAG 69501

Qy        605 ATATACATATATACACATATATACATATATACACACATATATACACATACATATACACAC 664
                ||    | |  | |||| | |   |  || |   ||   | |  |   | ||| |
Db      69502 TTAATATTTTGCAGACATTTCTGTTTTCATCCTTTTATTCTTGTATCTGTCTTTATATTT 69561

Qy        665 ACATAGATATACATATATATA 685
               | |  |   | | |||| |||
Db      69562 AAAGTGGGTTTCTTATAAATA 69582


RESULT 93
US-10-417-375-36
; Sequence 36, Application US/10417375
; Publication No. US20040219528A1
; GENERAL INFORMATION:
;  APPLICANT: David W. Morris
;  APPLICANT:  Marc Malandro
;  TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
;  FILE REFERENCE: 529452001600
;  CURRENT APPLICATION NUMBER: US/10/417,375
;  CURRENT FILING DATE:  2003-04-15
;  NUMBER OF SEQ ID NOS: 176
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 36
;   LENGTH: 91697
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-417-375-36

  Query Match             5.1%;  Score 237.8;  DB 18;  Length 91697;
  Best Local Similarity   71.6%;  Pred. No. 6.4e-38;
  Matches  353;  Conservative    0;  Mismatches  137;  Indels    3;  Gaps    3;

Qy        229 TATATATATATATATATATATATACACACACACACATATATATATACACATATATATA 288
               | ||||||| |||||| |||||||||| |    |    | |||| | |||| |||||||||| |
Db      19560 TCTATATATCTATATACATATATACATATGTATGTATATATTTCTATATACATATATACA 19619

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
                |  |||||| | ||  | | | || || | | |||| || | ||  || | ||||  ||
Db      19620 TATGTATATATAGATTGATATACATCTATATACATATGTATATATCTATGTATACATCTA 19679
```

```
Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              | |||||| | |||||||||  ||| || |||||  ||  | ||||||||||| || ||||
Db      19680 TGTATACATAGATGTATACATCTATGTATACATAGATGTATACATATATACATATGTATA 19739

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              ||   || |||||| |||||  | |  |||||||   |||  | |||  || |||||| ||
Db      19740 CATCTATGTATACATATATACATATGTATATACTATTATGTATATA-GTATACATATGTA 19798

Qy        469 CATATATACACATATATATAATATACACACATATATATACACATATATACACACATATAT 528
              ||||  ||  | ||||  |||  |||||||  | ||  ||||||   ||||||  | | |  |||
Db      19799 TATATGTATATATATCTAT-ATATACATATATGTATATATGTATATATCTATATACGTAT 19857

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACATATAT 588
              | |  |||||||| ||||| ||| |  |||||||| || |||| || |||| | ||||||
Db      19858 ATAGGTATATACATATATAAATATATCTATATACATATGTATATATCTATATAAATATAT 19917

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
              | ||||||  ||||||||  ||||| | | || |||| || |||| ||| ||| ||
Db      19918 GGATATATATCTACATATATCTATATACATATATCTATATATCTATATACATATAGATGT 19977

Qy        649 ACATACATATACA-CACACATAGATATACATATATATACACATATATATACGTATATATA 707
              | ||  || ||||   || ||  ||||| |||||||| |||| |||||   ||| ||||
Db      19978 ATATGTATCTACATGGCATATGTATATAGATATATATATACATCTATATGTATATCTATA 20037

Qy        708 TGTATATATATAT 720
              |  |||| |||||
Db      20038 TACATATCTATAT 20050
```

RESULT 94
US-10-437-963-90135
; Sequence 90135, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 90135
; LENGTH: 573
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_88833C.1
US-10-437-963-90135

Query Match 5.1%; Score 236.4; DB 18; Length 573;
Best Local Similarity 67.4%; Pred. No. 1.3e-38;
Matches 333; Conservative 0; Mismatches 161; Indels 0; Gaps 0;

```
Qy        227 TTTATATATATATATATATATATATACACACACACACATATATATATACACATATATA 286
              | | | | | | |||||||||||||||| | | | | | |||||||||||| | ||||| |
Db         75 TATGTGTGTGTCTATATATATATATATATATATATATATATATATATATATATACA 134

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||| ||| | ||||||| | | ||||||| | ||| |||| || |||
Db        135 TATATATATATATACATATATATATATATATACATATATATATATATATATATACA 194

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              || ||||| | |||| |||||| ||||| ||| ||| ||||| ||||||||||||||||
Db        195 TACATATATATATATATACATATATACACATATACATACATACATATATACACATATA 254
```

```
Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              | | | | | ||| ||||| ||||| ||||||||| ||||||| ||||| |
Db        255 CATATATACACATATACATACATACATATATATACATATATACATATATACATGGGAGCG 314

Qy        467 TACATATATACACATATATATAATATACACACATATATATACACATATATACACACATAT 526
              || |                |     |  |||||  | | |    |||| |          |
Db        315 TATACTATGCACACAGGCCCTCGCGTGTACACACCTTTTTGGTGATATTTTTTAGTTGGT 374

Qy        527 ATACACATATATACACATATATATACACATATATACACATATATACATATATACACATAT 586
              |||||  | |  |||||  |             |  |||          | |  |||||||| ||||
Db        375 GTACACGTGTGTACACGTGAGGGCCTGTGTGCATAGGATATGTTGCCTATATACATATAT 434

Qy        587 ATATACATATATACACATATATACATATATACACATATATACATATATACACACATATAT 646
              | |||  ||||||| |||||||||  |||||||  | |||||||  |||||||  | |  ||||
Db        435 ACATATATATATATACATATATATATATATATATATATATATATATATATATATGA 494

Qy        647 ACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATAT 706
                    || ||   |   | | |||  |||||  |||||||||  | |||||||||  |||||||
Db        495 GAGAATCCACTCATGTATATATATATATATATATATATATATATATATATATATATATAT 554

Qy        707 ATGTATATATATAT 720
              ||  ||||||||||||
Db        555 ATATATATATATAT 568
```

RESULT 95
US-10-437-963-100710
; Sequence 100710, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 100710
; LENGTH: 662
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_983C.1
US-10-437-963-100710

Query Match 5.1%; Score 236.4; DB 18; Length 662;
Best Local Similarity 69.7%; Pred. No. 1.4e-38;
Matches 347; Conservative 0; Mismatches 149; Indels 2; Gaps 2;

```
Qy        229 TATATATATATATATATATATATACACACACACACATATATATATACACATATATATA 288
              | |||||||||||||||||||||  | | |||   | || |||||| |   |||| ||||
Db        116 TTTATATATATATATATATATATGTATATACATGTATATGTATATACATGTATATGTATA 175

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              ||||  ||||  ||||||||  |||||  | ||||||  || |  ||  ||| ||||  ||
Db        176 TACATGTATATGTATATACATGTATATGTATATATATGTATATGTATGTATATACATGTA 235

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              ||| || | ||||||||||||  |||||  | ||||||  | | ||||||||   |||| ||
Db        236 TATGTATATATATGTATACATGTATATGTATATATATGTATATATATATATGTATATGTA 295

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              | | ||||  || | ||||||  | |||||||  | ||  ||  | |||||||  | |||||||
Db        296 TATATATATGTATATATATATGTATATATATATATATGTATATATATATATATATATATA 355

Qy        469 CATATATACACATATATATAATATACACACATATATATACACATATATAC-ACACATATA 527
```

```
Db        356 TATCTCAATCTTTAGAGAGCAATATCTCTCACATCATATCTTGTATTTCCAATCCGTTTG 415

Qy        528 TACACATATATACACATATATATACACATATATACACATATATACATATATACACATATA 587
              ||    | || | | | | | | | ||||| | | ||||||||||||| | || ||
Db        416 CACCATGAAGTTCATAAAAAAACATATACATATATATACATATACATATATATATGTA 475

Qy        588 TATACATATATACACATATATACATATATAC-ACATATATACATATATACACACATATAT 646
              | ||| ||||||||  |||| |  ||| |||   |||||||||| ||||| ||||| ||
Db        476 TGTACGTATATACATGTATACATGTATGTACGTTATATATACATACATACATACATACAT 535

Qy        647 ACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATAT 706
              ||| | | ||||| ||| | ||| ||||| |||||||| | ||||||||  |||||||
Db        536 ACATACATATATAGACATATATATATATATATATATATATATATATATATATATATAT 595

Qy        707 ATGTATATATATATGCTC 724
              ||||||||||||  | |
Db        596 ATGTATATATATACACAC 613


RESULT 96
US-10-017-161-2181/c
; Sequence 2181, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
;  FILE REFERENCE: 084335/0152
;  CURRENT APPLICATION NUMBER: US/10/017,161
;  CURRENT FILING DATE:  2002-12-18
;  PRIOR APPLICATION NUMBER: JP 2001/246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2430
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2181
;   LENGTH: 1192
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: source
;   LOCATION: (1)..(1192)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(992)
US-10-017-161-2181

  Query Match             5.1%;  Score 235.8;  DB 15;  Length 1192;
  Best Local Similarity   69.2%;  Pred. No. 2.4e-38;
  Matches  355;  Conservative    0;  Mismatches  147;  Indels   11;  Gaps    2;

Qy        214 AAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACATATATAT 273
              |||||| |     | |||| |||||||| | ||||| || | ||||||| |||| |||||
Db       1000 AAAAAATATTATATATATACATATATATGTTTATATGTATATACACACATACATGTATAT 941

Qy        274 ATACACATATATATACACATATATACATATATACACATATATACACATATATATACACAT 333
              ||  | ||||| || ||| |||||| | | ||| ||| || ||      ||||||||||
Db        940 ATGTATATATACATGCACGTATATATACACATATACACATGTATGTGCGTATATACACAT 881

Qy        334 ACATATGTACACATATATATACACATATGTATACACATATATACACATATATACACACAT 393
              |||| | || | | ||||||||||| | | | |  |||||||    ||||| ||||||
Db        880 ACATGTATATATACACATATACACATGTATGTGCGTATATATAC----ATATATACACAT 825

Qy        394 ATATACACATATATACACACATATATACACATATATACACATATATACACATATACACAT 453
              |||   | | |||| ||| ||||| |||  | | | |   |||||||| ||||||||||
Db        824 GTATGTGCGTGTATATACATATATACACATGTGTGTGCGTGTATATACATATATACACAT 765

Qy        454 ATATACACATATATACATATATACACATATATAT-------AATATACACACATATATAT 506
              | |   | | |||| | ||||| ||| | ||||        | || | |||||||||
Db        764 GTGTGTGCGTGTATATACATATATACACAGATATGTGTGTGCGTGTATATACATATATAC 705

Qy        507 ACACATATATACACACATATATACACATATATACACATATATATACACATATATACACAT 566
              ||| ||||  |      |||||||||| ||| | ||| |||| | |    |||| |||| ||
Db        704 ACAGATATGTGTGTGTATATATACATATACACACAGATATGTGTGTGTATATGTACATAT 645

Qy        567 ATATACATATATACACATATATATACATATATACACATATATACATATATACACATATAT 626
              | | |||| | |  | ||||| ||| ||| ||| ||| | ||   | |||| | | ||||
Db        644 ACACACATGTGTGTATATATACATATATACATATACACACATGTGTGTATATATACATAT 585

Qy        627 ACATATATACACACATATATACACATACATATACACACACATAGATATACATATATATAC 686
              | | |||||||||| | | || | ||||||||| ||| | | | |  |||||||||
Db        584 ATACATATACACACGTGTGTATATATACATATATACATATACACACGTGTGTATATATAC 525

Qy        687 ACATATATATACGTATATATATGTATATATATA 719
              | ||||| | || || || || || ||||||||
Db        524 ATATATACACACATACATGTACGTGTATATATA 492


RESULT 97
US-10-292-798-1827/c
; Sequence 1827, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
;  FILE REFERENCE: 084335/166
;  CURRENT APPLICATION NUMBER: US/10/292,798
;  CURRENT FILING DATE:  2002-11-13
;  PRIOR APPLICATION NUMBER: 10/017,161
;  PRIOR FILING DATE: 2001-12-18
;  PRIOR APPLICATION NUMBER: JP 2001-246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2070
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1827
;   LENGTH: 1192
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   LOCATION: source
;   FEATURE:
;   LOCATION: (1)..(1192)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(992)
US-10-292-798-1827

  Query Match             5.1%;  Score 235.8;  DB 17;  Length 1192;
  Best Local Similarity   69.2%;  Pred. No. 2.4e-38;
  Matches  355;  Conservative    0;  Mismatches  147;  Indels   11;  Gaps    2;

Qy        214 AAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACATATATAT 273
              |||||| |     | |||| |||||||| | ||||| || | ||||||| |||| |||||
Db       1000 AAAAAATATTATATATATACATATATATGTTTATATGTATATACACACATACATGTATAT 941

Qy        274 ATACACATATATATACACATATATACATATATACACATATATACACATATATATACACAT 333
              ||  | ||||| || ||| |||||| | | ||| ||| || ||      ||||||||||
Db        940 ATGTATATATACATGCACGTATATATACACATATACACATGTATGTGCGTATATACACAT 881

Qy        334 ACATATGTACACATATATATACACATATGTATACACATATATACACATATATACACACAT 393
              |||| | || | | ||||||||||| | | | |  |||||||    ||||| ||||||
Db        880 ACATGTATATATACACATATACACATGTATGTGCGTATATATAC----ATATATACACAT 825

Qy        394 ATATACACATATATACACACATATATACACATATATACACATATATACACATATACACAT 453
              |||   | | |||| ||| ||||| |||  | | | |   |||||||| ||||||||||
Db        824 GTATGTGCGTGTATATACATATATACACATGTGTGTGCGTGTATATACATATATACACAT 765

Qy        454 ATATACACATATATACATATATACACATATATAT-------AATATACACACATATATAT 506
              | |   | | |||| | ||||| ||| | ||||        | || | |||||||||
Db        764 GTGTGTGCGTGTATATACATATATACACAGATATGTGTGTGCGTGTATATACATATATAC 705
```

```
Qy        507 ACACATATATACACACATATATACACATATATACACATATATATACACATATATACACAT 566
              ||| |||| |       ||||||||| ||| | ||| |||| | |    |||| |||| ||
Db        704 ACAGATATGTGTGTGTATATATACATATACACACAGATATGTGTGTGTATATGTACATAT 645

Qy        567 ATATACATATATACACATATATATACATATATACACATATATACATATATACACATATAT 626
              | | |||| | |  | ||||| ||| ||| ||| ||| | ||   | |||| | | ||||
Db        644 ACACACATGTGTGTATATATACATATATACATATACACACATGTGTGTATATATACATAT 585

Qy        627 ACATATATACACACATATATACACATACATATACACACACATAGATATACATATATATAC 686
              | | |||||||||| | | || | ||||||||| ||| | | | |   |||||||||
Db        584 ATACATATACACACGTGTGTATATATACATATATACATATACACACGTGTGTATATATAC 525

Qy        687 ACATATATATACGTATATATATGTATATATATA 719
              | ||||| | || || || || || ||||||||
Db        524 ATATATACACACATACATGTACGTGTATATATA 492


RESULT 98
US-10-027-632-258173
; Sequence 258173, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 258173
;   LENGTH: 3673
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-258173

  Query Match             5.1%;  Score 235.2;  DB 13;  Length 3673;
  Best Local Similarity   67.2%;  Pred. No. 5.3e-38;
  Matches  347;  Conservative    0;  Mismatches  168;  Indels    1;  Gaps    1;

Qy        204 ATCTCAGAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACAC 263
              || | |  | |   | ||    | || | | || |||| |||||||||| | |||
Db       1857 ATATAAAGACATCTATAAATATATATAAAGACATCTATAAATATATATAAAGACATCTGT 1916

Qy        264 ACATATATATATACACATATATATACACATATATACATATATACACATATATACACATAT 323
              | |||||||||| | |||| | || | | ||||| | | || || | ||||||| | | |
Db       1917 AAATATATATAAAGACATCTGTAAATATATATAAAGACATCTATAAATATATATAAAGAC 1976

Qy        324 ATATACACATACATATGTACACATATATATACACATATGTATACACATATATACACATAT 383
              || || | ||| |||| | |    || |||| | |||| || | |||| |||| | ||||
Db       1977 ATCTATAAATATATATAAAGATGCATCTATAAATATATATAAAGACATCTATAAATATAT 2036

Qy        384 ATACACACATATATACACATATATACACACATATATACACATATATACACATATATACAC 443
              ||| | |||| |||| | ||| ||| | |||| |||| | ||||||| | | || || |
Db       2037 ATAAAGACATCTATAAATATACATAAAGACATCTATAAATATATATAAAAACATCTATAA 2096

Qy        444 ATATACACATATATACACATATATACATATATACACATATATATAATATACACACATATA 503
              ||||| | | | || ||  |||| | ||||| | |||| |||| |||||| | | ||  |
Db       2097 ATATATATAAAGATGCATCTATAAATATATAAAGACATCTATA-AATATATAAAGATGCA 2155
```

```
Qy        504 TATACACATATATACACACATATATACACATATATACACATATATATACACATATATACA 563
              | || | ||||||| |||  |||| | | ||| | |  ||| |||| | | ||| | |||
Db       2156 TCTATAAATATATAGACATCTATAAATATATAAAGATGCATCTATAAATATATAAAGACA 2215

Qy        564 CATATATACATATATACACATATATATACATATATACACATATATACATATATACACATA 623
              |||| | ||||||| | | || |||| ||||||| | | | | |  |||| | | |||
Db       2216 TCTATAAATATATATAAAGACATCTATAAATATATATAAAGACACATCTATAAATATATA 2275

Qy        624 TATACATATATACACACATATATACACATACATATACACACACATAGATATACATATATA 683
              || | | || || | | ||||||| | |  ||| || | | | ||| | |  ||| ||||
Db       2276 TAAAGACATCTATAAATATATATAAAGACGCATCTATAAATATATATAAAGGCATCTATA 2335

Qy        684 TACACATATATATACGTATATATATGTATATATATA 719
              | | ||||| | | | || ||||  |||||||| |
Db       2336 AATATATATAAAGATGCATCTATAAATATATATAAA 2371


RESULT 99
US-10-027-632-258173
; Sequence 258173, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 258173
;   LENGTH: 3673
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-258173

  Query Match             5.1%;  Score 235.2;  DB 17;  Length 3673;
  Best Local Similarity   67.2%;  Pred. No. 5.3e-38;
  Matches  347;  Conservative    0;  Mismatches  168;  Indels    1;  Gaps    1;

Qy        204 ATCTCAGAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACAC 263
              || | |  | |   | ||    | || | | || |||| |||||||||| | |||
Db       1857 ATATAAAGACATCTATAAATATATATAAAGACATCTATAAATATATATAAAGACATCTGT 1916

Qy        264 ACATATATATATACACATATATATACACATATATACATATATACACATATATACACATAT 323
              | |||||||||| | |||| | || | | ||||| | | || || | ||||||| | | |
Db       1917 AAATATATATAAAGACATCTGTAAATATATATAAAGACATCTATAAATATATATAAAGAC 1976

Qy        324 ATATACACATACATATGTACACATATATATACACATATGTATACACATATATACACATAT 383
              || || | ||| |||| | |    || |||| | |||| || | |||| |||| | ||||
Db       1977 ATCTATAAATATATATAAAGATGCATCTATAAATATATATAAAGACATCTATAAATATAT 2036

Qy        384 ATACACACATATATACACATATATACACACATATATACACATATATACACATATATACAC 443
              ||| | |||| |||| | ||| ||| | |||| |||| | ||||||| | | || || |
Db       2037 ATAAAGACATCTATAAATATACATAAAGACATCTATAAATATATATAAAAACATCTATAA 2096

Qy        444 ATATACACATATATACACATATATACATATATACACATATATATAATATACACACATATA 503
              ||||| | | | || ||  |||| | ||||| | |||| |||| |||||| | | ||  |
Db       2097 ATATATATAAAGATGCATCTATAAATATATAAAGACATCTATA-AATATATAAAGATGCA 2155
```

```
Db       2097 ATATATATAAAGATGCATCTATAAATATATAAAGACATCTATA-AATATATAAAGATGCA 2155

Qy        504 TATACACATATATACACACATATATACACATATATACACATATATATACACATATATACA 563
              | || | ||||||| |||  |||| | | ||| | |  ||| |||| | | ||| | |||
Db       2156 TCTATAAATATATAGACATCTATAAATATATAAAGATGCATCTATAAATATATAAAGACA 2215

Qy        564 CATATATACATATATACACATATATATACATATATACACATATATACATATATACACATA 623
              |||| | ||||||| | | || |||| ||||||| | | | | |  |||| | | |||
Db       2216 TCTATAAATATATATAAAGACATCTATAAATATATATAAAGACACATCTATAAATATATA 2275

Qy        624 TATACATATATACACACATATATACACATACATATACACACACATAGATATACATATATA 683
              || | | || || | | ||||||| | |  ||| || | | | ||| | |  ||| ||||
Db       2276 TAAAGACATCTATAAATATATATAAAGACGCATCTATAAATATATATAAAGGCATCTATA 2335

Qy        684 TACACATATATATACGTATATATATGTATATATATA 719
               | | ||||| | | | || ||||  |||||||| |
Db       2336 AATATATATAAAGATGCATCTATAAATATATATAAA 2371


RESULT 100
US-10-311-455-1971/c
; Sequence 1971, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1971
;   LENGTH: 21537
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1971

  Query Match             5.1%;  Score 235.2;  DB 15;  Length 21537;
  Best Local Similarity   69.6%;  Pred. No. 1.2e-37;
  Matches  348;  Conservative    0;  Mismatches  148;  Indels    4;  Gaps    2;

Qy        197 AAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACA 256
              ||||   | | || |||||  | |   | | |   | |       |||| ||  ||| | |
Db      16071 AAATTAAAACACACAAAAACTATATAACTTATCCAAAAACACACTATAAATCCATATATA 16012

Qy        257 CACACACACATATATATATACACATATATATACACATATATACATATATACACATATATA 316
              | | | | |||||||||||| | ||||| | | | ||| || |||||||| | |||||
Db      16011 TATATATATATATATATATATATATATACACATATATAAATCCATATATATATATATACG 15952

Qy        317 CACATATATATACACATACATATGTACACATATATATACACATATGTATACACATATATA 376
              || |||||||   | ||| ||||  ||||| |||||||||||||| || ||||||||||
Db      15951 CATATATATATACGTATATATATATACACACACATATATACACATATATACACACATATATA 15892

Qy        377 CACATATATATACACACATATATACACATATATACACACATATATACACATATATACACATA 436
              |||||||| ||| | ||||||| | |||||||||||||| | ||| | ||||||||||| |
Db      15891 CACATATACACATATATATATATATATATACACACACACATATATATATATACACACA 15832

Qy        437 TATACACATATACACATATATACACATATATACACATATATACACATATATATA---ATATA 493
              | | | ||||| | | | | |||||||||||| |||||||||||| | | |||   |||||
Db      15831 CACATATATATATATACACACACACATATATATATATATACACACACACATATATATATA 15772

Qy        494 CACACATATATATACACATATATACACACATATATACACATATATACACATATATATACA 553
               | | ||| | | |||||||||||| |    ||  | | | ||||| | ||||| | | |

Db      15771 TATATATACACACACACATATATATAATATAATCGATATAAATATATATATATAAAAAAA 15712

Qy        554 CATATATACACATATATACATATATACACATATATATACATATATACACATATATACATA 613
               |||||   |  |||| ||           |||  || || | || ||   |    ||   |
Db      15711 AATATACCTATTTATAAACTCTCTATTCCATTAATTTATACATCTAATAAACAATTATTT 15652

Qy        614 TATACACATATATACATATATACACACATATATACACATACATATACACACACATAGATA 673
              |||| | ||||||| |||||||||||||||||||| | ||| ||||| | | ||||| |||
Db      15651 TATATATATATATATATATATACACACATATATATA-ATATATATATATATACATACATA 15593

Qy        674 TACATATATATACACATATA 693
               | | ||| ||| |||||||
Db      15592 CATACATACATATACATATA 15573


Search completed: April 22, 2005, 07:11:26
Job time : 2049.07 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 20, 2005, 22:52:45 ; Search time 1795.26 Seconds
(without alignments)
13936.065 Million cell updates/sec

Title: US-09-939-209A-3_COPY_1_4121
Perfect score: 4121
Sequence: 1 agttcaagaccagcctgagc.....tgtgagggcaaattgaagtc 4121

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5633728 seqs, 3035525691 residues

Total number of hits satisfying chosen parameters: 11267456

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq:*
19: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
20: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:*
21: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
22: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4121	100.0	20300	10	US-09-939-209A-3	Sequence 3, Appli
c 2	356.6	8.7	121600	18	US-10-723-860-1125	Sequence 1125, Ap
3	345.8	8.4	109730	19	US-10-741-600-17809	Sequence 17809, A
c 4	327.4	7.9	188971	17	US-10-235-192A-27	Sequence 27, Appl
5	323.6	7.9	163382	18	US-10-367-094-179	Sequence 179, App
c 6	322.2	7.8	26997	17	US-10-221-714A-474	Sequence 474, App
7	320.6	7.8	3310	13	US-10-027-632-113621	Sequence 113621,
8	320.6	7.8	3310	13	US-10-027-632-113622	Sequence 113622,
9	320.6	7.8	3310	17	US-10-027-632-113621	Sequence 113621,
10	320.6	7.8	3310	17	US-10-027-632-113622	Sequence 113622,
c 11	311.2	7.6	26997	17	US-10-221-714A-473	Sequence 473, App
12	308	7.5	1990	15	US-10-017-161-2383	Sequence 2383, Ap
13	308	7.5	1990	17	US-10-292-798-2025	Sequence 2025, Ap
14	300.6	7.3	1800	18	US-10-437-963-12666	Sequence 12666, A
15	296	7.2	15418	9	US-09-783-203-1	Sequence 1, Appli
16	296	7.2	15418	9	US-09-994-427A-1	Sequence 1, Appli
17	296	7.2	15418	10	US-09-995-419A-1	Sequence 1, Appli
18	296	7.2	15418	14	US-10-141-220-1	Sequence 1, Appli
19	296	7.2	15418	14	US-10-023-969-1	Sequence 1, Appli
20	296	7.2	15418	14	US-10-206-447-1	Sequence 1, Appli
21	296	7.2	15418	17	US-10-674-836-1	Sequence 1, Appli
22	296	7.2	15418	18	US-10-811-012-1	Sequence 1, Appli
c 23	294.8	7.2	1486	17	US-10-292-798-2005	Sequence 2005, Ap
24	293.6	7.1	27067	10	US-09-841-158-5	Sequence 5, Appli
25	293.6	7.1	27067	18	US-10-801-837-5	Sequence 5, Appli
c 26	293.2	7.1	1486	15	US-10-017-161-2361	Sequence 2361, Ap
c 27	291.6	7.1	1672	15	US-10-017-161-2381	Sequence 2381, Ap
c 28	291.6	7.1	1672	17	US-10-292-798-2023	Sequence 2023, Ap
c 29	273.8	6.6	81924	18	US-10-741-601-5636	Sequence 5636, Ap
30	267.2	6.5	14603	19	US-10-741-600-17799	Sequence 17799, A
c 31	266.4	6.5	6971	15	US-10-311-455-1210	Sequence 1210, Ap
c 32	265.6	6.4	263852	18	US-10-812-232-6	Sequence 6, Appli
c 33	263.6	6.4	2522	17	US-10-108-260A-543	Sequence 543, App
c 34	263	6.4	227968	18	US-10-723-860-1357	Sequence 1357, Ap
c 35	262.6	6.4	109730	19	US-10-741-600-17809	Sequence 17809, A
c 36	262	6.4	1800	18	US-10-437-963-12666	Sequence 12666, A
37	262	6.4	1980090	18	US-10-719-993-6815	Sequence 6815, Ap
38	262	6.4	1980090	19	US-10-741-600-17676	Sequence 17676, A
c 39	261.4	6.3	12069	17	US-10-257-166-11	Sequence 11, Appl
c 40	259.6	6.3	1980090	18	US-10-719-993-6815	Sequence 6815, Ap
c 41	259.6	6.3	1980090	19	US-10-741-600-17676	Sequence 17676, A
42	258.8	6.3	1489	15	US-10-017-161-2363	Sequence 2363, Ap
43	258.8	6.3	1489	17	US-10-292-798-2007	Sequence 2007, Ap
c 44	258.6	6.3	163382	18	US-10-367-094-179	Sequence 179, App
45	258.4	6.3	82938	10	US-09-818-657-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-09-939-209A-3
; Sequence 3, Application US/09939209A
; Publication No. US20030113721A1
; GENERAL INFORMATION:
; APPLICANT: LEVITT, PAT R.
; APPLICANT: MIRNICS, KAROLY
; APPLICANT: KODAVALI, MENKATA CHOWDARI
; APPLICANT: NIMGAONKAR, VISHWAJIT L.
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR FACILITATING THE DIAGNOSIS AND TREATMENT (
; TITLE OF INVENTION: SCHIZOPHRENIA
; FILE REFERENCE: 00-539-US
; CURRENT APPLICATION NUMBER: US/09/939,209A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: 60/228,021
; PRIOR FILING DATE: 2000-08-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 20300
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: A genomic sequence containing RGS4 nucleic acid sequence and sequ
; OTHER INFORMATION: ences upstream and downstream to the RGS4 nucleic acid sequence
US-09-939-209A-3

Query Match 100.0%; Score 4121; DB 10; Length 20300;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4121; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAATTA 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

```
Db          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAATTA 60

Qy         61 GACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAAT 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAAT 120

Qy        121 CACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAGC 180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAGC 180

Qy        181 CCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATA 240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATA 240

Qy        241 TATATATATATACACACACACACATATATATATACACATATATATACACATATATATACA 300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TATATATATATACACACACACACATATATATATACACATATATATACACATATATATACA 300

Qy        301 TATATACACATATATACACATATATATACACATACATATGTACACATATATATACACATA 360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TATATACACATATATACACATATATATACACATACATATGTACACATATATATACACATA 360

Qy        361 TGTATACACATATATACACATATATACACACATATATACACATATATACACACATATATA 420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGTATACACATATATACACATATATACACACATATATACACATATATACACACATATATA 420

Qy        421 CACATATATACACATATATACACATATACACATATATACACATATATACATATATACACA 480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CACATATATACACATATATACACATATACACATATATACACATATATACATATATACACA 480

Qy        481 TATATATAATATACACACATATATATACACATATATACACACATATATACACATATATAC 540
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TATATATAATATACACACATATATATACACATATATACACACATATATACACATATATAC 540

Qy        541 ACATATATATACACATATATACACATATATACATATATACACATATATATACATATATAC 600
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ACATATATATACACATATATACACATATATACATATATACACATATATATACATATATAC 600

Qy        601 ACATATATACATATATACACATATATACATATATACACACATATATACACATACATATAC 660
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ACATATATACATATATACACATATATACATATATACACACATATATACACATACATATAC 660

Qy        661 ACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATAT 720
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATAT 720

Qy        721 GCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTT 780
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTT 780

Qy        781 TGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATAC 840
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATAC 840

Qy        841 ATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATATATATA 900
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 ATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATATATATA 900

Qy        901 TAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTATAGAAGAATGG 960
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTATAGAAGAATGG 960

Qy        961 CTAAGATGCCAAAGCTCAGTGTATGTGGCAGAGGCATGGTGTAGGGTGTGTCCAGGTTCA 1020
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTAAGATGCCAAAGCTCAGTGTATGTGGCAGAGGCATGGTGTAGGGTGTGTCCAGGTTCA 1020

Qy       1021 TATATTGCATTAAGTGTGAGAACACCCTGGAGTATGAACCAAGAAAATGCAAAAGCCAGA 1080
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TATATTGCATTAAGTGTGAGAACACCCTGGAGTATGAACCAAGAAAATGCAAAAGCCAGA 1080

Qy       1081 AGTGATGGAGGAAATGAGACACAATAATGAAGATATTGAGAGGAGGGTGTGGGCCTAGAG 1140
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AGTGATGGAGGAAATGAGACACAATAATGAAGATATTGAGAGGAGGGTGTGGGCCTAGAG 1140

Qy       1141 TGAAGCTTTTCGTGCCAGTACTTCTTTTGAAGGCCCAGTTCTCTTCTCTCTCGGGGGCTC 1200
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 TGAAGCTTTTCGTGCCAGTACTTCTTTTGAAGGCCCAGTTCTCTTCTCTCTCGGGGGCTC 1200

Qy       1201 CTTCATCTCTCATAGAGTCCACAGCTTTTAAGGGCCAACACTTGAGGTCAGCCTGGCTCT 1260
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 CTTCATCTCTCATAGAGTCCACAGCTTTTAAGGGCCAACACTTGAGGTCAGCCTGGCTCT 1260

Qy       1261 CTCATTTGAGCTGGATAGAACATTTTAGAGCACCATCTATTCTTCAAGAGGAAGTTTAAA 1320
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CTCATTTGAGCTGGATAGAACATTTTAGAGCACCATCTATTCTTCAAGAGGAAGTTTAAA 1320

Qy       1321 AATAAAAGAACCTTGAAGAGGAAAAAATGTAGACATTCAATCTAACCTTTTCATTTTACT 1380
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 AATAAAAGAACCTTGAAGAGGAAAAAATGTAGACATTCAATCTAACCTTTTCATTTTACT 1380

Qy       1381 AGCCAAAGCTAAATAGAATGCAGGTTACCTGTTTTTCAGCCAGGCACCATCATTTCCTAA 1440
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 AGCCAAAGCTAAATAGAATGCAGGTTACCTGTTTTTCAGCCAGGCACCATCATTTCCTAA 1440

Qy       1441 TTGTTATAAAATTTATTATTATTGTTGTTATTATTATTATTTGCCATAAGAAGTTTCCCA 1500
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 TTGTTATAAAATTTATTATTATTGTTGTTATTATTATTATTTGCCATAAGAAGTTTCCCA 1500

Qy       1501 TATCCTTTTAGTATAACAAAAACACAATTCACAAGCATTATAAAACCCATGGTGTCTAAC 1560
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 TATCCTTTTAGTATAACAAAAACACAATTCACAAGCATTATAAAACCCATGGTGTCTAAC 1560

Qy       1561 TATTAAAAAAAATTAAGTGGAACACACTTGTCCCAGCTACTGGGGAGGCTGAGGAGGGAGG 1620
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 TATTAAAAAAAATTAAGTGGAACACACTTGTCCCAGCTACTGGGGAGGCTGAGGAGGGAGG 1620

Qy       1621 ATCACGTGATCCCAGGGGGTCAAGGTTATGGAGAGCTATGATTGTGCCACTGCACTCCAG 1680
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 ATCACGTGATCCCAGGGGGTCAAGGTTATGGAGAGCTATGATTGTGCCACTGCACTCCAG 1680

Qy       1681 CCTGGGTGACAGGGAAAGACCCTGTCTCTAAAATTTTTTTTAAAAAAACTAAACTGGTTT 1740
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 CCTGGGTGACAGGGAAAGACCCTGTCTCTAAAATTTTTTTTAAAAAAACTAAACTGGTTT 1740

Qy       1741 TATTACAGAGATTCTGGAGACAGCTACACATAAAAGGGTGGTATGCCTCATATTAGCTAC 1800
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 TATTACAGAGATTCTGGAGACAGCTACACATAAAAGGGTGGTATGCCTCATATTAGCTAC 1800

Qy       1801 CCAGGGAGGTGGAATGCCAACTTAGGTGGTGTCACCACTATTAAAAATGCCCCAAAGCAA 1860
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 CCAGGGAGGTGGAATGCCAACTTAGGTGGTGTCACCACTATTAAAAATGCCCCAAAGCAA 1860

Qy       1861 TCAAAACTGAGAACTTCCTGGGAGCTTAGCATTGTGCAAAAGCAGCACAAAACACTTAAA 1920
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 TCAAAACTGAGAACTTCCTGGGAGCTTAGCATTGTGCAAAAGCAGCACAAAACACTTAAA 1920

Qy       1921 CAATTCACAGTTGTGTTGGAATGGGAAGGCCTGGAAATATAAACCAAAGAGTATATTGTC 1980
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 CAATTCACAGTTGTGTTGGAATGGGAAGGCCTGGAAATATAAACCAAAGAGTATATTGTC 1980

Qy       1981 TAAATTGATAGAGATTACAATTGCCTGAAAGAAAAAGTTGACTTTTAACTAGAATGTTCA 2040
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 TAAATTGATAGAGATTACAATTGCCTGAAAGAAAAAGTTGACTTTTAACTAGAATGTTCA 2040

Qy       2041 GAGTAGGTTTACAGAAGAAGCTCTTAAACTGGGCTCCAGTGGATTTGTCAATGCTTTGGA 2100
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 GAGTAGGTTTACAGAAGAAGCTCTTAAACTGGGCTCCAGTGGATTTGTCAATGCTTTGGA 2100

Qy       2101 AGCTGGTGGGGTGGGAGGGTTGGAGGGGGCATAAAAAGTCATGTTGGTATGCTCTGCTCA 2160
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 AGCTGGTGGGGTGGGAGGGTTGGAGGGGGCATAAAAAGTCATGTTGGTATGCTCTGCTCA 2160

Qy       2161 AGTCTCCATTCTGTTTCCTTTTCCTCTTTTCAATGTCATGTCCCATTATTTCATTATGGG 2220
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 AGTCTCCATTCTGTTTCCTTTTCCTCTTTTCAATGTCATGTCCCATTATTTCATTATGGG 2220
```

```
Qy       2221 CTTCCCTTTATCCAGGATCAATATGCCACCTCTTGGTTGTCTTTTACCTACTTCTCCACC 2280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 CTTCCCTTTATCCAGGATCAATATGCCACCTCTTGGTTGTCTTTTACCTACTTCTCCACC 2280

Qy       2281 TCACTATGGAATCGTCCTTGGGTAGCTCCTGTGCTTGGGAACCTGCACGGGCACTTTTCT 2340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 TCACTATGGAATCGTCCTTGGGTAGCTCCTGTGCTTGGGAACCTGCACGGGCACTTTTCT 2340

Qy       2341 GATGTCTTGATTCCAGCTTTACTCCTAAAACTTAAATGCTGAGGGGCCAACACCATGGCA 2400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 GATGTCTTGATTCCAGCTTTACTCCTAAAACTTAAATGCTGAGGGGCCAACACCATGGCA 2400

Qy       2401 GTGGTAGGGATGGGAATGGGGGTCTTGTAACACACTACATAAACTACACGAAATAAACTA 2460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 GTGGTAGGGATGGGAATGGGGGTCTTGTAACACACTACATAAACTACACGAAATAAACTA 2460

Qy       2461 CATGAAACTCAACATGTTTGCAAGACTCAGTTCACATCCATGAGGAGCTCATGCTTCTCC 2520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2461 CATGAAACTCAACATGTTTGCAAGACTCAGTTCACATCCATGAGGAGCTCATGCTTCTCC 2520

Qy       2521 CTCCTGCTCCCCTAGCACACATGATTATCTCTATTTGGAAATGTTTGGCATTTTTGGTGA 2580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2521 CTCCTGCTCCCCTAGCACACATGATTATCTCTATTTGGAAATGTTTGGCATTTTTGGTGA 2580

Qy       2581 AGTGAATGGTTCAATAACTTTCTCCACCATCAGAACAAAAGCTCTTTAAGGTTAGGGATG 2640
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2581 AGTGAATGGTTCAATAACTTTCTCCACCATCAGAACAAAAGCTCTTTAAGGTTAGGGATG 2640

Qy       2641 GGATCATACACACTTCCCTTGTCCAAGTCCCCATCACCCCTTATCTAGACAATTGCTACA 2700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2641 GGATCATACACACTTCCCTTGTCCAAGTCCCCATCACCCCTTATCTAGACAATTGCTACA 2700

Qy       2701 GTTTCCTACACACTCTTCTAACCTCTTGCAGTCTATTTTCATAAAACAGCTAGAGAACTT 2760
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2701 GTTTCCTACACACTCTTCTAACCTCTTGCAGTCTATTTTCATAAAACAGCTAGAGAACTT 2760

Qy       2761 TGAGATGTAAGTCAAAAAATAGAACATGTCGCTCTTTCCCATTGTTTTTGAAATAAAGTT 2820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2761 TGAGATGTAAGTCAAAAAATAGAACATGTCGCTCTTTCCCATTGTTTTTGAAATAAAGTT 2820

Qy       2821 CAACCCCCTTACCAGGGTCAACAAGGCCCTGCAATGATTTGGTCCTGTTAAAAATTCTTT 2880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2821 CAACCCCCTTACCAGGGTCAACAAGGCCCTGCAATGATTTGGTCCTGTTAAAAATTCTTT 2880

Qy       2881 AGCCTTAACTCATGCTGTTCTTCCTTACACTCACTGCATTCTAGCCATTGAGGTTTCTAT 2940
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2881 AGCCTTAACTCATGCTGTTCTTCCTTACACTCACTGCATTCTAGCCATTGAGGTTTCTAT 2940

Qy       2941 GCATCAAACTTTTTTTGGTCCCAGCACTGTGCACATCCTTCTGGGTAGAATGCCCCTTGA 3000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2941 GCATCAAACTTTTTTTGGTCCCAGCACTGTGCACATCCTTCTGGGTAGAATGCCCCTTGA 3000

Qy       3001 TTTGTATAATTAGCACCTCCTTCATCATTTAGGTCTTAGTATAACTACTACCTTCTTAGA 3060
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3001 TTTGTATAATTAGCACCTCCTTCATCATTTAGGTCTTAGTATAACTACTACCTTCTTAGA 3060

Qy       3061 GAAGCTCTGCTTCTTCATCCTATAAAAAAGTAAAATTCCTTACCCTGTTATTTTTTAAGT 3120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3061 GAAGCTCTGCTTCTTCATCCTATAAAAAAGTAAAATTCCTTACCCTGTTATTTTTTAAGT 3120

Qy       3121 CATCCGTGTTTCATTCTGTTAAAGTTCTTATCACAATTTATCATTATTTTATTTACAGTC 3180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3121 CATCCGTGTTTCATTCTGTTAAAGTTCTTATCACAATTTATCATTATTTTATTTACAGTC 3180

Qy       3181 ATGTGCCACATAACAATGTTTCAGTCAGGGATAGAACACAAATGTATCTGGCCCCATAAT 3240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3181 ATGTGCCACATAACAATGTTTCAGTCAGGGATAGAACACAAATGTATCTGGCCCCATAAT 3240

Qy       3241 ATTATAAGCTGAGAAATTTCTATTAACTAGTGATATCGCAGCCATCATAAGTGTAATGCA 3300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3241 ATTATAAGCTGAGAAATTTCTATTAACTAGTGATATCGCAGCCATCATAAGTGTAATGCA 3300

Qy       3301 GGACATTACCTTTTCTATGTTTAGATATGTTAGATACACAAATATATTTCATTGTGTTAT 3360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3301 GGACATTACCTTTTCTATGTTTAGATATGTTAGATACACAAATATATTTCATTGTGTTAT 3360

Qy       3361 AATTTCCTACAGTATTCAGTACAGTAACATGCTGTACAGGTTTGTAACCTAGGAGTAATA 3420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3361 AATTTCCTACAGTATTCAGTACAGTAACATGCTGTACAGGTTTGTAACCTAGGAGTAATA 3420

Qy       3421 GGCTATACCATACAGCTTAGGTGTGTAGTAGGCTATAACCATCTAGGTTTGTGTAAGTAC 3480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3421 GGCTATACCATACAGCTTAGGTGTGTAGTAGGCTATAACCATCTAGGTTTGTGTAAGTAC 3480

Qy       3481 ATTCTATGATATTCCCACAATGATGAAATCACCTAACTACACATTTCTCAGAATGTTTCA 3540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3481 ATTCTATGATATTCCCACAATGATGAAATCACCTAACTACACATTTCTCAGAATGTTTCA 3540

Qy       3541 CTGTTGTGAAGTGACCCATGACTATATTTTCCTATATACTTGATATTTTTGTGCATCTGC 3600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3541 CTGTTGTGAAGTGACCCATGACTATATTTTCCTATATACTTGATATTTTTGTGCATCTGC 3600

Qy       3601 CCATGAGAATGTAGTGTAAGATCAAAGGATGCAAGAATGGGTTCTATCCAGTATAGTACC 3660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3601 CCATGAGAATGTAGTGTAAGATCAAAGGATGCAAGAATGGGTTCTATCCAGTATAGTACC 3660

Qy       3661 CACTACACTGGTGGATGTCAATATGTATTTGTTAGATTAATATCTCAAGAATGAGCACCT 3720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3661 CACTACACTGGTGGATGTCAATATGTATTTGTTAGATTAATATCTCAAGAATGAGCACCT 3720

Qy       3721 TTCTCAGACACATAAAAGATGCTCAATATAAAAGTTTGTTGAACTGAACGTTATTGGCAA 3780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3721 TTCTCAGACACATAAAAGATGCTCAATATAAAAGTTTGTTGAACTGAACGTTATTGGCAA 3780

Qy       3781 ATGTAACATGATCGGATTTAAAGAGGAGCGAAACAGAGGTCTGGCTCAAACACCATACTT 3840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3781 ATGTAACATGATCGGATTTAAAGAGGAGCGAAACAGAGGTCTGGCTCAAACACCATACTT 3840

Qy       3841 CTAGAGTGCATAAGAGGTAGCAGTTGATTACCACTGGCGACAGGAGAAAAAAGAGCTTGA 3900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3841 CTAGAGTGCATAAGAGGTAGCAGTTGATTACCACTGGCGACAGGAGAAAAAAGAGCTTGA 3900

Qy       3901 CCGCAGGGTACTGTGAAGACATTTCAGGTTGATGGCACAGAACAGGGGAAATACATAAAT 3960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3901 CCGCAGGGTACTGTGAAGACATTTCAGGTTGATGGCACAGAACAGGGGAAATACATAAAT 3960

Qy       3961 GTGTGGGAATATTCAGTGGTCTGGGATGACTACATAGTAGAATATAATGAAGAAAAGAGT 4020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3961 GTGTGGGAATATTCAGTGGTCTGGGATGACTACATAGTAGAATATAATGAAGAAAAGAGT 4020

Qy       4021 GGAAGGGAAAGATGAAAAGTTGGAATGGGGATGAATTATGAAAGTACCAGAATGTTATGC 4080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4021 GGAAGGGAAAGATGAAAAGTTGGAATGGGGATGAATTATGAAAGTACCAGAATGTTATGC 4080

Qy       4081 TAAGGAATCTAGATTTTAAAATGTGAGGGCAAATTGAAGTC 4121
             |||||||||||||||||||||||||||||||||||||||||
Db       4081 TAAGGAATCTAGATTTTAAAATGTGAGGGCAAATTGAAGTC 4121
```

RESULT 2
US-10-723-860-1125/c
; Sequence 1125, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2

```
; SEQ ID NO 1125
;   LENGTH: 121600
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-723-860-1125

  Query Match             8.7%;  Score 356.6;  DB 18;  Length 121600;
  Best Local Similarity   74.7%;  Pred. No. 7.8e-62;
  Matches  543;  Conservative    0;  Mismatches  169;  Indels   15;  Gaps    7;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAA--- 57
              || || |||||| ||||    ||||  |||||||||| ||||||||||||||| |||||
Db      44776 AGATCGAGACCATCCTGGCTAACACAGTGAAACCCCGTCTCTACTAAAAATCCAAAAAAA 44717

Qy         58 --TTAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGG 115
                |||| | ||| |||||     | |||||| ||||||||||  |||||| ||||||||
Db      44716 AGTTAGCCGGGCGTGGTGGCGGGCACCTGTAGTCCCAGCTACTCTGGAGGCTGAGGCAGG 44657

Qy        116 AGAATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCAC 174
              ||||| | ||||| ||  || || |||| |||| | ||| | |||   |||| |||||
Db      44656 AGAATGGCGTGAACTTG--GGAGGCGGAGCTTGCAGTGAGCCGAGATAGCGCCACTGCAC 44599

Qy        175 TCCAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATA 234
              ||||| |  ||| || ||||||| |   ||||||| |||||||| |       | ||||||
Db      44598 TCCAGTCTGGGCGAC-AGAGCGAGACTCCATCTCAAAAAAAAAATA----TATATATATA 44544

Qy        235 TATATATATATATATATACACACACACACACATATATATATACACATATATATACACATA 294
              || ||||||||| |||||| | | | | | ||||||| ||| | | ||||||| | |||
Db      44543 TAGATATATATAGATATACAGATATATAGATATATATAGATATATAGATATATATAGATA 44484

Qy        295 TATACATATATACACATATATACACATATATATACACATACATATGTACACATATATATA 354
              |||| ||||||| | ||||||| | ||||||| | | ||| |||| || | |||||||
Db      44483 TATAGATATATATAGATATATAGATATATATAGATATATAGATATATATAGATATATAGA 44424

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              | ||||  ||| | | ||||| | | ||||| | | ||||| | | ||||| | | | |
Db      44423 TATATATAGATATATAGATATATATAGATATATAGATATATAGATATATATAGATATATA 44364

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              |||||  | | |||||  | |||||||  | |||||  | |||||||  | |||||||  |||||
Db      44363 GATATATATAGATATATAGATATATATAGATATATAGATATATAGATATATATAGATATA 44304

Qy        475 TACACATATATATAATATACACACATATATATACACATATATACACACATATATACACAT 534
              || |  |||| || ||||| | | ||||| ||| | ||| ||| | | |||||||  | ||
Db      44303 TAGATGTATAGAT-ATATATAGATATATAGATATATATAGATATATAGATATATAGATAT 44245

Qy        535 ATATACACATATATATACACATATATACACATATATACATATATACACATATATATACAT 594
              |||||  | |||||  |||  | |||  |||  | |||||  | |||||  | |  |||  |||||  |
Db      44244 ATATAGATATATAGATATATATAGATATATATATAGATATATAGATATATAGATATATAG 44185

Qy        595 ATATACACATATATACATATATACA-CATATATACATATATACACACATATATACACATA 653
              |||||  | |||||||  |||||||  |  |||||||  |||||||  |  |  |||||  |  |  |||
Db      44184 ATATATAGATATATAGATATATAGATGATATATAGATATATATAGATATATAGATATATA 44125

Qy        654 CATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATA 713
              |||||  |  |  |  |||||||||  |||  |||||  |  |||||||  |   ||||  ||||
Db      44124 GATATATAGATATATAGATATATATAGATATATAGATATATAGATATATAGATATATATA 44065

Qy        714 TATATAT 720
              |||||||
Db      44064 GATATAT 44058


RESULT 3
US-10-741-600-17809
; Sequence 17809, Application US/10741600
; Publication No. US20050026169A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001499
;  CURRENT APPLICATION NUMBER: US/10/741,600
;  CURRENT FILING DATE:  2003-12-22
;  NUMBER OF SEQ ID NOS: 73997
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17809
;   LENGTH: 109730
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-741-600-17809

  Query Match             8.4%;  Score 345.8;  DB 19;  Length 109730;
  Best Local Similarity   73.7%;  Pred. No. 1.2e-59;
  Matches  537;  Conservative    0;  Mismatches  177;  Indels   15;  Gaps    7;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAATTA 60
              ||||  |||||||||||| |||||||| |||||||| |||| ||||||||||||||||||
Db     107515 AGTTTGAGACCAGCCTGACCAACATGGAGAAACCCCGTCTCCACTAAAAATACAAAATTA 107574

Qy         61 GACAGGCATGGTGATACACGCCTGT-AATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              |   |||||||||   ||  ||||| ||||||||||||||| |||||| ||||||||||||
Db     107575 GCGGGGCATGGTGGCGCATCCCTGTAAATCCCAGCTACTTGGGAGGCTGAGGCAGGAGAA 107634

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCCA 178
              |||||||||  | | || |||||||||||| || ||| | ||||   |||||| |||||||
Db     107635 TCACTTGAA--TCCAGGAGGTGGAGGTTGTGGTGAGCCGAGATGGAGCCATTACACTCCA 107692

Qy        179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAA--------AAGGCATTTTA 230
              |||  |||||||||||| |||   |||||| ||||| | |          | | || |||
Db     107693 GCCTGGGCAACAAGAGCAAAACTCCATCTCAAAAAAATATATATATATGTATGTATATTA 107752

Qy        231 TATATATATATATATATATACACACACACACACATATATATATACACATATATATACA 290
              |||| ||||| ||||||||||  | | | | | | |||| |||||||| ||| ||||| |
Db     107753 TATAAATATAAATATATATATTTATATAAATATATATATTTATATACATATAAATATAAA 107812

Qy        291 CATATATACATATATACACATATATACACATATATATACACATACATATGTACACATATA 350
              |||||||  |||||||| |||||||   |||| |||| ||||  ||| || | |||||
Db     107813 TATATATATTTATATACATATATATATTTATATTTATATACATATGTATATAAATATATA 107872

Qy        351 TATACACATATGTATACACATATATACACATATATACACACATATATACACATATATACA 410
              |||  |  ||| || | | ||||||  | || |||| | | |||||||   || |||| |
Db     107873 TATTTATGTATATAAATATATATATTTATATGTATATAAATATATATATTTATGTATATA 107932

Qy        411 CACATATATACACATATATACACATATATACACATATACACATATATACACATATATACA 470
              | |||||||  |||| || | | ||||| | || || |  ||||||  | |||||
Db     107933 AATATATATATTTATATGTATATAAATATATATATCTATATGTATATAAATATATATTTT 107992

Qy        471 TATATACACATATATATAATATACACACATATATATACACATATATACACACATATATAC 530
              |||||  | ||| |||| ||||    | || ||||||  | |||  ||    | || ||||
Db     107993 TATATGTATATAAATAT-ATATTTTTATATGTATATAAATATAGATTTTTATATGTATAT 108051

Qy        531 ACATATATACACATATATATACACATATATACACATATATACATATATACACATATATAT 590
              | |||||||   |||| |||| | |||||||   |||| || ||| ||| | || |||||
Db     108052 AAATATATATTTATATGTATATAAATATATATTTATATGTATATAAATATATATTTATAT 108111

Qy        591 ACATATATACACATAT-ATACATATATACACATATATACATATATACACACATATATACA 649
              |||||  | | |||| | | |||| || | ||| ||| |||||||     ||||||  |
Db     108112 GTATATAAATATATATAACATATATTTATATATAAATATATATATATTTTTATATATGTA 108171

Qy        650 CATACATATACACACACATAGATATACATATATATACACATATATAT-ACGTATATATAT 708
              ||| |||||  | |   |||  ||||  | |||||||  | ||||||| || |||||  |||
Db     108172 TATAAATATATATATTTATATGTATATAAATATATATAAATATATATAACATATATTTAT 108231

Qy        709 GTATATATA 717
              |||||||||
Db     108232 GTATATATA 108240


RESULT 4
US-10-235-192A-27/c
; Sequence 27, Application US/10235192A
; Publication No. US20040043389A1
; GENERAL INFORMATION:
;  APPLICANT: McCarthy, Jeanette
;  TITLE OF INVENTION: Methods and Compositions for Identifying
```

```
;  TITLE OF INVENTION:  Risk Factors for Abnormal Lipid Levels and the Diseases
;  TITLE OF INVENTION:  and Disorders Associated Therewith
;  FILE REFERENCE: MMI-011
;  CURRENT APPLICATION NUMBER: US/10/235,192A
;  CURRENT FILING DATE:  2002-09-04
;  NUMBER OF SEQ ID NOS: 49
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 27
;   LENGTH: 188971
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 145329-145428, 187896-187995, 195894-195993
;   OTHER INFORMATION: N = any nucleotide
US-10-235-192A-27

  Query Match             7.9%;  Score 327.4;  DB 17;  Length 188971;
  Best Local Similarity   67.6%;  Pred. No. 8.9e-56;
  Matches  601;  Conservative    0;  Mismatches  261;  Indels   27;  Gaps    9;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC--AAAAT 58
              || ||||||||| ||||  |||||||||||||| || ||||||||||||||||  |||||
Db     104585 AGATCAAGACCATCCTGGCCAACATGGTGAAACTCCGTCTCTACTAAAAATACAAAAAAT 104526

Qy         59 TAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGA 118
              |||   ||| ||||| |   | ||||||||||||||||||| |||||| || ||||||||
Db     104525 TAGCTGGGCGTGGTGGTGGGCTCCTGTAATCCCAGCTACTTGGGAGGCTGAAGCAGGAGA 104466

Qy        119 ATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTG---CAC 174
              ||| ||||||||  | || || ||||||||| | ||| |||||||  |||||||    |||
Db     104465 ATCGCTTGAACC--CAGGAGGCGGAGGTTGCAGTGAGCCAAGATCGCGCCATTGCACCAC 104408

Qy        175 TCCAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATA 234
              |||||||  ||||| ||||| ||||     ||||| ||||||||||    |     | | |
Db     104407 TCCAGCCTGGGCAAAAAGAGAGAAACTCTGTCTCAAAAAAAAAAAAAAAAAAAAAAAAAA 104348

Qy        235 TATATATATATATATATACACACACACACATATATATATACACATATATATACACATA 294
              |||||||||||||||| | | | | ||||| | ||||||| |||||||        |
Db     104347 AATATATATATATATATATATATATACACACACATATATATTCATATATA-------A 104295

Qy        295 TATACATATATACACATATATACACATATATATACACATACATATGTACACATATATATA 354
              |  | |||||||  |||||||   ||  ||||| | || |||||  ||  ||||||||||
Db     104294 TTCATATATATATTCATATATA---ATTCATATATATATTCATATATATTCATATATATA 104238

Qy        355 CACATATGT-ATACACATATATACACATATATACACACATATATACACATATATACACAC 413
              |||||  |  ||  ||  |||||||  ||||||||  ||  |||||||  ||||||||  |  |
Db     104237 TTCATATATAATTCATATATATATTCATATATATTCATATATATATTCATATATATATTC 104178

Qy        414 ATATATACACATATATACACATATATACACATATACACATATATACACATATATACACATAT 473
              |||||||      ||  |  |  ||||||  ||  |||||  ||||||||  |   |||||  |||
Db     104177 ATATATA-----ATTCATATATATATTCATATATATTCATATATATATTCATATATATAT 104123

Qy        474 ATACACATATATATAATATACACACATATATATACACATATATACACACATATATACACA 533
              |  |  |||   |   |||||  |  |||||||||  ||  |||||||  ||  ||||||  ||  |
Db     104122 TCATATATAATTCATATATATATTCATATATATTCATATATATATTCATATATATTCATA 104063

Qy        534 TATATACACATATATATACACATATATACACATATATACATATATACACATATATATACA 593
              ||||||  ||||||||||  ||||||||  ||||||||  ||  |||  |  ||||  |||  |
Db     104062 TATATATTCATATATATATTCATATATATTCATATATATATTCATATATATATTCATATA 104003

Qy        594 TATATACACATATATACATATATACACATATATACATATATACACACATATATACACATA 653
              |||||  ||  |||||||   |||||  |  ||  |||  ||||||  ||  |  ||||  |  |  |||
Db     104002 TATATTCATATATATATTCATATATATTCATATATATATTCATATATATTCATATATA 103943

Qy        654 CATATACACACACATAGATATACA---TATATATACACATATATATACGTATATATATGT 710
              ||  |  |  |  |  ||  |||||  |     |||||||  |   ||||||   |||||
Db     103942 TATTCATATATATATTCATATATAATTCATATATATATTCATATATATATTCATATATAA 103883

Qy        711 ATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGA 770
              ||  ||||||  |  ||  |  |      |  |||  ||  ||    |  |    |||||
Db     103882 ATTCATATATAATTCATATATATATTCATATAGATATTCATATATAATTCATATAGATAT 103823

Qy        771 AAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTA 830
              |  |||  ||  |  |     |  ||    ||  |   |  ||  |      |     ||  |||  |
Db     103822 TCATAGATATTCATATATATTCATATAGATATTCATATATTCATATATATTCATATATTC 103763

Qy        831 GGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGA 879
              |  |  ||  ||||  |  |  |     |  |    ||  |   |   |    ||  |  |
Db     103762 ATGTATATTCATATAGATATTCATATATAGATATTCATAGATATTCATA 103714


RESULT 5
US-10-367-094-179
; Sequence 179, Application US/10367094
; Publication No. US20040170982A1
; GENERAL INFORMATION:
;  APPLICANT: David W. Morris
;  APPLICANT:  Marc Malandro
;  TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
;  FILE REFERENCE: 529452001500
;  CURRENT APPLICATION NUMBER: US/10/367,094
;  CURRENT FILING DATE:  2003-02-14
;  NUMBER OF SEQ ID NOS: 203
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 179
;   LENGTH: 163382
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(163382)
;   OTHER INFORMATION: n = A,T,C or G
US-10-367-094-179

  Query Match             7.9%;  Score 323.6;  DB 18;  Length 163382;
  Best Local Similarity   68.2%;  Pred. No. 5e-55;
  Matches  495;  Conservative    0;  Mismatches  224;  Indels    7;  Gaps    3;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              |||||||||||||||||  |||||||||||||||||||||||||||||||||| ||||||
Db     144405 AGTTCAAGACCAGCCTGGCCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAAATT 144464

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              || |||||  |||||  ||  ||||||||||||||||||   |||||| ||||||||||||
Db     144465 AGCCAGGCGTGGTGGCACGTGCCTGTAATCCCAGCTACCCAGGAGGCTGAGGCAGGAGAA 144524

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              |  ||||||  |     | ||  | ||| |    | |    | ||||||||||||||||||||
Db     144525 TTGCTTGAAGC-CAAGAGGCAGAAGTTGCAGTGAGCCGAGATCATGCCATTGCACTCCAG 144583

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  ||  |||||| ||| |  |||| ||| ||||||||||      |  |   |  |
Db     144584 CCTGGGGGACAAGAACGAGACTTCATGTCAAAAAAAAAAAATACACAGTCTTCCTTGGCT 144643

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
                  | | | | | | | || ||| |||| | |||| |||| | || |||
Db     144644 TAGCTGTATGGGAAATATATATATATTTATTTATATATATATTTATATATATTTATTT 144703

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACAT 359
              |||||||   || |||| | || ||| || | |||  ||| || | || ||| || | |
Db     144704 ATATATATTTATTTATATATATTTATTTATATATATTTATTTATATATTTATTTATATTT 144763

Qy        360 ATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATAT 419
              || | || | || |||  | ||||||    | ||||  ||   ||||||  |   ||||||
Db     144764 ATTTTTATATATTTATTTATATATATTTGTATATATTTATTTATATATTTATTTATATAT 144823

Qy        420 ACACATATATACACATATATACACATATACACATATATACACATATATACATA-----TA 474
              |  ||||||   |||||||     |  | | |||| || | || |||| |||     ||
Db     144824 TTATTTATATATTTATATATATTAGTTCATATATATTTATATATTTATATATATTAGTTA 144883

Qy        475 TACACATATATATAATATACACACATATATATACACATATATACACACATATATACACAT 534
              || | || ||||||  ||   | | |||| |||| | || | || | |  ||| || | ||
Db     144884 TATATATTTATATATTAGTTATATATATTTATATATATTTTTATATATTTATTTATATAT 144943

Qy        535 ATATACACATATATATACACATATATACACATATATACATATATACACATATATATACAT 594
```

```
Db     144944 ATGTATATATTAACATATTTATATATATTCATATATTTATATATAAATATATATATTAAT 145003

Qy        595 ATATACACATATATACATATATACACATATATACATATATACACACATATATACACATAC 654

Db     145004 ATATTTATATATATTTATATATTAATATATGTTTATATATTTATACATATTTATGTATAT 145063

Qy        655 ATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATAT 714

Db     145064 TAATATATATATTTATATATATTTATATATGTATTTATATATATTTATATATTTATATTT 145123

Qy        715 ATATAT 720

Db     145124 ATTTAT 145129


RESULT 6
US-10-221-714A-474/c
; Sequence 474, Application US/10221714A
; Publication No. US20040048254A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with
;  TITLE OF INVENTION:  tumor suppressor genes and oncogenes
;  FILE REFERENCE: 5013.1005
;  CURRENT APPLICATION NUMBER: US/10/221,714A
;  CURRENT FILING DATE:  2003-01-21
;  PRIOR APPLICATION NUMBER: PCT/EP01/02955
;  PRIOR FILING DATE: 2001-03-15
;  PRIOR APPLICATION NUMBER: DE 10013847.0
;  PRIOR FILING DATE: 2000-03-15
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 540
; SEQ ID NO 474
;   LENGTH: 26997
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-714A-474

  Query Match             7.8%;  Score 322.2;  DB 17;  Length 26997;
  Best Local Similarity   66.0%;  Pred. No. 4.4e-55;
  Matches  498;  Conservative    0;  Mismatches  253;  Indels    4;  Gaps    2;

Qy        207 TCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATATACACACACACACA 266

Db       6221 TAAATAAAAAAAATATAATATATATATATATATATATATATATACACACATATATATA 6162

Qy        267 TATATATATACACATATATATACACATATATACATATATACACATATATACACATATATA 326

Db       6161 AACATATATACACACACATATACATATATACACACATATATACATATATACACATATATA 6102

Qy        327 TACACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATA 386

Db       6101 TATATATATACATATATATACATATATATACACATATATATATATACATATATATA 6042

Qy        387 CACACATATATACACATATATACACACATATATACACATATATACACATATATACACATA 446

Db       6041 CATATACATATACACATATATATATATATACATATATATACATATACATATACACATATA 5982

Qy        447 TACACATATATACACATATATACATATATACACATATATATA-ATATACACACATATATA 505

Db       5981 TATATATATACATATACATACATATATATACACATATATATATATATACATATACATACA 5922

Qy        506 TACACATATATACACACATATATACACATATATACACATATATATACACATATATACACA 565

Db       5921 TATATATACACATATATATATATATACATATATACACGTATATACATATATACACACATA 5862

Qy        566 TATATACATATATACACATATATATACATATATACACATATATACATATATACACA---T 622

Db       5861 TTTATATATACGTATATACATATATATATATATACATATATATATATATATACATATT 5802

Qy        623 ATATACATATATACACACATATATACACATACATATACACACACATAGATATACATATAT 682

Db       5801 ATATATATATATATATACATATATAAATATACATATATATATATATATACATATATACAT 5742

Qy        683 ATACACATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTA 742

Db       5741 ATATATATATATAAATATATATATATATATACGTATATATATATACACACACACACCATA 5682

Qy        743 GCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAA 802

Db       5681 AAATACTATACACCCATAAAAAAAAAAAAAAATAACATTTACAACAACTTAAATAAAATTA 5622

Qy        803 GGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTAT 862

Db       5621 AAAACCATTATTCTAACCGAAATAACTCAAAAATAAAAAACCAAATATTATATATTCTCA 5562

Qy        863 CTGTAAGGTAGTTAAGAAGGTAACACTATATATATATATAGTGAAAGCAGTGTAAACCTA 922

Db       5561 CTTATAACTAAAAACTAAACTATAAAAACACAAAAACATAAAAATAATACAATAAACTTT 5502

Qy        923 AAGGATGGGCCAAGGATTTAAATGTTATAGAAGAA 957

Db       5501 AAAAACTCTAAAAAAAACATAAAAATAAATAAAAA 5467


RESULT 7
US-10-027-632-113621
; Sequence 113621, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 113621
;   LENGTH: 3310
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-113621

  Query Match             7.8%;  Score 320.6;  DB 13;  Length 3310;
  Best Local Similarity   65.7%;  Pred. No. 3.8e-55;
  Matches  559;  Conservative    0;  Mismatches  284;  Indels    8;  Gaps    6;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACA-AAATT 59
```

```
Db       1488 AGTTCAAGTCTAGTCTGACCAATATGGTGAAACCCCGTCTCTACTGGAAATACAGAAATT 1547

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              || ||||  ||||  | | |||||||| ||| |||||||  |||||| ||||||||||||
Db       1548 AGCCAGGTGTGGTAGTGCTCGCCTGTAGTCCTAGCTACTCGGGAGGCTGAGGCAGGAGAA 1607

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCCA 178
              || ||||||||||  || ||  |||||||| | ||| | ||||| ||||  |||||||||
Db       1608 TCGCTTGAACCTG--GGAGGCAGAGGTTGCAGTGAGCCGAGATCCTGCCCCTGCACTCCA 1665

Qy        179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATA 238
              |||   |  |  ||||||| ||    ||||| ||||||    | | | | || |||||||
Db       1666 GCC--TGGCAACAGAGCGAGATTCTGTCTCAAAAAAAACCCCATGTA-TATACATATATA 1722

Qy        239 TATATATATATATACACACACACACATATATATATACACATATATATACACATATATA 298
              ||||||||| | | | | |    | |||||||| |||| ||| ||||||||  |||||||
Db       1723 CGTATATATACACATATATATGTATACATATATGTATATACACATATACGTATATATA 1782

Qy        299 CATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACA 358
              | |||| |||  |||| |  |||| ||| |  || |||   | | | | || ||   |
Db       1783 TGTGTATATACACGTATATATGTATACATATATGTATATACACATATACACATGTATGTA 1842

Qy        359 TATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATA 418
              ||  |||| |  |||||||  ||||||  | | ||||    |||||||||  |||||
Db       1843 TACATATATATGTATATACGTGTATATATGTATACATATGTGTATATATACACGTATATA 1902

Qy        419 TACACATATATACACATATATACACATATACACATATATACACATATATACACATATACA 478
              |||||||| || || | | || | |||     ||||| |||||  |||||||  || |
Db       1903 TACACATATGTATACGTGTGTATATATACGTGTATATACACACACGTATACATACGTATA 1962

Qy        479 CATATATAT-AATATACACACATATATATACACATATATACACACATATATACACATATA 537
              |||||  |  ||||| ||||| || |||||| || | || |    ||||| ||| | |
Db       1963 TATATACGTGTATATATACACACATGTATACATATGTGTATATGTGTATATATACACACA 2022

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATACATATA 597
              |  | | ||||||    |||||||||||||| || | ||| ||   |||||| ||| |||
Db       2023 TGTATACATATATGTGTATATATACACATATGTATACATACACGTGTATATACACACATA 2082

Qy        598 TACACATATATACATATATACACATATATACATATATACACACATATATACACATACATA 657
              |  | | ||| || | |||| | | | |||  |||| | | |  | |||||| |  ||
Db       2083 TGTATACATACACGTGTATATATACACATATGTATACATATATGTGTATACATACGCACG 2142

Qy        658 TACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATA 717
              || |||||  ||  |||||||||| | || | |||   | |  || | |||||||| |||
Db       2143 TATACACATGTATGTATACATATTTGTATATATACTCACATATACACATATGTATACATA 2202

Qy        718 TATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGA 777
              |||          | |         ||  |   | |   |  | | | |   | | | |
Db       2203 CATGTGTATATACACGCACACATACACATATATGTGCATATACGCACATATATACATATA 2262

Qy        778 GTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAA 837
              |  |||   || |||    | |   | ||        ||   ||  ||         | |
Db       2263 TGTACACATATGTATATATACACACATATATACATATATGTACACATGTACGTACACATA 2322

Qy        838 TACATAAATGT 848
              |||||| ||||
Db       2323 TACATATATGT 2333


RESULT 8
US-10-027-632-113622
; Sequence 113622, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 113622
;   LENGTH: 3310
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-113622

  Query Match             7.8%;  Score 320.6;  DB 13;  Length 3310;
  Best Local Similarity   65.7%;  Pred. No. 3.8e-55;
  Matches  559;  Conservative    0;  Mismatches  284;  Indels    8;  Gaps    6;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACA-AAATT 59
              |||||||| | || |||| ||| |||||||||||||| ||||||||  ||||||| |||||
Db       1488 AGTTCAAGTCTAGTCTGACCAATATGGTGAAACCCCGTCTCTACTGGAAATACAGAAATT 1547

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              || ||||  ||||  | | |||||||| ||| |||||||  |||||| ||||||||||||
Db       1548 AGCCAGGTGTGGTAGTGCTCGCCTGTAGTCCTAGCTACTCGGGAGGCTGAGGCAGGAGAA 1607

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCCA 178
              || ||||||||||  || ||  |||||||| | ||| | ||||| ||||  |||||||||
Db       1608 TCGCTTGAACCTG--GGAGGCAGAGGTTGCAGTGAGCCGAGATCCTGCCCCTGCACTCCA 1665

Qy        179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATA 238
              |||   |  |  ||||||| ||    ||||| ||||||    | | | | || |||||||
Db       1666 GCC--TGGCAACAGAGCGAGATTCTGTCTCAAAAAAAACCCCATGTA-TATACATATATA 1722

Qy        239 TATATATATATATACACACACACACATATATATATACACATATATATACACATATATA 298
              ||||||||| | | | | |    | |||||||| |||| ||| ||||||||  |||||||
Db       1723 CGTATATATACACATATATATGTATACATATATGTATATACACATATACGTATATATA 1782

Qy        299 CATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACA 358
              | |||| |||  |||| |  |||| ||| |  || |||   | | | | || ||   |
Db       1783 TGTGTATATACACGTATATATGTATACATATATGTATATACACATATACACATGTATGTA 1842

Qy        359 TATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATA 418
              ||  |||| |  |||||||  ||||||  | | ||||    |||||||||  |||||
Db       1843 TACATATATATGTATATACGTGTATATATGTATACATATGTGTATATATACACGTATATA 1902

Qy        419 TACACATATATACACATATATACACATATACACATATATACACATATATACATATATACA 478
              |||||||| || || | | || | |||     ||||| |||||  |||||||  || |
Db       1903 TACACATATGTATACGTGTGTATATATACGTGTATATACACACACGTATACATACGTATA 1962

Qy        479 CATATATAT-AATATACACACATATATATACACATATATACACACATATATACACATATA 537
              |||||  |  ||||| ||||| || |||||| || | || |    ||||| ||| | |
Db       1963 TATATACGTGTATATATACACACATGTATACATATGTGTATATGTGTATATATACACACA 2022

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATACATATA 597
              |  | | ||||||    |||||||||||||| || | ||| ||   |||||| ||| |||
Db       2023 TGTATACATATATGTGTATATATACACATATGTATACATACACGTGTATATACACACATA 2082

Qy        598 TACACATATATACATATATACACATATATACATATATACACACATATATACACATACATA 657
              |  | | ||| || | |||| | | | |||  |||| | | |  | |||||| |  ||
Db       2083 TGTATACATACACGTGTATATATACACATATGTATACATATATGTGTATACATACGCACG 2142

Qy        658 TACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATA 717
              || |||||  ||  |||||||||| | || | |||   | |  || | |||||||| |||
Db       2143 TATACACATGTATGTATACATATTTGTATATATACTCACATATACACATATGTATACATA 2202

Qy        718 TATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGA 777
              |||          | |         ||  |   | |   |  | | | |   | | | |
```

```
Db       2203 CATGTGTATATACACGCACACATACACATATATGTGCATATACGCACATATATACATATA 2262

Qy        778 GTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAA 837
              |  |||   || |||    | |   | ||        ||    ||  ||       | |
Db       2263 TGTACACATATGTATATATACACACATATATACATATATGTACACATGTACGTACACATA 2322

Qy        838 TACATAAATGT 848
              |||||| ||||
Db       2323 TACATATATGT 2333


RESULT 9
US-10-027-632-113621
; Sequence 113621, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 113621
;   LENGTH: 3310
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-113621

  Query Match             7.8%;  Score 320.6;  DB 17;  Length 3310;
  Best Local Similarity   65.7%;  Pred. No. 3.8e-55;
  Matches  559;  Conservative    0;  Mismatches  284;  Indels    8;  Gaps    6;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACA-AAATT 59
              |||||||| | || |||| ||| |||||||||||| ||||||||  |||||||  |||||
Db       1488 AGTTCAAGTCTAGTCTGACCAATATGGTGAAACCCCGTCTCTACTGGAAATACAGAAATT 1547

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              || ||||  ||||  | | |||||||| ||| |||||||  ||||||  ||||||||||||
Db       1548 AGCCAGGTGTGGTAGTGCTCGCCTGTAGTCCTAGCTACTCGGGAGGCTGAGGCAGGAGAA 1607

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCCA 178
              || ||||||||||  || ||  |||||||| | ||| | |||||  ||||  |||||||||
Db       1608 TCGCTTGAACCTG--GGAGGCAGAGGTTGCAGTGAGCCGAGATCCTGCCCCTGCACTCCA 1665

Qy        179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATA 238
              |||   |  |  ||||||| ||    ||||| ||||||    | | | | || |||||||
Db       1666 GCC--TGGCAACAGAGCGAGATTCTGTCTCAAAAAAAACCCCATGTA-TATACATATATA 1722

Qy        239 TATATATATATATACACACACACACATATATATATACACATATATATACACATATATA 298
              ||||||||  | | | | |    | ||||||| |||| ||| ||||||||  |||||||
Db       1723 CGTATATATACACATATATATGTATACATATATGTATATACACATATATACGTATATATA 1782

Qy        299 CATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACA 358
              | |||| |||  |||| |  |||| ||| |  || |||   | | | | || ||   |
Db       1783 TGTGTATATACACGTATATATGTATACATATATGTATATACACATATACACATGTATGTA 1842

Qy        359 TATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATA 418
```

```
              ||  |||| |  |||||||  ||||||   | | ||||    ||||||||||  |||||
Db       1843 TACATATATATGTATATACGTGTATATATGTATACATATGTGTATATATACACGTATATA 1902

Qy        419 TACACATATATACACATATATACACATATACACATATATACACATATATACATATATACA 478
              |||||||||| || || | | || | |||      |||||  |||||  ||||||||  || |
Db       1903 TACACATATGTATACGTGTGTATATATACGTGTATATACACACACGTATACATACGTATA 1962

Qy        479 CATATATAT-AATATACACACATATATATACACATATATACACACATATATACACATATA 537
              |||||  |  ||||| |||||  || |||||| ||  | ||  |     |||||| ||| | |
Db       1963 TATATACGTGTATATATACACACATGTATACATATGTGTATATGTGTATATATACACACA 2022

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATACATATA 597
              |  | | ||||||    |||||||||||||| || | ||| ||   ||||||  ||| |||
Db       2023 TGTATACATATATGTGTATATATACACATATGTATACATACACGTGTATATACACACATA 2082

Qy        598 TACACATATATACATATATACACATATATACATATATACACACATATATACACATACATA 657
              |  | | ||| || | |||| | | | |||  |||| | | |  | ||||||  |  ||
Db       2083 TGTATACATACACGTGTATATATACACATATGTATACATATATGTGTATACATACGCACG 2142

Qy        658 TACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATA 717
              || |||||  ||  |||||||| | || | |||   | |  || | |||||||||  |||
Db       2143 TATACACATGTATGTATACATATTTGTATATATACTCACATATACACATATGTATACATA 2202

Qy        718 TATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGA 777
              |||          | |          ||  |  | |   |  | | | |  | | | |
Db       2203 CATGTGTATATACACGCACACATACACATATATGTGCATATACGCACATATATACATATA 2262

Qy        778 GTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAA 837
              |  |||   || |||    | |   | ||        ||    ||  ||       | |
Db       2263 TGTACACATATGTATATATACACACATATATACATATATGTACACATGTACGTACACATA 2322

Qy        838 TACATAAATGT 848
              |||||| ||||
Db       2323 TACATATATGT 2333


RESULT 10
US-10-027-632-113622
; Sequence 113622, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 113622
;   LENGTH: 3310
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-113622

  Query Match             7.8%;  Score 320.6;  DB 17;  Length 3310;
  Best Local Similarity   65.7%;  Pred. No. 3.8e-55;
  Matches  559;  Conservative    0;  Mismatches  284;  Indels    8;  Gaps    6;
```

```
Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACA-AAATT 59
              |||||||| | || |||| ||| |||||||||||||| |||||||| ||||||| |||||
Db       1488 AGTTCAAGTCTAGTCTGACCAATATGGTGAAACCCCGTCTCTACTGGAAATACAGAAATT 1547

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              || ||||  ||||  | | |||||||| ||| ||||||| |||||| ||||||||||||
Db       1548 AGCCAGGTGTGGTAGTGCTCGCCTGTAGTCCTAGCTACTCGGGAGGCTGAGGCAGGAGAA 1607

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCCA 178
              || |||||||||  || ||  |||||||| | ||| | ||||| |||| |||||||||
Db       1608 TCGCTTGAACCTG--GGAGGCAGAGGTTGCAGTGAGCCGAGATCCTGCCCCTGCACTCCA 1665

Qy        179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATA 238
              |||   |  | |||||||  ||    ||||| ||||||    | | | | || |||||||
Db       1666 GCC--TGGCAACAGAGCGAGATTCTGTCTCAAAAAAAACCCCATGTA-TATACATATATA 1722

Qy        239 TATATATATATATACACACACACACATATATATATACACATATATATACACATATATA 298
              ||||||||  | | | | |   | ||||||| |||| ||| ||||||||  |||||||
Db       1723 CGTATATATACACATATATATGTATACATATATGTATATACACATATATACGTATATATA 1782

Qy        299 CATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACA 358
              | |||| |||  |||| |  |||| ||| |  || |||   | | | | | ||  ||   |
Db       1783 TGTGTATATACACGTATATATGTATACATATATGTATATACACATATACACATGTATGTA 1842

Qy        359 TATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATA 418
              ||  |||| |  |||||||   ||||||   | | ||||    |||||||||  |||||
Db       1843 TACATATATATGTATATACGTGTATATATGTATACATATGTGTATATATACACGTATATA 1902

Qy        419 TACACATATATACACATATATACACATATACACATATATACACATATATACATATATACA 478
              |||||||||  ||  ||  | |  | |||     |||||  |||||  ||||||||  || |
Db       1903 TACACATATGTATACGTGTGTATATATACGTGTATATACACACACGTATACATACGTATA 1962

Qy        479 CATATATAT-AATATACACACATATATATACACATATATACACACATATATACACATATA 537
              |||||  |  ||||| ||||| || |||||| || | || |    ||||||  ||| | |
Db       1963 TATATACGTGTATATATACACACATGTATACATATGTGTATATGTGTATATATACACACA 2022

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATACATATA 597
              |  | | ||||||    |||||||||||||| || | ||| ||   ||||||  ||| |||
Db       2023 TGTATACATATATGTGTATATATACACATATGTATACATACACGTGTATATATACACACATA 2082

Qy        598 TACACATATATACATATATACACATATATACATATATACACACATATATACACATACATA 657
              |  | | ||| || | |||| | | | |||  |||| | | |  | |||||| |  ||
Db       2083 TGTATACATACACGTGTATATATACACATATGTATACATATATGTGTATACATACGCACG 2142

Qy        658 TACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATA 717
              || |||||  ||  |||||||||| | || | |||   | |  || | ||||||||| |||
Db       2143 TATACACATGTATGTATACATATTTGTATATATACTCACATATACACATATGTATACATA 2202

Qy        718 TATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGA 777
              |||          |  |          ||  |  | |   |  | | | |  | | | |
Db       2203 CATGTGTATATACACGCACACATACACATATATGTGCATATACGCACATATATACATATA 2262

Qy        778 GTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAA 837
              |  |||   || |||    | |   | ||       ||    ||  ||       | |
Db       2263 TGTACACATATGTATATATACACACATATATACATATATGTACACATGTACGTACACATA 2322

Qy        838 TACATAAATGT 848
              |||||| ||||
Db       2323 TACATATATGT 2333


RESULT 11
US-10-221-714A-473/c
; Sequence 473, Application US/10221714A
; Publication No. US20040048254A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with
;  TITLE OF INVENTION:  tumor suppressor genes and oncogenes
;  FILE REFERENCE: 5013.1005
;  CURRENT APPLICATION NUMBER: US/10/221,714A
;  CURRENT FILING DATE:  2003-01-21
;  PRIOR APPLICATION NUMBER: PCT/EP01/02955
;  PRIOR FILING DATE: 2001-03-15
;  PRIOR APPLICATION NUMBER: DE 10013847.0
;  PRIOR FILING DATE: 2000-03-15
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 540
; SEQ ID NO 473
;   LENGTH: 26997
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-714A-473

  Query Match             7.6%;  Score 311.2;  DB 17;  Length 26997;
  Best Local Similarity   76.9%;  Pred. No. 7.9e-53;
  Matches  409;  Conservative    0;  Mismatches  113;  Indels   10;  Gaps    2;

Qy        203 CATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACA 262
              | | ||| ||| | | ||         | ||||||||||||||||| | | || || | | |
Db      21339 CTTTTCATAAATACATAATATTCCATAATATATATATATATATACACACGCATATATA 21280

Qy        263 CACATATATATATACACATATATATACACATATATACATATATACACATATATACACATA 322
              ||||||||   | ||| ||| ||||| | |||||||||||||| ||| ||||||| |  ||
Db      21279 CACATATACCCACACATATACATATATATATATATACATATACACATATATATATCTA 21220

Qy        323 TATATACACATACATATGT-----ACACATATATATACACATATGTATACACATATATAC 377
              |||||||||  | |||  |     | | ||| | ||| |||||| |||| ||| ||||||
Db      21219 CATATACACACATATACATATATAATATATACACATATACATATATATATACACATATAC 21160

Qy        378 ACATATATACACACATATATACACATATATACACACATATATACACATATATACACATAT 437
              ||||||| | || || ||||| |||||||||| | |||||||||||| |||| ||| |||
Db      21159 ACATATACATACGCACATATAAACATATATATATACATATATACACGTATACACATATAC 21100

Qy        438 ATACACATATACACATATATACACATATATACATATATACACATATATATA-----ATAT 492
              ||| ||| ||| | |||||||||  ||||| ||||||| | | ||| |||||     |||
Db      21099 ATATACACATATATATATATACATATATACACATATACATATATACATATATATATAC 21040

Qy        493 ACACACATATATATACACATATATACACACATATATACACATATATACACATATATATAC 552
              | | | | | ||||||| ||||| | | | ||||||| |||||||||||||||||| |||
Db      21039 ATATATACACATATACATATATACATATATATATATATACATATATACACATATACATAT 20980

Qy        553 ACATATATACACATATATACATATATACACATATATATACATATATACACATATATACAT 612
              | | ||||| | |||||||||||||||||||||||||| ||| ||| ||| | ||||||| |
Db      20979 ATACATATATATATATATACATATATACACATATACATATATACATATATATATATAC 20920

Qy        613 ATATACACATATATACATATATACACACATATATACACATACATATACACACACATAGAT 672
              ||||| ||| |||||||||||||| | ||||||| | ||| ||||| | | | ||| ||
Db      20919 ATATATACACATATACATATATACATATATATATATATATATATATATATATACAT 20860

Qy        673 ATACATATATATACACATATATATACGTATATATATGTATATATATATGCTC 724
              ||| |||||||||| |||||| ||  |||||||| |||||||||| | |
Db      20859 ATATATATATATACATATATATCTATATATATATATATATATATATACAC 20808


RESULT 12
US-10-017-161-2383
; Sequence 2383, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
```

```
;  FILE REFERENCE: 084335/0152
;  CURRENT APPLICATION NUMBER: US/10/017,161
;  CURRENT FILING DATE:  2002-12-18
;  PRIOR APPLICATION NUMBER: JP 2001/246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2430
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2383
;   LENGTH: 1990
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: source
;   LOCATION: (1)..(1990)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(1790)
US-10-017-161-2383

  Query Match             7.5%;  Score 308;  DB 15;  Length 1990;
  Best Local Similarity   80.2%;  Pred. No. 1.2e-52;
  Matches  401;  Conservative    0;  Mismatches   90;  Indels    9;  Gaps    3;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | || | | |||| |||||||| | | | | ||||||| || ||||||||||||||||||
Db        826 TATACACACATATGTATATATACACATATATACACACATATGTATATATACACATATATA 885

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              |||||||| || ||||| ||| ||||| |||||||| |||| | | |||||| ||||||
Db        886 CACACATATGTATATATACACATATATACACACATATGTATATATACACATATATACACA 945

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||| || | |||   ||| | | |  || |  |||||| ||||||||||||||||||| |
Db        946 CATATGTATATATACACATATATACACGTATATGTATATATACACATATATACACATACA 1005

Qy        407 TACACA----CATATATACACATATATACACATATATACACATATACACATATATACACA 462
              ||||||    ||||| |||||||||||||||| |||||||||||  ||||||||||
Db       1006 TACACATATGTATATACACACATATATACACATACATACACATATACGTATATATACACA 1065

Qy        463 TATATACATATATACACATATATATAATATACACACATATATATACACATATATACACAC 522
              |||| ||||| |||||||||||  | ||||| |||||||||| ||| ||   || | ||
Db       1066 TATACACATACATACACATATACGT-ATATATACACATATATACACATATGCGTATATAC 1124

Qy        523 ATATATACACATATATACACATATATATACACATATATACACATATATACATATATACAC 582
              |||||||||||||||||||    ||||||  |||||||||||||||||||| ||||||
Db       1125 ATATATACACATATATACAC----ATATACGTATATATACACATATATACACATATACGT 1180

Qy        583 ATATATATACATATATACACATATATACATATATACACATATATACATATATACACACAT 642
              ||||||| |||||||||||||||||||| |||||   ||||| |||||||||||| ||
Db       1181 ATATATACACATATATACACATATATACACATATATGTATATACACATATATACACATAT 1240

Qy        643 ATATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTAT 702
              |||| | | ||||||| ||||| |||  |||| ||| ||||| |||  ||| ||  |||
Db       1241 GTATATATACACATATATACACATATATGTATATATACATATATACACGTATGTATATAT 1300

Qy        703 ATATATGTATATATATATGC 722
              | | || ||| |||||||
Db       1301 ACACATATATGTATATATAC 1320


RESULT 13
US-10-292-798-2025
; Sequence 2025, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
;  FILE REFERENCE: 084335/166
;  CURRENT APPLICATION NUMBER: US/10/292,798
;  CURRENT FILING DATE:  2002-11-13
;  PRIOR APPLICATION NUMBER: 10/017,161
;  PRIOR FILING DATE: 2001-12-18
;  PRIOR APPLICATION NUMBER: JP 2001-246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2070
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2025
;   LENGTH: 1990
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   LOCATION: source
;   FEATURE:
;   LOCATION: (1)..(1990)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(1790)
US-10-292-798-2025

  Query Match             7.5%;  Score 308;  DB 17;  Length 1990;
  Best Local Similarity   80.2%;  Pred. No. 1.2e-52;
  Matches  401;  Conservative    0;  Mismatches   90;  Indels    9;  Gaps    3;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | || | | |||| |||||||| | | | | ||||||| || ||||||||||||||||||
Db        826 TATACACACATATGTATATATACACATATATACACACATATGTATATATACACATATATA 885

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              |||||||| || ||||| ||| ||||| |||||||| |||| | | |||||| ||||||
Db        886 CACACATATGTATATATACACATATATACACACATATGTATATATACACATATATACACA 945

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||| || | |||   ||| | | |  || |  |||||| ||||||||||||||||||| |
Db        946 CATATGTATATATACACATATATACACGTATATGTATATATACACATATATACACATACA 1005

Qy        407 TACACA----CATATATACACATATATACACATATATACACATATACACATATATACACA 462
              ||||||    ||||| |||||||||||||||| |||||||||||  ||||||||||
Db       1006 TACACATATGTATATACACACATATATACACATACATACACATATACGTATATATACACA 1065

Qy        463 TATATACATATATACACATATATATAATATACACACATATATATACACATATATACACAC 522
              |||| ||||| |||||||||||  | ||||| |||||||||| ||| ||   || | ||
Db       1066 TATACACATACATACACATATACGT-ATATATACACATATATACACATATGCGTATATAC 1124

Qy        523 ATATATACACATATATACACATATATATACACATATATACACATATATACATATATACAC 582
              |||||||||||||||||||    ||||||  |||||||||||||||||||| ||||||
Db       1125 ATATATACACATATATACAC----ATATACGTATATATACACATATATACACATATACGT 1180

Qy        583 ATATATATACATATATACACATATATACATATATACACATATATACATATATACACACAT 642
              ||||||| |||||||||||||||||||| |||||   ||||| |||||||||||| ||
Db       1181 ATATATACACATATATACACATATATACACATATATGTATATACACATATATACACATAT 1240

Qy        643 ATATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTAT 702
              |||| | | ||||||| ||||| |||  |||| ||| ||||| |||  ||| ||  |||
Db       1241 GTATATATACACATATATACACATATATGTATATATACATATATACACGTATGTATATAT 1300

Qy        703 ATATATGTATATATATATGC 722
              | | || ||| |||||||
Db       1301 ACACATATATGTATATATAC 1320


RESULT 14
US-10-437-963-12666
; Sequence 12666, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
```

```
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 12666
;   LENGTH: 1800
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_18773C.1
US-10-437-963-12666

  Query Match             7.3%;  Score 300.6;  DB 18;  Length 1800;
  Best Local Similarity   78.2%;  Pred. No. 3.6e-51;
  Matches  387;  Conservative    0;  Mismatches  104;  Indels    4;  Gaps    2;

Qy        229 TATATATATATATATATATATATACACACACACACATATATATATACACATATATATA 288
              |||||||||||| |||||||||| | | | ||| | ||| |||||||||| ||||| |||
Db       1298 TATATATATATACATATATATATATATATACATATATACATATATACATATATACATA 1357

Qy        289 CACATATATACATATATACACATATATAC---ACATATATATACACATACATATGTACAC 345
              ||||||| | ||||| ||||| ||||||    ||||||| ||| | ||||||||| ||
Db       1358 TACATATACATATATACACACACATATACATAACATATACATATATATACATATATATGT 1417

Qy        346 ATATATATACACATATGTATACACATATATACACATATATACACACATATATACACATAT 405
              ||||||| | | |||| || | | ||| ||| | || |||| | |||||||| ||||||
Db       1418 ATATATACATATATATATACATATATACATATATGTATATATACATATATATACATAT 1477

Qy        406 ATACACACATATATACACATATATACACATATATACACATATACACATATATACACATAT 465
              ||| ||  |||||||| |||||||  |||||||| ||||| |||||  || | |  |
Db       1478 ATATACGTATATATACATATATATACGTATATATACATATATATACATACGTATATACGT 1537

Qy        466 ATACATATATACACATATATATAATATACACACATATATATACACATATATACACACATA 525
              |||||||  || | ||| |||| ||||||| || |||||||||| |||||||| || ||
Db       1538 ATACATACGTATATATACATAT-ATATACATACGTATATATACATATATATACATACGTA 1596

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| |||||||| |||||||||| | ||||| | | ||||| | |||||| |
Db       1597 TATATACATATATATACATATATATATATATATACATATATATACATATATACATTTTCG 1656

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| |||||||   ||||||| |||| || | | |||||
Db       1657 TATATATATATATATATATATATATATATATATGTATATATATATATGTATATATATATA 1716

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | |   ||| ||||| |||| |||| | |||||| ||  ||||||
Db       1717 TATATATATATATATATATATGTATATATATATATATGTATATATATATATGTATATATATA 1776

Qy        706 TATGTATATATATAT 720
              ||||||||||||||||
Db       1777 TATGTATATATATAT 1791


RESULT 15
US-09-783-203-1
; Sequence 1, Application US/09783203
; Patent No. US20020098582A1
; GENERAL INFORMATION:
;  APPLICANT: Geron Corporation
;  APPLICANT:  Gold, Joseph
;  APPLICANT:  Lebkowski, Jane
;  TITLE OF INVENTION: Tpacked stem cells
;  FILE REFERENCE: 096/003
;  CURRENT APPLICATION NUMBER: US/09/783,203
;  CURRENT FILING DATE:  2001-02-13
;  PRIOR APPLICATION NUMBER: 60/253,443
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/253,357
;  PRIOR FILING DATE: 2000-11-27
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn version 3.0

; SEQ ID NO 1
;   LENGTH: 15418
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-783-203-1

  Query Match             7.2%;  Score 296;  DB 9;  Length 15418;
  Best Local Similarity   69.2%;  Pred. No. 7.9e-50;
  Matches  479;  Conservative    0;  Mismatches  200;  Indels   13;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||| |||||||||||  ||| || | |||||||||||| |||||||| || ||||||
Db        646 AGTTCGAGACCAGCCTGGCCAATATAGCGAAACCCCATCTGTACTAAAAACACAAAAATT 705

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   ||  |||||| ||| |||||||||||||| ||||  ||||||  ||||||||||
Db        706 AGCTGGGTGTGGTGACACATGCCTGTAATCCCAGGTACTCAGGAGGCTAAGGCAGGAGAA 765

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              |  ||||||  |  ||  | || |||     |  ||||||    ||| ||||||||||
Db        766 TTGCTTGAACTGGGAGGCAGAGGTGGT-----GAGCCAAGATTGCACCAGTGCACTCCAG 820

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  ||  || |||| ||||   ||||||| ||| ||| |||   | |  ||||| |||
Db        821 CCTTGGTGAC-AGAGTGAAACTCCATCTCAAAAACAAACAAACAAAATACATATACATAA 879

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              |||||||  | | | | ||| | ||| ||||||| ||| |||||||  |||||||
Db        880 ATATATATGCACATATATATACATATATAAATATATATACACATATATAAATCTATATAC 939

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACAT-----ATATATA 354
              ||||||||| ||||| ||| |||-|| ||  | |||||||| || ||||     ||||||
Db        940 ATATATACATATATACACATATATAAATCTATATACATATATATACATATATAATATATT 999

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              |||||| || | | ||| ||| | | ||||||| | ||| ||||| ||||| | | |||
Db       1000 TACATATATAAATATATACATATATAAATATACATATATAAATACATATATAAATATACA 1059

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              |||||| | ||| ||| | | ||||||| ||||| | ||||| | | ||| ||| | |||
Db       1060 TATATAAATATACATATATAAATATACATATATAAATATATACATATATAAATATACATA 1119

Qy        475 TACACATATATATA-ATATACACACATATATATACACATATATACACACATATATACACA 533
              || | |||||||| ||||| | | ||||| ||| |||  |||| ||| |||||||| |
Db       1120 TATAAATATATATACATATATAAATATATAAATATACAAGTATATACAAATATATACATA 1179

Qy        534 TATATACACATATATATACACATATATACACATATATACATATATACACATATATATACA 593
              |||| |   |||||  || | | ||||| | ||| ||| ||| | | || | |   |
Db       1180 TATAAATGTATATACGTATATACATATATATATAAATATATAAAAAAACTTTTGGCTGGG 1239

Qy        594 TATATACACATATATACATATATACACATATATACATATATACACACATATATACACATA 653
              | |   || || |       | | | | |     || | || ||| | | |||||
Db       1240 CACCTTTCCAAATCTCATGGCACATATAAGTCTCATGGTAACCTCAAATAAAAAAACATA 1299

Qy        654 CATATACACACACATAGATATACATATATATA 685
              |     || | ||| | |   ||| |
Db       1300 TAACAGATACACCAAAAATAAAAACCAATAAA 1331


RESULT 16
US-09-994-427A-1
; Sequence 1, Application US/09994427A
; Patent No. US20020128221A1
; GENERAL INFORMATION:
;  APPLICANT: Geron Corporation
;  APPLICANT:  Schiff, J. Michael
;  TITLE OF INVENTION: GLYCOSYLTRANSFERASE VECTORS FOR TREATING CANCER
;  FILE REFERENCE: 083,002
;  CURRENT APPLICATION NUMBER: US/09/994,427A
;  CURRENT FILING DATE:  2002-02-26
;  PRIOR APPLICATION NUMBER: 60/253,395
;  PRIOR FILING DATE: 2000-11-27
;  NUMBER OF SEQ ID NOS: 17
```

```
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 15418
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-994-427A-1

  Query Match             7.2%;  Score 296;  DB 9;  Length 15418;
  Best Local Similarity   69.2%;  Pred. No. 7.9e-50;
  Matches  479;  Conservative    0;  Mismatches  200;  Indels   13;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||| |||||||||||  ||| || | ||||||||||| |||||||| || ||||||
Db        646 AGTTCGAGACCAGCCTGGCCAATATAGCGAAACCCCATCTGTACTAAAAACACAAAAATT 705

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   ||  |||||| ||| |||||||||||||| ||||  ||||||  ||||||||||
Db        706 AGCTGGGTGTGGTGACACATGCCTGTAATCCCAGGTACTCAGGAGGCTAAGGCAGGAGAA 765

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              | |||||||  |  ||  | || |||      |   ||||||    ||| ||||||||||
Db        766 TTGCTTGAACTGGGAGGCAGAGGTGGT-----GAGCCAAGATTGCACCAGTGCACTCCAG 820

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  ||  || |||| ||||   ||||||| ||| ||| |||   | |  ||||| |||
Db        821 CCTTGGTGAC-AGAGTGAAACTCCATCTCAAAAACAAACAAACAAAATACATATACATAA 879

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              ||||||||  | | | | ||| | ||| ||||||| ||| |||||||  |  |||||||
Db        880 ATATATATGCACATATATATACATATATAAATATATATACACATATATAAATCTATATAC 939

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACAT-----ATATATA 354
              ||||||||| ||||| ||| ||||| ||  | |||||||| || ||||      ||||||
Db        940 ATATATACATATATACACATATATAAATCTATATACATATATATACATATATAATATATT 999

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              ||||||  || | | ||| ||| | | ||||||| | ||| ||||| ||||| | | |||
Db       1000 TACATATATAAATATATACATATATAAATATACATATATAAATACATATATAAATATACA 1059

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              |||||| | ||| ||| | | ||||||| ||||| | ||||| | | ||| ||| | |||
Db       1060 TATATAAATATACATATATAAATATACATATATAAATATATACATATATAAATATACATA 1119

Qy        475 TACACATATATATA-ATATACACACATATATATACACATATATACACACATATATACACA 533
              || | |||||||||  ||||| | | ||||| ||| |||  |||| ||| |||||||| |
Db       1120 TATAAATATATATACATATATAAATATATAAATATACAAGTATATACAAATATATACATA 1179

Qy        534 TATATACACATATATATACACATATATACACATATATACATATATACACATATATATACA 593
              ||||  |   |||||  || | | | ||||| | ||| ||| ||| | | || | |   |
Db       1180 TATAAATGTATATACGTATATACATATATATATAAATATATAAAAAAACTTTTGGCTGGG 1239

Qy        594 TATATACACATATATACATATATACACATATATACATATATACACACATATATACACATA 653
              |  |   || || |      | | | |  | |      ||  | || ||| | | ||||||
Db       1240 CACCTTTCCAAATCTCATGGCACATATAAGTCTCATGGTAACCTCAAATAAAAAAACATA 1299

Qy        654 CATATACACACACATAGATATACATATATATA 685
              |           || | ||| | |   ||| |
Db       1300 TAACAGATACACCAAAAATAAAAACCAATAAA 1331


RESULT 17
US-09-995-419A-1
; Sequence 1, Application US/09995419A
; Publication No. US20030032187A1
; GENERAL INFORMATION:
;  APPLICANT: Geron Corporation
;  APPLICANT:  McWhir, Jim
;  APPLICANT:  Gold, Joseph D.
;  APPLICANT:  Schiff, J. Michael
;  TITLE OF INVENTION: 096,004 - SeqList
;  FILE REFERENCE: 096,004 - SeqList
;  CURRENT APPLICATION NUMBER: US/09/995,419A
;  CURRENT FILING DATE:  2001-11-26
;  PRIOR APPLICATION NUMBER: 60/253,357
;  PRIOR FILING DATE: 2000-11-27
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 15418
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-995-419A-1

  Query Match             7.2%;  Score 296;  DB 10;  Length 15418;
  Best Local Similarity   69.2%;  Pred. No. 7.9e-50;
  Matches  479;  Conservative    0;  Mismatches  200;  Indels   13;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||| |||||||||||  ||| || | ||||||||||| |||||||| || ||||||
Db        646 AGTTCGAGACCAGCCTGGCCAATATAGCGAAACCCCATCTGTACTAAAAACACAAAAATT 705

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   ||  |||||| ||| |||||||||||||| ||||  ||||||  ||||||||||
Db        706 AGCTGGGTGTGGTGACACATGCCTGTAATCCCAGGTACTCAGGAGGCTAAGGCAGGAGAA 765

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              | |||||||  |  ||  | || |||      |   ||||||    ||| ||||||||||
Db        766 TTGCTTGAACTGGGAGGCAGAGGTGGT-----GAGCCAAGATTGCACCAGTGCACTCCAG 820

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  ||  || |||| ||||   ||||||| ||| ||| |||   | |  ||||| |||
Db        821 CCTTGGTGAC-AGAGTGAAACTCCATCTCAAAAACAAACAAACAAAATACATATACATAA 879

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              ||||||||  | | | | ||| | ||| ||||||| ||| |||||||  |  |||||||
Db        880 ATATATATGCACATATATATACATATATAAATATATATACACATATATAAATCTATATAC 939

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACAT-----ATATATA 354
              ||||||||| ||||| ||| ||||| ||  | |||||||| || ||||      ||||||
Db        940 ATATATACATATATACACATATATAAATCTATATACATATATATACATATATAATATATT 999

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              ||||||  || | | ||| ||| | | ||||||| | ||| ||||| ||||| | | |||
Db       1000 TACATATATAAATATATACATATATAAATATACATATATAAATACATATATAAATATACA 1059

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              |||||| | ||| ||| | | ||||||| ||||| | ||||| | | ||| ||| | |||
Db       1060 TATATAAATATACATATATAAATATACATATATAAATATATACATATATAAATATACATA 1119

Qy        475 TACACATATATATA-ATATACACACATATATATACACATATATACACACATATATACACA 533
              || | |||||||||  ||||| | | ||||| ||| |||  |||| ||| |||||||| |
Db       1120 TATAAATATATATACATATATAAATATATAAATATACAAGTATATACAAATATATACATA 1179

Qy        534 TATATACACATATATATACACATATATACACATATATACATATATACACATATATATACA 593
              ||||  |   |||||  || | | | ||||| | ||| ||| ||| | | || | |   |
Db       1180 TATAAATGTATATACGTATATACATATATATATAAATATATAAAAAAACTTTTGGCTGGG 1239

Qy        594 TATATACACATATATACATATATACACATATATACATATATACACACATATATACACATA 653
              |  |   || || |      | | | |  | |      ||  | || ||| | | ||||||
Db       1240 CACCTTTCCAAATCTCATGGCACATATAAGTCTCATGGTAACCTCAAATAAAAAAACATA 1299

Qy        654 CATATACACACACATAGATATACATATATATA 685
              |           || | ||| | |   ||| |
Db       1300 TAACAGATACACCAAAAATAAAAACCAATAAA 1331


RESULT 18
US-10-141-220-1
; Sequence 1, Application US/10141220
; Publication No. US20030040111A1
; GENERAL INFORMATION:
;  APPLICANT: Geron Corporation
;  APPLICANT:  Gold, Joseph
;  APPLICANT:  Lebkowski, Jane
;  TITLE OF INVENTION: Tpacked stem cells
;  FILE REFERENCE: 096/003
```

```
;  CURRENT APPLICATION NUMBER: US/10/141,220
;  CURRENT FILING DATE:  2002-05-07
;  PRIOR APPLICATION NUMBER: US/09/783,203
;  PRIOR FILING DATE: 2001-02-13
;  PRIOR APPLICATION NUMBER: 60/253,443
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/253,357
;  PRIOR FILING DATE: 2000-11-27
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
;   LENGTH: 15418
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-141-220-1

  Query Match             7.2%;  Score 296;  DB 14;  Length 15418;
  Best Local Similarity   69.2%;  Pred. No. 7.9e-50;
  Matches  479;  Conservative    0;  Mismatches  200;  Indels   13;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||| |||||||||||  ||| || | ||||||||||||| |||||||| || ||||||
Db        646 AGTTCGAGACCAGCCTGGCCAATATAGCGAAACCCCATCTGTACTAAAAACACAAAAATT 705

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   ||  |||||| ||| |||||||||||||| ||||  ||||||  ||||||||||
Db        706 AGCTGGGTGTGGTGACACATGCCTGTAATCCCAGGTACTCAGGAGGCTAAGGCAGGAGAA 765

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              |  |||||||  |  ||  | || |||     |   ||||||    ||| ||||||||||
Db        766 TTGCTTGAACTGGGAGGCAGAGGTGGT-----GAGCCAAGATTGCACCAGTGCACTCCAG 820

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||   ||  || |||| ||||   ||||||| ||| ||| |||   | |  ||||| |||
Db        821 CCTTGGTGAC-AGAGTGAAACTCCATCTCAAAAACAAACAAACAAAATACATATACATAA 879

Qy        240 ATATATATATATACACACACACACACATATATATATACACATATATATACACATATATAC 299
              ||||||||  | | | | | ||| | ||| ||||||| ||| ||||||| |  |||||||
Db        880 ATATATATGCACATATATATACATATATAAATATATATACACATATATAAATCTATATAC 939

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACAT-----ATATATA 354
              ||||||||| ||||| ||| ||||| ||  | |||||||| || ||||      ||||||
Db        940 ATATATACATATATACACATATATAAATCTATATACATATATATACATATATAATATATT 999

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              ||||||  || | | ||| ||| | | ||||||| | ||| ||||| ||||| | | |||
Db       1000 TACATATATAAATATATACATATATAAATATACATATATAAATACATATATAAATATACA 1059

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              |||||| | ||| ||| | | ||||||| ||||| | ||||| | | ||| ||| | |||
Db       1060 TATATAAATATACATATATAAATATACATATATAAATATATACATATATAAATATACATA 1119

Qy        475 TACACATATATATA-ATATACACACATATATATACACATATATACACACATATATACACA 533
              || | |||||||||  |||||  | |  |||||  ||| |||   ||||  ||| |||||||||  |
Db       1120 TATAAATATATATACATATATAAATATATAAATATACAAGTATATACAAATATATACATA 1179

Qy        534 TATATACACATATATATACACATATATACACATATATACATATATACACATATATATACA 593
              ||||  |   |||||   ||  | |  |||||  | |||  |||  | |  ||  | |  ||
Db       1180 TATAAATGTATATACGTATATACATATATATATAAATATATAAAAAAACTTTTGGCTGGG 1239

Qy        594 TATATACACATATATACATATATACACATATATACATATATACACACATATATACACATA 653
              |  |   || ||  |      | | | |  | |      ||  | || ||| | | |||||
Db       1240 CACCTTTCCAAATCTCATGGCACATATAAGTCTCATGGTAACCTCAAATAAAAAAACATA 1299

Qy        654 CATATACACACACATAGATATACATATATATA 685
              |          || | ||| | |   ||| |
Db       1300 TAACAGATACACCAAAAATAAAAACCAATAAA 1331


RESULT 19
US-10-023-969-1
; Sequence 1, Application US/10023969
; Publication No. US20030095989A1
; GENERAL INFORMATION:
;  APPLICANT: Geron Corporation
;  APPLICANT:  Irving, John
;  APPLICANT:  Lebkowski, Jane
;  TITLE OF INVENTION: Chimeric Cytolytic Viruses for Cancer Treatment
;  FILE REFERENCE: 084,002
;  CURRENT APPLICATION NUMBER: US/10/023,969
;  CURRENT FILING DATE:  2001-12-17
;  PRIOR APPLICATION NUMBER: 60/256,418
;  PRIOR FILING DATE: 2000-12-18
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 15418
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-023-969-1

  Query Match             7.2%;  Score 296;  DB 14;  Length 15418;
  Best Local Similarity   69.2%;  Pred. No. 7.9e-50;
  Matches  479;  Conservative    0;  Mismatches  200;  Indels   13;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||| |||||||||||  ||| || | ||||||||||||| |||||||| || ||||||
Db        646 AGTTCGAGACCAGCCTGGCCAATATAGCGAAACCCCATCTGTACTAAAAACACAAAAATT 705

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   ||  |||||| ||| |||||||||||||| ||||  ||||||  ||||||||||
Db        706 AGCTGGGTGTGGTGACACATGCCTGTAATCCCAGGTACTCAGGAGGCTAAGGCAGGAGAA 765

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              |  |||||||  |  ||  | || |||     |   ||||||    ||| ||||||||||
Db        766 TTGCTTGAACTGGGAGGCAGAGGTGGT-----GAGCCAAGATTGCACCAGTGCACTCCAG 820

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||   ||  || |||| ||||   ||||||| ||| ||| |||   | |  ||||| |||
Db        821 CCTTGGTGAC-AGAGTGAAACTCCATCTCAAAAACAAACAAACAAAATACATATACATAA 879

Qy        240 ATATATATATATACACACACACACACATATATATATACACATATATATACACATATATAC 299
              ||||||||  | | | | | ||| | ||| ||||||| ||| ||||||| |  |||||||
Db        880 ATATATATGCACATATATATACATATATAAATATATATACACATATATAAATCTATATAC 939

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACAT-----ATATATA 354
              ||||||||| ||||| ||| ||||| ||  | |||||||| || ||||      ||||||
Db        940 ATATATACATATATACACATATATAAATCTATATACATATATATACATATATAATATATT 999

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              ||||||  || | | ||| ||| | | ||||||| | ||| ||||| ||||| | | |||
Db       1000 TACATATATAAATATATACATATATAAATATACATATATAAATACATATATAAATATACA 1059

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              |||||| | ||| ||| | | ||||||| ||||| | ||||| | | ||| ||| | |||
Db       1060 TATATAAATATACATATATAAATATACATATATAAATATATACATATATAAATATACATA 1119

Qy        475 TACACATATATATA-ATATACACACATATATATACACATATATACACACATATATACACA 533
              || | |||||||||  |||||  | |  |||||  ||| |||   ||||  ||| |||||||||  |
Db       1120 TATAAATATATATACATATATAAATATATAAATATACAAGTATATACAAATATATACATA 1179

Qy        534 TATATACACATATATATACACATATATACACATATATACATATATACACATATATATACA 593
              ||||  |   |||||   ||  | |  |||||  | |||  |||  | |  ||  | |  ||
Db       1180 TATAAATGTATATACGTATATACATATATATATAAATATATAAAAAAACTTTTGGCTGGG 1239

Qy        594 TATATACACATATATACATATATACACATATATACATATATACACACATATATACACATA 653
              |  |   || ||  |      | | | |  | |      ||  | || ||| | | |||||
Db       1240 CACCTTTCCAAATCTCATGGCACATATAAGTCTCATGGTAACCTCAAATAAAAAAACATA 1299

Qy        654 CATATACACACACATAGATATACATATATATA 685
              |          || | ||| | |   ||| |
Db       1300 TAACAGATACACCAAAAATAAAAACCAATAAA 1331


RESULT 20
US-10-206-447-1
```

```
; Sequence 1, Application US/10206447
; Publication No. US20030099616A1
; GENERAL INFORMATION:
;  APPLICANT: Geron Corporation
;  APPLICANT:  Irving, John
;  APPLICANT:  Karpf, David
;  APPLICANT:  Schiff, Michael
;  TITLE OF INVENTION: DUAL SPECIFICITY TUMOR KILLING VECTORS DRIVEN BY THE TELOMERASE P
;  FILE REFERENCE: 085/002
;  CURRENT APPLICATION NUMBER: US/10/206,447
;  CURRENT FILING DATE:  2002-07-25
;  PRIOR APPLICATION NUMBER: 60/308,029
;  PRIOR FILING DATE: 2001-07-25
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 15418
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-206-447-1

  Query Match             7.2%;  Score 296;  DB 14;  Length 15418;
  Best Local Similarity   69.2%;  Pred. No. 7.9e-50;
  Matches  479;  Conservative    0;  Mismatches  200;  Indels   13;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||| |||||||||||  ||| || | ||||||||||||| ||||||||| || ||||||
Db        646 AGTTCGAGACCAGCCTGGCCAATATAGCGAAACCCCATCTGTACTAAAAACACAAAAATT 705

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   ||  |||||| ||| ||||||||||||| ||||  ||||||  ||||||||||
Db        706 AGCTGGGTGTGGTGACACATGCCTGTAATCCCAGGTACTCAGGAGGCTAAGGCAGGAGAA 765

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              |  |||||||  |  ||  | || |||     |   ||||||    ||| ||||||||||
Db        766 TTGCTTGAACTGGGAGGCAGAGGTGGT-----GAGCCAAGATTGCACCAGTGCACTCCAG 820

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  ||  || |||| ||||   ||||||| ||| ||| |||   | |  ||||| |||
Db        821 CCTTGGTGAC-AGAGTGAAACTCCATCTCAAAAACAAACAAACAAAATACATATACATAA 879

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              ||||||||  | | | | | ||| | ||| ||||||| ||| ||||||| |  |||||||
Db        880 ATATATATGCACATATATATACATATATAAATATATATACACATATATAAATCTATATAC 939

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACAT-----ATATATA 354
              |||||||| ||||| ||| ||||| ||  | |||||||| || ||||     ||||||
Db        940 ATATATACATATATACACATATATAAATCTATATACATATATATACATATATAATATATT 999

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              ||||||  || |  | ||| ||| | | ||||||| | ||| ||||| ||||| | | |||
Db       1000 TACATATATAAATATATACATATATAAATATACATATATAAATACATATATAAATATACA 1059

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              ||||||  | ||| ||| | | ||||||| ||||| | ||||| | | ||| ||| | |||
Db       1060 TATATAAATATACATATATAAATATACATATATAAATATATACATATATAAATATACATA 1119

Qy        475 TACACATATATATA-ATATACACACATATATATACACATATATACACACATATATACACA 533
              || | |||||||||| ||||| | | ||||| ||| |||  |||| ||| ||||||||| |
Db       1120 TATAAATATATATACATATATAAATATATAAATATACAAGTATATACAAATATATACATA 1179

Qy        534 TATATACACATATATATACACATATATACACATATATACATATATACACATATATATACA 593
              ||||  |   |||||  || | | ||||| | ||| ||| ||| | | || | |   |
Db       1180 TATAAATGTATATACGTATATACATATATATATAAATATATAAAAAAACTTTTGGCTGGG 1239

Qy        594 TATATACACATATATACATATATACACATATATACATATATACACACATATATACACATA 653
              | |  || || |     | | | | | |     ||  | || ||| | | |||||
Db       1240 CACCTTTCCAAATCTCATGGCACATATAAGTCTCATGGTAACCTCAAATAAAAAAACATA 1299

Qy        654 CATATACACACACATAGATATACATATATATA 685
              |          || | ||| | |   ||| |
Db       1300 TAACAGATACACCAAAAATAAAAACCAATAAA 1331
```

```
RESULT 21
US-10-674-836-1
; Sequence 1, Application US/10674836
; Publication No. US20040072787A1
; GENERAL INFORMATION:
;  APPLICANT: Morin, Gregg B.
;  APPLICANT:  Lichtsteiner, Serge
;  APPLICANT:  Vasserot, Alain
;  APPLICANT:  Adams, Robert R.
;  APPLICANT:  Geron Corporation
;  TITLE OF INVENTION: Telomerase Reverse Transcriptase Transcriptional
;  TITLE OF INVENTION:  Regulatory Sequences and Methods of Using
;  FILE REFERENCE: 019/246P
;  CURRENT APPLICATION NUMBER: US/10/674,836
;  CURRENT FILING DATE:  2003-09-29
;  PRIOR APPLICATION NUMBER: US/09/244,438
;  PRIOR FILING DATE: 1999-02-04
;  NUMBER OF SEQ ID NOS: 23
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 15418
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: Human TERT promoter
US-10-674-836-1

  Query Match             7.2%;  Score 296;  DB 17;  Length 15418;
  Best Local Similarity   69.2%;  Pred. No. 7.9e-50;
  Matches  479;  Conservative    0;  Mismatches  200;  Indels   13;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||| |||||||||||  ||| || | ||||||||||||| ||||||||| || ||||||
Db        646 AGTTCGAGACCAGCCTGGCCAATATAGCGAAACCCCATCTGTACTAAAAACACAAAAATT 705

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   ||  |||||| ||| ||||||||||||| ||||  ||||||  ||||||||||
Db        706 AGCTGGGTGTGGTGACACATGCCTGTAATCCCAGGTACTCAGGAGGCTAAGGCAGGAGAA 765

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              |  |||||||  |  ||  | || |||     |   ||||||    ||| ||||||||||
Db        766 TTGCTTGAACTGGGAGGCAGAGGTGGT-----GAGCCAAGATTGCACCAGTGCACTCCAG 820

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  ||  || |||| ||||   ||||||| ||| ||| |||   | |  ||||| |||
Db        821 CCTTGGTGAC-AGAGTGAAACTCCATCTCAAAAACAAACAAACAAAATACATATACATAA 879

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              ||||||||  | | | | | ||| | ||| ||||||| ||| ||||||| |  |||||||
Db        880 ATATATATGCACATATATATACATATATAAATATATATACACATATATAAATCTATATAC 939

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACAT-----ATATATA 354
              ||||||||| |||||  ||| ||||| ||  | ||||||||  || ||||     ||||||
Db        940 ATATATACATATATACACATATATAAATCTATATACATATATATACATATATAATATATT 999

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              ||||||  || |  | ||| ||| | | ||||||| | ||| ||||| ||||| | | |||
Db       1000 TACATATATAAATATATACATATATAAATATACATATATAAATACATATATAAATATACA 1059

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              ||||||  | ||| ||| | | ||||||| ||||| | ||||| | | ||| ||| | |||
Db       1060 TATATAAATATACATATATAAATATACATATATAAATATATACATATATAAATATACATA 1119

Qy        475 TACACATATATATA-ATATACACACATATATATACACATATATACACACATATATACACA 533
              || | |||||||||| ||||| | | ||||| ||| |||  |||| ||| ||||||||| |
Db       1120 TATAAATATATATACATATATAAATATATAAATATACAAGTATATACAAATATATACATA 1179

Qy        534 TATATACACATATATATACACATATATACACATATATACATATATACACATATATATACA 593
              ||||  |   |||||  || | | ||||| | ||| ||| ||| | | || | |   |
Db       1180 TATAAATGTATATACGTATATACATATATATATAAATATATAAAAAAACTTTTGGCTGGG 1239

Qy        594 TATATACACATATATACATATATACACATATATACATATATACACACATATATACACATA 653
```

```
Db       1240 CACCTTTCCAAATCTCATGGCACATATAAGTCTCATGGTAACCTCAAATAAAAAAACATA 1299

Qy        654 CATATACACACACATAGATATACATATATATA 685
              |           || | ||| | |   ||| |
Db       1300 TAACAGATACACCAAAAATAAAAACCAATAAA 1331
```

RESULT 22
US-10-811-012-1
; Sequence 1, Application US/10811012
; Publication No. US20040152189A1
; GENERAL INFORMATION:
; APPLICANT: Geron Corporation
; APPLICANT: McWhir, Jim
; APPLICANT: Gold, Joseph D.
; APPLICANT: Schiff, J. Michael
; TITLE OF INVENTION: Selective Antibody Targeting of Undifferentiated Stem Cells
; FILE REFERENCE: 096,006D- SeqList
; CURRENT APPLICATION NUMBER: US/10/811,012
; CURRENT FILING DATE: 2004-03-26
; PRIOR APPLICATION NUMBER: 09/995,419
; PRIOR FILING DATE: 2001-11-26
; PRIOR APPLICATION NUMBER: 60/253,357
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/253,443
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/253,395
; PRIOR FILING DATE: 2000-11-27
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 15418
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-811-012-1

Query Match 7.2%; Score 296; DB 18; Length 15418;
Best Local Similarity 69.2%; Pred. No. 7.9e-50;
Matches 479; Conservative 0; Mismatches 200; Indels 13; Gaps 5;

```
Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||| |||||||||||  ||| || | ||||||||||| |||||||| || ||||||
Db        646 AGTTCGAGACCAGCCTGGCCAATATAGCGAAACCCCATCTGTACTAAAAACACAAAAATT 705

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   ||  |||||| ||| |||||||||||||| ||||  ||||||  ||||||||||
Db        706 AGCTGGGTGTGGTGACACATGCCTGTAATCCCAGGTACTCAGGAGGCTAAGGCAGGAGAA 765

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              | |||||||  |  ||  | || |||     |   ||||||    ||| ||||||||||
Db        766 TTGCTTGAACTGGGAGGCAGAGGTGGT-----GAGCCAAGATTGCACCAGTGCACTCCAG 820

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  ||  || |||| ||||   ||||||| ||| ||| |||   | |  ||||| |||
Db        821 CCTTGGTGAC-AGAGTGAAACTCCATCTCAAAAACAAACAAACAAAATACATATACATAA 879

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              ||||||||  | | | | | ||| | ||| ||||||| ||| ||||||| |  |||||||
Db        880 ATATATATGCACATATATATACATATATAAATATATATACACATATATAAATCTATATAC 939

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACAT-----ATATATA 354
              |||||||||  ||||| ||| ||||| ||  | |||||||| || ||||      |||||
Db        940 ATATATACATATATACACATATATAAATCTATATACATATATATACATATATAATATATT 999

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              ||||||  || | | ||| ||| | | ||||||| | ||| ||||| ||||| | | |||
Db       1000 TACATATATAAATATATACATATATAAATATACATATATAAATACATATATAAATATACA 1059

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              |||||| | ||| ||| | | ||||||| ||||| | ||||| | | | ||| ||| | |||
Db       1060 TATATAAATATACATATATAAATATACATATATAAATATATACATATATAAATATACATA 1119
```

```
Qy        475 TACACATATATATA-ATATACACACATATATATACACATATATACACACATATATACACA 533
              || | |||||||||  |||||  | |  |||||  |||  |||   ||||  |||  |||||||||  |
Db       1120 TATAAATATATATACATATATAAATATATAAATATACAAGTATATACAAATATATACATA 1179

Qy        534 TATATACACATATATATACACATATATACACATATATACATATATACACATATATATACA 593
              ||||  |   |||||  ||  | |  |||||  | |||  |||  |||  | |  ||  | |    |
Db       1180 TATAAATGTATATACGTATATACATATATATATAAATATATAAAAAAACTTTTGGCTGGG 1239

Qy        594 TATATACACATATATACATATATACACATATATACATATATACACACATATATACACATA 653
              |  |    ||  || |       | | | |  | |       ||   | ||| ||| | | |||||
Db       1240 CACCTTTCCAAATCTCATGGCACATATAAGTCTCATGGTAACCTCAAATAAAAAAACATA 1299

Qy        654 CATATACACACACATAGATATACATATATATA 685
              |           || | ||| | |   ||| |
Db       1300 TAACAGATACACCAAAAATAAAAACCAATAAA 1331
```

RESULT 23
US-10-292-798-2005/c
; Sequence 2005, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
; FILE REFERENCE: 084335/166
; CURRENT APPLICATION NUMBER: US/10/292,798
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: 10/017,161
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: JP 2001-246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2070
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2005
; LENGTH: 1486
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; LOCATION: source
; FEATURE:
; LOCATION: (1)..(1486)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (201)..(1286)
US-10-292-798-2005

Query Match 7.2%; Score 294.8; DB 17; Length 1486;
Best Local Similarity 74.9%; Pred. No. 5.1e-50;
Matches 397; Conservative 0; Mismatches 127; Indels 6; Gaps 2;

```
Qy        197 AAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACA 256
              | || |   | || | || | | |      || || | | ||  | ||||||| |||||||
Db        999 ATATATACACACATATAATATATACATATTTATACACACATGCACATATATAAATACACA 940

Qy        257 CACACACACATATATATATACACATATATATACACATATATACATATATACACATATATA 316
              | | ||| ||||| | | | | || |||||| |||||||||||||||||||||||| ||
Db        939 TATATACATATATACACACATATATGTATATATACATATATACATATATACACATATGTA 880

Qy        317 CACATATATATACACATACATATGTACACATATATATACACATATGTATACACATATATA 376
              | ||| ||||| ||||| |||   | |  |||||||||||||||||||| | | | |||
Db        879 TATATACATATATACATATATACACATATGTATATATACACATATGTATATATACACATA 820

Qy        377 CACATATATACACACATATATACACATATATACACACATATATACACATATATACACATA 436
              ||||||| || | | | | |||||||||| || | | | |||||||||| || | | |||
Db        819 CACATATGTATATATACACATACACATATGTATATATACACATACACATATGTATATATA 760

Qy        437 TATACACATATACACATATATACACATATATACATATATACACATATATATA-----ATA 491
              | | |||  |||   |||||||||||| ||  |||||||||||||||| ||      | |
Db        759 CACATACACATATGTATATATACACATACATGTATATATACACATATATGTATATATACA 700
```

```
Qy        492 TACACACATATATATACACATATATACACACATATATACACATATATACACATATATATA 551
              | |||||| |||| ||| | | ||| |    |||||||||||||| | | | ||| | |||
Db        699 CATACACATGTATACACATACACATATATGTATATATACACATACACATATATACACATA 640

Qy        552 CACATATATACACATATATACATATATACACATATATATACATATATACACATATATACA 611
              ||| ||||||||||||||||| ||||||||||||||    |||||||||||||| |||
Db        639 TACACATATACACATATATACACATATACACATATATAAGTATATATACACATATACACA 580

Qy        612 TATATACACATATATACATATATACACACATATATACACATACATAT-ACACACACATAG 670
              ||||   | ||||| |||||||||||| |||  || | ||||| || | ||||| |||
Db        579 TATACGTATATATACACATATATACACATATACGTATATATACACATAATACACATATAC 520

Qy        671 ATATACATATATATACACATATATATACGTATATATATGTATATATATAT 720
              |||| ||| | ||| | | | |||||||||||||||| ||||||| | ||
Db        519 GTATAAATACACATATATACACATATACGTATATATACGTATATACACAT 470


RESULT 24
US-09-841-158-5
; Sequence 5, Application US/09841158
; Publication No. US20030166046A1
; GENERAL INFORMATION:
;  APPLICANT: BEASLEY, Ellen M.
;  TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
;  TITLE OF INVENTION:  NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND
;  TITLE OF INVENTION:  USES THEREOF
;  FILE REFERENCE: CL001229
;  CURRENT APPLICATION NUMBER: US/09/841,158
;  CURRENT FILING DATE:  2001-04-25
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
;   LENGTH: 27067
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(27067)
;   OTHER INFORMATION: n = A,T,C or G
US-09-841-158-5

  Query Match             7.1%;  Score 293.6;  DB 10;  Length 27067;
  Best Local Similarity   75.8%;  Pred. No. 3.1e-49;
  Matches  391;  Conservative    0;  Mismatches  119;  Indels    6;  Gaps    2;

Qy        230 ATATATATATATATATATATATACACACACACACACATATAT-ATATACACATATATATA 288
              || ||||  |  | |  |  ||| |      | ||| |||||| | ||| | ||||| | |
Db      26509 ATGTATAACTTAAAAATTTCATATATGTGTATACATATATATGAAATATATATATACAGA 26568

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              |||| ||||| || |||||| |||||||||||||||||  | | | ||||  |||||||
Db      26569 CACACATATATATGTATACATATATATACACATATATATGTATACATATATACACACATA 26628

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              ||| || |||||| |||||||| |||||||||||||||||||| ||| | | ||| | |
Db      26629 TATGTATACATATATATACACACATATACACATATATACACACACATACATATATACACA 26688

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              |||| ||||| |||||||||||||||| ||  |||||||||  |  ||||  | | ||| | |
Db      26689 CACATATATACACACATATATACACACATGCACACATATATATGTATACATATATACACA 26748

Qy        469 CATATATACACATATATATA-----ATATACACACATATATATACACATATATACACACA 523
              ||| |||||  ||||||| |     |||||||||||||||||||||| ||||||||||
Db      26749 CATGTATACGTATATATACACACATATATACACACATATATATACACACATATACACACA 26808

Qy        524 TATATACACATATATACACATATATATACACATATATACACATATATACATATATACACA 583
              || | |||||||||||||| |||||||||| ||||| ||| | ||| |||| || |||
Db      26809 TACACACACATATATACACACATATATACACACATATATACACACATATATATGTATACA 26868

Qy        584 TATATATACATATATACACATATATACATATATACACATATATACATATATACACACATA 643
              |||||||||| | ||| | | | ||||| | ||||| ||||| ||||||| ||| | | |
Db      26869 TATATATACACACATATATACACATACACACATACATATATACACATATACACATATACA 26928

Qy        644 TATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATA 703
              | ||| ||||| | |    | ||||| ||||||| | || || |||||| |||||  | |
Db      26929 CACACATATACACACATGTATACATATATATACACACATGTATACATATGTATACACACA 26988

Qy        704 TATATGTATATATATATGCTCCAGAGTTCATAAGAG 739
              |||| | |||| ||||    |     | |||| | |
Db      26989 CATATATGTATACATATATACACACATACATATGTG 27024


RESULT 25
US-10-801-837-5
; Sequence 5, Application US/10801837
; Publication No. US20040152167A1
; GENERAL INFORMATION:
;  APPLICANT: BEASLEY, Ellen M. et al.
;  TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
;  TITLE OF INVENTION:  NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND
;  TITLE OF INVENTION:  USES THEREOF
;  FILE REFERENCE: CL001229-DIV
;  CURRENT APPLICATION NUMBER: US/10/801,837
;  CURRENT FILING DATE:  2004-03-17
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
;   LENGTH: 27067
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(27067)
;   OTHER INFORMATION: n = A,T,C or G
US-10-801-837-5

  Query Match             7.1%;  Score 293.6;  DB 18;  Length 27067;
  Best Local Similarity   75.8%;  Pred. No. 3.1e-49;
  Matches  391;  Conservative    0;  Mismatches  119;  Indels    6;  Gaps    2;

Qy        230 ATATATATATATATATATATATACACACACACACACATATAT-ATATACACATATATATA 288
              || ||||  |  | |  |  ||| |      | ||| |||||| | ||| | ||||| | |
Db      26509 ATGTATAACTTAAAAATTTCATATATGTGTATACATATATATGAAATATATATATACAGA 26568

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              |||| ||||| || |||||| |||||||||||||||||  | | | ||||  |||||||
Db      26569 CACACATATATATGTATACATATATATACACATATATATGTATACATATATACACACATA 26628

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              ||| || |||||| |||||||| |||||||||||||||||||| ||| | | ||| | |
Db      26629 TATGTATACATATATATACACACATATACACATATATACACACACATACATATATACACA 26688

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              |||| ||||| |||||||||||||||| ||  |||||||||  |  ||||  | | ||| | |
Db      26689 CACATATATACACACATATATACACACATGCACACATATATATGTATACATATATACACA 26748

Qy        469 CATATATACACATATATATA-----ATATACACACATATATATACACATATATACACACA 523
              ||| |||||  ||||||| |     |||||||||||||||||||||| ||||||||||
Db      26749 CATGTATACGTATATATACACACATATATACACACATATATATACACACATATACACACA 26808

Qy        524 TATATACACATATATACACATATATATACACATATATACACATATATACATATATACACA 583
              || | |||||||||||||| |||||||||| ||||| ||| | ||| |||| || |||
Db      26809 TACACACACATATATACACACATATATACACACATATATACACACATATATATGTATACA 26868

Qy        584 TATATATACATATATACACATATATACATATATACACATATATACATATATACACACATA 643
              |||||||||| | ||| | | | ||||| | ||||| ||||| ||||||| ||| | | |
Db      26869 TATATATACACACATATATACACATACACACATACATATATACACATATACACATATACA 26928

Qy        644 TATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATA 703
              | ||| ||||| | |    | ||||| ||||||| | || || |||||| |||||  | |
Db      26929 CACACATATACACACATGTATACATATATATACACACATGTATACATATGTATACACACA 26988

Qy        704 TATATGTATATATATATGCTCCAGAGTTCATAAGAG 739
              |||| | |||| ||||    |     | |||| | |
Db      26989 CATATATGTATACATATATACACACATACATATGTG 27024
```

```
RESULT 26
US-10-017-161-2361/c
; Sequence 2361, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
;  FILE REFERENCE: 084335/0152
;  CURRENT APPLICATION NUMBER: US/10/017,161
;  CURRENT FILING DATE:  2002-12-18
;  PRIOR APPLICATION NUMBER: JP 2001/246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2430
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2361
;   LENGTH: 1486
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: source
;   LOCATION: (1)..(1486)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(1286)
US-10-017-161-2361

  Query Match             7.1%;  Score 293.2;  DB 15;  Length 1486;
  Best Local Similarity   74.7%;  Pred. No. 1.1e-49;
  Matches  396;  Conservative    0;  Mismatches  128;  Indels    6;  Gaps    2;

Qy        197 AAATGTCATCTCAGAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACA 256
              | || |   | || | || | | |     || || | | ||  | || |||| |||||||
Db        999 ATATATACACACATATAATATATACATATTTATACACACATGCACATGTATAAATACACA 940

Qy        257 CACACACACATATATATATACACATATATATACACATATATACATATATACACATATATA 316
              | | ||| ||||| | | | | || |||||| ||||||||||||||||||||||||| ||
Db        939 TATATACATATATACACACATATATGTATATATACATATATACATATATACACATATGTA 880

Qy        317 CACATATATATACACATACATATGTACACATATATATACACATATGTATACACATATATA 376
              | ||| ||||| ||||| |||   | |  |||||||||||||||||||||| | | | |||
Db        879 TATATACATATATACATATATACACATATGTATATATACACATATGTATATATACACATA 820

Qy        377 CACATATATACACACATATATACACATATATACACACATATATACACATATATACACATA 436
              ||||||| || | | | | |||||||||| || | | | | |||||||||| || | |||
Db        819 CACATATGTATATATACACATACACATATGTATATATACACATACACATATGTATATATA 760

Qy        437 TATACACATATACACATATATACACATATATACATATATACACATATATATA-----ATA 491
              | | ||| |||   |||||||||||||| ||  |||||||||||||||| ||      | |
Db        759 CACATACACATATGTATATATACACATACATGTATATATACACATATATGTATATATACA 700

Qy        492 TACACACATATATATACACATATATACACACATATATACACATATATACACATATATATA 551
              | |||||| |||| ||| | | ||| |   |||||||||||||| | | | ||| | |||
Db        699 CATACACATGTATACACATACACATATATGTATATATACACATACACATATATACACATA 640

Qy        552 CACATATATACACATATATACATATATACACATATATATACATATATACACATATATACA 611
              ||| |||||||||||||||||| ||||||||||||||||   ||||||||||||||| |||
Db        639 TACACATATACACATATATACACATATACACATATAAGTATATATACACATATACACA 580

Qy        612 TATATACACATATATACATATATACACACATATATACACATACATAT-ACACACACATAG 670
              ||||   | ||||| |||||||||||||| |||   || | ||||| || | ||||| |||
Db        579 TATACGTATATATACACATATATACACATATACGTATATATACACATAATACACATATAC 520

Qy        671 ATATACATATATATACACATATATATACGTATATATATGTATATATATAT 720
              |||| ||| | ||| | | | |||||||||||||||| ||||||| | ||
Db        519 GTATAAATACACATATATACACATATACGTATATATACGTATATACACAT 470


RESULT 27
US-10-017-161-2381/c
; Sequence 2381, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
;  FILE REFERENCE: 084335/0152
;  CURRENT APPLICATION NUMBER: US/10/017,161
;  CURRENT FILING DATE:  2002-12-18
;  PRIOR APPLICATION NUMBER: JP 2001/246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2430
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2381
;   LENGTH: 1672
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: source
;   LOCATION: (1)..(1672)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(1472)
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (208)..(307)
;   OTHER INFORMATION: a, t, c, g, unknown or other
US-10-017-161-2381

  Query Match             7.1%;  Score 291.6;  DB 15;  Length 1672;
  Best Local Similarity   75.5%;  Pred. No. 2.4e-49;
  Matches  376;  Conservative    0;  Mismatches  119;  Indels    3;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACATATATATATACACATATATA 286
              | |||||||||||||||||||||||||| | | ||||||||| | ||||||||||||||||
Db       1390 TATATATATATATATATATATATATATATACACACACACATATATACACATATATA 1331

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              |||||| | | | ||||| ||| ||||| ||||||||||||||||||||| |||   | | |
Db       1330 TACACACACACATATATACACATATATACACACATATATATACACATATATACACATATA 1271

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||| | | | ||||  | | | |||| |||||||||||||||||||| |||| || ||||
Db       1270 TATACACATATATATACACATATATATGTACACATATATACACACATGTATATACGTATA 1211

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              |     |||||||||||||| | ||| |  ||| |   | ||||||||||||||  | |||||
Db       1210 TGTGTACATATATACACACACATATATGTATGTGTATATATACACATATATGTATATATA 1151

Qy        467 TACATATATACACATATATATA---ATATACACACATATATATACACATATATACACACA 523
              |||| | | | |||| ||||||    |||| | ||| | ||| | | || |||| |
Db       1150 TACACACACAGACATGTATATATGTGTATATATACACACATAGATATATGTATATATGTG 1091

Qy        524 TATATACACATATATACACATATATATACACATATATACACATATATACATATATACACA 583
              |||||||||| | ||| |  | |||||||||| |||||    |||||||| | | | | |
Db       1090 TATATACACACACATATATGTGTATATACACACATATATGTGTATATACACACACATATA 1031

Qy        584 TATATATACATATATACACATATATACATATATACACATATATACATATATACACACATA 643
              | | |||| | | | ||| |||| |  ||||| ||||| |||||  | |||| |||||
Db       1030 TGTGTATATACACACACATATATGTGTATATACACACACATATATGTGTATATACACACA 971

Qy        644 TATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATA 703
              ||| |  |  |||||||||||||||  |||||  | |||||||||||||| ||  |||
Db        970 CATATATGTGTATATACACACACATGTATATATGTGTATATACACATATATGTATATATG 911

Qy        704 TATATGTATATATATATG 721
              | ||| || | |||||||
Db        910 TGTATATACACATATATG 893


RESULT 28
```

```
US-10-292-798-2023/c
; Sequence 2023, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
;  FILE REFERENCE: 084335/166
;  CURRENT APPLICATION NUMBER: US/10/292,798
;  CURRENT FILING DATE:  2002-11-13
;  PRIOR APPLICATION NUMBER: 10/017,161
;  PRIOR FILING DATE: 2001-12-18
;  PRIOR APPLICATION NUMBER: JP 2001-246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2070
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2023
;   LENGTH: 1672
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   LOCATION: source
;   FEATURE:
;   LOCATION: (1)..(1672)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(1472)
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (208)..(307)
;   OTHER INFORMATION: a, t, c, g, unknown or other
US-10-292-798-2023

  Query Match             7.1%;  Score 291.6;  DB 17;  Length 1672;
  Best Local Similarity   75.5%;  Pred. No. 2.4e-49;
  Matches  376;  Conservative    0;  Mismatches  119;  Indels    3;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||||| | | ||||||||| | ||||||||||||||||
Db       1390 TATATATATATATATATATATATATATATACACACACACACATATATACACATATATA 1331

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              |||||| | | | ||||| ||| ||||| ||||||||||||||||||||| |||    | | |
Db       1330 TACACACACACATATATACACATATATACACACATATATATACACATATATACACATATA 1271

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||| | | | ||||  | | | |||| ||||||||||||||||||| |||| || ||||
Db       1270 TATACACATATATATACACATATATATGTACACATATATACACACATGTATATACGTATA 1211

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              |    |||||||||||||| | ||| |  ||| |  | ||||||||||||||  | |||||
Db       1210 TGTGTACATATATACACACACATATATGTATGTGTATATATACACATATATGTATATATA 1151

Qy        467 TACATATATACACATATATATA---ATATACACACATATATATACACATATATACACACA 523
              |||| | | | |||| ||||||    |||| | ||| | ||| | | || |||| |
Db       1150 TACACACACAGACATGTATATATGTGTATATATACACACATAGATATATGTATATATGTG 1091

Qy        524 TATATACACATATATACACATATATATACACATATATACACATATATACATATATACACA 583
              |||||||||| | ||| |  | |||||||||| |||||    |||||||| | | | | |
Db       1090 TATATACACACACATATATGTGTATATACACACATATATGTGTATATACACACACATATA 1031

Qy        584 TATATATACATATATACACATATATACATATATACACATATATACATATATACACACATA 643
              | | |||| | | | ||| |||| |  ||||| ||||| |||||  | |||| ||||| |
Db       1030 TGTGTATATACACACACATATATGTGTATATACACACACATATATGTGTATATACACACA 971

Qy        644 TATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATA 703
              |||  |  |  |||||||||||||||  |||||  | ||||||||||||||||  |||
Db        970 CATATATGTGTATATACACACACATGTATATATGTGTATATACACATATATGTATATATG 911

Qy        704 TATATGTATATATATATG 721
              | ||| || | |||||||
```

```
Db        910 TGTATATACACATATATG 893


RESULT 29
US-10-741-601-5636/c
; Sequence 5636, Application US/10741601
; Publication No. US20040166519A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  STENOSIS, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001500
;  CURRENT APPLICATION NUMBER: US/10/741,601
;  CURRENT FILING DATE:  2003-12-22
;  NUMBER OF SEQ ID NOS: 26415
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5636
;   LENGTH: 81924
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(81924)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-2
US-10-741-601-5636

  Query Match             6.6%;  Score 273.8;  DB 18;  Length 81924;
  Best Local Similarity   65.9%;  Pred. No. 5.7e-45;
  Matches  458;  Conservative    0;  Mismatches  232;  Indels    5;  Gaps    4;

Qy         42 TACTAAAAATACAAAATTAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCG 101
              ||  ||||| | ||| ||||   |||||||||  |||| ||||||||||||||||| || |
Db      38267 TAAAAAAAAAAAAAAGTTAGTGGGGCATGGTGGCACACACCTGTAATCCCAGCTATTTGG 38208

Qy        102 GAGGCCGAGGCAGGAGAATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGC-AAGA 160
              ||||| |||||||||||||  ||||||||  | || ||||||||||||| | ||||  | |
Db      38207 GAGGCTGAGGCAGGAGAATTGCTTGAACC--CAGGAGGTGGAGGTTGCAGTGAGCTGAAA 38150

Qy        161 TCATGCCATTGCACTCCAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAA 220
              | ||| ||||||||||||||    |  || |||| || |      ||||| || | | | | |
Db      38149 TTGTGCTATTGCACTCCAGCCTGAGTGAC-AGAGTGAGACTCTGTCTCAAAATATATATA 38091

Qy        221 AGGCATTTTATATATATATATATATATACACACACACACATATATATATACACA 280
                  | |||| | | | | |||||| | |||||||| | || | | | || | |
Db      38090 TACACACACACATATGTGTGTGTGTATATACATACACACACATATATGTGTGTGTATATA 38031

Qy        281 TATATATACACATATATACATATATACACATATATACACATATATATACACATACATATG 340
              ||| ||| ||| ||| ||| | |||     |||||||| |||| | | | ||   |||
Db      38030 CATACATATACACATACACACACATATGTGTATATACATGTATACACATATATGTGTATA 37971

Qy        341 TACACATATATATACACATATGTATACACATATATACACATATATACACACATATATACA 400
              ||||| | | | || | | | |||||||||||  |     |||||    | ||| | | |
Db      37970 TACACGTGTGTGTATATGTGTATATACACATATGTGTGTGTATATGTGTATATACACATA 37911

Qy        401 CATATATACACACATATATACACATATATACACATATATACACATATACACATATATACA 460
              | | ||  |    ||||||||||||| |     |||||    |||||||||||| |
Db      37910 TGTGTGTATATGTGTATATACACATATGTGTGTGTATATGTGTATATACACATATGTGTG 37851

Qy        461 CATATATACATATATACACATATATATAATATACACACATATATATACACATATATACAC 520
              |||||   |||||||||||| | |  | || |    ||||||||   |||||||||
Db      37850 TGTATATGTGTATATACACATATGTGTG-TGTATATGTGTATATATATGTATATATACAC 37792

Qy        521 ACATATATACACATATATACACATATATATACACATATATACACATATATACATATATAC 580
              | | | | ||| |||| |     | ||| || | | | | | | | ||  |||||||
Db      37791 ATACACACACATATATGTTTGTGTGTATGTATATACACACATATGTGTGTATGTATATAC 37732

Qy        581 ACATATATATACATATATACACATATATACATATATACACATATATACATATATACACAC 640
              ||| |||| |   |||| ||| |||| |||  ||| ||||| |||  | |  |||||||| ||
Db      37731 ACACATATGTGTGTATACACATATATGCACACATATATACACGTGTGTATATATACCCAT 37672

Qy        641 ATATATACACATACATATACACACACATAGATATACATATATATACACATATATATACGT 700
              |||| || | ||   |||| | |||  |  ||||||||  | |  | ||||| ||  ||
Db      37671 ATATGTATATATGTGTATATATACATGTGTGTATACATACGTGTGTATATATACATGTGT 37612
```

```
Qy        701 ATATATATGTATATATATATGCTCCAGAGTTCATA 735
              |||| ||   | | ||||| |    | ||  | |
Db      37611 GTATACATACGTGTGTATATACATGTGTGTATACA 37577


RESULT 30
US-10-741-600-17799
; Sequence 17799, Application US/10741600
; Publication No. US20050026169A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001499
;  CURRENT APPLICATION NUMBER: US/10/741,600
;  CURRENT FILING DATE:  2003-12-22
;  NUMBER OF SEQ ID NOS: 73997
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17799
;   LENGTH: 14603
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-741-600-17799

  Query Match             6.5%;  Score 267.2;  DB 19;  Length 14603;
  Best Local Similarity   71.5%;  Pred. No. 6e-44;
  Matches  449;  Conservative    0;  Mismatches  168;  Indels   11;  Gaps    7;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC--AAAAT 58
              || ||||||||| ||||  |||||||||||||| || ||||||||||||||  |||||
Db      10438 AGATCAAGACCATCCTGGCCAACATGGTGAAACTCCGTCTCTACTAAAAATACAAAAAT 10497

Qy         59 TAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGA 118
              |||   ||| ||||| |   | ||||||||||||||||||| |||||| || ||||||||
Db      10498 TAGCTGGGCGTGGTGGTGGGCTCCTGTAATCCCAGCTACTTGGGAGGCTGAAGCAGGAGA 10557

Qy        119 ATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTG---CAC 174
              ||| ||||||||  | || || |||||||| | ||| |||||||  |||||||   |||
Db      10558 ATCGCTTGAACC--CAGGAGGCGGAGGTTGCAGTGAGCCAAGATCGCGCCATTGCACCAC 10615

Qy        175 TCCAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATA 234
              |||||||  |||||  ||||     |||||  ||||||||||||    |   | | |
Db      10616 TCCAGCCTGGGCAAAAAGAGAGAAACTCTGTCTCAAAAAAAAAAAAAAAAAAAAAAAAA 10675

Qy        235 TATATATATATATATATACACACACACACATATATATATACACATATATATACACATA 294
              |||||||||||||||| | ||||||||||||||||||  | | ||| ||  | | |||
Db      10676 AATATATATATATATATATATACACACACATATATATTCATATATA-ATTCATATATA 10734

Qy        295 TATACATATATACACATATATACACATATATATACACATACATATGTACACATATATATA 354
              ||| |||||||          |  |||||||  |   | |||  ||  |||||||| |
Db      10735 TATTCATATATAATTCATATATAATTCATATATATATTCATATAGATATTCATATATA-A 10793

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              |||||  |||  |||| |||  ||||||||  || ||| |||  |||||||  || | |
Db      10794 TTCATATAGATATTCATAGATATTCATATATATTCATATAGATATTCATATATTCATATA 10853

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              |||  | | ||  ||  | ||  ||| | |||| || ||||| | |   ||| | ||  |
Db      10854 TATTCATATATTCATGTATATTCATATAGATATTCATATATAGATATTCATAGATATTCA 10913

Qy        475 TACACATATATATAATATACACACATATATATACACATATATACACACATATATACACAT 534
              || | |||| | ||||| | |   ||||||| | ||  || | | ||  ||  | |
Db      10914 TATATAGATAT-TCATATAGATATTCATATATAGATATTCATATAGATATTCATGTATAG 10972

Qy        535 ATATACACATATATATACACATATATACACATATATACATATATACACATATATATACAT 594
              |||| || |||||||| || ||||| | |   ||||| |||||| || || |||||  ||
Db      10973 ATATTCATATATATATTCATATATAGATATTCATATAGATATATTCATATCTATATGAAT 11032

Qy        595 ATATACACATATATACATATATACACAT 622
              ||||| |   |||||  ||||||   ||
Db      11033 ATATATATGAATATATCTATATATGAAT 11060
```

```
RESULT 31
US-10-311-455-1210/c
; Sequence 1210, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determi
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1210
;   LENGTH: 6971
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: 2741
;   OTHER INFORMATION: n is a or g or c or t
US-10-311-455-1210

  Query Match             6.5%;  Score 266.4;  DB 15;  Length 6971;
  Best Local Similarity   71.1%;  Pred. No. 6.4e-44;
  Matches  367;  Conservative    0;  Mismatches  146;  Indels    3;  Gaps    1;

Qy        212 AAAAAAAAAAGGCATTTTATATATATATATATATATACACACACACACATATAT 271
              | | | | |   | | ||||||||  |||||||||| ||||||| | | | | ||||||
Db       6472 ATACATATATACCTATATATATATACCTATATATATAAATACACATATATATATATAT 6413

Qy        272 ATATACACATATATATACACATATATACATATATACACATATATACACATATATATACAC 331
              |||   | |||||||||| |||  |||| ||||||||||  ||||| |||||||||    |
Db       6412 ATACCAATATATATATATACACGTATATATATATACACACTTATATACATATATACCTAT 6353

Qy        332 ATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACACAC 391
              ||| |||  |||||||||||||| | |||  |||| | ||| |||||||| |||| | |
Db       6352 ATATATAAATACACATATATATATATATACCTATATATATAAATACACATTTATATATAT 6293

Qy        392 ATATATACACATATATACACACATATATACACATATATACACATATATACACATATACAC 451
              | ||| |||  |||||| | | |    || | | ||| |||  |||||| | ||||||
Db       6292 AAATACACATTTATATATATATACGCCTATATAAATACACATTTATATATATATATACCT 6233

Qy        452 ATATATACACATATATACATATATACACATATATAATATACACACATATATATACACA 511
              ||||||| || |||||| ||| ||||  |||||||||| | |||   |||||||| | |
Db       6232 ATATATATACCTATATATATAAATACCTATATATATAAAACACATTTATATATATATATA 6173

Qy        512 TATATACACACATATATACACATATATACACATATATATACACATATATACACATATATA 571
              ||||  | | ||| | |   ||||||| || | |||||| | |  |||| | | | |||
Db       6172 CCTATATATAAATACACATTTATATATATACCTCTATATATATACCTATATATAAAAATA 6113

Qy        572 CATATATACACATATATATACATATATACACATATATACATATATACACATATATACATA 631
              || || || | |||     || ||||| | | | | ||  |||||||| | |||||| |||
Db       6112 CACATTTATATATACGCCTATATATAAAAATACACATTTATATATACGCCTATATATATA 6053

Qy        632 TATACACACATATATACACATACATATACACACACATAGATATACATATATAT---ACAC 688
              |||||||  |||||| | ||||   || | | | | | | | || ||||||||    ||||
Db       6052 AATACACATTTATATATATATACGCCTATATATATAAATACCTATATATATATAAAACAC 5993

Qy        689 ATATATATACGTATATATATGTATATATATATGCTC 724
              || ||||||  ||||||||  |||||||| || | |
Db       5992 ATTTATATATATATATATACCTATATATAAATACAC 5957
```

```
RESULT 32
US-10-812-232-6/c
; Sequence 6, Application US/10812232
; Publication No. US20040265961A1
; GENERAL INFORMATION:
;  APPLICANT: Liu, Wei
;  APPLICANT:  Wu, Leeying
;  APPLICANT:  Channavajhala, Padma L,
;  APPLICANT:  Lin, Lih-Ling
;  APPLICANT:  Zhang, Yuhua
;  TITLE OF INVENTION: Novel Proteins Homologous to Kinase Suppressor of Ras
;  FILE REFERENCE: 01997.026700
;  CURRENT APPLICATION NUMBER: US/10/812,232
;  CURRENT FILING DATE:  2004-03-29
;  PRIOR APPLICATION NUMBER: US 60/457,928
;  PRIOR FILING DATE: 2003-03-28
;  PRIOR APPLICATION NUMBER: US 60/491,283
;  PRIOR FILING DATE: 2003-07-31
;  NUMBER OF SEQ ID NOS: 254
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6
;   LENGTH: 263852
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3814)..(4319)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (6552)..(6571)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (17206)..(17225)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (19680)..(19699)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (24320)..(24339)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (26572)..(26591)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (32986)..(33005)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (56103)..(56724)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (59672)..(59691)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (61525)..(61544)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (71035)..(71054)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (79771)..(79807)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (91343)..(91801)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (141749)..(141768)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (144739)..(144758)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (164171)..(164340)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (173689)..(173708)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (175051)..(175070)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (192924)..(192993)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (202010)..(202029)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (247837)..(247856)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (251852)..(252421)
;   OTHER INFORMATION: n is a, c, g, or t
US-10-812-232-6

  Query Match             6.4%;  Score 265.6;  DB 18;  Length 263852;
  Best Local Similarity   73.7%;  Pred. No. 4.5e-43;
  Matches  367;  Conservative    0;  Mismatches  124;  Indels    7;  Gaps    2;

Qy        229 TATATATATATATATATATATACACACACACACATATATATATACACATATATATA 288
              || |||||||||||| ||| ||| | ||| ||| ||| || |  | ||||||| ||  |||
Db      94187 TACATATATATATACATACATACATACATACATACATATGTCCACATACACACATGCATA 94128

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              | ||||||| | ||||| | ||||||| ||| | | | |  ||| ||||  |||| | |
Db      94127 TATATATATATACATATATATATATATACACACACACATTCATGCATAAATACAAACA 94068

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACAT------ATATACACA 402
              |||| |||||||  |||| ||| ||| ||| | |||| |||||||      ||| |||||
Db      94067 TATACACACATACATATAAACACATACACAGAGATATGCACACATACATACATACACACA 94008

Qy        403 TATATACACACATATATACACATATATACACATATATACACATATACACATATATACACA 462
              || ||| | ||||  |||||||| ||||||| | | ||||||  | || |||||| |||
Db      94007 TACATATATACATGCATACACATAAATACACACAGACACACATGCATACGTATATATACA 93948

Qy        463 TATATACATATATACACATATATATAATATACACACATATATATACACATATATACACAC 522
              |||||  ||| ||| ||||||| |  | ||||| | | | ||||||||| ||||||||||
Db      93947 TATATGTATACATATACATATACA-CACATACATAAACACATATACACAGATATACACAC 93889

Qy        523 ATATATACACATATATACACATATATATACACATATATACACATATATACATATATACAC 582
              ||| | | ||| | | | |||||| |||||| ||| | | ||||||  | | | |||  |
Db      93888 ATACACAAACACACAGAGACATATGTATACATATACACATACATATGCAAACACATAAGC 93829

Qy        583 ATATATATACATATATACACATATATACATATATACACATATATACATATATACACACAT 642
              | ||| |||||||||||| | ||||||| ||| | ||||| ||||||| ||||| | |
Db      93828 ACATACATACATATATACATACATATACACATACATACATACATACATAAATACATAGAG 93769
```

```
Qy        643 ATATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTAT 702
              | | |||   ||||||| |||||   | | ||||| ||| ||||||| | | | |||| |
Db      93768 ACACACATGCACATATATACACATGCATACATACACATAAATACACACAGACACACGTGT 93709

Qy        703 ATATATGTATATATATAT 720
              |||||   | ||| | ||
Db      93708 ATATACACACATACACAT 93691
```

RESULT 33
US-10-108-260A-543/c
; Sequence 543, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20040005560A1el full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 543
; LENGTH: 2522
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-108-260A-543

Query Match 6.4%; Score 263.6; DB 17; Length 2522;
Best Local Similarity 68.9%; Pred. No. 1.5e-43;
Matches 376; Conservative 0; Mismatches 169; Indels 1; Gaps 1;

```
Qy        236 ATATATATATATATATACACACACACACATATATATATACACATATATATACACATAT 295
              ||  |  | | |||||| | | ||||||| |  |||||||    ||||||||||||  |
Db       2217 ATTCAACTGTGTATATATATATACACACATACGTATATATGTGTATATATATACACACGT 2158

Qy        296 ATACATATATACACATATATACACATATATATACACATACATATGTACACATATATATAC 355
              ||| ||| | | | || |||||| | | |    | | ||||| || ||   ||| ||| ||
Db       2157 ATATATACACATATATGTATACATACACACGCATATATACACATATATGTATACATACAC 2098

Qy        356 ACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACAT 415
              ||·||| ||||||||||||||  | | ||| |||| ||||||||||||||||||  | ||||
Db       2097 ACGTATATATACACATATATGTATACATACACACGCATATATACACATATATGTATACAT 2038

Qy        416 ATATACACATATATACACATATATACACATATACACATATATACACATATATACATATAT 475
              | | ||  |||||||||||||||| || | |||||||   ||| | | | ||| ||  ||
Db       2037 ACACACGTATATATACACATATATGCATACATACACACGCATATATACACATATATGCAT 1978

Qy        476 ACACATATATATAATATACACACATATATATACACATATATACACACATATATACACATA 535
              ||| | |||  | ||||| |||||||||| || ||||| ||||  | | ||| ||| |||
Db       1977 ACATACATACGT-ATATATACACATATATGTATACATACATACGTATACATACACATATA 1919

Qy        536 TATACACATATATATACACATATATACACATATATACATATATACACATATATATACATA 595
              | || ||||| ||   || ||||| ||| |||| || | ||| | || ||||||||||| |
Db       1918 TGTATACATACATGCGCATATATACACATATATGTATACATACATACGTATATATACACA 1859

Qy        596 TATACACATATATACATATATACACATATATACATATATACACACATATATACACATACA 655
              |||·|   ||| |||||||  || ||||| | | || |||||| ||||  || | |||||
Db       1858 TATATGTATACATACATACGTATACATACACATATGTATACATACATACGTATATATACA 1799

Qy        656 TATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATA 715
              ||| |   | | | || | |||||||||||||||||| ||   ||| ||  |||||||
Db       1798 CATATATGTATACATACATGCGTATATATACACATATATGTATACATACATGCGTATATA 1739

Qy        716 TATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGA 775
              || |       |  | ||||      | ||  |   ||    || | ||| |   | | |
Db       1738 TACACATATATGCATACATACATATGAACACATATATATGTGTGTGTATATATACATATA 1679

Qy        776 GAGTTT 781
              |||
Db       1678 AGATTT 1673

RESULT 34
US-10-723-860-1357/c
; Sequence 1357, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1357
; LENGTH: 227968
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-723-860-1357

Query Match 6.4%; Score 263; DB 18; Length 227968;
Best Local Similarity 71.6%; Pred. No. 1.4e-42;
Matches 373; Conservative 0; Mismatches 145; Indels 3; Gaps 2;

```
Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||  |||||||||||  ||||||||||||||||||||||||||||||||||| ||||||
Db       5103 AGTTTGAGACCAGCCTGGCCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAAATT 5044

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              || ||||||||||||  ||| ||||||||||||||||||| | |||||| || | |||||||
Db       5043 AGCCAGGCATGGTGGCACATGCCTGTAATCCCAGCTACCTGGGAGGCTGAAGTAGGAGAA 4984

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              || || ||||| |  || ||  |||| ||| | |||| |          | | |   ||
Db       4983 TCTCTGGAACCAG--GGAGGCAGAGGCTGCAGTGAGCCAAGATCATTCCTGGGTGTCAGA 4926

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              || |  ||     |     |  | || ||||||||||||   |    | | | | |
Db       4925 GAGAGACTCCATCCCTGCCCGTCCCCCGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4866

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              | | | ||||||| | | | | | | |||||||||||| | ||||| | ||||| | |||
Db       4865 AAAAAAATATATATATATATATATATATATATATATATATATACACACACACACATAT 4806

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACAT 359
              |||| | | ||||| |||||||||||||||| ||| |||   || | ||| ||||| | ||
Db       4805 ATATACATATATATACACACATATATATACATATATATACATATATATACATATATATAT 4746

Qy        360 ATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATAT 419
              | ||||| | ||| ||| | ||||| ||| | ||||| ||| ||||||||| | ||||||
Db       4745 ACGTATATATATACATATATATATACACATATATATACACATATATATACATATATATAT 4686

Qy        420 ACACATATATACACATATATACACATATACACATATATACACATATATACATATATACAC 479
              |||||||||| ||| ||||| | | ||| | ||||| ||| ||||||||||||||||| |
Db       4685 ACACATATATATACACATATATATACATATATATATACACATATATATACATATATATAT 4626

Qy        480 ATATATATAATATACACACATATATATACACATATATACAC 520
              || |||||  |||  | |      || |    || || |||
Db       4625 ATGTATATTTTATTGAGATCACATTAAATGTGTACATTCAC 4585
```

RESULT 35
US-10-741-600-17809/c
; Sequence 17809, Application US/10741600
; Publication No. US20050026169A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001499
```

```
;  CURRENT APPLICATION NUMBER: US/10/741,600
;  CURRENT FILING DATE:  2003-12-22
;  NUMBER OF SEQ ID NOS: 73997
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17809
;    LENGTH: 109730
;    TYPE: DNA
;    ORGANISM: Homo sapiens
US-10-741-600-17809

  Query Match             6.4%;  Score 262.6;  DB 19;  Length 109730;
  Best Local Similarity   72.0%;  Pred. No. 1.3e-42;
  Matches  384;  Conservative    0;  Mismatches  144;  Indels    5;  Gaps    3;

Qy        204 ATCTCAGAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACAC 263
              || | | | | | | |||   || |||||||||| |||||||| |||| ||| | | |
Db     108244 ATGTTATATATACATAAATATATGTTATATATATTTATATATATTTATATACATATAAAT 108185

Qy        264 ACATATATATATACACATATATATACACATATATACATATATACACATATATA-CACATA 322
              | |||||| |||| ||||||||| | | ||||||| || |||| | | |||||    |||
Db     108184 ATATATATTTATATACATATATAAAAATATATATATATTTATATATAAATATATGTTATA 108125

Qy        323 TATATACACATACATATGTACACATATATATACACATATGTATACACATATATACACATA 382
              |||||  | ||||||||  | | |||| |||| ||||||  ||| | || |||| |||||
Db     108124 TATATTTATATACATATAAATATATATTTATATACATATAAATATATATTTATATACATA 108065

Qy        383 TATACACACATATATACACATATATACACACATATATACACATATATACACATATATACA 442
              || | | || |||| ||||||| | |   |||| || | | ||||| | |||||||
Db     108064 TAAATATATATTTATATACATATAAAAATCTATATTTATATACATATAAAAATATATATT 108005

Qy        443 CATATACACATATATACACATATATACATATATACACATATATATA---ATATACACACA 499
              ||||||| ||| | | | |||| || ||| ||| | |||||||||    ||||||| | |
Db     108004 TATATACATATAAAAATATATATTTATATACATATAGATATATATATTTATATACATATA 107945

Qy        500 TATATATACACATATATACACACATATATACACATATATACACATATATATACACATATA 559
              |||||||  |  |||||||| | |||||||  |  |||||||| ||| ||||| | || ||
Db     107944 AATATATATATTTATATACATAAATATATATATTTATATACATATAAATATATATATTTA 107885

Qy        560 TACACATATATACATATATACACATATATATACATATATACACATATATACATATATACA 619
              || ||||| ||| ||||||  | ||| ||||  ||||| | | | ||||| |||| || |
Db     107884 TATACATAAATATATATATTTATATACATATGTATATAAATATAAATATATATATGTATA 107825

Qy        620 CATATATACATATATACACACATATATACACATACATATACACACACATAGATATACATA 679
              | |||||  |||| || |   |||| || | | | |||||  |   | ||| ||||| |||
Db     107824 TAAATATATATATTTATATTTATATGTATATAAATATATATATTTATATAAATATATATA 107765

Qy        680 TATATACACATAT-ATATACGTATATATATGTATATATATATGCTCCAGAGTT 731
              | ||||   |||| |||||| || |||||| ||||| | | ||     |||||
Db     107764 TTTATATTTATATAATATACATACATATATATATATTTTTTTGAGATGGAGTT 107712


RESULT 36
US-10-437-963-12666/c
; Sequence 12666, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 12666
;    LENGTH: 1800
;    TYPE: DNA
;    ORGANISM: Oryza sativa
;    FEATURE:
;    OTHER INFORMATION: Clone ID: PAT_MRT4530_18773C.1
US-10-437-963-12666

  Query Match             6.4%;  Score 262;  DB 18;  Length 1800;
  Best Local Similarity   72.0%;  Pred. No. 2.8e-43;
  Matches  370;  Conservative    0;  Mismatches  140;  Indels    4;  Gaps    2;

Qy        229 TATATATATATATATATATATATACACACACACACACATATATATATACACATATATATA 288
              |||||||||||||||| ||||||| | | ||| | | ||||||| ||| | |||||||||
Db       1796 TATATATATATATATACATATATATATATACATATATATATATACATATATATATATATA 1737

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              || ||||||| ||||||| | ||||||| | | ||||||| | |||||||| || | |||
Db       1736 CATATATATATATATATATATATATATATACATATATATATATACATATATATATATA 1677

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              |||||| | |||| |||| ||  | ||  | |||| || | | || |||| | |||||||
Db       1676 TATATATATATATATATATATACGAAAATGTATATATGTATATATATGTATATATATATA 1617

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              |   |||||||  ||||||||  || |||| | || || | |||  ||   |||||||
Db       1616 TATGTATATATATGTATATATACGTATGTATATATATGTATATATACGTATGTATATATA 1557

Qy        469 CATATATACACATATATATAATATACACACATATATATACACATATATACACACATATAT 528
              ||||||  || ||| ||||   || | || ||| |||| | || |||| | || |||||
Db       1556 TGTATATATACGTATGTATACG-TATATACGTATGTATATATATGTATATATACGTATAT 1498

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACATATAT 588
              | |  |||||| || ||||||||   |||||||   ||||||||||||||   ||||||
Db       1497 ATATGTATATATACGTATATATATGTATATATATGTATATATACATATATATGTATATAT 1438

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATA---TACACACATATA 645
              || |||||||   |||||||||||||||   |||||||  ||||   ||   | || |
Db       1437 GTATATATATATGTATATATACATATATATGTATATATATGTATATGTTATGTATATGTG 1378

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              |     ||| || || |   | ||  |||||  |||||||  | |||| ||||  ||||||
Db       1377 TGTGTATATATGTATATGTATATGTATATATGTATATATGTATATATGTATATATATATA 1318

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAG 739
              ||| ||| |||||||      |  |  ||||| |
Db       1317 TATATATGTATATATATATATGTATATATAAGCG 1284


RESULT 37
US-10-719-993-6815
; Sequence 6815, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001496
;  CURRENT APPLICATION NUMBER: US/10/719,993
;  CURRENT FILING DATE:  2003-11-24
;  NUMBER OF SEQ ID NOS: 55342
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6815
;    LENGTH: 1980090
;    TYPE: DNA
;    ORGANISM: Homo sapiens
;    FEATURE:
;    NAME/KEY: misc_feature
;    LOCATION: (1)...(1980090)
;    OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-2
US-10-719-993-6815

  Query Match             6.4%;  Score 262;  DB 18;  Length 1980090;
  Best Local Similarity   70.7%;  Pred. No. 5.8e-42;
  Matches  420;  Conservative    0;  Mismatches  165;  Indels    9;  Gaps    5;
```

```
Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC--AAAAT 58
              ||||| ||||||||||  |||||||| ||||||||||||||||||||||||  |||||
Db    1507113 AGTTCGAGACCAGCCTGGTCAACATGGTAAAACCCCATCTCTACTAAAAATACAAAAAAT 1507172

Qy         59 TAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGA 118
              ||| | ||  |||||      ||||||||||||||||||  |||||| ||| || ||||
Db    1507173 TAGCCGGGTGTGGTGGCGAGTGCCTGTAATCCCAGCTACTCTGGAGGCTGAGACATGAGA 1507232

Qy        119 ATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCC 177
              ||| ||||||||||  || || |||||||| | ||| | |||||  |||| | ||||||
Db    1507233 ATCGCTTGAACCTG--GGAGGCGGAGGTTGCAGTGAGCCGAGATCGCGCCACTACACTCC 1507290

Qy        178 AGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATAT 237
              ||||  ||||||||||||||||     ||||| ||||||||||   |    | | | ||
Db    1507291 AGCCTGGGCAACAAGAGCGAAACTCTGTCTCAAAAAAAAAAAAA---AAAAAAAAAAAAT 1507347

Qy        238 ATATATATATATATACACACACACACATATATATATACACATATATATACACATATATAT 297
              |||||||||||||| | | | | |   || ||||||||    |||||||  | |||| |
Db    1507348 ATATATATATATATATATATATATATGTATGTATATATATGTGTATATATGTATATATGT 1507407

Qy        298 ACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACAC 357
              ||||||||   |||||||    || |||||   ||| || |||| |   | ||||||
Db    1507408 GCATATATATGTATATATATGTGTAGATATATGTATATATGTGTATATGTGTATATATGT 1507467

Qy        358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417
              |||| | || | || | || | || |||| |    | |||| |  |||||| |    |||
Db    1507468 ATATATGTATATATGTGTATATATGTATATATGTGTGTATATATGTATATATATGTGTAT 1507527

Qy        418 ATACACATATATACACATATATACACATATACACATATATACACATATATACATATATAC 477
              |||   |||||||    ||||||   ||||     ||||||   |||||||  | | |
Db    1507528 ATATGTATATATATGTGTATATATGTATATGTGTGTATATATGTATATATATGTGTGTGT 1507587

Qy        478 ACATATATATAATATACACACATATATATACACATATATACACACATATATACACATATA 537
              | |||| ||| ||||      |||||||  | ||||||  | | |||| |      ||||
Db    1507588 ATATATGTAT-ATATGTGTGTATATATATGTATATATATGTATATATATGTGTGTGTATA 1507646

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATA 591
              || |  ||||||||   |||||||    | |||| ||||||  | |||||||||
Db    1507647 TATATGTATATATATGTATATATATGTGTGTATATATATATGTATATATATATA 1507700
```

RESULT 38
US-10-741-600-17676
; Sequence 17676, Application US/10741600
; Publication No. US20050026169A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001499
; CURRENT APPLICATION NUMBER: US/10/741,600
; CURRENT FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 73997
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17676
; LENGTH: 1980090
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1980090)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-741-600-17676

Query Match 6.4%; Score 262; DB 19; Length 1980090;
Best Local Similarity 70.7%; Pred. No. 5.8e-42;
Matches 420; Conservative 0; Mismatches 165; Indels 9; Gaps 5;

```
Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC--AAAAT 58
              ||||| ||||||||||  |||||||| ||||||||||||||||||||||||  |||||
Db    1507113 AGTTCGAGACCAGCCTGGTCAACATGGTAAAACCCCATCTCTACTAAAAATACAAAAAAT 1507172

Qy         59 TAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGA 118
              ||| | ||  |||||      ||||||||||||||||||  |||||| ||| || ||||
Db    1507173 TAGCCGGGTGTGGTGGCGAGTGCCTGTAATCCCAGCTACTCTGGAGGCTGAGACATGAGA 1507232

Qy        119 ATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCC 177
              ||| ||||||||||  || || |||||||| | ||| | |||||  |||| | ||||||
Db    1507233 ATCGCTTGAACCTG--GGAGGCGGAGGTTGCAGTGAGCCGAGATCGCGCCACTACACTCC 1507290

Qy        178 AGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATAT 237
              ||||  ||||||||||||||||     ||||| ||||||||||   |    | | | ||
Db    1507291 AGCCTGGGCAACAAGAGCGAAACTCTGTCTCAAAAAAAAAAAAA---AAAAAAAAAAAAT 1507347

Qy        238 ATATATATATATATACACACACACACATATATATATACACATATATATACACATATATAT 297
              |||||||||||||| | | | | |   || ||||||||    |||||||  | |||| |
Db    1507348 ATATATATATATATATATATATATATGTATGTATATATATGTGTATATATGTATATATGT 1507407

Qy        298 ACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACAC 357
              ||||||||   |||||||    || |||||   ||| || |||| |   | ||||||
Db    1507408 GCATATATATGTATATATATGTGTAGATATATGTATATATGTGTATATGTGTATATATGT 1507467

Qy        358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417
              |||| | || | || | || | || |||| |    | |||| |  |||||| |    |||
Db    1507468 ATATATGTATATATGTGTATATATGTATATATGTGTGTATATATGTATATATATGTGTAT 1507527

Qy        418 ATACACATATATACACATATATACACATATACACATATATACACATATATACATATATAC 477
              |||   |||||||    ||||||   ||||     ||||||   |||||||  | | |
Db    1507528 ATATGTATATATATGTGTATATATGTATATGTGTGTATATATGTATATATATGTGTGTGT 1507587

Qy        478 ACATATATATAATATACACACATATATATACACATATATACACACATATATACACATATA 537
              | |||| ||| ||||      |||||||  | ||||||  | | |||| |      ||||
Db    1507588 ATATATGTAT-ATATGTGTGTATATATATGTATATATATGTATATATATGTGTGTGTATA 1507646

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATA 591
              || |  ||||||||   |||||||    | |||| ||||||  | |||||||||
Db    1507647 TATATGTATATATATGTATATATATGTGTGTATATATATATGTATATATATATA 1507700
```

RESULT 39
US-10-257-166-11/c
; Sequence 11, Application US/10257166
; Publication No. US20040023230A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Method and Nucleic Acids for Analysing the Methylation of
; TITLE OF INVENTION: Genes Implicated in Pharmacogenomics
; FILE REFERENCE: 5013.1011
; CURRENT APPLICATION NUMBER: US/10/257,166
; CURRENT FILING DATE: 2002-10-07
; PRIOR APPLICATION NUMBER: PCT/EP01/07470
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-06-29
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 178
; SEQ ID NO 11
; LENGTH: 12069
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (390, 398)
US-10-257-166-11

Query Match 6.3%; Score 261.4; DB 17; Length 12069;
Best Local Similarity 70.5%; Pred. No. 8.5e-43;
Matches 363; Conservative 0; Mismatches 151; Indels 1; Gaps 1;

```
Qy        204 ATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACAC 263
              || ||   | | | | ||     | |||||||||||| |||| |||||| | | | | |
```

```
Db       9846 ATATCTATATATATATAAATATATCTATATATATATAAATATCTATATATATATAAATAT 9787

Qy        264 ACATATATATATACACATATATATACACATATATACATATATACACATATATACACATAT 323
              |||||  |||| | || ||||||| | ||| |||  ||||| ||| |||||  | ||||
Db       9786 CTATATACGTATAAATATCTATATATATATAAATATCTATATCCACGTATATCTATATAT 9727

Qy        324 ATATACACATACATATGTACACATATATATACACATATGTATACACATATATACACATAT 383
              ||||| | ||  |||| || | | ||||||| |    | |||||   ||||||| | | ||
Db       9726 ATATAAATATCTATATATATATAAATATATATATCCACGTATATCTATATATATATAAAT 9667

Qy        384 ATACACACATATATACACATATATACACACATATATACACATATATACACATATATACAC 443
              ||  | | ||||||| | |||||||  ||| |||||  | ||||||| | |||| |||
Db       9666 ATCTATATATATATAAATATATATATCCACGTATATCTATATATATATAAATATCTACGT 9607

Qy        444 ATATACACATATATACACATATATACATATATACACATATATATAATATACACACATATA 503
              ||| | | |||||| ||| |||||  |||||||  | | |||| | |  || | | |||||
Db       9606 ATAAATATATATATCCACGTATATCTATATATATATAAATATCT-ACGTATAAATATATA 9548

Qy        504 TATACACATATATACACACATATATACACATATATACACATATATATACACATATATACA 563
              ||| ||| |||||  | | ||||||| | || |||| | ||||| ||| | ||||  ||
Db       9547 TATCCACGTATATCTATATATATATAAATATCTATATATATATAAATATATATATCCACG 9488

Qy        564 CATATATACATATATACACATATATATACATATATACACATATATACATATATACACATA 623
              ||||  || ||||||| | |||| |||| ||||||| | ||  | ||  ||||||| | |||
Db       9487 TATATCTATATATATATAAATATCTATATATATATAAATATCTCTATATATATAAATATA 9428

Qy        624 TATACATATATACACACATATATACACATACATATACACACACATAGATATACATATATA 683
              ||||   | ||| |   ||||||| | | | || || | | | ||| ||||  |||||||
Db       9427 TATATCCACATATATCTATATATATATAAATATCTATATATATATAAATATCTATATATA 9368

Qy        684 TACACATATATATACGTATATATATGTATATATAT 718
              || | | |||| ||  ||||||||  ||| |||||
Db       9367 TATATAAATATCTATATATATATAAATATCTATAT 9333


RESULT 40
US-10-719-993-6815/c
; Sequence 6815, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001496
;  CURRENT APPLICATION NUMBER: US/10/719,993
;  CURRENT FILING DATE:  2003-11-24
;  NUMBER OF SEQ ID NOS: 55342
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6815
;   LENGTH: 1980090
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1980090)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-719-993-6815

  Query Match             6.3%;  Score 259.6;  DB 18;  Length 1980090;
  Best Local Similarity   84.6%;  Pred. No. 1.8e-41;
  Matches  303;  Conservative    0;  Mismatches   54;  Indels    1;  Gaps    1;

Qy        270 ATATATACACATATATATACACATATATACATATATACACATATATACACATATATATAC 329
              |||||||  | |||  |||||  | |||  | |||||||||  |||||||||  ||||||||||||
Db    1507701 ATATATATATATATACATATATATATACACACATATATATACATATATATACATATATATAC 1507642

Qy        330 ACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACAC 389
              |||  ||||||  ||  ||||||||||||| ||||  ||  |||||||||||||  |||||  ||||
Db    1507641 ACACACATATATATACATATATATACATATATATACACACATATATACATATATACACAC 1507582

Qy        390 ACATATATACACATATATACACACATATATACACATATATACACATATATACACATATAC 449
              |||||||||  |||||||||||||||||||  | |  |||  | |  | |||||  | |  |||  ||
Db    1507581 ACATATATATACATATATACACACATATACATATATACACATATATATACATATATACAC 1507522

Qy        450 ACATATATACACATATATACATATATACACATATATATA-ATATACACACATATATATAC 508
              | |||||||||  |||||  |||  |||||  |||||||||  |  |||||  |||  |||||  |||
Db    1507521 ATATATATACATATATACACACATATATACATATATACACATATATACATATATACATAT 1507462

Qy        509 ACATATATACACACATATATACACATATATACACATATATATACACATATATACACATAT 568
              |  |  |  |||  ||||||||||||||  ||||  ||||||||||||||||  ||||||  |||||||
Db    1507461 ATACACATATACACATATATACATATATCTACACATATATATACATATATATGCACATAT 1507402

Qy        569 ATACATATATACACATATATATACATATATACACATATATACATATATACACATATAT 626
              |||||||||||||||||||||||||||||||  |||  |  |||||||  |||||||  |  ||||  |
Db    1507401 ATACATATATACACATATATATACATACATATATATATATATATATATATATTT 1507344


RESULT 41
US-10-741-600-17676/c
; Sequence 17676, Application US/10741600
; Publication No. US20050026169A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001499
;  CURRENT APPLICATION NUMBER: US/10/741,600
;  CURRENT FILING DATE:  2003-12-22
;  NUMBER OF SEQ ID NOS: 73997
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17676
;   LENGTH: 1980090
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1980090)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-2
US-10-741-600-17676

  Query Match             6.3%;  Score 259.6;  DB 19;  Length 1980090;
  Best Local Similarity   84.6%;  Pred. No. 1.8e-41;
  Matches  303;  Conservative    0;  Mismatches   54;  Indels    1;  Gaps    1;

Qy        270 ATATATACACATATATATACACATATATACATATATACACATATATACACATATATATAC 329
              |||||||  | |||  |||||  | |||  | |||||||||  |||||||||  ||||||||||||
Db    1507701 ATATATATATATATACATATATATATACACACATATATATACATATATATACATATATATAC 1507642

Qy        330 ACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACAC 389
              |||  ||||||  ||  ||||||||||||| ||||  ||  |||||||||||||  |||||  ||||
Db    1507641 ACACACATATATATACATATATATACATATATATACACACATATATACATATATACACAC 1507582

Qy        390 ACATATATACACATATATACACACATATATACACATATATACACATATATACACATATAC 449
              |||||||||  |||||||||||||||||||  | |  |||  | |  | |||||  | |  |||  ||
Db    1507581 ACATATATATACATATATACACACATATACATATATACACATATATATACATATATACAC 1507522

Qy        450 ACATATATACACATATATACATATATACACATATATATA-ATATACACACATATATATAC 508
              | |||||||||  |||||  |||  |||||  |||||||||  |  |||||  |||  |||||  |||
Db    1507521 ATATATATACATATATACACACATATATACATATATACACATATATATACATATATACATAT 1507462

Qy        509 ACATATATACACACATATATACACATATATACACATATATATACACATATATACACATAT 568
              |  |  |  |||  ||||||||||||||  ||||  ||||||||||||||||  ||||||  |||||||
Db    1507461 ATACACATATACACATATATACATATATCTACACATATATATACATATATATGCACATAT 1507402

Qy        569 ATACATATATACACATATATATACATATATACACATATATACATATATACACATATAT 626
              |||||||||||||||||||||||||||||||  |||  |  |||||||  |||||||  |  ||||  |
Db    1507401 ATACATATATACACATATATATACATACATATATATATATATATATATATATTT 1507344


RESULT 42
US-10-017-161-2363
; Sequence 2363, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
```

```
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
;  FILE REFERENCE: 084335/0152
;  CURRENT APPLICATION NUMBER: US/10/017,161
;  CURRENT FILING DATE:  2002-12-18
;  PRIOR APPLICATION NUMBER: JP 2001/246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2430
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2363
;   LENGTH: 1489
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: source
;   LOCATION: (1)..(1489)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(1289)
US-10-017-161-2363

  Query Match             6.3%;  Score 258.8;  DB 15;  Length 1489;
  Best Local Similarity   72.9%;  Pred. No. 1.2e-42;
  Matches  360;  Conservative    0;  Mismatches  132;  Indels    2;  Gaps    2;

Qy        229 TATATATATATATATATATATATACACACACACACACATATATATATACACATATATATA 288
              | ||||||||| ||| |||||||| ||| | | |||||||| |||||| || ||   |||
Db        519 TGTATATATATGTATGTATATATATACATATATACACATATGTATATATACGTACC-ATA 577

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              |  |||||   |||||| || |||||| |||| |||||| || || ||||  ||| |||
Db        578 TATGTATATGTGTATATATACGTATATATACATGTATATATACGTATATATACACATATA 637

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              ||||| |||| | |||| || |||||| || |||| | | | ||| |    | |||||
Db        638 CATATATACATTTATATATACGTATATATACGTATACATATATATACACGTATATATACA 697

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              ||| ||||| ||   |||||| || |||| | | ||| ||    |||||| || |||| |
Db        698 TACATATATACACGTGTATATATACGTATACATATATACACGTGTATATATACGTATACA 757

Qy        469 CATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATATA 527
              |||||||||| | |||||| |  || ||| ||||| |      ||||||||  | | |||
Db        758 TATATATACACGTGTATATATACGTATACATATATACACGTGTATATATACGTATACATA 817

Qy        528 TACACATATATACACATATATATACACATATATACACATATATACATATATACACATATA 587
              || |||  | || | |||  |||||| ||||| ||   ||||||  |  |||||| |||||
Db        818 TATACACGTGTATATATACGTATACATATATACACGTGTATATATACGTATACATATATA 877

Qy        588 TATACATATATACACATATATACATATATACACATATATACATATATACACACATATATA 647
              |    |||||||| | ||||| | |   | | ||| |||  |||||||||    ||||||
Db        878 CACGTGTATATACATACATATATACACGTGTATATACATATGTATATACACGTGTATATA 937

Qy        648 CACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATA 707
                |||  |||||| | || || |||||||| |||||||| |||||||||| ||||||
Db        938 TGTATATGTATACATATACGTATATATACATGTATATACACGTATATATACGTGTATATA 997

Qy        708 TGTATATATATATG 721
              |||||| ||||| |
Db        998 TGTATACATATACG 1011


RESULT 43
US-10-292-798-2007
; Sequence 2007, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
;  FILE REFERENCE: 084335/166
;  CURRENT APPLICATION NUMBER: US/10/292,798
;  CURRENT FILING DATE:  2002-11-13
;  PRIOR APPLICATION NUMBER: 10/017,161
;  PRIOR FILING DATE: 2001-12-18
;  PRIOR APPLICATION NUMBER: JP 2001-246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2070
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2007
;   LENGTH: 1489
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   LOCATION: source
;   FEATURE:
;   LOCATION: (1)..(1489)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(1289)
US-10-292-798-2007

  Query Match             6.3%;  Score 258.8;  DB 17;  Length 1489;
  Best Local Similarity   72.9%;  Pred. No. 1.2e-42;
  Matches  360;  Conservative    0;  Mismatches  132;  Indels    2;  Gaps    2;

Qy        229 TATATATATATATATATATATATACACACACACACACATATATATATACACATATATATA 288
              | ||||||||| ||| |||||||| ||| | | |||||||| |||||| || ||   |||
Db        519 TGTATATATATGTATGTATATATATACATATATACACATATGTATATATACGTACC-ATA 577

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              |  |||||   |||||| || |||||| |||| |||||| || || ||||  ||| |||
Db        578 TATGTATATGTGTATATATACGTATATATACATGTATATATACGTATATATACACATATA 637

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              ||||| |||| | |||| || |||||| || |||| | | | ||| |    | |||||
Db        638 CATATATACATTTATATATACGTATATATACGTATACATATATATACACGTATATATACA 697

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              ||| ||||| ||   |||||| || |||| | | ||| ||    |||||| || |||| |
Db        698 TACATATATACACGTGTATATATACGTATACATATATACACGTGTATATATACGTATACA 757

Qy        469 CATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATATA 527
              |||||||||| | |||||| |  || ||| ||||| |      ||||||||  | | |||
Db        758 TATATATACACGTGTATATATACGTATACATATATACACGTGTATATATACGTATACATA 817

Qy        528 TACACATATATACACATATATATACACATATATACACATATATACATATATACACATATA 587
              || |||  | || | |||  |||||| ||||| ||   ||||||  |  |||||| |||||
Db        818 TATACACGTGTATATATACGTATACATATATACACGTGTATATATACGTATACATATATA 877

Qy        588 TATACATATATACACATATATACATATATACACATATATACATATATACACACATATATA 647
              |    |||||||| | ||||| | |   | | ||| |||  |||||||||    ||||||
Db        878 CACGTGTATATACATACATATATACACGTGTATATACATATGTATATACACGTGTATATA 937

Qy        648 CACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATA 707
                |||  |||||| | || || |||||||| |||||||| |||||||||| ||||||
Db        938 TGTATATGTATACATATACGTATATATACATGTATATACACGTATATATACGTGTATATA 997

Qy        708 TGTATATATATATG 721
              |||||| ||||| |
Db        998 TGTATACATATACG 1011


RESULT 44
US-10-367-094-179/c
; Sequence 179, Application US/10367094
; Publication No. US20040170982A1
; GENERAL INFORMATION:
;  APPLICANT: David W. Morris
;  APPLICANT:  Marc Malandro
;  TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
;  FILE REFERENCE: 529452001500
;  CURRENT APPLICATION NUMBER: US/10/367,094
```

```
;  CURRENT FILING DATE:  2003-02-14
;  NUMBER OF SEQ ID NOS: 203
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 179
;   LENGTH: 163382
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(163382)
;   OTHER INFORMATION: n = A,T,C or G
US-10-367-094-179

  Query Match             6.3%;  Score 258.6;  DB 18;  Length 163382;
  Best Local Similarity   70.0%;  Pred. No. 9.9e-42;
  Matches  362;  Conservative    0;  Mismatches  154;  Indels    1;  Gaps    1;

Qy        204 ATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATATACACACACACAC 263
              || | |  | | | |||        |||| ||||||| |||| |||| | ||| | | |
Db     145185 ATATAAATATATATAAATAAATAAATATAAATATATAAATATGTATAAATACATATATAA 145126

Qy        264 ACATATATATATACACATATATATACACATATATACATATATACACATATATACACATAT 323
              | | ||||| ||| | | ||||||| | ||| ||| ||| ||| | ||| ||| | ||||
Db     145125 ATAAATATAAATATATAAATATATATAAATACATATATAAATATATATAAATATATATAT 145066

Qy        324 ATATACACATACATATGTACACATATATATACACATATGTATACACATATATACACATAT 383
                ||| ||||| ||||||| | ||||||| ||| ||||  ||| | | ||||| | | ||
Db     145065 TAATATACATAAATATGTATAAATATATAAACATATATTAATATATAAATATATATAAAT 145006

Qy        384 ATACACACATATATACACATATATACACACATATATACACATATATACACATATATACAC 443
              |||   | |||||||   ||||||| | | ||  ||| | ||| |||     ||||| ||
Db     145005 ATATTAATATATATATTTATATATAAATATATGAATATATATAAATATGTTAATATATAC 144946

Qy        444 ATATACACATATATACACATATATACATATATACACATATATATAATATACACACATATA 503
              ||||| | | ||| | | ||| | | |||||||| | |||||||    ||| | | |||||
Db     144945 ATATATATAAATAAATATATAAAAATATATATAAATATATATAACTAATATATAAATATA 144886

Qy        504 TATACACATATATACACACATATATACACATATATACACATATATATACACATATATACA 563
              |||| ||  ||||| | | ||||||| | ||||||  ||  ||||||| | ||||||| |
Db     144885 TATA-ACTAATATATATAAATATATAAATATATATGAACTAATATATATAAATATATAAA 144827

Qy        564 CATATATACATATATACACATATATATACATATATACACATATATACATATATACACATA 623
              | ||||| | ||| | | ||| ||| | ||||| | | ||||||| ||| ||| | | |
Db     144826 TAAATATATAAATAAATATATAAATAAATATATACAAATATATATAAATAAATATATAAA 144767

Qy        624 TATACATATATACACACATATATACACATACATATACACACACATAGATATACATATATA 683
               ||| ||||| | | | ||||| | | ||| ||||| | | | ||| ||| | | |||||
Db     144766 AATAAATATAAATAAATATATAAATAAATATATATAAATAAATATATATAAATAAATATA 144707

Qy        684 TACACATATATATACGTATATATATGTATATATATAT 720
              || | ||| |||||  || |||||| |||| ||| ||
Db     144706 TATAAATAAATATATATAAATATATATATAAATAAAT 144670


RESULT 45
US-09-818-657-3
; Sequence 3, Application US/09818657
; Publication No. US20030077773A1
; GENERAL INFORMATION:
;  APPLICANT: BRANDON, Rhonda et al.
;  TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
;  TITLE OF INVENTION:  NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
;  TITLE OF INVENTION:  AND USES THEREOF
;  FILE REFERENCE: CL001006-CIP
;  CURRENT APPLICATION NUMBER: US/09/818,657
;  CURRENT FILING DATE:  2001-03-28
;  PRIOR APPLICATION NUMBER: 09/730,002
;  PRIOR FILING DATE: 2000-12-06
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 82938
;   TYPE: DNA
;   ORGANISM: Human
US-09-818-657-3

  Query Match             6.3%;  Score 258.4;  DB 10;  Length 82938;
  Best Local Similarity   70.5%;  Pred. No. 8.1e-42;
  Matches  361;  Conservative    0;  Mismatches  146;  Indels    5;  Gaps    1;

Qy        214 AAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACACATATATAT 273
              | | | | |    | |||| | | | |||| |||| |||||||||    | |||||| |||
Db       6946 ATACATATGTATATATATACACACACATATGTATACATACACACATGTATACATATGTAT 7005

Qy        274 ATACACATATATATACACATATATACATATATACACATATATACACATATATATACACAT 333
              |||   ||||| |||||||| |||||||| || | || |||| | ||| | ||  | |
Db       7006 ATATGTATATAGATACACATGTATACATATGTATATATGTATATAGATACACATGTATAC 7065

Qy        334 ACATATGTACACATATATATACACATATGTATACACATATATACACATATATACACACAT 393
              | || | || |  |||||| | ||| |||||||||| || |||| |  |||||| | |||
Db       7066 ATATGTATATATGTATATAGATACACATGTATACATATGTATATATGTATATAGATACAC 7125

Qy        394 ATATACACATATATACACACATATATACACATATATACACATATATACACATATACACAT 453
              || || |||||| || | |  |||||| | | | ||  | | ||||  | ||||  | ||
Db       7126 ATGTATACATATGTATATATGTATATAGATACACATGTATACATATGTATATATGTATAT 7185

Qy        454 ATATACACATATATACATATATACACATATATATA-----ATATACACACATATATATAC 508
              | |||||||| |||||||| || | || ||||||      | ||  | |||||| ||||
Db       7186 AGATACACATGTATACATATGTATATATGTATATAGATACACATGTATACATATGTATAT 7245

Qy        509 ACATATATACACACATATATACACATATATACACATATATATACACATATATACACATAT 568
              |  |||||| | ||| || || |||||| || | || |||||||||  ||| || | || |
Db       7246 ATGTATATAGATACACATGTATACATATGTATATATGTATATACATGTATGTATATATGT 7305

Qy        569 ATACATATATACACATATATATACATATATACACATATATACATATATACACATATATAC 628
              |||  ||||||||  ||| ||||  | ||||  ||||||  |||| || | | || ||
Db       7306 ATATGTATATACATGTATGTATATGTGTATATGTATATATGTATATGTATATACATGTAT 7365

Qy        629 ATATATACACACATATATACACATACATATACACACACATAGATATACATATATATACAC 688
              ||||||  |||  |||||||| |  ||||| | |||  || |||   ||||| ||| |
Db       7366 GTATATATACAGCTATATACATGTGTATATATATACAGCTATATACGTATATACATATAT 7425

Qy        689 ATATATATACGTATATATATGTATATATATAT 720
              || |||||| ||||| |||||| ||||||||
Db       7426 ATGTATATATGTATACATATGTGTATATATAT 7457


RESULT 46
US-10-017-161-2361
; Sequence 2361, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
;  FILE REFERENCE: 084335/0152
;  CURRENT APPLICATION NUMBER: US/10/017,161
;  CURRENT FILING DATE:  2002-12-18
;  PRIOR APPLICATION NUMBER: JP 2001/246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2430
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2361
;   LENGTH: 1486
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: source
;   LOCATION: (1)..(1486)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(1286)
US-10-017-161-2361
```

Query Match 6.3%; Score 258; DB 15; Length 1486;
Best Local Similarity 70.4%; Pred. No. 1.7e-42;
Matches 345; Conservative 0; Mismatches 145; Indels 0; Gaps 0;

```
Qy       227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATA 286
            | ||||||||| || ||| | |||||| ||| |     |  | |||||||||| || | ||
Db       718 TGTATATATACATGTATGTGTATATATACATATGTGTATGTGTATATATACATATGTGTA 777

Qy       287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
            |    |||||||||||| |   | |||| | | |||| |   |   |  |||| |||| |
Db       778 TGTGTATATATACATATGTGTATGTGTATATATACATATGTGTATGTGTATATATACATA 837

Qy       347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
            | | |||| | | || | |  | |||| || | || |||| | |||||| |  | ||||
Db       838 TGTGTATATATACATATGTGTATATATGTATATATGTATATATACATATGTGTATATATG 897

Qy       407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
            || | |  |||||| ||||||||     ||||||   |||||    ||| || |||| |
Db       898 TATATATGTATATATACATATATGTGTGTATATATGTATATATGTGTATTTATACATGTG 957

Qy       467 TACATATATACACATATATATAATATACACACATATATATACACATATATACACACATAT 526
            |  | | || | |||| ||||   || |    | |||||| |||||||| |   ||||
Db       958 CATGTGTGTATAAATATGTATATATTATATGTGTGTATATATACATATATGTATGTATAT 1017

Qy       527 ATACACATATATACACATATATATACACATATATACACATATATACATATATACACACATAT 586
            ||||| ||||| | |||||||||| | | ||| | ||| |||   |||||| ||| ||||
Db       1018 ATACATATATACATACATATATACATATATACACACATATACGCACATATAGACATATAT 1077

Qy       587 ATATACATATATACACATATATACATATATACACATATATACATATATACACACATATAT 646
            || || ||||| |||||||||||||||||||   |||||||| | ||| | ||||||||
Db       1078 ATGTATATATACACACATATATACATATATGTATATATACACACATATATACATATAT 1137

Qy       647 ACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATAT 706
            |   ||| ||| ||||| | ||| |||||  ||||||||||||||||||   ||||||
Db       1138 ATGTATATATACACACATATATACATATATGTATATATACACATATATATGTATATATAC 1197

Qy       707 ATGTATATAT 716
            |  |||||||
Db       1198 ACATATATAT 1207
```

RESULT 47
US-10-257-166-11
; Sequence 11, Application US/10257166
; Publication No. US20040023230A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Method and Nucleic Acids for Analysing the Methylation of
; TITLE OF INVENTION: Genes Implicated in Pharmacogenomics
; FILE REFERENCE: 5013.1011
; CURRENT APPLICATION NUMBER: US/10/257,166
; CURRENT FILING DATE: 2002-10-07
; PRIOR APPLICATION NUMBER: PCT/EP01/07470
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-06-29
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 178
; SEQ ID NO 11
; LENGTH: 12069
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (390, 398)
US-10-257-166-11

Query Match 6.3%; Score 257.6; DB 17; Length 12069;
Best Local Similarity 62.9%; Pred. No. 5.1e-42;
Matches 433; Conservative 0; Mismatches 249; Indels 6; Gaps 2;

```
Qy       226 TTTTATATATATATATATATATATATACACACACACACACATATATATATACACATATAT 285
            || |||||||||| |||| |||||||| | | | | ||    |||||||||   ||||||
Db       9239 TTATATATATATAGATATTTATATATATATAGATATACGTGGATATATATATTTATATAT 9298

Qy       286 ATACACATATATACATATATACACATATATACACATATATATACACATACATATGTACAC 345
            ||| | |||| || ||||||| | |||||      |||||  |||   |||  |||| || |
Db       9299 ATATAGATATTTATATATATATAGATATACGTGGATATAGATATTTATATATATATAGAT 9358

Qy       346 ATATATATACACATATGTATACACATATATACACATATATACACACATATATACACATAT 405
            || |||||| | ||||   |||    |||||||  | | ||||   | | |||||||  | ||||
Db       9359 ATTTATATATATATATAGATATTTATATATATATAGATATTTATATATATATAGATATAT 9418

Qy       406 ATACACACATATATACACATATATACACATATATACACATATACACATATATACACATAT 465
            |   | | ||||||  | |||||||  | ||||  || | |||||  | ||||  || | ||||
Db       9419 GTGGATATATATATTTATATATATAGAGATATTTATATATATATAGATATTTATATATAT 9478

Qy       466 ATACATATATACACATATATATA-----ATATACACACATATATATACACATATATACAC 520
            ||| |||||     |||||||||     |||||  | | |||| |||| | |||||  | |
Db       9479 ATAGATATACGTGGATATATATATTTATATATATATAGATATTTATATATATATAGATAT 9538

Qy       521 ACATATATACACATATATACACATATATATACACATATATACACATATATACATATATAC 580
            ||  | ||| | |||| || || |||| | |||||||  | |||||      |||||
Db       9539 ACGTGGATATATATATTTATACGTAGATATTTATATATATATAGATATACGTGGATATAT 9598

Qy       581 ACATATATATACATATATACACATATATACATATATACACATATATACATATATACACAC 640
            | || ||||   | ||||  | |||||| | ||||||    |||| |||| || | |
Db       9599 ATATTTATACGTAGATATTTATATATATATAGATATACGTGGATATATATATTTATATAT 9658

Qy       641 ATATATACACATACATATACACACACATAGATATACATATATATACACATATATATACGT 700
            ||||| | |  || ||||| | | | |||  |  | ||||||||  | |||||||||  |
Db       9659 ATATAGATATTTATATATATATAGATATACGTGGATATATATATTTATATATATATAGAT 9718

Qy       701 ATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGG 760
            || ||||| |||||| |||| |        | |||         |  |    ||    |
Db       9719 ATTTATATATATATAGATATACGTGGATATAGATATTTATATATATATAGATATTTATAC 9778

Qy       761 GGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATG 820
            | ||| ||  |   | |  |  | ||   | |||    | |  || || |     |
Db       9779 GTATATAGATATTTATATATATATAGATATTTATATATATATAGATATATTTATATATAT 9838

Qy       821 GCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAA 880
              | |||     ||| ||| |||  | | | ||| |   ||| | ||      | | |
Db       9839 ATAGATATATATGGATATATATATTTATAT-ATATATAATATTTATATATAGATATATAT 9897

Qy       881 GGTAACACTATATATATATATAGTGAAA 908
            ||  | | |||  ||||||||||  || |
Db       9898 GGATATAGTATTTATATATATATAGATA 9925
```

RESULT 48
US-10-292-798-2005
; Sequence 2005, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
; FILE REFERENCE: 084335/166
; CURRENT APPLICATION NUMBER: US/10/292,798
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: 10/017,161
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: JP 2001-246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2070
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2005
; LENGTH: 1486

```
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   LOCATION: source
;   FEATURE:
;   LOCATION: (1)..(1486)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(1286)
US-10-292-798-2005

  Query Match             6.2%;  Score 256.4;  DB 17;  Length 1486;
  Best Local Similarity   70.2%;  Pred. No. 3.6e-42;
  Matches  344;  Conservative    0;  Mismatches  146;  Indels    0;  Gaps    0;

Qy        227 TTTATATATATATATATATATATATACACACACACACATATATATATACACATATATA 286
              | ||||||||| || ||| | |||||| ||| |     |  | |||||||||| || | ||
Db        718 TGTATATATACATGTATGTGTATATATACATATGTGTATGTGTATATATACATATGTGTA 777

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              |    ||||||||||||| |   |  | |||| | | |||| |   |   |  |||| |||| |
Db        778 TGTGTATATATACATATGTGTATGTGTATATATACATATGTGTATGTGTATATATACATA 837

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              | | |||| | | || | |  | |||| || | || |||| | |||||| |  | ||||
Db        838 TGTGTATATATACATATGTGTATATATGTATATATGTATATATACATATGTGTATATATG 897

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | |  |||||| ||||||||     ||||||  |||||    ||| || | || |
Db        898 TATATATGTATATATACATATATGTGTGTATATATGTATATATGTGTATTTATATATGTG 957

Qy        467 TACATATATACACATATATATAATATACACACATATATATACACATATATACACACATAT 526
              |  | | || | |||| ||||   || |     | |||||| |||||||| |   ||||
Db        958 CATGTGTGTATAAATATGTATATATTATATGTGTGTATATATACATATATGTATGTATAT 1017

Qy        527 ATACACATATATACACATATATATACACATATATACACATATATACATATATACACACATAT 586
              ||||| ||||| | |||||||| | | ||| | ||| |||   ||||||| ||| ||||
Db       1018 ATACATATATACATACATATATACATATATACACACATATACGCACATATAGACATATAT 1077

Qy        587 ATATACATATATACACATATATACATATATACACATATATACATATATACACACATATAT 646
              || || ||||| |||||||||||||||||   |||||||| | ||| | ||||||||
Db       1078 ATGTATATATACACACATATATACATATATGTATATATACACACATATATACATATAT 1137

Qy        647 ACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATAT 706
              |   ||| ||| ||||| | ||| |||||  ||||||||||||||||||   ||||||
Db       1138 ATGTATATATACACACATATATACATATATGTATATATACACATATATATGTATATATAC 1197

Qy        707 ATGTATATAT 716
              |  |||||||
Db       1198 ACATATATAT 1207


RESULT 49
US-10-437-963-46068
; Sequence 46068, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 46068
;   LENGTH: 610
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_48971C.1
US-10-437-963-46068

  Query Match             6.2%;  Score 255.8;  DB 18;  Length 610;
  Best Local Similarity   69.6%;  Pred. No. 3.3e-42;
  Matches  364;  Conservative    0;  Mismatches  152;  Indels    7;  Gaps    1;

Qy        197 AAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACA 256
              ||||  || |  | | |   || |||  | || |||  ||  ||   || ||||||   |
Db         39 AAATAACACCGAAAATATTTAATAAGAAAATTAATACGTAGGTACGGATGTATATATGAA 98

Qy        257 CACACACACATATATATATACACATATATATACACATATATACATATATACACATATATA 316
              |   ||  | |||||||   || ||||||     ||| || || ||||   || |
Db         99 AGTATGGACGGAAATATATATGTATGTATATAATAGTATGTATATGTATATCTATGTGAG 158

Qy        317 CACATATATATACACATACATATGTACACATATATATACACATATGTATACACATATATA 376
              |||||||||   ||| |||| || | |||||||| | |||| |||| | |||||||
Db        159 TGTATATATATATGTATATATATATATATATATATATATATATATATATATATATATATA 218

Qy        377 CACATATATACACACATATATACACATATATACACACATATATACACATATATACACATA 436
              | ||||||| | | ||||||| | ||||||| | |  ||| |  | ||||||| | |||
Db        219 TATATATATATATATATATATATATATATATATATATATGTATGTGTATATATATATATA 278

Qy        437 TATACACATATACACATATATACACATATATACATATATACACATATATATAATATACAC 496
              ||||   | |  | ||||||| | ||||||| |||||| | ||||||||       |
Db        279 TATATGTGTGTGTATATATATATATATATATATATATATATATATATATATAT-------AT 331

Qy        497 ACATATATATACACATATATACACACATATATACACATATATACACATATATATACACAT 556
              | |||||||| ||| | | ||||||| || || | |||||||||| | || || | ||
Db        332 ATATATATATATACACACACACACACATGTATATATATATACACACACATGTATATAT 391

Qy        557 ATATACACATATATACATATATACACATATATATACATATATACACATATATACATAT 616
              |||||||||| || | ||| ||| |||||||||| ||||||| ||| || |||||||||
Db        392 ATATACACATGTACATATACATATACATATATATATATATATATACACATGTACATATAT 451

Qy        617 ACACATATATACATATATACACACATATATACACATACATATACACACACATAGATATAC 676
              |||  || ||| ||||||||||||||||||| ||| | ||||| | | | | | |||
Db        452 ACATGTACATATATATATACACACATATATATACACATATATATATATACACACATAT 511

Qy        677 ATATATATACACATATATATACGTATATATATGTATATATATA 719
              |||||||| ||||    |  | | | || |  || || ||||
Db        512 ATATATATATACATGACGACTCTTCTGTACACCTAGATGTATA 554


RESULT 50
US-10-017-161-2363/c
; Sequence 2363, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
;  FILE REFERENCE: 084335/0152
;  CURRENT APPLICATION NUMBER: US/10/017,161
;  CURRENT FILING DATE:  2002-12-18
;  PRIOR APPLICATION NUMBER: JP 2001/246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2430
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2363
;   LENGTH: 1489
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: source
;   LOCATION: (1)..(1489)
;   FEATURE:
;   NAME/KEY: CDS
```

```
;   LOCATION: (201)..(1289)
US-10-017-161-2363

  Query Match             6.2%;  Score 255.4;  DB 15;  Length 1489;
  Best Local Similarity   63.8%;  Pred. No. 5.8e-42;
  Matches  438;  Conservative    0;  Mismatches  241;  Indels    8;  Gaps    3;

Qy        236 ATATATATATATATATACACACACACACACATATATATATACACATATATATACACATAT 295
              | || | |||||||||||| | | ||  | |||| | |||| | | ||| |||   ||||
Db       1197 ACATGTGTATATATATACATATATACGTATATATGTGTATATATACATACATAGGTATAT 1138

Qy        296 ATACATATATACACATATATACACATATATATACACATACATATGTACACATATATATAC 355
              | | ||||| ||| | ||||| || |||||||  | |||  ||| || |||| ||||||
Db       1137 ACATATATACACATACATATATACGTATATATGTATATATGTATATATACATGTATATAT 1078

Qy        356 ACATATG-TATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              ||||| | || |||||||||||| |||||| ||  |||| || |||||||||||| |
Db       1077 GTATATGTTGTATACATATATACACGTATATATACGTATATGTATACATATATACACGTA 1018

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              |||||||  |||| || |||||||||||| |||| | |  | || | | || || |||||
Db       1017 TATATACGTATATGTATACATATATACACGTATATATACGTGTATATACATGTATATATA 958

Qy        475 TACACATATATATAATATACACACATATATATACACATATATACACACATATATACACAT 534
              | || |     ||| | | |||||||||| ||   |||||||| |  |||||||||  |
Db        957 CGTATATGT-----ATACATATACATATATACACGTGTATATACATATGTATATACACGT 903

Qy        535 ATATACACATATATATACACATATATACACATATATACATATATACACATATATATACAT 594
              |||| |  ||| |||| |||  | || | || ||||| |||||| |||  | |||| ||
Db        902 GTATATATGTATGTATATACACGTGTATATATGTATACGTATATATACACGTGTATATAT 843

Qy        595 ATATACACATATATACATATATACACATATATACATATATACACACATATATACACATAC 654
              |||||  |||||||||  | || | || ||||| |||||| |||| | |||| |  ||
Db        842 GTATACGTATATATACACGTGTATATATGTATACGTATATATACACGTGTATATATGTAT 783

Qy        655 ATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATAT 714
              |  || | | |||     |||| |||| |||||  ||||||| |||| ||||||||||||
Db        782 ACGTATATATACACGTGTATATATATGTATACGTATATATACACGTGTATATATGTATAC 723

Qy        715 ATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAG 774
              |||||   | | ||  ||| |       |  |   |      ||| | | | |
Db        722 GTATATATACACGTGTATATATGTATGTATATATACGTGTATATATATGTATACGTATAT 663

Qy        775 AGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGG 834
              | |  |  | |  | |||||    |     | | |       |     | |||   |
Db        662 ATACGTATATATAAATGTAT--ATATGTATATGTGTATATATACGTATATATACATGTAT 605

Qy        835 AAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATA 894
              | |||| || || |   | || ||| ||  | |||  ||   | |      |||||
Db        604 ATATACGTATATATACACATATACATATATGGTACGTATATATACATATGTGTATATATG 545

Qy        895 TATATATAGTGAAAGCAGTGTAAACCT 921
              ||||||||   | |    | || || |
Db        544 TATATATATACATACATATATATACAT 518


RESULT 51
US-10-292-798-2007/c
; Sequence 2007, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
;  FILE REFERENCE: 084335/166
;  CURRENT APPLICATION NUMBER: US/10/292,798
;  CURRENT FILING DATE:  2002-11-13
;  PRIOR APPLICATION NUMBER: 10/017,161
;  PRIOR FILING DATE: 2001-12-18
;  PRIOR APPLICATION NUMBER: JP 2001-246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2070
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2007
;   LENGTH: 1489
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   LOCATION: source
;   FEATURE:
;   LOCATION: (1)..(1489)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(1289)
US-10-292-798-2007

  Query Match             6.2%;  Score 255.4;  DB 17;  Length 1489;
  Best Local Similarity   63.8%;  Pred. No. 5.8e-42;
  Matches  438;  Conservative    0;  Mismatches  241;  Indels    8;  Gaps    3;

Qy        236 ATATATATATATATATACACACACACACACATATATATATACACATATATATACACATAT 295
              | || | |||||||||||| | | ||  | |||| | |||| | | ||| |||   ||||
Db       1197 ACATGTGTATATATATACATATATACGTATATATGTGTATATATACATACATAGGTATAT 1138

Qy        296 ATACATATATACACATATATACACATATATATACACATACATATGTACACATATATATAC 355
              | | ||||| ||| | ||||| || |||||||  | |||  ||| || |||| ||||||
Db       1137 ACATATATACACATACATATATACGTATATATGTATATATGTATATATACATGTATATAT 1078

Qy        356 ACATATG-TATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              ||||| | || |||||||||||| |||||| ||  |||| || |||||||||||| |
Db       1077 GTATATGTTGTATACATATATACACGTATATATACGTATATGTATACATATATACACGTA 1018

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              |||||||  |||| || |||||||||||| |||| | |  | || | | || || |||||
Db       1017 TATATACGTATATGTATACATATATACACGTATATATACGTGTATATACATGTATATATA 958

Qy        475 TACACATATATATAATATACACACATATATATACACATATATACACACATATATACACAT 534
              | || |     ||| | | |||||||||| ||   |||||||| |  |||||||||  |
Db        957 CGTATATGT-----ATACATATACATATATACACGTGTATATACATATGTATATACACGT 903

Qy        535 ATATACACATATATATACACATATATACACATATATACATATATACACATATATATACAT 594
              |||| |  ||| |||| |||  | || | || ||||| |||||| |||  | |||| ||
Db        902 GTATATATGTATGTATATACACGTGTATATATGTATACGTATATATACACGTGTATATAT 843

Qy        595 ATATACACATATATACATATATACACATATATACATATATACACACATATATACACATAC 654
              |||||  |||||||||  | || | || ||||| |||||| |||| | |||| |  ||
Db        842 GTATACGTATATATACACGTGTATATATGTATACGTATATATACACGTGTATATATGTAT 783

Qy        655 ATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATAT 714
              |  || | | |||     |||| |||| |||||  ||||||| |||| ||||||||||||
Db        782 ACGTATATATACACGTGTATATATATGTATACGTATATATACACGTGTATATATGTATAC 723

Qy        715 ATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAG 774
              |||||   | | ||  ||| |       |  |   |      ||| | | | |
Db        722 GTATATATACACGTGTATATATGTATGTATATATACGTGTATATATATGTATACGTATAT 663

Qy        775 AGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGG 834
              | |  |  | |  | |||||    |     | | |       |     | |||   |
Db        662 ATACGTATATATAAATGTAT--ATATGTATATGTGTATATATACGTATATATACATGTAT 605

Qy        835 AAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATA 894
              | |||| || || |   | || ||| ||  | |||  ||   | |      |||||
Db        604 ATATACGTATATATACACATATACATATATGGTACGTATATATACATATGTGTATATATG 545

Qy        895 TATATATAGTGAAAGCAGTGTAAACCT 921
              ||||||||   | |    | || || |
Db        544 TATATATATACATACATATATATACAT 518


RESULT 52
US-10-322-281-123/c
; Sequence 123, Application US/10322281
; Publication No. US20040126762A1
; GENERAL INFORMATION:
```

```
;  APPLICANT: David W. Morris
;  APPLICANT:  Marc S. Malandro
;  TITLE OF INVENTION: Novel Compositions and Methods in Cancer
;  FILE REFERENCE: 529452001000
;  CURRENT APPLICATION NUMBER: US/10/322,281
;  CURRENT FILING DATE:  2002-12-17
;  NUMBER OF SEQ ID NOS: 866
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 123
;   LENGTH: 104399
;   TYPE: DNA
;   ORGANISM: Mus musculus
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(104399)
;   OTHER INFORMATION: n = A,T,C or G
US-10-322-281-123

  Query Match           6.2%;  Score 255.4;  DB 18;  Length 104399;
  Best Local Similarity   73.3%;  Pred. No. 3.7e-41;
  Matches  367;  Conservative    0;  Mismatches  131;  Indels    3;  Gaps    3;

Qy        230 ATATATATATATATATATATATACACACACACACACATATATATATACACATATATATAC 289
              ||| ||| ||| ||||||| |||| ||| | | ||||| ||| ||||  | ||||||||
Db      13721 ATACATACATACATATATACATACGTACATATATACATACATACATACGTACATATATAC 13662

Qy        290 ACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATAT 349
              | |  || | ||||| | ||||| ||||| ||||| ||||  | | ||||  | |||||
Db      13661 ATACGTACATATATACATACATACATACATATATATACATACGTACATATATACATACATAC 13602

Qy        350 ATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATAC 409
              |||   |||||| || | || || ||| | | ||| | ||| | ||||| |||||  |||
Db      13601 ATACGTACATATATACATACGTACATATATACATACATACATACATATATACATACGTAC 13542

Qy        410 ACACATATATACACATATATACACATATATACACATATACACATATATACACATATATAC 469
              | | ||| ||||| | |||   ||||||||||| | ||||| | ||||| ||||| ||||
Db      13541 ATATATACATACATACATACGTACATATATACATACATACATACATATATACATACATAC 13482

Qy        470 ATATATACACATATATATAATATACACACATATATATACACATATATACACACATATATA 529
              |||||||| | ||| |||||| | | | | ||  ||| | | ||| | ||| ||||| |
Db      13481 ATATATACATACATACATAATACATATATACATTCATATATACATA-ATACATATATACA 13423

Qy        530 CACATATATACACATATATAT-ACACATATATACACATAT-ATACATATATACACATATA 587
              |||||||||||| | ||| || | |||||||||||| | || ||||||| || || |||||
Db      13422 TACATATATACATACATACATAATACATATATACATACATAATACATACATGCATATATA 13363

Qy        588 TATACATATATACACATATATACATATATACACATATATACATATATACACACATATATA 647
              || |||| ||| | | ||| ||| | ||||| | ||| | | ||||| ||| |  || |
Db      13362 CATGCATACATATATACATACACACACATACATACATACATATATATACATACGTACA 13303

Qy        648 CACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATA 707
              | ||||||| | ||| || || ||||| | ||| ||||| ||||| |||| || |||||
Db      13302 TATATACATACATACATACGTACATATATACATACATACATATATACATACATACATATA 13243

Qy        708 TGTATATATATATGCTCCAGA 728
              |  ||| ||| ||  | || |
Db      13242 TACATACATACATAATACATA 13222


RESULT 53
US-10-292-798-607/c
; Sequence 607, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
;  FILE REFERENCE: 084335/166
;  CURRENT APPLICATION NUMBER: US/10/292,798
;  CURRENT FILING DATE:  2002-11-13
;  PRIOR APPLICATION NUMBER: 10/017,161
;  PRIOR FILING DATE: 2001-12-18
;  PRIOR APPLICATION NUMBER: JP 2001-246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2070
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 607
;   LENGTH: 84510
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   LOCATION: source
;   FEATURE:
;   LOCATION: (1)..(84510)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(272)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (29879)..(29999)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (45509)..(45616)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (50871)..(50974)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (58541)..(58638)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (60172)..(60304)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (62427)..(62580)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (71080)..(71140)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (72348)..(72417)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (75647)..(75738)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (77542)..(77668)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (82828)..(82869)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (84170)..(84310)
US-10-292-798-607

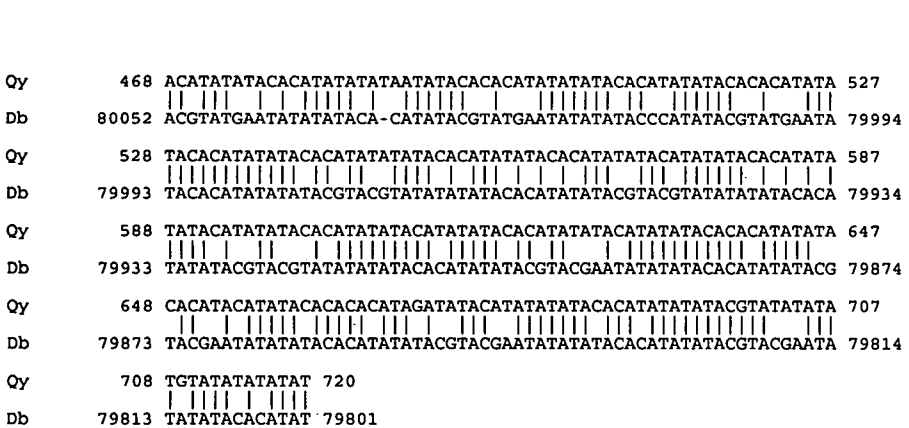
  Query Match           6.2%;  Score 255;  DB 17;  Length 84510;
  Best Local Similarity   72.6%;  Pred. No. 4e-41;
  Matches  358;  Conservative    0;  Mismatches  130;  Indels    5;  Gaps    2;

Qy        232 ATATATATATATATATATATACACACACACACACATATATATATACACATATATATAC 291
              ||||||||| ||||   |||| | | | ||||| || |||||||||| ||  ||||| ||
Db      80292 ATATATATACATATTCTTATATATATATACACATATGTATATATACATATGAATATATAC 80233

Qy        292 ATATATACATATATACACATATATACACATATATATACACATACATATGTACACATATAT 351
              | ||||  |||||||| ||  ||| | | | ||||||| ||  |||| ||||||||||
Db      80232 ACATATGTATATATACATATGAATATATACACATATACATATGAATATATACACATATAC 80173

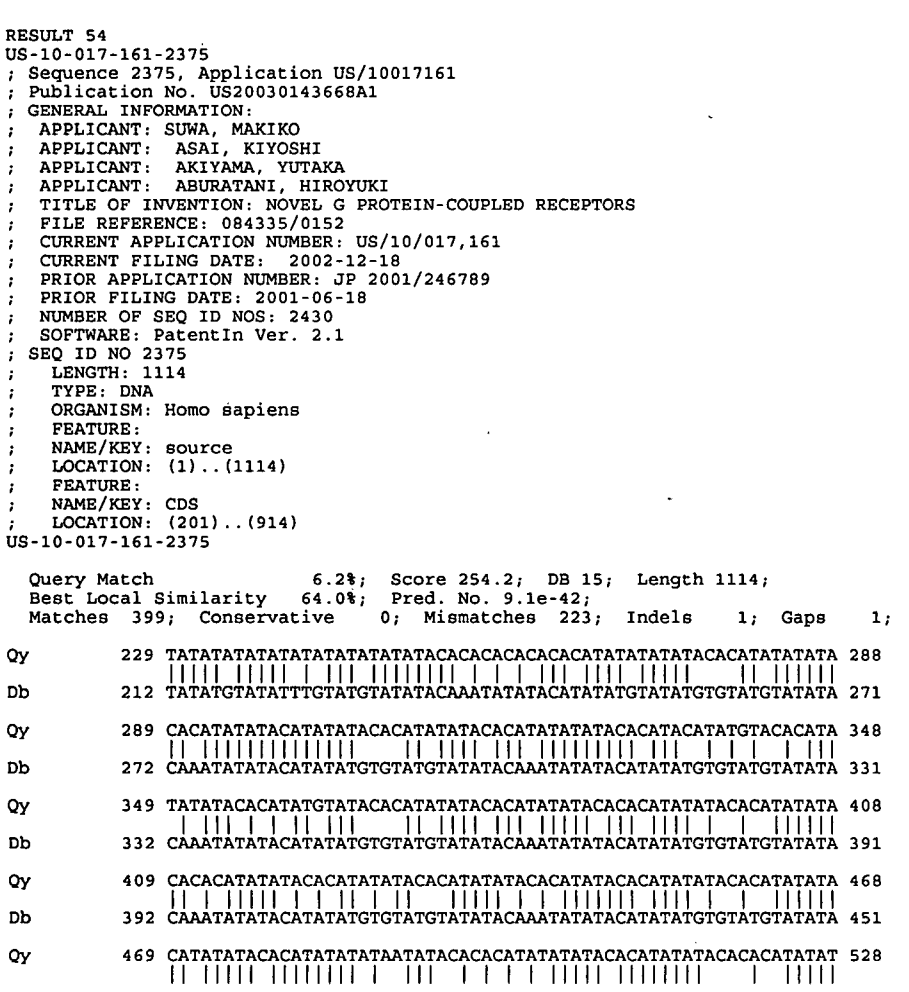
Qy        352 ATACACATATGTATACACATATATACACATATATACACACATATATACACATATATACAC 411
              |||   |||| || | | ||| |||   |||||||||| |||||   |   ||||| ||
Db      80172 ATATGAATATATACATATATACATATGAATATATACACATATATACGTATGAATATATAC 80113

Qy        412 ACATATATAC----ACATATATACACATATATACACATATACACATATATACACATATAT 467
              ||||||||||      ||||||| ||| ||||| || |||  | ||||||||||||||||
Db      80112 ACATATATACGTATGAATATATATACACATATATACGTATGAATATATATACACATATAT 80053
```



RESULT 54
US-10-017-161-2375
; Sequence 2375, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 084335/0152
; CURRENT APPLICATION NUMBER: US/10/017,161
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: JP 2001/246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2430
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2375
; LENGTH: 1114
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: source
; LOCATION: (1)..(1114)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (201)..(914)
US-10-017-161-2375

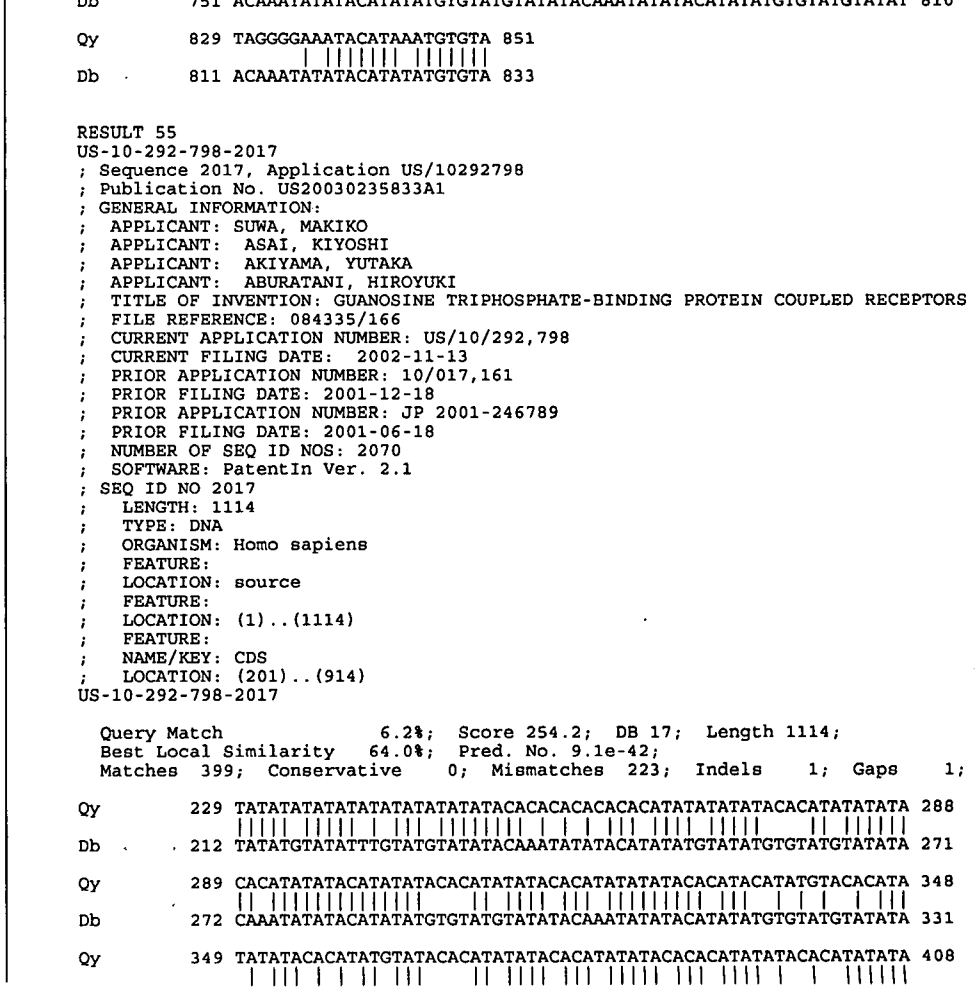
Query Match 6.2%; Score 254.2; DB 15; Length 1114;
Best Local Similarity 64.0%; Pred. No. 9.1e-42;
Matches 399; Conservative 0; Mismatches 223; Indels 1; Gaps 1;





RESULT 55
US-10-292-798-2017
; Sequence 2017, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
; FILE REFERENCE: 084335/166
; CURRENT APPLICATION NUMBER: US/10/292,798
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: 10/017,161
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: JP 2001-246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2070
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2017
; LENGTH: 1114
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; LOCATION: source
; FEATURE:
; LOCATION: (1)..(1114)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (201)..(914)
US-10-292-798-2017

Query Match 6.2%; Score 254.2; DB 17; Length 1114;
Best Local Similarity 64.0%; Pred. No. 9.1e-42;
Matches 399; Conservative 0; Mismatches 223; Indels 1; Gaps 1;



```
Db        332 CAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATATA 391

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              || | ||||| | | || | ||   |||| | | ||||||| |||| |   |  ||||||
Db        392 CAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATATA 451

Qy        469 CATATATACACATATATATAATATACACACATATATATACACATATATACACACATATAT 528
              || ||||| |||||||| |  |||  | | | | ||||| ||||||||     |  |||||
Db        452 CAAATATATACATATATGT-GTATGTATATACAAATATATACATATATGTGTATGTATAT 510

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACATATAT 588
              |||| |||||||||| |||| | ||   ||||| | | ||||| | |||| |   |  |||||
Db        511 ACAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATAT 570

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
              | | |||||||||| |||| |  || |||| ||| ||||| | |||||      || ||||
Db        571 ACAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATAT 630

Qy        649 ACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATAT 708
              ||| | ||||||| | |   |  || || ||| | ||| | | ||||||  |||| |||||
Db        631 ACAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATAT 690

Qy        709 GTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAG 768
                | |||||||       | ||   ||       |  |  |   ||    ||    || |
Db        691 ACAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATAT 750

Qy        769 GAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAG 828
              |||  | |     | |   || | |    | ||  | ||       ||        |||
Db        751 ACAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATAT 810

Qy        829 TAGGGGAAATACATAAATGTGTA 851
                  | |||||| |||||||
Db        811 ACAAATATATACATATATGTGTA 833


RESULT 56
US-09-728-552-3/c
; Sequence 3, Application US/09728552
; Publication No. US20030096398A1
; GENERAL INFORMATION:
;  APPLICANT: Choo, Kong-Hong Andy
;  APPLICANT:  Du Sart, Desiree
;  APPLICANT:  Cancilla, Michael R.
;  TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
;  FILE REFERENCE: Davies Col
;  CURRENT APPLICATION NUMBER: US/09/728,552
;  CURRENT FILING DATE:  2000-12-02
;  PRIOR APPLICATION NUMBER: 09/078,294
;  PRIOR FILING DATE: 1998-05-13
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 80595
;   TYPE: DNA
;   ORGANISM: Nucleotide sequence of HC-contig
US-09-728-552-3

  Query Match             6.2%;  Score 253.6;  DB 10;  Length 80595;
  Best Local Similarity   71.9%;  Pred. No. 7.7e-41;
  Matches  361;  Conservative    0;  Mismatches  134;  Indels    7;  Gaps    2;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | |||| |||| ||| | |||||| ||| | | ||| |  ||| | |||| ||| |||||
Db      16026 TATATACATATTTATGTCTATATACACATATATACATATTTATGTCTATATACACATATA 15967

Qy        287 TACACATATATACATATATACACATATATACACA------TATATATACACATACATATG 340
              |||| || |||   |||||||||||||||||| |      | ||||||||||| |||
Db      15966 TACATATTTATGTCTATATACACATATATACATATTTATGTCTATATACACATATATACA 15907

Qy        341 TACACATATATATACACATATGTATACACATATATACACATATATACACACATATATACA 400
              ||   || | |||| ||| || |||||| || |||     ||||| ||| | ||| |||
Db      15906 TATTTATGTCTATATACACATATATACATATTTATGTCTATATACACATATATACATATT 15847

Qy        401 CATATATACACACATATATACACATATATACACATATATACACATATACACATATATACA 460
              || | || | ||| ||||| |||||| ||    ||||||||||||||| |||||| ||
Db      15846 TATGTCTATATACACATATATACATATTTATGTCTATATACACATATATACATATTTATG 15787

Qy        461 CATATATACATATATACACATATATAT-AATATACACACATATATATACACATATATACA 519
              ||||||||| ||||| |||||| |||    |||| ||||||||||| | |  ||| | |
Db      15786 TCTATATACACATATATACATATTTATGTCTATATACACATATATACATATTTATGTCTA 15727

Qy        520 CACATATATACACATATATACACATATATATACACATATATACACATATATACATATATA 579
              | | | ||| | | ||||  |  | ||||||||||||||||||| || |||   ||||||
Db      15726 TATACACATATATACATATTTATGTCTATATACACATATATACATATTTATGTCTATATA 15667

Qy        580 CACATATATATACATATATACACATATATACATATATACACATATATACATATATACACA 639
              |||||||||| | || |||  | ||||| |||||||||| || |||   ||||||||||
Db      15666 CACATATATACATATTTATGTATATATACACATATATACATATTTATGTCTATATACACA 15607

Qy        640 CATATATACACATACATATACACACACATAGATATACATATATATACACATATATATACG 699
              |||||  |  |  ||  |||| |    |  || ||||    | |||||  |||| |||||
Db      15606 TATATACATATTTATGTATATATTTATGTATATATGTGTGTATATGTACATACATATATA 15547

Qy        700 TATATATATGTATATATATATG 721
              |||  |||  ||  |||||||||
Db      15546 CATACATAAATACATATATATG 15525


RESULT 57
US-10-856-888-1/c
; Sequence 1, Application US/10856888
; Publication No. US20040235037A1
; GENERAL INFORMATION:
;  APPLICANT: Blumenfeld, Marta
;  APPLICANT:  Bougueleret, Lydie
;  APPLICANT:  Chumakov, Ilya
;  TITLE OF INVENTION: Polymorphic Markers of Prostate Carcinoma Tumor Antigen-1 (PCTA-1)
;  FILE REFERENCE: GEN-T112XC1
;  CURRENT APPLICATION NUMBER: US/10/856,888
;  CURRENT FILING DATE:  2004-05-27
;  PRIOR APPLICATION NUMBER: US/09/326,402
;  PRIOR FILING DATE: 1999-06-04
;  PRIOR APPLICATION NUMBER: 60/088,187
;  PRIOR FILING DATE: 1998-06-05
;  PRIOR APPLICATION NUMBER: 60/102,324
;  PRIOR FILING DATE: 1998-09-28
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 106746
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 1..68647
;   OTHER INFORMATION: 5'regulation  region
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 66647..68647
;   OTHER INFORMATION: promoter
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 97156..106746
;   OTHER INFORMATION: 3'regulation  region
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 68648..68741
;   OTHER INFORMATION: exon0
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 70647..70794
;   OTHER INFORMATION: exon1
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 82208..82296
;   OTHER INFORMATION: exon2
```

; FEATURE:
; NAME/KEY: exon
; LOCATION: 83613..83823
; OTHER INFORMATION: exon3
; FEATURE:
; NAME/KEY: exon
; LOCATION: 85298..85417
; OTHER INFORMATION: exon4
; FEATURE:
; NAME/KEY: exon
; LOCATION: 86389..86445
; OTHER INFORMATION: exon5
; FEATURE:
; NAME/KEY: exon
; LOCATION: 87496..87522
; OTHER INFORMATION: exon6
; FEATURE:
; NAME/KEY: exon
; LOCATION: 87650..87775
; OTHER INFORMATION: exon6bis
; FEATURE:
; NAME/KEY: exon
; LOCATION: 88295..88383
; OTHER INFORMATION: exon7
; FEATURE:
; NAME/KEY: exon
; LOCATION: 89484..89649
; OTHER INFORMATION: exon8
; FEATURE:
; NAME/KEY: exon
; LOCATION: 92749..97155
; OTHER INFORMATION: exon9
; FEATURE:
; NAME/KEY: exon
; LOCATION: 92749..92883
; OTHER INFORMATION: exon9bis
; FEATURE:
; NAME/KEY: exon
; LOCATION: 95821..97155
; OTHER INFORMATION: exon9ter
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 70647..70794
; OTHER INFORMATION: homology with genset EST : A241850
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 68648..68741
; OTHER INFORMATION: homology with genset EST : A241850
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 82208..82229
; OTHER INFORMATION: homology with genset EST : A241850
; FEATURE:
; NAME/KEY: allele
; LOCATION: 278
; OTHER INFORMATION: 99-1601-278 : polymorphic base A or C
; FEATURE:
; NAME/KEY: allele
; LOCATION: 402
; OTHER INFORMATION: 99-1601-402 : polymorphic base A or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 472
; OTHER INFORMATION: 99-1601-472 : polymorphic base A or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 2955
; OTHER INFORMATION: 99-13801-100 : polymorphic base T or C
; FEATURE:
; NAME/KEY: allele
; LOCATION: 12167
; OTHER INFORMATION: 99-13806-166 : polymorphic base G or A
; FEATURE:
; NAME/KEY: allele
; LOCATION: 12536
; OTHER INFORMATION: 99-13799-376 : polymorphic base T or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 17593
; OTHER INFORMATION: 99-13798-297 : polymorphic base T or C
; FEATURE:
; NAME/KEY: allele
; LOCATION: 17606
; OTHER INFORMATION: 99-13798-284 : polymorphic base T or C
; FEATURE:
; NAME/KEY: allele
; LOCATION: 22079
; OTHER INFORMATION: 99-1602-200 : polymorphic base G or C
; FEATURE:
; NAME/KEY: allele
; LOCATION: 28964
; OTHER INFORMATION: 99-13794-186 : polymorphic base T or C
; FEATURE:
; NAME/KEY: allele
; LOCATION: 29003
; OTHER INFORMATION: 99-13794-147 : polymorphic base C or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 31077
; OTHER INFORMATION: 99-13812-384 : polymorphic base T or C
; FEATURE:
; NAME/KEY: allele
; LOCATION: 31766
; OTHER INFORMATION: 99-13805-313 : polymorphic base T or C
; FEATURE:
; NAME/KEY: allele
; LOCATION: 34791
; OTHER INFORMATION: 99-1587-281 : polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 45751
; OTHER INFORMATION: 99-1582-430 : polymorphic base C or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 49847
; OTHER INFORMATION: 99-1585-465 : polymorphic base T or C
; FEATURE:
; NAME/KEY: allele
; LOCATION: 49855
; OTHER INFORMATION: 99-1585-457 : polymorphic base T or C
; FEATURE:
; NAME/KEY: allele
; LOCATION: 49886
; OTHER INFORMATION: 99-1585-426 : polymorphic base G or A
; FEATURE:
; NAME/KEY: allele
; LOCATION: 49900
; OTHER INFORMATION: 99-1585-412 : polymorphic base G or A
; FEATURE:
; NAME/KEY: allele
; LOCATION: 49906
; OTHER INFORMATION: 99-1585-406 : polymorphic base C or A
; FEATURE:
; NAME/KEY: allele
; LOCATION: 49921
; OTHER INFORMATION: 99-1585-391 : polymorphic base C or A
; FEATURE:
; NAME/KEY: allele
; LOCATION: 49939
; OTHER INFORMATION: 99-1585-373 : polymorphic base G or A
; FEATURE:
; NAME/KEY: allele
; LOCATION: 50256
; OTHER INFORMATION: 99-1585-55 : polymorphic base C or A
; FEATURE:
; NAME/KEY: allele

```
;   LOCATION: 54955
;   OTHER INFORMATION: 99-1607-373 : polymorphic base T or C
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 64239
;   OTHER INFORMATION: 99-1577-105 : polymorphic base A or G
;   FEATURE:
;   NAME/KEY: allele

  Query Match             6.1%;  Score 253;  DB 18;  Length 106746;
  Best Local Similarity   69.4%;  Pred. No. 1.1e-40;
  Matches  416;  Conservative    0;  Mismatches  175;  Indels    8;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              |||||||||||||||||  |||||||||||||||||| |||||||||||||||  ||||||
Db      38717 AGTTCAAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAAAATT 38658

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   ||||||||| | ||||||||||||||||||||  |||| | ||||||||||||
Db      38657 AGCTGGGCATGGTGGTGCACGCCTGTAATCCCAGCTACTCAGGAGACTGAGGCAGGAGAA 38598

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCCA 178
              ||||||||||||  || |||||| ||| || ||| ||| |    | ||| |||||||||
Db      38597 TCACTTGAACCTG--GGAGGTGGAGTTTGTGGTGAGCCAAAACTTTCCCACTGCACTCCA 38540

Qy        179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCAT-TTTATATATAT 237
              |||      |   | |   | || |   | |  | |     | |   ||    | ||||
Db      38539 GCCAGTATATATATATATATATATACACACGTACACGTGTACACATATACACACATATGC 38480

Qy        238 ATATATATATATATACACACACACACATATATATATACACATATATATACACATATAT 297
              | |  | || ||| |||||| |||||   |.||| || ||||||| || |  | || ||
Db      38479 ACACGTGTACATACACACACATACACATGCACATACATGCACATATGTACATGCGTACAT 38420

Qy        298 ACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACAC 357
              ||| | ||  |||||||||||||| || || | |  ||||    |||||| || || | |
Db      38419 ACACACATGTACATATATACACACATGTACATATGTACACCCATACACACATGTACATAT 38360

Qy        358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417
               ||      | ||| | ||  |||||| ||||| |||||| | |  || |||   |||
Db      38359 GTACACCCATACACACATGTACATATGTACACCCATATACATATGTACATATGTACACCC 38300

Qy        418 ATACACATATATACACATATATACACATATACACATATATACACATATATACATATATAC 477
              |||||||||| |||| || || || |||| |||  || |||   | |   | | | ||
Db      38299 ATACACATATGTACATATGTACACCCATACACATGTACATATGTACACCCATACACATGT 38240

Qy        478 ACATATATATAATATACACACATATATATACACATATATACACACATATATACACATATA 537
              |||||| |   | || |||| || || |||   ||||  | ||||||| |||| |||||
Db      38239 ACATATGT---ACATCCACATATGTACATATGTATATCCATACACATATGTACATATATA 38183

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATACATAT 596
              ||||||| || | |  || ||| | || | |  |||| ||   |||||||||   ||
Db      38182 CACACATATGTACATATGTACATATATATTTGTGTATATCTATGTATATATATAAGAAT 38124


RESULT 58
US-10-856-888-12/c
; Sequence 12, Application US/10856888
; Publication No. US20040235037A1
; GENERAL INFORMATION:
;  APPLICANT: Blumenfeld, Marta
;  APPLICANT:  Bougueleret, Lydie
;  APPLICANT:  Chumakov, Ilya
;  TITLE OF INVENTION: Polymorphic Markers of Prostate Carcinoma Tumor Antigen-1 (PCTA-1
;  FILE REFERENCE: GEN-T112XC1
;  CURRENT APPLICATION NUMBER: US/10/856,888
;  CURRENT FILING DATE:  2004-05-27
;  PRIOR APPLICATION NUMBER: US/09/326,402
;  PRIOR FILING DATE: 1999-06-04
;  PRIOR APPLICATION NUMBER: 60/088,187
;  PRIOR FILING DATE: 1998-06-05
;  PRIOR APPLICATION NUMBER: 60/102,324
;  PRIOR FILING DATE: 1998-09-28
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
;   LENGTH: 106746
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 1..68647
;   OTHER INFORMATION: 5'regulation  region
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 66647..68647
;   OTHER INFORMATION: promoter
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 97156..106746
;   OTHER INFORMATION: 3'regulation  region
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 68648..68741
;   OTHER INFORMATION: exon0
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 70647..70794
;   OTHER INFORMATION: exon1
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 82208..82296
;   OTHER INFORMATION: exon2
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 83613..83823
;   OTHER INFORMATION: exon3
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 85298..85417
;   OTHER INFORMATION: exon4
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 86389..86445
;   OTHER INFORMATION: exon5
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 87496..87522
;   OTHER INFORMATION: exon6
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 87650..87775
;   OTHER INFORMATION: exon6bis
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 88295..88383
;   OTHER INFORMATION: exon7
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 89484..89649
;   OTHER INFORMATION: exon8
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 92749..97155
;   OTHER INFORMATION: exon9
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 92749..92883
;   OTHER INFORMATION: exon9bis
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 95821..97155
;   OTHER INFORMATION: exon9ter
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 70647..70794
;   OTHER INFORMATION: homology with genset EST : A241850
```

```
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 68648..68741
;   OTHER INFORMATION: homology with genset EST : A241850
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 82208..82229
;   OTHER INFORMATION: homology with genset EST : A241850
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 278
;   OTHER INFORMATION: 99-1601-278 : polymorphic base A or C
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 402
;   OTHER INFORMATION: 99-1601-402 : polymorphic base w= A or T
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 472
;   OTHER INFORMATION: 99-1601-472 : polymorphic base A or T
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 2955
;   OTHER INFORMATION: 99-13801-100 : polymorphic base T or C
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 12167
;   OTHER INFORMATION: 99-13806-166 : polymorphic base G or A
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 12536
;   OTHER INFORMATION: 99-13799-376 : polymorphic base T or G
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 17593
;   OTHER INFORMATION: 99-13798-297 : polymorphic base T or C
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 17606
;   OTHER INFORMATION: 99-13798-284 : polymorphic base T or C
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 22079
;   OTHER INFORMATION: 99-1602-200 : polymorphic base G or C
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 28964
;   OTHER INFORMATION: 99-13794-186 : polymorphic base T or C
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 29003
;   OTHER INFORMATION: 99-13794-147 : polymorphic base C or G
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 31077
;   OTHER INFORMATION: 99-13812-384 : polymorphic base T or C
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 31766
;   OTHER INFORMATION: 99-13805-313 : polymorphic base T or C
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 34791
;   OTHER INFORMATION: 99-1587-281 : polymorphic base A or G
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 45751
;   OTHER INFORMATION: 99-1582-430 : polymorphic base C or T
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 49847
;   OTHER INFORMATION: 99-1585-465 : polymorphic base T or C
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 49855
;   OTHER INFORMATION: 99-1585-457 : polymorphic base T or C
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 49886
;   OTHER INFORMATION: 99-1585-426 : polymorphic base G or A
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 49900
;   OTHER INFORMATION: 99-1585-412 : polymorphic base G or A
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 49906
;   OTHER INFORMATION: 99-1585-406 : polymorphic base C or A
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 49921
;   OTHER INFORMATION: 99-1585-391 : polymorphic base C or A
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 49939
;   OTHER INFORMATION: 99-1585-373 : polymorphic base G or A
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 50256
;   OTHER INFORMATION: 99-1585-55 : polymorphic base C or A
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 54955
;   OTHER INFORMATION: 99-1607-373 : polymorphic base T or C
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 64239
;   OTHER INFORMATION: 99-1577-105 : polymorphic base A or G
;   FEATURE:
;   NAME/KEY: allele

  Query Match             6.1%;  Score 253;  DB 18;  Length 106746;
  Best Local Similarity   69.4%;  Pred. No. 1.1e-40;
  Matches  416;  Conservative    0;  Mismatches  175;  Indels    8;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              |||||||||||||||||  |||||||||||||||| |||||||||||||||| ||||||
Db      38717 AGTTCAAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAAAATT 38658

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   ||||||||| | ||||||||||||||||||||||  |||| | ||||||||||||
Db      38657 AGCTGGGCATGGTGGTGCACGCCTGTAATCCCAGCTACTCAGGAGACTGAGGCAGGAGAA 38598

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCCA 178
              |||||||||||||  || ||||||| ||| || ||| ||| |   | ||| |||||||||
Db      38597 TCACTTGAACCTG--GGAGGTGGAGTTTGTGGTGAGCCAAAACTTTCCCACTGCACTCCA 38540

Qy        179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCAT-TTTATATATAT 237
              |||    |   | |   | || |   | |  | |     | |   ||    | ||||
Db      38539 GCCAGTATATATATATATATATACACACGTACACGTGTACACATATACACACATATGC 38480

Qy        238 ATATATATATATATACACACACACACATATATATATACACATATATATACACATATAT 297
              | |  | || ||| ||||||| |||||   | ||| || ||||||| || |  | || ||
Db      38479 ACACGTGTACATACACACACATACACATGCACATACATGCACATATGTACATGCGTACAT 38420

Qy        298 ACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACAC 357
              ||| | ||  |||||||||||||| || || | |  ||||    ||||| || || | |
Db      38419 ACACACATGTACATATATACACACATGTACATATGTACACCCATACACACATGTACATAT 38360

Qy        358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417
              ||    | ||| | ||  |||||| ||||| |||||| | |   || |||   |||
Db      38359 GTACACCCATACACACATGTACATATGTACACCCATATACATATGTACATATGTACACCC 38300

Qy        418 ATACACATATATACACATATATACACATATACACATATATACACATATATACATATATAC 477
              |||||||||| |||| || || || |||| |||  || |||   | |   | | | ||
Db      38299 ATACACATATGTACATATGTACACCCATACACATGTACATATGTACACCCATACACATGT 38240
```

```
Qy        478 ACATATATATAATATACACACATATATATACACATATATACACACATATATACACATATA 537
              |||||| |   | || |||| || || |||   ||||  | |||||||| |||| |||||
Db      38239 ACATATGT---ACATCCACATATGTACATATGTATATCCATACACATATGTACATATATA 38183

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATACATAT 596
              |||||||| || | |  || ||| | || | |  |||| ||   |||||||||   ||
Db      38182 CACACATATGTACATATGTACATATATATTTGTGTATATCTATGTATATATATAAGAAT 38124


RESULT 59
US-09-728-552-4/c
; Sequence 4, Application US/09728552
; Publication No. US20030096398A1
; GENERAL INFORMATION:
;  APPLICANT: Choo, Kong-Hong Andy
;  APPLICANT:  Du Sart, Desiree
;  APPLICANT:  Cancilla, Michael R.
;  TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
;  FILE REFERENCE: Davies Col
;  CURRENT APPLICATION NUMBER: US/09/728,552
;  CURRENT FILING DATE:  2000-12-02
;  PRIOR APPLICATION NUMBER: 09/078,294
;  PRIOR FILING DATE: 1998-05-13
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 80246
;   TYPE: DNA
;   ORGANISM: Nucleotide sequence of NC-contig
US-09-728-552-4

  Query Match             6.1%;  Score 252.6;  DB 10;  Length 80246;
  Best Local Similarity   71.9%;  Pred. No. 1.2e-40;
  Matches  360;  Conservative    0;  Mismatches  134;  Indels    7;  Gaps    2;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | || |||| ||| | |||||| | ||| | | ||| |  ||| | |||| ||| | |||
Db      15787 TATACATATTTATGTCTATATACACACATATATACATATTTATGTCTATATACACACATA 15728

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              |||| || |||   |||||||||||||||||| || ||| |  | ||||| || || |||
Db      15727 TACATATTTATGTCTATATACACATATATACATATTTATGTCTATATACACATATATACA 15668

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              ||| |||    ||||  | | | ||| |||   || | || | ||| |||||  ||||||
Db      15667 TATTTATGTCTATATACACATATATACATATTTATGTCTATATACACATATATACATATT 15608

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              |  |    ||||||||||||||||||| || |||    ||||||||||||||||| || ||
Db      15607 T--ATGTCTATATACACATATATACATATTTATGTCTATATACACATATATACATATTTA 15550

Qy        467 TACATATATACACATATATA-----TAATATACACACATATATATACACATATATACACA 521
              |   ||||||||||||||||     | || |  | | | | ||||| ||||||  ||
Db      15549 TGTCTATATACACATATATACATATTTATGTCTATATACACATATATACATATTTATGTC 15490

Qy        522 CATATATACACATATATACACATATATATACACATATATACACATATATACATATATACA 581
              |||||  ||| ||| |||   || | |||| ||| ||||| ||||||  ||  | |||| |
Db      15489 TATATACACATATACATATTTATGTCTATATACACATATATACATATTTATGTCTATATA 15430

Qy        582 CATATATATACATATATACACATATATACATATATACACATATATACATATATACACACA 641
              || ||||||||||||  ||    |||||||| |||||  ||||||  ||   ||||||  |||||
Db      15429 CACATATATACATATTTATGTCTATATACACATATATACATATTTATGTATATATACACA 15370

Qy        642 TATATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTA 701
              ||||||||  ||  ||  |  | ||| | ||||||||||  |||  | ||||  |||   ||
Db      15369 TATATACATATTTATGTCTATATACACATATATACATATTTATGTATATATTTATGTATA 15310

Qy        702 TATATATGTATATATATATGC 722
              ||| | ||||||| || || |
Db      15309 TATGTGTGTATATGTACATAC 15289


RESULT 60
US-09-818-657-3/c
; Sequence 3, Application US/09818657
; Publication No. US20030077773A1
; GENERAL INFORMATION:
;  APPLICANT: BRANDON, Rhonda et al.
;  TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
;  TITLE OF INVENTION:  NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
;  TITLE OF INVENTION:  AND USES THEREOF
;  FILE REFERENCE: CL001006-CIP
;  CURRENT APPLICATION NUMBER: US/09/818,657
;  CURRENT FILING DATE:  2001-03-28
;  PRIOR APPLICATION NUMBER: 09/730,002
;  PRIOR FILING DATE: 2000-12-06
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 82938
;   TYPE: DNA
;   ORGANISM: Human
US-09-818-657-3

  Query Match             6.1%;  Score 252.6;  DB 10;  Length 82938;
  Best Local Similarity   70.7%;  Pred. No. 1.2e-40;
  Matches  350;  Conservative    0;  Mismatches  144;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | |||| | |||| |||| |||||||||| | | |   | |||  ||||||    ||||||
Db      7454 TATATACACATATGTATACATATATACATATATATGTATATACGTATATAGCTGTATATA 7395

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || ||| || || |||   |  | ||||| | |||| ||||||||| |||||||| |||| |
Db      7394 TATACACATGTATATAGCTGTATATATACATACATGTATATACATATACATATATACATA 7335

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              || | |||  |||||   | || | | ||||||| ||||| | |||  |||||||| |||||
Db      7334 TACACATATACATACATGTATATACATATACATATATACATACATGTATATACATATATA 7275

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              | |   ||| ||    || |||| ||||||||||| || || |||| | ||   |||||
Db      7274 CATATGTATACATGTGTATCTATATACATATATACATATGTATACATGTGTATCTATATA 7215

Qy        467 TACATATATACACATATATAT-AATATACACACATATATATACACATATATACACACATA 525
              | ||||| | |  |||| ||   |||  | | | |||||||||| || ||||||    ||
Db      7214 CATATATACATATGTATACATGTGTATCTATATACATATATACATATGTATACATGTGTA 7155

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              | || | | ||||| |||||| ||||||   | |||  | |||| ||| ||| |||   |||
Db      7154 TCTATATACATATATACATATGTATACATGTGTATCTATATACATATATACATATGTATA 7095

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              || |  || |||| |||||||||||||| || |||| | ||  |||||||| | ||| |
Db      7094 CATGTGTATCTATATACATATATACATATGTATACATGTGTATCTATATACATATATACA 7035

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              ||   ||||||  |   |  ||| ||||| | |||| || |||| | | || ||||| |
Db      7034 TATGTATACATGTGTATCTATATACATATATACATATGTATACATGTGTGTATGTATACA 6975

Qy        706 TATGTATATATATAT 720
              ||||| | | |||||
Db      6974 TATGTGTGTGTATAT 6960


RESULT 61
US-10-417-375-36/c
; Sequence 36, Application US/10417375
; Publication No. US20040219528A1
; GENERAL INFORMATION:
;  APPLICANT: David W. Morris
;  APPLICANT:  Marc Malandro
;  TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
;  FILE REFERENCE: 529452001600
;  CURRENT APPLICATION NUMBER: US/10/417,375
```

```
;  CURRENT FILING DATE:  2003-04-15
;  NUMBER OF SEQ ID NOS: 176
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 36
;    LENGTH: 91697
;    TYPE: DNA
;    ORGANISM: Homo sapiens
US-10-417-375-36

  Query Match             6.1%;  Score 252.2;  DB 18;  Length 91697;
  Best Local Similarity   69.6%;  Pred. No. 1.6e-40;
  Matches  357;  Conservative    0;  Mismatches  153;  Indels    3;  Gaps    1;

Qy        210 GAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACACATAT 269
              | | | | | |  ||  | || |||| |||||| ||||| ||| | |   | | | ||||
Db      20066 GCATATATATACAGCTATATAGATATGTATATAGATATACATATAGATGTATATATATAT 20007

Qy        270 ATATATACACAT---ATATATACACATATATACATATATACACATATATACACATATATA 326
              |||||||| ||   || || | ||||||| |  |||||  | ||| ||| | | |||||
Db      20006 CTATATACATATGCCATGTAGATACATATACATCTATATGTATATAGATATATAGATATA 19947

Qy        327 TACACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATA 386
              |  | ||| |||| |  | |||||||| || |||| |||| | ||||||||| || ||||
Db      19946 TGTATATAGATATATGTAGATATATATCCATATATTTATATAGATATATACATATGTATA 19887

Qy        387 CACACATATATACACATATATACACACATATATACACATATATACACATATATACACATA 446
              | | |||| || | || |||| ||  |||||    | ||| ||| | | ||||| |||||
Db      19886 TAGATATATTTATATATGTATATACCTATATACGTATATAGATATATACATATATACATA 19827

Qy        447 TACACATATATACACATATATACATATATACACATATATATAATATACACACATATATAT 506
              ||   ||||||| | ||||||||||||||||| || ||||  |||||||  |     |||||
Db      19826 TATGTATATATAGATATATATACATATATACATATGTATACTATATACATAATAGTATAT 19767

Qy        507 ACACATATATACACACATATATACACATATATACACATATATATACACATATATACACAT 566
              ||| || |||| |   ||| ||| |  |||| | |  ||||||  | | || ||   ||
Db      19766 ACATATGTATATATGTATACATAGATGTATACATATGTATATATGTATACATCTATGTAT 19707

Qy        567 ATATACATATATACACATATATATACATATATACACATATATACATATATACACATATAT 626
              | ||| || ||||||  ||| |||||||| ||  | | ||| | ||||| | |  |||||
Db      19706 ACATAGATGTATACATCTATGTATACATAGATGTATACATAGATATATACATATGTATAT 19647

Qy        627 ACATATATACACACATATATACACATACATATACACACACATAGATATACATATATATAC 686
              | || ||||   | |||||| |||||  |||| |   | ||||| ||| ||| ||| |
Db      19646 AGATGTATATCAATCTATATATACATATGTATATATGTATATAGAAATATATACATACAT 19587

Qy        687 ACATATATATACGTATATATATGTATATATATA 719
              |  |||||||   |||| |||| || || ||||
Db      19586 ATGTATATATGTATATAGATATATAGATGTATA 19554


RESULT 62
US-10-017-161-2375/c
; Sequence 2375, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
;  FILE REFERENCE: 084335/0152
;  CURRENT APPLICATION NUMBER: US/10/017,161
;  CURRENT FILING DATE:  2002-12-18
;  PRIOR APPLICATION NUMBER: JP 2001/246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2430
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2375
;    LENGTH: 1114
;    TYPE: DNA
;    ORGANISM: Homo sapiens
;    FEATURE:
;    NAME/KEY: source
;    LOCATION: (1)..(1114)
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (201)..(914)
US-10-017-161-2375

  Query Match             6.1%;  Score 251;  DB 15;  Length 1114;
  Best Local Similarity   63.7%;  Pred. No. 4.1e-41;
  Matches  397;  Conservative    0;  Mismatches  225;  Indels    1;  Gaps    1;

Qy        230 ATATATATATATATATATATATACACACACACACACATATATATATACACATATATATAC 289
              ||| | |||||| ||||||| |  | | ||| ||||||||||| |||| |  | |||||||
Db        864 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 805

Qy        290 ACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATAT 349
              | | | ||| || |||| |  | |||| ||||| | ||| |  || |||| |  | |||
Db        804 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 745

Qy        350 ATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATAC 409
              ||| ||| |||||||||  |  | |||| ||||| | | |  | ||||||    |||||||
Db        744 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 685

Qy        410 ACACATATATACACATATATACACATATATACACATATACACATATATACACATATATAC 469
              | ||| |||||    ||||||    ||||| | ||| ||| |  ||||||    |||||||
Db        684 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 625

Qy        470 ATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATATAT 528
              ||| | | || |||||  | || | ||||| | ||| |  ||||||     |||||
Db        624 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 565

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACATATAT 588
              | ||| |||||    |||||| |  | ||| ||||||||||||  ||||||     |||||
Db        564 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 505

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
              ||||| |||||    ||||||   |||||||| | | ||| || |||| |    |||||||
Db        504 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 445

Qy        649 ACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATAT 708
              | | |||||||    | | ||   |||| | ||| | | | || ||||||  | ||||||
Db        444 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 385

Qy        709 GTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAG 768
              || | |||||||      | |||        || |  ||   |  | | ||| |
Db        384 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 325

Qy        769 GAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAG 828
              | | | |  | | |    |||||     | |   | || |      |  | | | |||
Db        324 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 265

Qy        829 TAGGGGAAATACATAAATGTGTA 851
                  | |||||||  |||| ||
Db        264 ATACACATATACATATATGTATA 242


RESULT 63
US-10-292-798-2017/c
; Sequence 2017, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
;  FILE REFERENCE: 084335/166
;  CURRENT APPLICATION NUMBER: US/10/292,798
;  CURRENT FILING DATE:  2002-11-13
;  PRIOR APPLICATION NUMBER: 10/017,161
;  PRIOR FILING DATE: 2001-12-18
;  PRIOR APPLICATION NUMBER: JP 2001-246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2070
```

```
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2017
;   LENGTH: 1114
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   LOCATION: source
;   FEATURE:
;   LOCATION: (1)..(1114)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(914)
US-10-292-798-2017

  Query Match             6.1%;  Score 251;  DB 17;  Length 1114;
  Best Local Similarity   63.7%;  Pred. No. 4.1e-41;
  Matches  397;  Conservative    0;  Mismatches  225;  Indels    1;  Gaps    1;

Qy        230 ATATATATATATATATATATATACACACACACACATATATATATACACATATATATAC 289
              ||| | |||||| ||||||| |  | | ||| |||||||||| |||| |  | |||||||
Db        864 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 805

Qy        290 ACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATAT 349
              | | | ||| || |||| |  | |||| ||||| | ||| |  || |||| |  | |||
Db        804 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 745

Qy        350 ATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATAC 409
              ||| ||| |||||||||  |  | |||| ||||| | | |  |||||    |||||||
Db        744 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 685

Qy        410 ACACATATATACACATATATACACATATATACACATATACACATATATACACATATATAC 469
              | ||| |||||   ||||||    ||||| | ||| ||| |  ||||||    |||||||
Db        684 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 625

Qy        470 ATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATATAT 528
              ||| | | | || ||||||  | ||| | ||||| | ||| |  ||||||     |||||
Db        624 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 565

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACACATATAT 588
              | ||| |||||   ||||||  |  | ||| ||||||||||||  ||||||   |||||
Db        564 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 505

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
              ||||| |||||   ||||||  ||||||| | | ||| || |||| |    |||||||
Db        504 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 445

Qy        649 ACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATAT 708
              | | |||||||   | | ||   |||| | ||| | | | || ||||||  | ||||||
Db        444 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 385

Qy        709 GTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAG 768
              || | |||||||     |  |||       ||  |   ||   | | ||| |
Db        384 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 325

Qy        769 GAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAG 828
              | | | |   | | |    |||||     | |   | || |      |  | | | |||
Db        324 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 265

Qy        829 TAGGGGAAATACATAAATGTGTA 851
                  | ||||||| |||| ||
Db        264 ATACACATATACATATATGTATA 242


RESULT 64
US-10-052-482-232
; Sequence 232, Application US/10052482
; Publication No. US20040072264A1
; GENERAL INFORMATION:
;  APPLICANT: Engelhard, Eric
;  APPLICANT:  Morris, David
;  TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
;  FILE REFERENCE: A-71087/RMS/DCF
;  CURRENT APPLICATION NUMBER: US/10/052,482
;  CURRENT FILING DATE:  2002-08-15
;  PRIOR APPLICATION NUMBER: US 09/747,377
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: US 09/798,586
;  PRIOR FILING DATE: 2001-03-02
;  NUMBER OF SEQ ID NOS: 241
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 232
;   LENGTH: 96595
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (49183)..(49621)
;   OTHER INFORMATION: "n" at positions 49183 to 49621 can be any base
US-10-052-482-232

  Query Match             6.1%;  Score 251;  DB 17;  Length 96595;
  Best Local Similarity   69.5%;  Pred. No. 2.8e-40;
  Matches  354;  Conservative    0;  Mismatches  152;  Indels    3;  Gaps    1;

Qy        229 TATATATATATATATATATATATACACACACACACATATATATATACACATATATATA 288
              || ||| | |||| | |||||||||| ||| | |||||||| | | |||| ||| ||| |
Db      90982 TACATACACATATGTGTATATATACATACATATACACATATGTGTGTACATATACATACA 91041

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              || | | |||  | | |||| ||| ||||||| | ||| ||| ||   ||  |  |||||
Db      91042 CATACACATATGTGTGTACATATACATACACACACATACACATATGTGTACATGTACATA 91101

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              |||  |||  | |  |||    | || |||| | ||| ||| ||| ||||     | || ||
Db      91102 CATACACATACACATATGTGTACATGTACATACATACACATACACATATGTGTACATGTA 91161

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              || ||||| | | ||| ||||    | || |||| |||||||||||||| ||| || | ||
Db      91162 CATACATACACATACACATATGTGTACATGTACATATATACACATATACACATATGTGTA 91221

Qy        469 CATATATACACATATATATAATATACACACATATATATACACATATATACACACATATAT 528
              ||| || | | ||| |    |||||||| || | || |    | ||||| |||||||||
Db      91222 CATGTACATATATACAC---ATATACACATATGTGTACATGTACATATATACACATATAC 91278

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACATATAT 588
              ||| || | |||| || |||||||||||| ||| | || | ||||| || | | | ||
Db      91279 ACATATGTGTACATGTACATATACACATATACACATATGTGTACATGTACATATACACAT 91338

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
              ||||| ||||    | ||  ||||||| |||  | || |    | ||| |||||||| |
Db      91339 ATACACATATGTGTACATGCACATATACACATGTGTACATGCACATATACACATATGTGT 91398

Qy        649 ACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATAT 708
              ||||||| ||| ||||| ||    | || ||| ||||||||| |||| |  ||||||| ||
Db      91399 ACATACACATATACACATATGTGTGTATATACATATACACACATATGTGTGTATATACAT 91458

Qy        709 GTATATATATATGCTCCAGAGTTCATAAG 737
              |||| |||||| |       | |||| |
Db      91459 ATATACATATATACATATATATACATATG 91487


RESULT 65
US-10-311-455-1972/c
; Sequence 1972, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determi
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
```

```
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1972
;   LENGTH: 21537
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1972

  Query Match             6.1%;  Score 250;  DB 15;  Length 21537;
  Best Local Similarity   68.1%;  Pred. No. 2.4e-40;
  Matches  391;  Conservative    0;  Mismatches  180;  Indels    3;  Gaps    3;

Qy        181 CCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATA 240
              |||     | || |     || |  |||||      | | |         | ||||||||||||
Db       6005 CCAAATTAAAATACAATAACGCGATCTCGATTCACTACACTATATATATATATATATATATA 5946

Qy        241 TATATATATATACACACACACACATATATATATACACATATATATACACATATATACA 300
              |||||||||||| ||||| |||| |||||||||||| | ||||||||||| | |||||||
Db       5945 TATATATATATATACACATACAC-CATATATATATATATATATATACATACATATACA 5887

Qy        301 TATATACACATATATACACATATATATACACATACATATGTACACATATATATACACATA 360
              |   |    |   || | | | || | | |    |   | |||| | | ||| ||
Db       5886 AAACAATTACTTACTAAATACACAAATCAATAAAATAAAAAATTCATAAACAAACATATC 5827

Qy        361 TGTATACACATATATACACATATATACACACATATATACACATATATACACACATATATA 420
              | | |  |||||||||  |  |||||  |        |||| | ||||||| | | |||||||
Db       5826 TCTCTTTACATATATATATCTATATCGATCACACTATATATATATATATATATATATATA 5767

Qy        421 CACATATATACACATATATACACATATACACATATATACACATATATACATATATACACA 480
              |||| |||||  | |||||||  | |||  |||||||||||  | |||||||  | | | | | |
Db       5766 CACACATATATATATATATATATACACACATATATATATATATATATACACACATATA 5707

Qy        481 TATATATAATATACACACATATATATACACATATATACACACATATATACACATATATAC 540
              |||||||  | |||||||||||||||||  | |||||||  |||||  |||||  | |||||||
Db       5706 TATATAT-ACATACACACATATATATATATATATATACACACATATATATATATATAT 5648

Qy        541 ACATATATATACACATATATACACATATATACATATATACACATATATATACATATATAC 600
              | |||||||||  | |||||||  | |||||||  |||||||  |||||  |  |||||||||
Db       5647 ATATATATATATATATATATATATATATATATATATATATATATATATACACGCATATATAT 5588

Qy        601 ACATATATACATATATACACATATATACATATATACACACATATATACACATACATATAC 660
              || |||||||||||||  | |   ||| |||||||  | |||||||||  | |||  |||||
Db       5587 ACGTATATACATATATATAAACTCATATATATATATATACATATATATATATATATATAT 5528

Qy        661 ACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATAT 720
              | | | ||| |   |||  |||||| |  | | | ||      || | | ||||  |  ||
Db       5527 ATATATATAAACTCACA-ATATATCCTTAAACAAATTATATAATCTCTATATACCTCAAT 5469

Qy        721 GCTCCAGAGTTCATAAGAGGTAGCAGTTGATTAC 754
                ||   | | | || |   |  | | |  | |||
Db       5468 TTTCTCAAATATAAAACAAAAATAATTATAATAC 5435


RESULT 66
US-09-997-722-196
; Sequence 196, Application US/09997722
; Publication No. US20040072154A1
; GENERAL INFORMATION:
;  APPLICANT: Morris, David
;  APPLICANT:  Engelhard, Eric
;  TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
;  FILE REFERENCE: A-71171/RMS/DCF
;  CURRENT APPLICATION NUMBER: US/09/997,722
;  CURRENT FILING DATE:  2001-11-30
;  PRIOR APPLICATION NUMBER: US 09/747,377
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: US 09/798,586
;  PRIOR FILING DATE: 2001-03-02
;  NUMBER OF SEQ ID NOS: 301
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 196
;   LENGTH: 96596
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-997-722-196

  Query Match             6.1%;  Score 249.6;  DB 11;  Length 96596;
  Best Local Similarity   70.5%;  Pred. No. 5.4e-40;
  Matches  347;  Conservative    0;  Mismatches  144;  Indels    1;  Gaps    1;

Qy        229 TATATATATATATATATATATATACACACACACACACATATATATATACACATATATATA 288
              | | || |||||| || || || ||    | | |||  | || ||||| |||| || ||
Db      57110 TGTGTACATATATACATGTAGATGCATGTAGATACATGTGTACATATATACATGTAGATG 57169

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              ||  || |||||| ||||||  || |||||  || |||  |   |||||| |  | ||||
Db      57170 CATGTAGATACATGTATACATGTAGATACATGTAGATACATGTATACATGTAGATACATG 57229

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              || |||||  |||  ||| | | || || | | || || | |||| |||||| ||||| |
Db      57230 TAGATACATGTATACATATATACATGTAGATACATGTAGATACATGTATACATATATACA 57289

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              | ||| ||  | | ||||| |||| || | |||| || ||||||||||  || |||
Db      57290 TGTAGATACATGTATACATATATACATGTAGATACATGTATACATATATACATGTAGATA 57349

Qy        469 CATATATACACATATATATAATATACACACATATATATACACATATATACACACATATAT 528
              ||| |||||| ||||| ||  || | |||  || ||| |   ||| ||| | |||| ||
Db      57350 CATGTATACATATATACAT-GTAGATACATGTAGATACATGTATACATATATACATGTAG 57408

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACATATAT 588
              | |||| || | |||| |||| | | ||| ||  | ||| ||  ||||| | ||||| ||
Db      57409 ATACATGTAGATACATGTATACATATATACATGTAGATACATGTATATAGATACATACAT 57468

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
              | | ||||| | ||||| ||||||||||  | |||| |||||||||  ||| ||||||
Db      57469 AGATATATAGATACATAGATACATATATTTATATATGTACATATATCTACATATATATCT 57528

Qy        649 ACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATAT 708
              | ||| | ||| |   ||||| |||| |||||||||  |  || |||||  | |||||||
Db      57529 ATATATACATATATCTACATATATATCTATATATATATATCTACATATATATCTATATAT 57588

Qy        709 GTATATATATAT 720
              ||| | ||||||
Db      57589 GTAGAAATATAT 57600


RESULT 67
US-10-437-963-61122
; Sequence 61122, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 61122
;   LENGTH: 623
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
```

```
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_62586C.1
US-10-437-963-61122

  Query Match             6.0%;  Score 249;  DB 18;  Length 623;
  Best Local Similarity   73.2%;  Pred. No. 8.2e-41;
  Matches  361;  Conservative    0;  Mismatches  125;  Indels    7;  Gaps    3;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db        122 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 181

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | |||||||| | ||| |||| || | |
Db        182 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 241

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        242 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 301

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | |      | |  |||
Db        302 TATATATATATATATATATATATATATATATATATATATATATATATAAGATGAAATGTATG 361

Qy        467 TACATATATACACATATATATAATATACACACAT---ATATATACACATATATACACACA 523
                ||   ||  ||||||||||||    || || ||   | |||| ||  | |  | | |
Db        362 GTCAGTGATCAACATATATATATGGCACGCAAATTTAAAATATTCATCTGTTGATAAAAT 421

Qy        524 TATA---TACACATATATACACATATATATACACATATATACACATATATACATATATAC 580
              | ||   |||| || ||| || ||||||||||||||| ||   |||| | |   | ||||
Db        422 TTTACACTACAAATGTATGCATATATATATACACATGTAACTACATGTGTGTGTCTATAT 481

Qy        581 ACATATATATACATATATACACATATATACATATATACACATATATACATATATACACAC 640
              | ||||||||| ||||||| | ||||||| ||||||| | ||||||| |||||| |
Db        482 ATATATATATATATATATATATATATATATATATATATATATATATATATATATAT-CGTGT 540

Qy        641 ATATATACACATACATATACACACACATAGATATACATATATATACACATATATATACGT 700
              |||||||| | | |   | |  ||| | | ||||| |||||||| | |||||||||  |
Db        541 ATATATACCCTTTCTCGTGCCTACATACATATATATATATATATATATATATATATAT 600

Qy        701 ATATATATGTATA 713
              ||||||||| ||||
Db        601 ATATATATATATA 613


RESULT 68
US-10-437-963-61122/c
; Sequence 61122, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 61122
;   LENGTH: 623
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_62586C.1
US-10-437-963-61122

  Query Match             6.0%;  Score 249;  DB 18;  Length 623;
  Best Local Similarity   72.7%;  Pred. No. 8.2e-41;
  Matches  364;  Conservative    0;  Mismatches  130;  Indels    7;  Gaps    3;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACAC-ATATAT 285
              | |||||||||||||||||||| ||| |  |||   |  ||||||||||| ||||||
Db        589 TATATATATATATATATATATGTATGTAGGCACGAGAAAGGGTATATATACACGATATAT 530

Qy        286 ATACACATATATACATATATACACATATATACACATATATATACACATACATATGTACAC 345
              ||| | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||   ||||
Db        529 ATATATATATATATATATATATATATATATATATATATATATATATATATAGACACAC 470

Qy        346 ATATATATACACATATGTATACACATATATACACAT-----ATATACACACATATATACA 400
              | || ||   |||| |||||| | ||||  | ||||      | ||| |    |||  |||
Db        469 ACATGTAGTTACATGTGTATATATATATGCATACATTTGTAGTGTAAAATTTTATCAACA 410

Qy        401 CATATATACACACATATATACACATATATACACATATATACACATATACACATATATACA 460
              ||  |||     |  |       ||||| | || |  |   |   || | || | |
Db        409 GATGAATATTTTAAATTTGCGTGCCATATATATATGTTGATCACTGACCATACATTTCAT 350

Qy        461 CATATATACATATATACACATATATATA-ATATACACACATATATATACACATATATACA 519
              | |||||| ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| |
Db        349 CTTATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 290

Qy        520 CACATATATACACATATATACACATATATATACACATATATACACATATATACATATATA 579
              | ||||||| | ||||||| | ||||||||| | ||||||| | ||||||| |||||||
Db        289 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 230

Qy        580 CACATATATATACATATATACACATATATACATATATACACATATATACATATATACACA 639
              | ||||||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | |
Db        229 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 170

Qy        640 CATATATACACATACATATACACACACATAGATATACATATATATACACATATATATACG 699
              ||||||| | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||
Db        169 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 110

Qy        700 TATATATATGTATATATATAT 720
              |||||||| ||||||||||||
Db        109 TATATATATATATATATATAT 89


RESULT 69
US-10-741-600-17651/c
; Sequence 17651, Application US/10741600
; Publication No. US20050026169A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001499
;  CURRENT APPLICATION NUMBER: US/10/741,600
;  CURRENT FILING DATE:  2003-12-22
;  NUMBER OF SEQ ID NOS: 73997
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17651
;   LENGTH: 141121
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-741-600-17651

  Query Match             6.0%;  Score 248.8;  DB 19;  Length 141121;
  Best Local Similarity   67.1%;  Pred. No. 9.3e-40;
  Matches  365;  Conservative    2;  Mismatches  174;  Indels    3;  Gaps    1;

Qy        181 CCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATA 240
              |||    ||  || |  | | | ||      ||||| | |       | |||||||| |
Db      77519 CCAATGCACCCGACCTCAGTATAATTCTTAAAAAATATTATATATATACACATATATATGTT 77460

Qy        241 TATATATATATACACACACACACACATATATATATACACATATATATACACATATATACA 300
              ||||| |||||||||||| |||   |||||| |:|| | | ||  |:  | ||||| |||
Db      77459 TATATGTATATACACACATACATGTATATATGTRTATATACATGCAYGTATATATACACA 77400

Qy        301 TATATACACATATATACACATATATATACACATACATATGTACACATATATATACACATA 360
              |||| |||  ||| | |  ||||| | | ||||  |||| ||||||||||| | |   ||
Db      77399 TATACACATGTATGTGCGTATATACACATACATGTATATATACACATATACACATGTATG 77340
```

```
Qy        361 TGTATACACATATATACACATATATACACACATATATACACATATATACACACATATATA 420
              ||  || | ||| ||| ||| || ||    | | |||| ||||||||||||||  | | |
Db      77339 TGCGTATATATACATATACACATGTATGTGCGTGTATATACATATATACACATGTGTGTG 77280

Qy        421 CACATATATACACATATATACACATATACACATATATACACATATATACATATATACACA 480
              |   |||||||| ||||| |||  | |   | | |||| |||||||||||||||||||||
Db      77279 CGTGTATATACATATATACACATGTGTGTGCGTGTATATACATATATACATATATACACA 77220

Qy        481 TATATAT---AATATACACACATATATATACACATATATACACACATATATACACATATA 537
              |||| |    |||| ||| |||| ||| ||| | ||      |||||||| ||||
Db      77219 GATATGTGTGTATATATACATATATACATATACACACATGTGTGTATATATACATATATA 77160

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATACATATA 597
              | | | | || |    |||||||||||||||||| | |||||  ||  | || |||||
Db      77159 CATATACACATGTGTGTATATATACACATATATACACACATACATGTACGTGTATATATA 77100

Qy        598 TACACATATATACATATATACACATATATACATATATACACACATATATACACATACATA 657
              | | ||| | |||  | |  | ||||| | ||||| |||||| | | || | |||||||
Db      77099 CATATATACACACACGTGTGTATATATACATATATACACACACGTGTGTATATATACATA 77040

Qy        658 TACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATA 717
              || ||||| ||| | | || | | | |  | | ||||||||  |||||||   |  | |
Db      77039 TATACACATATATACACACGTGTGTGTGTATACATATATACACATATATACACACGTGTG 76980

Qy        718 TATG 721
              | ||
Db      76979 TGTG 76976


RESULT 70
US-10-017-161-2365/c
; Sequence 2365, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
;  FILE REFERENCE: 084335/0152
;  CURRENT APPLICATION NUMBER: US/10/017,161
;  CURRENT FILING DATE:  2002-12-18
;  PRIOR APPLICATION NUMBER: JP 2001/246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2430
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2365
;   LENGTH: 1342
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: source
;   LOCATION: (1)..(1342)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(1142)
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (64)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (66)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (74)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (413)..(434)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (956)..(1055)
;   OTHER INFORMATION: a, t, c, g, unknown or other
US-10-017-161-2365

  Query Match             6.0%;  Score 248.4;  DB 15;  Length 1342;
  Best Local Similarity   69.9%;  Pred. No. 1.5e-40;
  Matches  348;  Conservative    0;  Mismatches  143;  Indels    7;  Gaps    2;

Qy        229 TATATATATATATATATATATACACACACACACATATATATATACACATATATATA 288
              || | | | |||| | |||||||||| | ||||||  | | | |||||  |  | ||||||
Db        891 TACACACACATATGTGTATATATACATATACACACCTACACACATATATATGTGTATATA 832

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              ||||||| ||| | ||| |  | |||| | | ||||| | |||||  | | |||||||
Db        831 TACATATACACACACATATATGTGTATATATACATATACATACATATGTGTATACACATG 772

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              |||||   || || | || || ||||| |||||||| | ||| |||| || |||||||||
Db        771 TATATGTGCACATATGTATACGTATATGCACATATACATACATATATGTATACATATATA 712

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              |||| |||| | ||  ||| || || ||| | |||||||  |||| | ||  |||| |||
Db        711 CACATATATGTGCATGTATGTATACGTATGTGCACATATGTACATGTGTATGCATAAATA 652

Qy        469 CATATATACACATATATATA-ATATACACACATATATATACA------CATATATACACA 521
              |||||||  |||| | |||| ||| | ||||||||||| ||||        | ||| | |||
Db        651 CATATATGTACATGTGTATACATAAATACACATATATGTACATGCGTGTACATAAATACA 592

Qy        522 CATATATACACATATATACACATATATATACACATATATACACATATATACATATATACA 581
              |||||||  ||||   || ||||| ||| ||| |||| ||||  | ||||||||||||||
Db        591 CATATATGTACATGCGTATACATAAATACACATATATGTACATGTGTATACATATATACA 532

Qy        582 CATATATATACATATATACACATATATACATATATACACATATATACATATATACACACA 641
              ||||||| |||||   || |||||||||| |||| |||| | |  | ||| | |||||
Db        531 CATATATGTACATGCGTATACATATATACACATATGTACATGTGTGTACATAAATACACA 472

Qy        642 TATATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTA 701
              ||| |||| ||   || | | | ||| | || |||||
Db        471 TATGTACATATGTGTACATAAATACACATATGTACATNNNNNNNNNNNNNNNNNNNNNG 412

Qy        702 TATATATGTATATATATA 719
              || || | ||||||||||
Db        411 TACATGTATATATATATA 394


RESULT 71
US-10-292-798-2009/c
; Sequence 2009, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
;  FILE REFERENCE: 084335/166
;  CURRENT APPLICATION NUMBER: US/10/292,798
;  CURRENT FILING DATE:  2002-11-13
;  PRIOR APPLICATION NUMBER: 10/017,161
;  PRIOR FILING DATE: 2001-12-18
;  PRIOR APPLICATION NUMBER: JP 2001-246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2070
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2009
;   LENGTH: 1342
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   LOCATION: source
;   FEATURE:
```

```
;   LOCATION: (1)..(1342)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(1142)
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (64)..(64)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (66)..(66)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (74)..(74)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (413)..(434)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (956)..(1055)
;   OTHER INFORMATION: a, t, c, g, unknown or other
US-10-292-798-2009

  Query Match             6.0%;  Score 248.4;  DB 17;  Length 1342;
  Best Local Similarity   69.9%;  Pred. No. 1.5e-40;
  Matches  348;  Conservative    0;  Mismatches  143;  Indels    7;  Gaps    2;

Qy        229 TATATATATATATATATATATATACACACACACACATATATATATACACATATATATA 288
              || | | | |||| | |||||||||| | ||||||  | | | ||||| |  | ||||||
Db        891 TACACACACATATGTGTATATATACATATACACACCTACACACATATATATGTGTATATA 832

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              ||||||| ||| | ||| |  | |||| | | ||||| | |||||  | | |||||||
Db        831 TACATATACACACACATATATGTGTATATATACATATACATACATATGTGTATACACATG 772

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              |||||   || || | || || ||||| |||||||| | ||| |||| || |||||||||
Db        771 TATATGTGCACATATGTATACGTATATGCACATATACATACATATGTATACATATATA 712

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              |||| |||| | ||  ||| || || ||| | |||||||  |||| | ||  |||| |||
Db        711 CACATATATGTGCATGTATGTATACGTATGTGCACATATGTACATGTGTATGCATAAATA 652

Qy        469 CATATATACACATATATATA-ATATACACACATATATATACA------CATATATACACA 521
              |||||||  |||| | |||| ||| | |||||||||| ||||       | ||| | |||
Db        651 CATATATGTACATGTGTATACATAAATACACATATATGTACATGCGTGTACATAAATACA 592

Qy        522 CATATATACACATATATACACATATATATACACATATATACACATATATACATATATACA 581
              |||||||  ||||   || ||||| ||| ||| |||| ||||  | ||||||||||||||
Db        591 CATATATGTACATGCGTATACATAAATACACATATATGTACATGTGTATACATATATACA 532

Qy        582 CATATATATACATATATACACATATATACATATATACACATATATACATATATACACACA 641
              ||||||| |||||   || |||||||||| |||| |||| | |  | ||| | |||||
Db        531 CATATATGTACATGCGTATACATATATACACATATGTACATGTGTGTACATAAATACACA 472

Qy        642 TATATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTA 701
              ||| |||| ||   || | | | ||| | || |||||
Db        471 TATGTACATATGTGTACATAAATACACATATGTACATNNNNNNNNNNNNNNNNNNNNNNG 412

Qy        702 TATATATGTATATATATA 719
              || || | ||||||||||
Db        411 TACATGTATATATATATA 394


RESULT 72
US-10-017-161-2127/c
; Sequence 2127, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT: ASAI, KIYOSHI
;  APPLICANT: AKIYAMA, YUTAKA
;  APPLICANT: ABURATANI, HIROYUKI
;  TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
;  FILE REFERENCE: 084335/0152
;  CURRENT APPLICATION NUMBER: US/10/017,161
;  CURRENT FILING DATE:  2002-12-18
;  PRIOR APPLICATION NUMBER: JP 2001/246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2430
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2127
;   LENGTH: 1363
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: source
;   LOCATION: (1)..(1363)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(1163)
US-10-017-161-2127

  Query Match             6.0%;  Score 248.2;  DB 15;  Length 1363;
  Best Local Similarity   71.7%;  Pred. No. 1.7e-40;
  Matches  367;  Conservative    0;  Mismatches  139;  Indels    6;  Gaps    3;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATA 286
              | || | ||||||| |||||| || | | | || |||| |   |||||||   | |||||
Db       1136 TGTACACATATATACATATATGTACATATATACGCACATACGCATATATATGTACATATA 1077

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              |||||| |||  |||||||  ||||||||||||||||  |||  |  ||||||| ||||||
Db       1076 TACACACATACGCATATATGTACATATATACACATATGCATATATGTACATATATACACA 1017

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||  |||  |  ||  |||| ||| |||| || |||| |||| | ||| | || ||||||
Db       1016 TATGCATATATGTACATATATACACATATGCATATATGTACATATATACACACGCATATA 957

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              |  ||| ||||| ||||||||    || |||||     |||||||| ||| | |  | ||
Db        956 TGTACATATATACACACATAT----ACGTATATGTGTGTATACACACATACATATGTGTA 901

Qy        467 TACATATATACACATATATATAATATACACACATATATATACACATATATACACACATAT 526
              |||| | ||||  |||||  | || |||||| ||| ||||| || |||||| ||||||||
Db        900 TACACACATACGTATATACGT-ATGTACACATATACATATATACGTATATATACACATAT 842

Qy        527 ATACACATATATACACATATATATACACATATATACACATATATACATATATACACATAT 586
              | ||  |||||  |||||||||| ||| |  || ||| ||  | ||| | |||||    ||
Db        841 ACACCTATATACACACATATACATATATGTACATATACGCACATATACATATATGTGTAC 782

Qy        587 ATATACATATATACACATATATACATATATACACATATATACA-TATATACACACATATA 645
              ||||| | | ||||| ||||  |||||||||||| || |||| ||||||||||||| ||
Db        781 ATATATACACATACATATATGCACATATATACACACATGTACACTATATACACACATGTA 722

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              ||| || || ||| | ||||| |   | |||  || | ||||  | | ||| ||| | |
Db        721 CACATATGCACATATATACACACATGCACACACGTACACACACGCACACATATGTACACA 662

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAG 737
              ||||| | || | | | |    |  |||| |
Db        661 CATGTACACATGTGTACACATATGCCCATATG 630


RESULT 73
US-10-292-798-1773/c
; Sequence 1773, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT: ASAI, KIYOSHI
;  APPLICANT: AKIYAMA, YUTAKA
;  APPLICANT: ABURATANI, HIROYUKI
```

```
;  TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
;  FILE REFERENCE: 084335/166
;  CURRENT APPLICATION NUMBER: US/10/292,798
;  CURRENT FILING DATE:  2002-11-13
;  PRIOR APPLICATION NUMBER: 10/017,161
;  PRIOR FILING DATE: 2001-12-18
;  PRIOR APPLICATION NUMBER: JP 2001-246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2070
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1773
;   LENGTH: 1363
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   LOCATION: source
;   FEATURE:
;   LOCATION: (1)..(1363)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(1163)
US-10-292-798-1773

  Query Match             6.0%;  Score 248.2;  DB 17;  Length 1363;
  Best Local Similarity   71.7%;  Pred. No. 1.7e-40;
  Matches  367;  Conservative    0;  Mismatches  139;  Indels    6;  Gaps    3;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | || | ||||||| |||||| || | | | || |||| |   |||||||   | |||||
Db       1136 TGTACACATATATACATATATGTACATATATACGCACATACGCATATATATGTACATATA 1077

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              |||||| |||  |||||||  ||||||||||||||||  ||| |  ||||||| ||||||
Db       1076 TACACACATACGCATATATGTACATATATACACATATGCATATATGTACATATATACACA 1017

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||  ||| |  ||  |||| ||| |||| || |||| |||| | ||| | || ||||||
Db       1016 TATGCATATATGTACATATATACACATATGCATATATGTACATATATACACACGCATATA 957

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              |  ||| ||||| ||||||||    || |||||     |||||||| ||| | |  | ||
Db        956 TGTACATATATACACACATAT----ACGTATATGTGTGTATACACACATACATATGTGTA 901

Qy        467 TACATATATACACATATATATAATATACACACATATATATACACATATATACACACATAT 526
              |||| | ||||  |||||  | || |||||| ||| ||||| || |||||| ||||||||
Db        900 TACACACATACGTATATACGT-ATGTACACATATACATATATACGTATATATACACATAT 842

Qy        527 ATACACATATATACACATATATATACACATATATACACATATATACATATATACACATAT 586
              | ||  ||||| |||||||| ||| |  || ||| ||  | ||| | |||||    ||
Db        841 ACACCTATATACACACATATACATATATGTACATATACGCACATATACATATATGTGTAC 782

Qy        587 ATATACATATATACACATATATACATATATACACATATATACA-TATATACACACATATA 645
              ||||| | | ||||| ||||  |||||||||||||| || |||| ||||||||||||| ||
Db        781 ATATATACACATACATATATGCACATATATACACACATGTACACTATATACACACATGTA 722

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              ||| || || ||| | ||||| |   | |||  || | ||||  | | ||| ||| | |
Db        721 CACATATGCACATATATACACACATGCACACACGTACACACACGCACACATATGTACACA 662

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAG 737
              ||||| | || | | | |     |  |||| |
Db        661 CATGTACACATGTGTACACATATGCCCATATG 630


RESULT 74
US-10-437-963-57702
; Sequence 57702, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 57702
;   LENGTH: 614
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_59488C.1
US-10-437-963-57702

  Query Match             6.0%;  Score 247.8;  DB 18;  Length 614;
  Best Local Similarity   71.2%;  Pred. No. 1.4e-40;
  Matches  341;  Conservative    0;  Mismatches  137;  Indels    1;  Gaps    1;

Qy        226 TTTTATATATATATATATATATATATACACACACACACACATATATATATACACATATAT 285
              ||  | |||||||||||||||||||||| | | | | | |||||||||||| | ||||||
Db        136 TTGAACATATATATATATATATATATATATATATATATATATATATATATATATATATAT 195

Qy        286 ATACACATATATACATATATACACATATATACACATATATATACACATACATATGTACAC 345
              ||| | |||||||   |  ||  || ||     |      || | |  |||     | || |
Db        196 ATATATATATATAATTTGATTTACGTACTCCCGGGAGTACAGAAGCATTTTCCTATATAT 255

Qy        346 ATATATATACACATATGTATACACATATATACACATATATACACACATATATACACATAT 405
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | ||||
Db        256 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 315

Qy        406 ATACACACATATATACACATATATACACATATATACACATATACACATATATACACATAT 465
              ||| | | |||||||| | |||||||| | |||||||| | ||||| | ||||||| | ||||
Db        316 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 375

Qy        466 ATACATATATACACATATATATAATATACACACATATATATACACATATATACACACATA 525
              ||| |||||||| | |||||||| ||||| | | |||||||||| | ||||||| | | |||
Db        376 ATATATATATATATATATATAT-ATATATATATATATATATATATATATATATATATATA 434

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db        435 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 494

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||||  |   |   |  |          ||
Db        495 TATATATATATATATATATATATATATATATATACGGAGGGAGTATTTGACTGTTCGTCTTA 554

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATAT 704
              | || | || | |     | | ||  | |   |||| ||||   ||| ||||| ||||||
Db        555 TTCAAAAACTTTTTGTGAAAAATGTAAAACTATATGTATATGTATACATATAAGTATAT 613


RESULT 75
US-10-723-860-1125
; Sequence 1125, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
;  APPLICANT: Aziz, Natasha
;  APPLICANT:  Ginsburg, Wendy M.
;  APPLICANT:  Zlotnik, Albert
;  TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
;  TITLE OF INVENTION:  Methods for Screening for Soft Tissue Sarcoma Modulators
;  FILE REFERENCE: 05882.0193.NPUS01
;  CURRENT APPLICATION NUMBER: US/10/723,860
;  CURRENT FILING DATE:  2003-11-26
;  PRIOR APPLICATION NUMBER: 60/429,739
;  PRIOR FILING DATE: 2002-11-26
;  NUMBER OF SEQ ID NOS: 8393
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1125
```

```
;   LENGTH: 121600
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-723-860-1125

  Query Match             6.0%;  Score 247.4;  DB 18;  Length 121600;
  Best Local Similarity   68.6%;  Pred. No. 1.7e-39;
  Matches  341;  Conservative    0;  Mismatches  156;  Indels    0;  Gaps    0;

Qy        224 CATTTTATATATATATATATATATATATACACACACACACACATATATATATACACATAT 283
              ||   |||||||||| ||||||| ||||||  |   |. | |   ||||||  |||| | || |
Db      43894 CAGCCTATATATATCTATATATCTATATATATCTATATATCTATATATCTATATATATCT 43953

Qy        284 ATATACACATATATACATATATACACATATATACACATATATATACACATACATATGTAC 343
              |||||   |||||||  ||||||   ||||||  | || | |||| |  || |||| ||
Db      43954 ATATATCTATATATATCTATATATCTATATATCTATATCTCTATATATCTATATATCTAT 44013

Qy        344 ACATATATATACACATATGTATACACATATATACACATATATACACACATATATACACAT 403
              | | ||||||   || | ||||    |||||    ||||||  | | || |||| | ||
Db      44014 ATCTCTATATATCTATCTATATATCTCTATATCTCTATATATCTATATATCTATATATAT 44073

Qy        404 ATATACACACATATATACACATATATACACATATATACACATATACACATATATACACAT 463
              |||| |   ||||||  | |||| || | |||| || | || || | || |||| |  |
Db      44074 CTATATATCTATATATCTATATATCTATATATATCTATATATCTATATATCTATATATCT 44133

Qy        464 ATATACATATATACACATATATATAATATACACACATATATATACACATATATACACACA 523
              |||||  |||||| |  ||||||| || || | |  |||||||  | |||| || | |
Db      44134 ATATATCTATATATATCTATATATCATCTATATATCTATATATCTATATATCTATATATC 44193

Qy        524 TATATACACATATATACACATATATATACACATATATACACATATATACATATATACACA 583
              ||||||   ||||||  | ||||||||  | ||||||  | ||||  || ||||||  | |
Db      44194 TATATATCTATATATCTATATATATATCTATATATATCTATATATCTATATATATCTATA 44253

Qy        584 TATATATACATATATACACATATATACATATATACACATATATACATATATACACACATA 643
              ||| |||| || |||| | || |||| || |||| | || |||||||  |||| |   |||
Db      44254 TATCTATATATCTATATATATCTATATATCTATATATATCTATACATCTATATATCTATA 44313

Qy        644 TATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATA 703
              ||||   ||| || || | |   ||| ||||  |||||| || | |||| ||||  | ||
Db      44314 TATATCTATATATCTATATATCTATATATATCTATATATCTATATATATCTATATATCTA 44373

Qy        704 TATATGTATATATATAT 720
              ||||| | ||||||| |
Db      44374 TATATATCTATATATCT 44390


RESULT 76
US-10-221-714A-313/c
; Sequence 313, Application US/10221714A
; Publication No. US20040048254A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with
;  TITLE OF INVENTION:  tumor suppressor genes and oncogenes
;  FILE REFERENCE: 5013.1005
;  CURRENT APPLICATION NUMBER: US/10/221,714A
;  CURRENT FILING DATE:  2003-01-21
;  PRIOR APPLICATION NUMBER: PCT/EP01/02955
;  PRIOR FILING DATE: 2001-03-15
;  PRIOR APPLICATION NUMBER: DE 10013847.0
;  PRIOR FILING DATE: 2000-03-15
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 540
; SEQ ID NO 313
;   LENGTH: 6286
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-714A-313

  Query Match             5.9%;  Score 245;  DB 17;  Length 6286;
  Best Local Similarity   57.3%;  Pred. No. 1.5e-39;
  Matches  443;  Conservative    0;  Mismatches  330;  Indels    0;  Gaps    0;

Qy        174 CTCCAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATAT 233
              | ||  ||||  |      |  | || |  | | | | |||  || |   |  | |||
Db       4589 CACCCACCCATCCCCATTCATAAACATTTACTATAAAATAAATCAATAAATAAATAATAA 4530

Qy        234 ATATATATATATATATATACACACACACACATATATATATACACATATATATACACAT 293
              ||  ||| ||| ||| ||| | | | | | ||||| ||| ||| | | ||| ||| | |
Db       4529 ATTAATAAATAAATAAATAAATAAATACATAAATAAATAAATAAATAAATAAATAA 4470

Qy        294 ATATACATATATACACATATATACACATATATATACACATACATATGTACACATATATAT 353
              ||| | | ||| | | ||| ||||| | ||| ||| | | | | ||  | | ||| |||
Db       4469 ATAAATAAATAAATAAATAAATACATAAATAAATAAATAAATAAATAAATAAATAA 4410

Qy        354 ACACATATGTATACACATATATACACATATATACACACATATATACACATATATACACAC 413
              | | |||  || | ||||| ||| | | ||| | | | | ||| | | ||| ||| | |
Db       4409 ATAAATAAATAAATACATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 4350

Qy        414 ATATATACACATATATACACATATATACACATATACACATATATACACATATATACATAT 473
              ||| ||||| | ||| | | ||| ||| | | ||| | | ||| | | ||| ||| |||
Db       4349 ATAAATACATAAATAAATAAATAAATAAATAAATAAATAAATACATAAATAAATAAATAA 4290

Qy        474 ATACACATATATATAATATACACACATATATATACACATATATACACACATATATACACA 533
              ||| | | ||| ||||    |  ||| ||| ||| | | ||| | | | | ||| | | |
Db       4289 ATAAATAAATAAATAAATAAACAACAAATAAATAAATAAATAAATAAATAAATAAA 4230

Qy        534 TATATACACATATATATACACATATATACACATATATACATATATACACATATATATACA 593
              || ||| | | ||| ||| | | ||| | | ||| ||| ||| ||| | | ||| ||| |
Db       4229 TAAATAAATACATAAATAAATAAATAAATAAATACATAAATAAATAAATAAATAAATAAA 4170

Qy        594 TATATACACATATATACATATATACACATATATACATATATACACACATATATACACATA 653
              || ||| | | ||| | | ||| | | ||| ||| ||| ||| | | ||| ||| | | |
Db       4169 TAAATAAATAAATACATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 4110

Qy        654 CATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATA 713
              | ||| | ||| | | | ||| ||| ||| | | ||| ||| |   ||| |||  || |
Db       4109 TAAATAAATACATAAATAAATAAATAAATAAATAAATAAATAAATAAATACATAAATAAA 4050

Qy        714 TATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAA 773
              || |||       | |  ||||          |     |  | |   | | |   |||
Db       4049 TAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATACATAAATAAATAAA 3990

Qy        774 GAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGG 833
              | |      | |  | |  ||    | |   | |  | |          |||||
Db       3989 TAAATAAATAAATAAATAAATAAATAAATACATAAATAAATAAATAAATACATAAA 3930

Qy        834 GAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATAT 893
              ||||| ||||||   |||  |||    ||   |||    | | || ||    | | || ||
Db       3929 TAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAACAACAAATAAAT 3870

Qy        894 ATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATG 946
              | ||| |||   |||  | |  | |  ||||  |      ||  |  |||| |
Db       3869 AAATAAATAAATAAATAAATAAATAAATAAATACATAAATAAATAAACG 3817


RESULT 77
US-10-437-963-66535/c
; Sequence 66535, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
```

```
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 66535
;   LENGTH: 575
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_67477C.1
US-10-437-963-66535

  Query Match             5.9%;  Score 243.6;  DB 18;  Length 575;
  Best Local Similarity   69.9%;  Pred. No. 1e-39;
  Matches  362;  Conservative    0;  Mismatches  144;  Indels   12;  Gaps    2;

Qy        215 AAAAAAAGGCATTTTATATATATATATATATATATACACACACACACATATATATA 274
              |||   |  ||  | |||||||||||||||||||||||||| | | | | | |||||||||
Db        550 AAAGGATTGCTATATATATATATATATATATATATATATATATATATATATATATATA 491

Qy        275 TACACATATATATACACATATATACATATATACACATATATACACATATATATACACATA 334
              || | |||||||||| | ||||||| ||||||| | ||||||| | | |    | |  || |
Db        490 TATATATATATATATATATATATATATATATATATATATATATATACAGGCAAATGCAAA 431

Qy        335 CATATGTACACATATATATACACATATGTATACACATATATACACATATATACACACATA 394
              ||    ||||     |   |  |       |     |   |     | |||  | || |||
Db        430 CAATATTACAAGATCAACAAGCCTACAAGACGATATTCCAGTTGGAAATAGTCCCAAATA 371

Qy        395 TATACACATATATACACACATATATACACATATATACACATATATACACATATACACATA 454
               |||   || ||     | | ||| | |  |||||| |    |      |||| |  ||
Db        370 AATATGGATGTAGGTGGAGAAATAAAAAGCTATATAAAGCAAGTGGAGCATAAAGGCAGC 311

Qy        455 TATAC-----------ACATATATACATATATACACATATATATA-ATATACACACATAT 502
              |||||           | ||||||| ||||||| | |||||||||| ||||| | | ||||
Db        310 TATACAATGTAGTAGTAGATATATATATATATATATATATATATATATATATATATATAT 251

Qy        503 ATATACACATATATACACACATATATACACATATATACACATATATATACACATATATAC 562
              ||||| | ||||||| | | ||||||| | ||||||| | ||||||||| | |||||||
Db        250 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 191

Qy        563 ACATATATACATATATACACATATATATACATATATACACATATATACATATATACACAT 622
              | |||||||| ||||||| | ||||||||| ||||||| | ||||||| ||||||| | ||
Db        190 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 131

Qy        623 ATATACATATATACACACATATATACACATACATATACACACACATAGATATACATATAT 682
              ||||| ||||||| | | ||||||| | ||| |||||||||||||| | ||||| ||||||
Db        130 ATATATATATATATATATATATATATATATATATATACACACACACATATATATATAT 71

Qy        683 ATACACATATATATACGTATATATATGTATATATATAT 720
              ||| | |||||||||  ||||||||| ||||||||||
Db         70 ATATATATATATATATATATATATATATATATATAT 33


RESULT 78
US-10-017-161-2385/c
; Sequence 2385, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
;  FILE REFERENCE: 084335/0152
;  CURRENT APPLICATION NUMBER: US/10/017,161
;  CURRENT FILING DATE:  2002-12-18
;  PRIOR APPLICATION NUMBER: JP 2001/246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2430
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2385
;   LENGTH: 1309
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: source
;   LOCATION: (1)..(1309)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(1109)
US-10-017-161-2385

  Query Match             5.9%;  Score 243.2;  DB 15;  Length 1309;
  Best Local Similarity   70.7%;  Pred. No. 1.7e-39;
  Matches  355;  Conservative    0;  Mismatches  138;  Indels    9;  Gaps    2;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | || |||| || || ||| ||||||| || |    || |||| ||||||| ||||  ||
Db        553 TATACATATGTACATCTATGTATATACGCATATGTGCATATATGTATATACGCATACGTA 494

Qy        287 TACACATATATACATATATACACATATATACACATATATATACAC------ATACATATG 340
              ||     |||||||  ||||| |||||| | || |||| |||| ||       |||||||||
Db        493 TATGTGTATATACGCATATATACATATGTGCATATATGTATATACGCATATATACATATG 434

Qy        341 TACACATATATATACACATATGTATACACATATATACACATATATACACACATATATACA 400
              | || |||| |||| ||  || |||||| || |  | | || |||| || ||||||||||
Db        433 TGCATATATGTATATACGCATATATACATATGTGCATATATGTATATACGCATATATACA 374

Qy        401 CATATATACACACATATATACACATATATACACATATATACACATATACACATATATACA 460
              || |  | | |  ||||||| |||||||||| | || | || || | || ||||  | |
Db        373 TATGTGCATATATGTATATACGCATATATACATACATGTGCATATGTGCATATATTCATA 314

Qy        461 CATATATACATATATACACATATATATAATATACACACATATATATACACATATATACAC 520
              || ||  || ||||    | ||||    || |||| ||| |||| |||| |||| ||||
Db        313 AGTACATCTATGTATATGTACATAT---ATGTACATACACATATGTACATATATGTACAT 257

Qy        521 ACATATATACACATATATACACATATATATACACATATATACACATATATACATATATAC 580
              ||| ||||  |||||||  ||||| | ||| |||||||  |||||||  | | |||||
Db        256 ACACATATGTACATATATGTACATACACATATACATATATGTACATATACATACATATAT 197

Qy        581 ACATATATATACATATATACACATATATACATATATACACATATATACATATATACACAC 640
              ||| ||| || ||||  ||| |||  || ||||| | | || |||| ||| | | | ||
Db        196 GCATGTATGTATATATGCACATATACCTATATATACATATATGTATATATACACATATAC 137

Qy        641 ATATATACACATACATATACACACACATAGATATACATATATATACACATATATATACGT 700
              ||||||  |||||  | || | |||||  | ||  | ||||  ||||| |||||||  | |
Db        136 ATATATGTACATATGTGTATATACACACATATGCATATATGTATACGCATATATGCATAT 77

Qy        701 ATATATATGTATATATATATGC 722
              || |||||  | | ||||||||
Db         76 ATGTATATACACACATATATGC 55


RESULT 79
US-10-292-798-2027/c
; Sequence 2027, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
;  FILE REFERENCE: 084335/166
;  CURRENT APPLICATION NUMBER: US/10/292,798
;  CURRENT FILING DATE:  2002-11-13
;  PRIOR APPLICATION NUMBER: 10/017,161
;  PRIOR FILING DATE: 2001-12-18
;  PRIOR APPLICATION NUMBER: JP 2001-246789
;  PRIOR FILING DATE: 2001-06-18
```

```
;  NUMBER OF SEQ ID NOS: 2070
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2027
;    LENGTH: 1309
;    TYPE: DNA
;    ORGANISM: Homo sapiens
;    FEATURE:
;    LOCATION: source
;    FEATURE:
;    LOCATION: (1)..(1309)
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (201)..(1109)
US-10-292-798-2027

  Query Match             5.9%;  Score 243.2;  DB 17;  Length 1309;
  Best Local Similarity   70.7%;  Pred. No. 1.7e-39;
  Matches  355;  Conservative    0;  Mismatches  138;  Indels    9;  Gaps    2;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | || |||| || || ||| ||||||| || |     || |||| ||||||| ||||  ||
Db        553 TATACATATGTACATCTATGTATATACGCATATGTGCATATATGTATATACGCATACGTA 494

Qy        287 TACACATATATACATATATACACATATATACACATATATATACAC------ATACATATG 340
              ||    |||||||  ||||| ||||||  | || |||| |||| ||      |||||||||
Db        493 TATGTGTATATACGCATATATACATATGTGCATATATGTATATACGCATATATACATATG 434

Qy        341 TACACATATATATACACATATGTATACACATATATACACATATATACACACATATATACA 400
              | || |||| |||| ||  || |||||| || |  | | || |||| || ||||||||||
Db        433 TGCATATATGTATATACGCATATATACATATGTGCATATATGTATATACGCATATATACA 374

Qy        401 CATATATACACACATATATACACATATATACACATATATACACATATACACATATATACA 460
              || |  | | |  ||||||| ||||||||||  | || | || || | || ||||  | |
Db        373 TATGTGCATATATGTATATACGCATATATACATACATGTGCATATGTGCATATATTCATA 314

Qy        461 CATATATACATATATACACATATATATAATATACACACATATATATACACATATATACAC 520
              || ||  || ||||   | ||||   || |||| ||| |||| |||| |||| ||||
Db        313 AGTACATCTATGTATATGTACATAT---ATGTACATACACATATGTACATATATGTACAT 257

Qy        521 ACATATATACACATATATACACATATATATACACATATATACACATATATACATATATAC 580
              ||| ||||  |||||||| ||||| | ||| |||||||| |||||| | | |||||
Db        256 ACACATATGTACATATATGTACATACACATATACATATATGTACATATACATACATATAT 197

Qy        581 ACATATATATACATATATACACATATATACATATATACACATATATACATATATACACAC 640
              ||| ||| || ||||  ||| |||  || ||||| | | || |||| ||| | | | ||
Db        196 GCATGTATGTATATATGCACATATACCTATATATACATATATGTATATATACACATATAC 137

Qy        641 ATATATACACATACATATACACACACATAGATATACATATATATACACATATATATACGT 700
              ||||||  |||||  | || | ||||| | ||  | |||| ||||| |||||||  |  |
Db        136 ATATATGTACATATGTGTATATACACACATATGCATATATGTATACGCATATATGCATAT 77

Qy        701 ATATATATGTATATATATATGC 722
              || |||||  | | ||||||||
Db         76 ATGTATATACACACATATATGC 55


RESULT 80
US-10-437-963-18717/c
; Sequence 18717, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 18717
;    LENGTH: 566
;    TYPE: DNA
;    ORGANISM: Oryza sativa
;    FEATURE:
;    OTHER INFORMATION: Clone ID: PAT_MRT4530_24249C.1
US-10-437-963-18717

  Query Match             5.9%;  Score 242.6;  DB 18;  Length 566;
  Best Local Similarity   68.5%;  Pred. No. 1.6e-39;
  Matches  354;  Conservative    0;  Mismatches  154;  Indels    9;  Gaps    1;

Qy        228 TTATATATATATATATATATATATACACACACACACACATATATATATACACATATATAT 287
              |||| | | | ||| ||| ||  |||  |   ||  |  ||||||| |||||| | | |
Db        555 TTATGTGTGTGTATGTATGTACGTACGTATGTACGTATGTATATATGTACACACACACAC 496

Qy        288 ACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACAT 347
              ||||||||||| ||||||| | ||||||| | |||||||| |  || |||| || | ||
Db        495 ACACATATATATATATATATATATATATATATATATATATAAAAGTATATATATATATAT 436

Qy        348 ATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATAT 407
              |||||||  | |||| |||| | ||||||| | ||||||| | | |||||||  | ||||||
Db        435 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 376

Qy        408 ACACACATATATACACATATATACACATATATACACATATACACATATATACACATATAT 467
              | | | ||          ||   |   |||||  |  ||              |    |  |||
Db        375 ATATATATTATGTGTTATGAGTTGCCATATGAATGCAGGGCATGGCGAAACTAGCTATGC 316

Qy        468 ACATATATACAC---------ATATATATAATATACACACATATATATACACATATATAC 518
              |  |  | |||         || | | |       | | |||||||| | |||||||
Db        315 TCGCAACTCCACAGTTCAGGTATGTGTGTGTGTGTTATATATATATATATATATATATAT 256

Qy        519 ACACATATATACACATATATACACATATATATACACATATATACACATATATACATATAT 578
              | | |||| || | ||||||||||||||||||  | | ||||||| |||||||| ||||||
Db        255 ATATATATGTATATATATATACACATATATACATATATATATATACATATATATATATAT 196

Qy        579 ACACATATATATACATATATACACATATATACATATATACACATATATACATATATACAC 638
              ||| ||||||||||||||||| | ||| ||| ||||||| | | | ||| |||||||  |
Db        195 ACATATATATATACATATATATATATACATATATATATATATACACATATATATATATAT 136

Qy        639 ACATATATACACATACATATACACACACATAGATATACATATATATACACATATATATAC 698
              | | | ||| | ||| ||||| | ||||||| ||||| ||| ||||| | |||||||||
Db        135 ATACACATATATATATATATATACACATATATATATATACATATATATATATATAT 76

Qy        699 GTATATATATGTATATATATATGCTCCAGAGTTCATA 735
              |||||||||| ||| | ||  |  |||    || |||
Db         75 GTATATATATATATGTGTAATTCGTCCGATATTAATA 39


RESULT 81
US-10-437-963-57702/c
; Sequence 57702, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 57702
;    LENGTH: 614
;    TYPE: DNA
```

```
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_59488C.1
US-10-437-963-57702

  Query Match             5.9%;  Score 242.2;  DB 18;  Length 614;
  Best Local Similarity   70.7%;  Pred. No. 2e-39;
  Matches  336;  Conservative    0;  Mismatches  138;  Indels    1;  Gaps    1;

Qy        208 CAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACACAT 267
              ||| ||| |      |  | ||||||||||||||||||||||||| | | | | | ||
Db        545 CAGTCAAATACTCCCTCCGTATATATATATATATATATATATATATATATATATATATAT 486

Qy        268 ATATATATACACATATATATACACATATATACATATATACACATATATACACATATATAT 327
              ||||||||| | |||||||||  | |||||||  |||||||  | |||||||  | ||||||||
Db        485 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 426

Qy        328 ACACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATAC 387
              | | ||| |||| || | |||||||||  | ||||  ||||  | |||||||  | |||||||
Db        425 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 366

Qy        388 ACACATATATACACATATATACACACATATATACACATATATACACATATATACACATAT 447
              | | |||||||  | |||||||  | | |||||||  | |||||||  | |||||||  | ||||
Db        365 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 306

Qy        448 ACACATATATACACATATATACATATATACACATATATATA-ATATACACACATATATAT 506
              | | |||||||  | |||||||  |||||||  | |||||||||  |||||  | |  |||||
Db        305 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAGGA 246

Qy        507 ACACATATATACACACATATATACACATATATACACATATATATACACATATATACACAT 566
              | |.    | |  || |          || || ||   |  |||||||| | |||||||  | ||
Db        245 AAATGCTTCTGTACTCCCGGGAGTACGTAAATCAAATTATATATATATATATATATATAT 186

Qy        567 ATATACATATATACACATATATATACATATATACACATATATACATATATACACATATAT 626
              |||||  |||||||  | |||||||||  |||||||  | |||||||   |  | ||| |
Db        185 ATATATATATATATATATATATATATATATATATATATATATATATATGTTCAAACAAACTCCA 126

Qy        627 ACATATATACACACATATATACACATACATATACACACACATAGATATACATATA 681
                  |  ||      | ||  || |||||  |||||    | |    | ||  | |||||
Db        125 CAGTGAATTTCATCTTAACTAGACATATATATAACAATAATAAAATTAAAATATA 71


RESULT 82
US-10-067-514-1
; Sequence 1, Application US/10067514
; Publication No. US20030054531A1
; GENERAL INFORMATION:
;  APPLICANT: Gretarsdottir, Solveig
;  APPLICANT:  Jonsdottir, Sif
;  APPLICANT:  Reynisdottir, Sigridur Th.
;  TITLE OF INVENTION: HUMAN STROKE GENE
;  FILE REFERENCE: 2345.2010-003
;  CURRENT APPLICATION NUMBER: US/10/067,514
;  CURRENT FILING DATE:  2002-02-04
;  PRIOR APPLICATION NUMBER: US 09/811/352
;  PRIOR FILING DATE: 2001-03-19
;  NUMBER OF SEQ ID NOS: 84
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 1691139
;   TYPE: DNA
;   ORGANISM: Human
US-10-067-514-1

  Query Match             5.9%;  Score 242;  DB 14;  Length 1691139;
  Best Local Similarity   70.0%;  Pred. No. 6.7e-38;
  Matches  368;  Conservative    1;  Mismatches  151;  Indels    6;  Gaps    3;

Qy        196 GAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATATACAC 255
              ||  ||| |     | | | |||| | |     | ||||||| ||||||||  |||||
Db     867636 GAGTTGTGAATATATATATAAAATATGTGTGTGTATATATACATATATATGCATATATGT 867695

Qy        256 ACACACACACATATATATATACACATATATATACACATATATACATATATACACATATAT 315
              ||  | | |||||||| |||| | |   |||||  || ||||||||||||||||||||||
Db     867696 ACCTATATACATATAT-TATATATACCTATATGAACCTATATACATATATACACATATAT 867754

Qy        316 ACACATATATATACACATACATATGTACACATATAT----ATACACATATGTATACACAT 371
              || |||||| ||| ||||| || :  || |||||     ||| ||| || | |  ||
Db     867755 GAACCTATATACACATATACACATAYGTACCTATATACACATATACACATATGTGTACCC 867814

Qy        372 ATATACACATATATACACACATATATACACATATATACACACATATATACACATATATAC 431
              |||||||||||||| ||| |  ||  || | | | ||| |||||||| |  ||  ||||||
Db     867815 ATATACACATATACACATATGTACCTATATACACATATACACATATGTGTACCCATATAC 867874

Qy        432 ACATATATACACATATACACATATATACACATATATACATATATACACATATATATAATA 491
              ||||||| ||| || |  ||  |||||||||||||| ||||||  |  ||  |||||  | |
Db     867875 ACATATACACATATGTGTACCCATATACACATATACACATATGTGTACCCATATA-CACA 867933

Qy        492 TACACACATATATATACACATATATACACACATATATACACATATATACACATATATATA 551
              || |||||||| | ||| |||||| ||| | | | |||     ||   | | | | |||||
Db     867934 TATACACATATGTGTACCCATATACACATATACACATATGTGTACCCATATACACATATA 867993

Qy        552 CACATATATACACATATATACATATATACACATATATATACATATATACACATATATACA 611
              |||||||  |  ||  |||||||  |||||||||||| | |||  ||||||||||||| |||
Db     867994 CACATATGTGTACCCATATACACATATACACATATGTGTACCCATATACACATATACACA 868053

Qy        612 TATATACACATATATACATATATACACACATATATACACATACATATACACACACATAGA 671
              | | |||  ||||||  ||||||| |||     ||    | | | ||||||||||     ||
Db     868054 TGTGTACCCATATACACATATACACATGTGTACCCATATACACATATACACATGTGTACC 868113

Qy        672 TATACATATATATACACATATATATACGTATATATATGTATATATA 717
              |||  | |  |||||  |||||  | |||  |||||    |||  | ||
Db     868114 CATATACACATATATGCATATGTGTACCCATATACGCATATGTGTA 868159


RESULT 83
US-10-419-723-1
; Sequence 1, Application US/10419723
; Publication No. US20040014099A1
; GENERAL INFORMATION:
;  APPLICANT: Gretarsdottir, Solveig
;  APPLICANT:  Jonsdottir, Sif
;  APPLICANT:  Reynisdottir, Sigridur Th.
;  APPLICANT:  Thorleifsson, Gudmar
;  APPLICANT:  Gulcher, Jeffrey
;  TITLE OF INVENTION: SUSCEPTIBILITY GENE FOR HUMAN STROKE;
;  TITLE OF INVENTION:  METHODS OF TREATMENT
;  FILE REFERENCE: 2345.2010-005
;  CURRENT APPLICATION NUMBER: US/10/419,723
;  CURRENT FILING DATE:  2003-04-18
;  PRIOR APPLICATION NUMBER: 10/255,120
;  PRIOR FILING DATE: 2002-09-25
;  PRIOR APPLICATION NUMBER: 10/067,514
;  PRIOR FILING DATE: 2002-02-04
;  PRIOR APPLICATION NUMBER: 09/811,352
;  PRIOR FILING DATE: 2001-03-19
;  NUMBER OF SEQ ID NOS: 87
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 1691139
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-419-723-1

  Query Match             5.9%;  Score 242;  DB 17;  Length 1691139;
  Best Local Similarity   70.0%;  Pred. No. 6.7e-38;
  Matches  368;  Conservative    1;  Mismatches  151;  Indels    6;  Gaps    3;

Qy        196 GAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATATACAC 255
              ||  ||| |     | | | |||| | |     | ||||||| ||||||||  |||||
Db     867636 GAGTTGTGAATATATATATAAAATATGTGTGTGTATATATACATATATATGCATATATGT 867695

Qy        256 ACACACACACATATATATATACACATATATATACACATATATACATATATACACATATAT 315
              ||  | | |||||||| |||| | |   |||||  || ||||||||||||||||||||||
Db     867696 ACCTATATACATATAT-TATATATACCTATATGAACCTATATACATATATACACATATAT 867754
```

```
Qy        316 ACACATATATATACACATACATATGTACACATATAT----ATACACATATGTATACACAT 371
              || |||||| ||| ||||| || :  || |||||    ||| ||| || | |  ||
Db     867755 GAACCTATATACACATATACACATAYGTACCTATATACACATATACACATATGTGTACCC 867814

Qy        372 ATATACACATATATACACACATATATACACATATATACACACATATATACACATATATAC 431
              |||||||||||||| ||| |  ||  || | | | ||| |||||||| |  ||  ||||||
Db     867815 ATATACACATATACACATATGTACCTATATACACATATACACATATGTGTACCCATATAC 867874

Qy        432 ACATATATACACATATACACATATATACACATATATACATATATACACATATATATAATA 491
              ||||||| ||| || |  ||  |||||||||||| |||||| |  ||  |||||  | |
Db     867875 ACATATACACATATGTGTACCCATATACACATATACACATATGTGTACCCATATA-CACA 867933

Qy        492 TACACACATATATATACACATATATACACACATATATACACATATATACACATATATATA 551
              || |||||||| | ||| |||||| ||| | | | |||    ||   | | | |||||
Db     867934 TATACACATATGTGTACCCATATACACATATACACATATGTGTACCCATATACACATATA 867993

Qy        552 CACATATATACACATATATACATATATACACATATATATACATATATACACATATATACA 611
              ||||||| |  ||  ||||||| |||||||||||| | |||  ||||||||||||| |||
Db     867994 CACATATGTGTACCCATATACACATATACACATATGTGTACCCATATACACATATACACA 868053

Qy        612 TATATACACATATATACATATATACACACATATATACACATACATATACACACACATAGA 671
              | | ||| |||||| |||||| |||    ||   | | | ||||||||||    ||
Db     868054 TGTGTACCCATATACACATATACACATGTGTACCCATATACACATATACACATGTGTACC 868113

Qy        672 TATACATATATATACACATATATATACGTATATATATGTATATATA 717
              ||| | | |||||  ||||| | |||  |||||    ||| | ||
Db     868114 CATATACACATATATGCATATGTGTACCCATATACGCATATGTGTA 868159


RESULT 84
US-10-017-161-2211/c
; Sequence 2211, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
;  FILE REFERENCE: 084335/0152
;  CURRENT APPLICATION NUMBER: US/10/017,161
;  CURRENT FILING DATE:  2002-12-18
;  PRIOR APPLICATION NUMBER: JP 2001/246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2430
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2211
;   LENGTH: 1441
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: source
;   LOCATION: (1)..(1441)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(1241)
US-10-017-161-2211

  Query Match             5.9%;  Score 241.2;  DB 15;  Length 1441;
  Best Local Similarity   69.5%;  Pred. No. 4.6e-39;
  Matches  357;  Conservative    0;  Mismatches  153;  Indels    4;  Gaps    2;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | || |||||||| ||| | |||| |  | | | |||||||||| |||||||||| |||
Db        702 TATACATATATATAGATACACATATGCGTATATATACACATATATGTATACACATACATA 643

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              | |  ||||| |||||||| | | | ||||| || | |||||| ||| ||| |||
Db        642 TGCGTATATACACATATATGTATACACATACATATGTGTATACATATATGTATACACATG 583

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||| |  | |||| |||  |||||||||   ||||| ||| |    |||| |||||||
Db        582 CATATGTGTATATATTTATGAACATATATATGTATATACACATATGCGTATATACATATA 523


Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              |  | ||| ||||  | |||| ||   | | ||| |  ||||||||||| |||    ||||
Db        522 TGTATACACATATGTATATATTTATGTACACATATATGTATACACATACATATGTGTATA 463

Qy        467 TACATATATACACATATATATAATATACACACATATATATACACATATATACACACATAT 526
              ||  ||| ||||||||||||   |||||||| | |||| |  | |||| |||||| || |
Db        462 TATTTATGTACACATATAT---GTATACACATACATATGTGTATATATTTACACATATGT 406

Qy        527 ATACACATATATACACATATATATACACATATATACACATATATACATATATACACATAT 586
              ||||||||| ||| |  | ||||||   | | | ||| || |||||| ||| | |  | |
Db        405 GTACACATACATATATGTGTATATATTTACACACACATATGTATACACATACATATGTGT 346

Qy        587 ATATACATATATACACATATATACATATATACACATATATACATA-TATACACACATATA 645
              |||||  |||  |||||||| || | | ||||  || | || ||| |  ||||| || ||
Db        345 ATATATTTATGCACACATATGTATACACATACGTATGTGTATATATTGCACACATATGTA 286

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              ||||||||||||    | | ||  ||  ||| |||| || ||| ||| ||| || ||||
Db        285 TACACATACATATGTGTATATATTTATGCACACATATGTATACACATACATATGTGTATA 226

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAG 739
              | | |||||| | |||         | |||| | |
Db        225 TGTTTATATACACATGTATACACATGCATATGTG 192


RESULT 85
US-10-292-798-1857/c
; Sequence 1857, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
;  FILE REFERENCE: 084335/166
;  CURRENT APPLICATION NUMBER: US/10/292,798
;  CURRENT FILING DATE:  2002-11-13
;  PRIOR APPLICATION NUMBER: 10/017,,161
;  PRIOR FILING DATE: 2001-12-18
;  PRIOR APPLICATION NUMBER: JP 2001-246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2070
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1857
;   LENGTH: 1441
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   LOCATION: source
;   FEATURE:
;   LOCATION: (1)..(1441)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(1241)
US-10-292-798-1857

  Query Match             5.9%;  Score 241.2;  DB 17;  Length 1441;
  Best Local Similarity   69.5%;  Pred. No. 4.6e-39;
  Matches  357;  Conservative    0;  Mismatches  153;  Indels    4;  Gaps    2;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | || |||||||| ||| | |||| |  | | | |||||||||| |||||||||| |||
Db        702 TATACATATATATAGATACACATATGCGTATATATACACATATATGTATACACATACATA 643

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              | |  ||||| |||||||| | | | ||||| || | |||||| ||| ||| |||
Db        642 TGCGTATATACACATATATGTATACACATACATATGTGTATACATATATGTATACACATG 583

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||| |  | |||| |||  |||||||||   ||||| ||| |    |||| |||||||
Db        582 CATATGTGTATATATTTATGAACATATATATGTATATACACATATGCGTATATACATATA 523
```

```
Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              |  | ||| ||||  | |||| ||   | | ||| |  |||||||||| |||    ||||
Db        522 TGTATACACATATGTATATATTTATGTACACATATATGTATACACATACATATGTGTATA 463

Qy        467 TACATATATACACATATATATAATATACACACATATATATACACATATATACACACATAT 526
              ||  ||| ||||||||||    |||||||| | |||| |  | |||| |||||| || |
Db        462 TATTTATGTACACATATAT---GTATACACATACATATGTGTATATATTTACACATATGT 406

Qy        527 ATACACATATATACACATATATATACACATATATACACATATATACATATATACACATAT 586
              |||||||| ||| |  | ||||||   | | | ||| || |||||| ||| | |  | |
Db        405 GTACACATACATATATGTGTATATATTTACACACACATATGTATACACATACATATGTGT 346

Qy        587 ATATACATATATACACATATATACATATATACACATATATACATA-TATACACACATATA 645
              |||||  |||  |||||||| || | | ||||  || | || ||| |  |||| || ||
Db        345 ATATATTTATGCACACATATGTATACACATACGTATGTGTATATATTGCACACATATGTA 286

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              ||||||||||||     | | ||  ||  ||| |||| || ||| ||| ||| || ||||
Db        285 TACACATACATATGTGTATATATTTATGCACACATATGTATACACATACATATGTGTATA 226

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAG 739
              | | |||||| | |||        | |||| | |
Db        225 TGTTTATATACACATGTATACACATGCATATGTG 192


RESULT 86
US-10-017-161-2387
; Sequence 2387, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
;  FILE REFERENCE: 084335/0152
;  CURRENT APPLICATION NUMBER: US/10/017,161
;  CURRENT FILING DATE:  2002-12-18
;  PRIOR APPLICATION NUMBER: JP 2001/246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2430
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2387
;   LENGTH: 1276
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: source
;   LOCATION: (1)..(1276)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(1076)
US-10-017-161-2387

  Query Match             5.8%;  Score 240.8;  DB 15;  Length 1276;
  Best Local Similarity   69.9%;  Pred. No. 5.3e-39;
  Matches  341;  Conservative    0;  Mismatches  142;  Indels    5;  Gaps    1;

Qy        230 ATATATATATATATATATATATACACACACACACATATATATATACACATATATATAC 289
              | ||||| |||| || |||||||||| |   ||| | |||  |||||||| || || ||
Db        793 ACATATACATATGTACATATATACATATGTACATATATACGTATATACATATGTACATCT 852

Qy        290 ACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATAT 349
              |  |||||||  ||||  |||||||  | |||   ||||  |||  | | || ||  ||
Db        853 ATGTATATACGCATATGTGCATATATGTATATACGCATACGTATATGTGTATATACGCAT 912

Qy        350 ATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATAC 409
              |||||||  || ||  |||  |  |||||||  |||||||||| |  |  |||  |  |||||||
Db        913 ATATACATATGTGCATATATGTATATACGCATATATACATATGTGCATATATGTATATAC 972

Qy        410 ACACATATATACACATATATACACATATATACACATATACACATATATACACATATATAC 469
               || ||||| | |  |  ||| |  ||||||| |||||| |||||| | || |||| ||
```

```
Db        973 GCATATATACATATGTGCATATATGTATATACGCATATATACATATGTGCATATATGTAT 1032

Qy        470 ATATATACACATATATATAATATACACACATATATATACACATATATACACACATATATA 529
              |||    | | |||       ||| |     ||||| |  | | ||  |||   |||| |||
Db       1033 ATACGCATATATA-----CATACATGTGCATATGTGCATATATTCATAAGTACATCTATG 1087

Qy        530 CACATATATACACATATATATACACATATATACACATATATACATATATACACATATATA 589
              | || || | | || || |||||||||| |||| |||| |||||| | | |  || |||
Db       1088 TATATGTACATATATGTACATACACATATGTACATATATGTACATACACATATGTACATA 1147

Qy        590 TACATATATACACATATATACATATATACACATATATACATATATACACACATATATACA 649
              ||  || ||||||||||| | |||| |||| ||| |||||||||| ||   || |||| |
Db       1148 TATGTACATACACATATACATATATGTACATATACATACATATATGCATGTATGTATATA 1207

Qy        650 CATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATATG 709
              |||||||| | | ||| |||||  ||||||||||||||||| |||  | || |||||
Db       1208 TGCACATATACCTATATATACATATATGTATATATACACATATACATATATGTACATATG 1267

Qy        710 TATATATA 717
              | ||||||
Db       1268 TGTATATA 1275


RESULT 87
US-10-292-798-2029
; Sequence 2029, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
;  FILE REFERENCE: 084335/166
;  CURRENT APPLICATION NUMBER: US/10/292,798
;  CURRENT FILING DATE:  2002-11-13
;  PRIOR APPLICATION NUMBER: 10/017,161
;  PRIOR FILING DATE: 2001-12-18
;  PRIOR APPLICATION NUMBER: JP 2001-246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2070
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2029
;   LENGTH: 1276
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   LOCATION: source
;   FEATURE:
;   LOCATION: (1)..(1276)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(1076)
US-10-292-798-2029

  Query Match             5.8%;  Score 240.8;  DB 17;  Length 1276;
  Best Local Similarity   69.9%;  Pred. No. 5.3e-39;
  Matches  341;  Conservative    0;  Mismatches  142;  Indels    5;  Gaps    1;

Qy        230 ATATATATATATATATATATATACACACACACACATATATATATACACATATATATAC 289
              | ||||| |||| || |||||||||| |   ||| | |||  |||||||| || || ||
Db        793 ACATATACATATGTACATATATACATATGTACATATATACGTATATACATATGTACATCT 852

Qy        290 ACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATAT 349
              |  |||||||  ||||  |||||||  | |||   ||||  |||  | | || ||  ||
Db        853 ATGTATATACGCATATGTGCATATATGTATATACGCATACGTATATGTGTATATACGCAT 912

Qy        350 ATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATAC 409
              |||||||  || ||  |||  |  |||||||  |||||||||| |  |  |||  |  |||||||
Db        913 ATATACATATGTGCATATATGTATATACGCATATATACATATGTGCATATATGTATATAC 972

Qy        410 ACACATATATACACATATATACACATATATACACATATACACATATATACACATATATAC 469
               || ||||| | |  |  ||| |  ||||||| |||||| |||||| | || |||| ||
```

```
Db        973 GCATATATACATATGTGCATATATGTATATACGCATATATACATATGTGCATATATGTAT 1032

Qy        470 ATATATACACATATATATAATATACACACATATATATACACATATATACACACATATATA 529
              |||   | | |||      ||| |    ||||| |  | | ||  |||   |||| |||
Db       1033 ATACGCATATATA-----CATACATGTGCATATGTGCATATATTCATAAGTACATCTATG 1087

Qy        530 CACATATATACACATATATATACACATATATACACATATATACATATATACACATATATA 589
              | || || | | || || |||||||||| |||| |||| |||||| | | |  || |||
Db       1088 TATATGTACATATATGTACATACACATATGTACATATATGTACATACACATATGTACATA 1147

Qy        590 TACATATATACACATATATACATATATACACATATATACATATATACACACATATATACA 649
              ||  || |||||||||| | |||| |||| ||| |||||||||| ||   || |||| |
Db       1148 TATGTACATACACATATACATATATGTACATATACATACATATATGCATGTATGTATATA 1207

Qy        650 CATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATATG 709
              ||||||||  | | ||| |||||  ||||||||||||||| |||  | || |||||
Db       1208 TGCACATATACCTATATATACATATATGTATATATACACATATACATATATGTACATATG 1267

Qy        710 TATATATA 717
              | ||||||
Db       1268 TGTATATA 1275


RESULT 88
US-10-437-963-76607
; Sequence 76607, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 76607
;   LENGTH: 534
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_76588C.1
US-10-437-963-76607

  Query Match             5.8%;  Score 240.6;  DB 18;  Length 534;
  Best Local Similarity   76.4%;  Pred. No. 4e-39;
  Matches  308;  Conservative    0;  Mismatches   94;  Indels    1;  Gaps    1;

Qy        293 TATATACATATATACACATATATACACATATATATACACATACATATGTACACATATATA 352
              ||| |    ||    | | ||||| ||||| ||||| | ||| |||| || | |||||||
Db         17 TATTTCACAATTATAAGAAATATATACATACATATATATATATATATATATATATATA 76

Qy        353 TACACATATGTATACACATATATACACATATATACACACATATATACACATATATACACA 412
              || | |||| |||| |||||||||| | |||||||||| ||||| ||||| |||||||||| |
Db         77 TATATATATATATATACATATATATATATATATACATACATACATACATATATATACATA 136

Qy        413 CATATATACACATATATACACATATATACACATATACACATATATACACATATATACATA 472
              |||| ||||| | ||||| | ||||||||| | ||||| ||||||| | ||||||| |||
Db        137 CATACATACATACATATATATATATATATACATACATACATATATATATATATATATATA 196

Qy        473 TATACACATATATATA-ATATACACACATATATATACACATATATACACACATATATACA 531
              |||| | ||||||||| ||||| | | |||||||||| | ||||||| ||| | ||||| |
Db        197 TATATATATATATATATATATATATATATATATATATATATATATATACATACATATATA 256

Qy        532 CATATATACACATATATATACACATATATACACATATATACATATATACACACATATATATA 591
              |||||||||||||||||||||| ||||| ||| |||||||||||||||| | |||| ||||
Db        257 CATATATACACATATATATACATATATACACATATATATACATATATATATATGTATA 316

Qy        592 CATATATACACATATATACATATATACACATATATACATATATACACACATATATACACA 651
              |||||||||||||||||| | ||||| | ||||||| | ||| |||||||| | | | |
Db        317 CATATATACACATATATATACATATATATATATATATACATACACACACACACATATC 376

Qy        652 TACATATACACACACATAGATATACATATATATACACATATAT 694
                |||   |   || | | || | |  |   |    ||| ||
Db        377 ATGATAGTAAGTTACTTGGCTAGAGAATTTGTTTTGAATACAT 419


RESULT 89
US-10-741-601-5636
; Sequence 5636, Application US/10741601
; Publication No. US20040166519A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  STENOSIS, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001500
;  CURRENT APPLICATION NUMBER: US/10/741,601
;  CURRENT FILING DATE:  2003-12-22
;  NUMBER OF SEQ ID NOS: 26415
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5636
;   LENGTH: 81924
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(81924)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-2
US-10-741-601-5636

  Query Match             5.8%;  Score 240.6;  DB 18;  Length 81924;
  Best Local Similarity   69.3%;  Pred. No. 3.5e-38;
  Matches  343;  Conservative    0;  Mismatches  149;  Indels    3;  Gaps    1;

Qy        233 TATATATATATATATATATACACACACACACATATATATATACACATATATATACACA 292
              ||| |||| | |  ||| || ||||||||  |  |||| | | |   ||||| | ||  |
Db      37545 TATGTATACACACGTATGTATACACACACGTATGTATACACACATGTATATACACACGTA 37604

Qy        293 TATATACATATATACACATATATACACATATATATACACATACATATGTACACATATATA 352
              | |||||| | ||  | ||||| |||| ||| |||||||| |  ||| || ||| |||||
Db      37605 TGTATACACACATGTATATATACACACGTATGTATACACACATGTATATATACACATATA 37664

Qy        353 TACACATATGTATACACATATATACACATATATACACACATATATACACATATATACACA 412
              |||| |||||  || | ||| | ||  ||||||       |||||    |||| |||||
Db      37665 TACATATATGGGTATATATACACACGTGTATATATGTGTGCATATATGTGTATACACACA 37724

Qy        413 CATATATACACATATATACACATATATACACATATACACATATATACACATATATACATA 472
              || | |  | ||| ||||||| ||||      |||| ||||| | ||| | |||||  |
Db      37725 TATGTGTGTATATACATACACACATATGTGTGTATATACATACACACAAACATATATGTG 37784

Qy        473 TATACACATATATATA---ATATACACACATATATATACACATATATACACACATATATA 529
              | |  |  | ||||||   ||||| | ||| ||||| ||||| ||||    | ||| | |
Db      37785 TGTGTATGTGTATATATACATATATATACACATATACACACACATATGTGTATATACACA 37844

Qy        530 CACATATATACACATATATATACACATATATACACATATATACATATATACACATATATA 589
              | | | | ||| || | |||| ||| |||||||||| | |||  | |||| ||| |||||
Db      37845 TATACACACACATATGTGTATATACACATATACACACACATATGTGTATATACACATATA 37904

Qy        590 TACATATATACACATATATACATATATACACATATATACATATATACACACATATATACA 649
              ||| ||||    ||||| ||||||| ||||| ||||   |||||||||| ||| | |||
Db      37905 CACACATATGTGTATATACACATATACACACACATATGTGTATATACACATATACACACA 37964

Qy        650 CATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATATG 709
              | |  |||||||| | ||   |||||| |||||||||||||| | |  |||| | |||
Db      37965 CGTGTATATACACATATATGTGTATACATGTATATACACATATGTGTGTGTATGTGTATA 38024

Qy        710 TATATATATATGCTC 724
              | |||  ||||| | |
Db      38025 TGTATGTATATACAC 38039
```

```
RESULT 90
US-10-437-963-46068/c
; Sequence 46068, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 46068
;   LENGTH: 610
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_48971C.1
US-10-437-963-46068

  Query Match             5.8%;  Score 240.4;  DB 18;  Length 610;
  Best Local Similarity   69.2%;  Pred. No. 4.7e-39;
  Matches  342;  Conservative    0;  Mismatches  151;  Indels    1;  Gaps    1;

Qy        226 TTTTATATATATATATATATATATACACACACACACACATATATATATACACATATAT 285
              ||  |||| || ||  | || | |          ||    | ||||||||||||    | |||
Db        559 TTGAGTATACATCTAGGTGTACAGAAGAGTCGTCATGTATATATATATATATGTGTGTAT 500

Qy        286 ATACACATATATACATATATACACATATATACACATATATATACACATACATATGTACAC 345
              ||| | ||||||   |||||| |  | | || | |||||| ||||  || |||||||||
Db        499 ATATATATATATGTGTATATATATGTGTGTATATATATATGTACATGTATATATGTACAT 440

Qy        346 ATATATATACACATATGTATACACATATATACACATATATACACACATATATACACATAT 405
              | ||||||| | |||| ||||    |||| || |  || ||     | ||||||| |||| |
Db        439 GTGTATATATATATATATATATGTATATGTATATGTACATGTGTATATATATATACATGT 380

Qy        406 ATACACACATATATACACATATATACACATATATACACATATACACATATATACACACAT 465
              |    | ||||||| |||| | |      | | |  | ||||| | ||||||| | ||||
Db        379 GTGTGTATATATATATACATGTGTGTGTGTGTGTGTATATATATATATATATATATATAT 320

Qy        466 ATACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACAT 524
              ||| ||||||| | ||||||||| |||||||||| ||||||| | ||||||| | |||
Db        319 ATATATATATATATATATATATATATATACACACATATATATATATATATATATACAC 260

Qy        525 ATATACACATATATACACATATATATACACATATATACACATATATACATATATACACAT 584
              ||| | | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | ||
Db        259 ATACATATATATATATATATATATATATATATATATATATATATATATATATATATAT 200

Qy        585 ATATATACATATATACACATATATACATATATACACATATATACATATATACACACAT 644
              ||||||| ||||||| | ||||||| | ||||| | ||| |  || ||| | | ||||||
Db        199 ATATATATATATATATATATATATATATACATATATATATACACTCACATAGATATACATAT 140

Qy        645 ATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATAT 704
              | | ||    ||||||| ||| ||| || | | | ||| |    |||||||| |    ||
Db        139 ACATACTATTATATACATACATATATATTTCCGTCCATACTTTCATATATACATCCGTAC 80

Qy        705 ATATGTATATATAT 718
              || ||||  || |
Db         79 CTACGTATTAATTT 66


RESULT 91
US-10-311-455-1842/c
; Sequence 1842, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determi
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1842
;   LENGTH: 13573
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1842

  Query Match             5.8%;  Score 240.4;  DB 15;  Length 13573;
  Best Local Similarity   69.2%;  Pred. No. 1.8e-38;
  Matches  537;  Conservative    0;  Mismatches  206;  Indels   33;  Gaps   14;

Qy          4 TCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAA--TACAAAATTAG 61
              | || |||| ||| | ||| ||    ||| ||| ||||||| |||||     |||||||
Db       6406 TTAAAACCAACCTAAACAATATAACAAAATCCCGTCTCTACAAAAAAATACAAAAATTAA 6347

Qy         62 ACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAATC 121
              ||  | |  | | |||| ||| |||||||| |||||||    |  |  |    |  | ||||
Db       6346 CCAAACTTAATAACACACACCTATAATCCCAACTACTTAAAAAACTAAAATAAAAAAATC 6287

Qy        122 ACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAGCC 181
              |||| ||||          | |       | ||  ||| | ||| | ||||||| ||
Db       6286 ACTTAAACCCAAAAAATAAAAAC-TACAATAAACCATAATCGTACCACTACACTCCAACC 6228

Qy        182 CAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATAT---A 238
              | ||| || |   ||   | |     | |||||| | |       | |||||||||   |
Db       6227 TAAACAATAAAACAAAACCCTAACAAAAAAAAAAATATATATATATATATATATAATAA 6168

Qy        239 TATATATATATATACACACACACACATATATATAT-ACACATATATATACACATATAT 297
              || |||||||||||  | | |||   ||||||||| | |||||||||| || |||||
Db       6167 TACATATATATATACGTATATACATTTATATATATATAATACATATATATATACGTATAT 6108

Qy        298 ACATATATACACATAT-ATACACATATATATACACATACATATGTACACATATATATACA 356
              |||| |||| | |||| ||||| ||||||||||  ||| | || || | ||||||| | |
Db       6107 ACATTTATATATATATAATACATATATATATACGTATATACATTTATATATATATA-ATA 6049

Qy        357 CATATGTATACACATATATACACATATAT-------ACACACATATATACACATATATAC 409
              ||||| |||| || ||||||||  |||||       | ||| ||||||| || |||||||
Db       6048 CATATATATATACGTATATACATTTATATATATATAATACATATATATATACGTATATAC 5989

Qy        410 ACACATAT-------ATACACATATATACACATATATACACATATACACATATATACACA 462
              |    ||||      ||||| |||||||| || ||||||||  |||| | ||||| | |||
Db       5988 ATTTATATATATATAATACATATATATATACGTATATACATTTATATATATATA-ATACA 5930

Qy        463 TATATACATATATACACATATATATAATATACACACATATATATACACATATATACACAC 522
              |||||| |||||||| ||||||||     | ||| |||||||||| ||||| |    |
Db       5929 TATATATATATATACATTTATATATATATAATACATATATATATACATATATACATTTAT 5870

Qy        523 ATATAT---ACACATATATACACATATATATACACATATATACACATATATACATATATA 579
              ||||||   | |||||||| | | |||||||||  |||||| | ||| ||||||||||
Db       5869 ATATATATAATACATATATATATACATATATACATTTATATATATATA-ATACATATATA 5811

Qy        580 CACATATATATACATATATACACAT---ATATACATATATACACATATATACATATATAC 636
              | | |||||||||| |||| | ||    ||| | ||||||| ||||||||||| ||||
Db       5810 TATACATATATACATTTATATATATATAATACATATATATATACATATATACATTTATAT 5751

Qy        637 ACACATATATACACATACATATACACACACATAGATATACATATAT-ATACACATATATA 695
```

```
                | | ||| ||||| ||| ||||||| | | | |  |||| |||||| ||||| |||||||
Db       5750 ATATATA-ATACATATATATATATACATATATACATTTATATATATATAATACATATATATA 5692

Qy        696 TACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGAT 751
              ||| |||||| || |||||||||||  | || |  |    | |  ||    || ||
Db       5691 TACATATATACATTTATATATATATAATACATATATATATACATATATACATTTAT 5636


RESULT 92
US-10-719-993-6943
; Sequence 6943, Application US/10719993
; Publication No. US20040265849A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001496
;  CURRENT APPLICATION NUMBER: US/10/719,993
;  CURRENT FILING DATE:  2003-11-24
;  NUMBER OF SEQ ID NOS: 55342
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6943
;   LENGTH: 145806
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-719-993-6943

  Query Match             5.8%;  Score 239;  DB 18;  Length 145806;
  Best Local Similarity   64.2%;  Pred. No. 9.6e-38;
  Matches  437;  Conservative    3;  Mismatches  228;  Indels   13;  Gaps    5;

Qy          7 AGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATTAGACAG 65
              ||  || ||||   |||||||||||  ||||||||||||||||||| ||||||||
Db      68913 AGCTCATCCTGGCTAACATGGTGAAAAGCCATCTCTACTAAAAATACAAAAATTAGCTGA 68972

Qy         66 GCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAATCACTT 125
              ||||||||  ||  ||||||| || ||||||||  |||||| |||||||||||||| |||
Db      68973 GCATGGTGGCACGTGCCTGTAGTCTCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCTT 69032

Qy        126 GAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCCAGCCCAG 184
              |||||||  || |||||||||| | | ||| |||||||||||||| ||||||||||||  |
Db      69033 GAACCTG--GGAGGTGGAGGTTACAGTGAGCCAAGATCATGCCACTGCACTCCAGCCTGG 69090

Qy        185 GCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATA 244
              || |||       | | |   || | |||||||| |      | ||||||||||: | |
Db      69091 GCGACAGAGCAAGACTCTTGCCTAAAAAAAAAAAATA------TATATATATATAWGTGTG 69144

Qy        245 TATATATACACACACACACATATATATATACACATATATATACACATATATACATATA 304
              | | | |                | |||||||   | |  |||| | ||||| |||||||
Db      69145 TGTGTGTGTGTGTGTGTGTACATATATACGTATACGTATACATACATACATACATACG 69204

Qy        305 TACACATATATACACATATATATACACATACATATGTACACATATATATACACATATGTA 364
              || ||||| ||||| || |||| | |||||| |||   || | ||| ||||| ||| | |
Db      69205 TATACATACATACATATGTATACATACATACGTATACGCATACATACATACATATACGCA 69264

Qy        365 TACACATATATACACATATATACACACATATATACACATATATACACACATATATACACA 424
              |||| | |||||| |||| ||||||| |||  || ||  ||| | | |  || |  ||
Db      69265 TACATACATATACGCATACATACACATATACGTATACGCATACATATATACGTACATGCA 69324

Qy        425 TATATACACATATATACACATATACACATATATACACATATATACATATATACACATATA 484
              || ||||| ||||| ||| | ||  |  |||||||| |:|||  |||||||||| |||||
Db      69325 TACATACATATATACACATACATGTATGTATATACATAYATACGCATATATACATATATA 69384

Qy        485 TATAATATACACACATATATATACACATATATACACACATATATACACATATATACACAT 544
                |||| |   | |||||||  | |||   :  |   || |||| |||| |||    ||
Db      69385 ---CGTATATATGCGTATATATGTATATACGCRTATGTATGTATATACATGTATGTGTAT 69441

Qy        545 ATATATACACATATATACACATATATACATATATACACATATATATACATATATACACAT 604
              ||||| ||| ||||||  |  | |   |  | |    |  |         | |   ||
Db      69442 ATATACACATATATATGAAGTTTTTATCTGACTTTATTAGCCACTAGAGCTTGTTTGCAG 69501

Qy        605 ATATACATATATACACATATATACATATATACACACATATATACACATACATATACACAC 664
               ||    | |  | |||| | |   |  || |    ||   |  |  |   | || |
```

```
Db      69502 TTAATATTTTGCAGACATTTCTGTTTTCATCCTTTTATTCTTGTATCTGTCTTTATATTT 69561

Qy        665 ACATAGATATACATATATATA 685
              | |  |   | | |||| |||
Db      69562 AAAGTGGGTTTCTTATAAATA 69582


RESULT 93
US-10-417-375-36
; Sequence 36, Application US/10417375
; Publication No. US20040219528A1
; GENERAL INFORMATION:
;  APPLICANT: David W. Morris
;  APPLICANT:  Marc Malandro
;  TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
;  FILE REFERENCE: 529452001600
;  CURRENT APPLICATION NUMBER: US/10/417,375
;  CURRENT FILING DATE:  2003-04-15
;  NUMBER OF SEQ ID NOS: 176
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 36
;   LENGTH: 91697
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-417-375-36

  Query Match             5.8%;  Score 237.8;  DB 18;  Length 91697;
  Best Local Similarity   71.6%;  Pred. No. 1.4e-37;
  Matches  353;  Conservative    0;  Mismatches  137;  Indels    3;  Gaps    3;

Qy        229 TATATATATATATATATATATATACACACACACACATATATATATACACATATATATA 288
              | ||||||| |||||| |||||||||| |    |    | |||| | |||| |||||||||| |
Db      19560 TCTATATATCTATATACATATATACATATGTATGTATATATTTCTATATACATATATACA 19619

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              |  |||||| | ||  | | | || || | | |||| || | ||  || | ||||  ||
Db      19620 TATGTATATATAGATTGATATACATCTATATACATATGTATATATCTATGTATACATCTA 19679

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              | |||||| | |||||||||  ||| || ||||| ||  | ||||||||||| || ||||
Db      19680 TGTATACATAGATGTATACATCTATGTATACATAGATGTATACATATATACATATGTATA 19739

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              ||   || |||||| ||||| | |  |||||||   |||  | |||  || ||||||  ||
Db      19740 CATCTATGTATACATATATACATATGTATATACTATTATGTATATA-GTATACATATGTA 19798

Qy        469 CATATATACACATATATATAATATACACACATATATATACACATATATACACACATATAT 528
              ||-|| || | |||| ||| ||||||| | || ||||||   ||||||  | | |  |||
Db      19799 TATATGTATATATATCTAT-ATATACATATATGTATATATGTATATATCTATATACGTAT 19857

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACATATAT 588
              | |  |||||||| ||||| ||| |  |||||||| || |||| || |||| | ||||||
Db      19858 ATAGGTATATACATATATAAATATATCTATATACATATGTATATATCTATATAAATATAT 19917

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
              | ||||||  ||||||||  ||||| | | || |||| || |||| ||| ||| ||
Db      19918 GGATATATATCTACATATATCTATATACATATATCTATATATCTATATACATATAGATGT 19977

Qy        649 ACATACATATACA-CACACATAGATATACATATATATACACATATATATACGTATATATA 707
              | ||  || ||||   || ||  ||||| ||||||||| |||| |||||   ||| ||||
Db      19978 ATATGTATCTACATGGCATATGTATATAGATATATATATACATCTATATGTATATCTATA 20037

Qy        708 TGTATATATATAT 720
              |  |||| |||||
Db      20038 TACATATCTATAT 20050


RESULT 94
US-10-437-963-90135
; Sequence 90135, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
```

```
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 90135
;   LENGTH: 573
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_88833C.1
US-10-437-963-90135

  Query Match             5.7%;  Score 236.4;  DB 18;  Length 573;
  Best Local Similarity   67.4%;  Pred. No. 3e-38;
  Matches  333;  Conservative    0;  Mismatches  161;  Indels    0;  Gaps    0;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | | | | | | |||||||||||||||| | | | | | | |||||||||||| | ||||| |
Db         75 TATGTGTGTGTCTATATATATATATATATATATATATATATATATATATATATATATACA 134

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||| ||| | |||||||| | | ||||||| | ||| |||| || |||
Db        135 TATATATATATATATACATATATATATATATATACATATATATATATATATATATACA 194

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              || ||||| | |||| |||||| ||||| ||| ||| ||||| ||||||||||||||||||
Db        195 TACATATATATATATATATACATATATACACATATACATACATACATATATACACATATA 254

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              | | | | | ||| ||||| ||||| ||||||||| ||||||| ||||| |
Db        255 CATATATACACATATACATACATACATATATATACATATATACATATATACATGGGAGCG 314

Qy        467 TACATATATACACATATATATAATATACACACATATATATACACATATATACACACATAT 526
              || |                  |    | |||||  | | |    ||||          |
Db        315 TATACTATGCACACAGGCCCTCGCGTGTACACACCTTTTTGGTGATATTTTTTAGTTGGT 374

Qy        527 ATACACATATATACACATATATATACACATATATACACATATATACATATATACACACATAT 586
              |||||  | | |||||  |             |  |||        | | ||||||||  ||||
Db        375 GTACACGTGTGTACACGTGAGGGCCTGTGTGCATAGGATATGTTGCCTATATACATATAT 434

Qy        587 ATATACATATATACACATATATACATATATACACATATATACATATATACACACATATAT 646
              | ||| ||||||| |||||||| ||||||| | ||||||| ||||||| | | ||||
Db        435 ACATATATATATATACATATATATATATATATATATATATATATATATATATATATGA 494

Qy        647 ACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATAT 706
                || ||    |   | | ||| ||||| ||||||||| | ||||||||||  |||||||
Db        495 GAGAATCCACTCATGTATATATATATATATATATATATATATATATATATATATATATAT 554

Qy        707 ATGTATATATATAT 720
              || ||||||||||
Db        555 ATATATATATATAT 568


RESULT 95
US-10-437-963-100710
; Sequence 100710, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 100710
;   LENGTH: 662
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_983C.1
US-10-437-963-100710

  Query Match             5.7%;  Score 236.4;  DB 18;  Length 662;
  Best Local Similarity   69.7%;  Pred. No. 3.2e-38;
  Matches  347;  Conservative    0;  Mismatches  149;  Indels    2;  Gaps    2;

Qy        229 TATATATATATATATATATATATACACACACACACACATATATATATACACATATATATA 288
              | |||||||||||||||||||||||  | | |||   | || |||||| |   |||| ||||
Db        116 TTTATATATATATATATATATATGTATATACATGTATATGTATATACATGTATATGTATA 175

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              ||||  ||||  |||||||| |||||  | |||||| || |  ||  ||| ||||  ||
Db        176 TACATGTATATGTATATACATGTATATGTATATATATGTATATGTATGTATATACATGTA 235

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              ||| || | ||||||||||| |||||  | ||||||  | | |||||||   |||| ||
Db        236 TATGTATATATATGTATACATGTATATGTATATATATGTATATATATATATGTATATGTA 295

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              | | |||| || | |||||||  | |||||||  | || || | |||||||  | |||||||
Db        296 TATATATATGTATATATATATGTATATATATATATATGTATATATATATATATATATA 355

Qy        469 CATATATACACATATATATAATATACACACATATATATACACATATATAC-ACACATATA 527
              || |  |    || | |  |    | | || ||      |   ||| | | |  | | |
Db        356 TATCTCAATCTTTAGAGAGCAATATCTCTCACATCATATCTTGTATTTCCAATCCGTTTG 415

Qy        528 TACACATATATACACATATATATACACATATATACACATATATACATATATACACATATA 587
              ||    |  | || | | | | | | | ||||| | | ||||||||||||| | || ||
Db        416 CACCATGAAGTTCATAAAAAAACATATACATATATATACATATACATATATATATGTA 475

Qy        588 TATACATATATACACATATATACATATATAC-ACATATATACATATATACACACATATAT 646
              | ||| ||||||||  |||| |  ||| |||    |||||||||| ||||| ||||| ||
Db        476 TGTACGTATATACATGTATACATGTATGTACGTTATATATACATACATACATACATACAT 535

Qy        647 ACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATAT 706
              ||| | | ||||| ||| | ||| ||||| |||||||| | |||||||||  |||||||
Db        536 ACATACATATATAGACATATATATATATATATATATATATATATATATATATATATAT 595

Qy        707 ATGTATATATATATGCTC 724
              |||||||||||||  | |
Db        596 ATGTATATATATACACAC 613


RESULT 96
US-10-017-161-2181/c
; Sequence 2181, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
;  FILE REFERENCE: 084335/0152
;  CURRENT APPLICATION NUMBER: US/10/017,161
;  CURRENT FILING DATE:  2002-12-18
;  PRIOR APPLICATION NUMBER: JP 2001/246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2430
```

```
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2181
;   LENGTH: 1192
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: source
;   LOCATION: (1)..(1192)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(992)
US-10-017-161-2181

  Query Match             5.7%;  Score 235.8;  DB 15;  Length 1192;
  Best Local Similarity   69.2%;  Pred. No. 5.4e-38;
  Matches  355;  Conservative    0;  Mismatches  147;  Indels   11;  Gaps    2;

Qy        214 AAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACACATATATAT 273
              |||||| |      | |||| |||||||| | ||||| || | ||||||| |||| |||||
Db       1000 AAAAAATATTATATATATACATATATATGTTTATATGTATATACACACATACATGTATAT 941

Qy        274 ATACACATATATATACACATATATACATATATACACATATATACACATATATATACACAT 333
              ||  | ||||| || ||| |||||| | | ||| ||| || ||      ||||||||||
Db        940 ATGTATATATACATGCACGTATATATACACATATACACATGTATGTGCGTATATACACAT 881

Qy        334 ACATATGTACACATATATATACACATATGTATACACATATATACACATATATACACACAT 393
              |||| | || | | | |||||||||| | | | |  ||||||||    ||||| ||||||
Db        880 ACATGTATATATACACATATACACATGTATGTGCGTATATATAC----ATATATACACAT 825

Qy        394 ATATACACATATATACACACATATATACACATATATACACATATATACACATATACACAT 453
              |||   | | |||| ||| ||||| |||  | | | |   |||||||| ||||||||||
Db        824 GTATGTGCGTGTATATACATATATACACATGTGTGTGCGTGTATATACATATATACACAT 765

Qy        454 ATATACACATATATACATATATACACATATATAT-------AATATACACACATATATAT 506
              | |   | | |||| | ||||| ||| | ||||         | || | |||||||||
Db        764 GTGTGTGCGTGTATATACATATATACACAGATATGTGTGTGCGTGTATATACATATATAC 705

Qy        507 ACACATATATACACACATATATACACATATATACACATATATATACACATATATACACAT 566
              ||| |||| |       |||||||| ||| | ||| |||| | |     |||| |||| ||
Db        704 ACAGATATGTGTGTGTATATATACATATACACACAGATATGTGTGTGTATATGTACATAT 645

Qy        567 ATATACATATATACACATATATATACATATATACACATATATACATATATACACATATAT 626
              | | |||| | |  | ||||| ||| ||| ||| ||| | ||   | |||| | | ||||
Db        644 ACACACATGTGTGTATATATACATATATACATATACACACATGTGTGTATATATACATAT 585

Qy        627 ACATATATACACACATATATACACATACATATACACACACATAGATATACATATATATAC 686
              | | |||||||||| | | || | |||||||| ||| | | | |   |||||||||
Db        584 ATACATATACACACGTGTGTATATATACATATATACATATACACACGTGTGTATATATAC 525

Qy        687 ACATATATATACGTATATATATGTATATATATA 719
              | ||||| | || || || || || ||||||||
Db        524 ATATATACACACATACATGTACGTGTATATATA 492


RESULT 97
US-10-292-798-1827/c
; Sequence 1827, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
;  FILE REFERENCE: 084335/166
;  CURRENT APPLICATION NUMBER: US/10/292,798
;  CURRENT FILING DATE:  2002-11-13
;  PRIOR APPLICATION NUMBER: 10/017,161
;  PRIOR FILING DATE: 2001-12-18
;  PRIOR APPLICATION NUMBER: JP 2001-246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2070
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1827
;   LENGTH: 1192
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   LOCATION: source
;   FEATURE:
;   LOCATION: (1)..(1192)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(992)
US-10-292-798-1827

  Query Match             5.7%;  Score 235.8;  DB 17;  Length 1192;
  Best Local Similarity   69.2%;  Pred. No. 5.4e-38;
  Matches  355;  Conservative    0;  Mismatches  147;  Indels   11;  Gaps    2;

Qy        214 AAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACACATATATAT 273
              |||||| |      | |||| |||||||| | ||||| || | ||||||| |||| |||||
Db       1000 AAAAAATATTATATATATACATATATATGTTTATATGTATATACACACATACATGTATAT 941

Qy        274 ATACACATATATATACACATATATACATATATACACATATATACACATATATATACACAT 333
              ||  | ||||| || ||| |||||| | | ||| ||| || ||      ||||||||||
Db        940 ATGTATATATACATGCACGTATATATACACATATACACATGTATGTGCGTATATACACAT 881

Qy        334 ACATATGTACACATATATATACACATATGTATACACATATATACACATATATACACACAT 393
              |||| | || | | | |||||||||| | | | |  ||||||||    ||||| ||||||
Db        880 ACATGTATATATACACATATACACATGTATGTGCGTATATATAC----ATATATACACAT 825

Qy        394 ATATACACATATATACACACATATATACACATATATACACATATATACACATATACACAT 453
              |||   | | |||| ||| ||||| |||  | | | |   |||||||| ||||||||||
Db        824 GTATGTGCGTGTATATACATATATACACATGTGTGTGCGTGTATATACATATATACACAT 765

Qy        454 ATATACACATATATACATATATACACATATATAT-------AATATACACACATATATAT 506
              | |   | | |||| | ||||| ||| | ||||         | || | |||||||||
Db        764 GTGTGTGCGTGTATATACATATATACACAGATATGTGTGTGCGTGTATATACATATATAC 705

Qy        507 ACACATATATACACACATATATACACATATATACACATATATATACACATATATACACAT 566
              ||| |||| |       |||||||| ||| | ||| |||| | |     |||| |||| ||
Db        704 ACAGATATGTGTGTGTATATATACATATACACACAGATATGTGTGTGTATATGTACATAT 645

Qy        567 ATATACATATATACACATATATATACATATATACACATATATACATATATACACATATAT 626
              | | |||| | |  | ||||| ||| ||| ||| ||| | ||   | |||| | | ||||
Db        644 ACACACATGTGTGTATATATACATATATACATATACACACATGTGTGTATATATACATAT 585

Qy        627 ACATATATACACACATATATACACATACATATACACACACATAGATATACATATATATAC 686
              | | |||||||||| | | || | |||||||| ||| | | | |   |||||||||
Db        584 ATACATATACACACGTGTGTATATATACATATATACATATACACACGTGTGTATATATAC 525

Qy        687 ACATATATATACGTATATATATGTATATATATA 719
              | ||||| | || || || || || ||||||||
Db        524 ATATATACACACATACATGTACGTGTATATATA 492


RESULT 98
US-10-027-632-258173
; Sequence 258173, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
```

```
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 258173
;   LENGTH: 3673
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-258173

  Query Match             5.7%;  Score 235.2;  DB 13;  Length 3673;
  Best Local Similarity   67.2%;  Pred. No. 1.2e-37;
  Matches  347;  Conservative    0;  Mismatches  168;  Indels    1;  Gaps    1;

Qy        204 ATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATATACACACACACAC 263
              || | |  | |    | ||      | || | | || |||| |||||||||| | |||
Db       1857 ATATAAAGACATCTATAAATATATATAAAGACATCTATAAATATATATAAAGACATCTGT 1916

Qy        264 ACATATATATATACACATATATATACACATATATACATATATACACATATATACACATAT 323
              | |||||||||| | |||| | || | | ||||| | | || || | ||||||| | | |
Db       1917 AAATATATATAAAGACATCTGTAAATATATATAAAGACATCTATAAATATATATAAAGAC 1976

Qy        324 ATATACACATACATATGTACACATATATATACACATATGTATACACATATATACACATAT 383
              || || | ||| |||| | |   || |||| | |||| || | |||| |||| | ||||
Db       1977 ATCTATAAATATATATAAAGATGCATCTATAAATATATATAAAGACATCTATAAATATAT 2036

Qy        384 ATACACACATATATACACATATATACACACATATATACACATATATACACATATATACAC 443
              ||| | |||| |||| | ||| ||| | |||| |||| | ||||||| | | || || |
Db       2037 ATAAAGACATCTATAAATATACATAAAGACATCTATAAATATATATAAAAACATCTATAA 2096

Qy        444 ATATACACATATATACACATATATACATATATACACATATATATAATATACACACATATA 503
              ||||| | | | || ||  |||| | ||||| | |||| |||| |||||| | | ||  |
Db       2097 ATATATATAAAGATGCATCTATAAATATATAAAGACATCTATA-AATATATAAAGATGCA 2155

Qy        504 TATACACATATATACACACATATATACACATATATACACATATATATACACATATATACA 563
              | || | ||||||| |||  |||| | | ||| | |   ||| |||| | | ||| | |||
Db       2156 TCTATAAATATATAGACATCTATAAATATATAAAGATGCATCTATAAATATATAAAGACA 2215

Qy        564 CATATATACATATATACACATATATATACATATATACACATATATACATATATACACATA 623
                |||| | ||||||| | | || |||| ||||||| | | | | |  |||| | | |||
Db       2216 TCTATAAATATATATAAAGACATCTATAAATATATATAAAGACACATCTATAAATATATA 2275

Qy        624 TATACATATATACACACATATATACACATACATATACACACACATAGATATACATATATA 683
              || | | || || | | ||||||| | |  ||| || | | | ||| | |  ||| ||||
Db       2276 TAAAGACATCTATAAATATATATAAAGACGCATCTATAAATATATATAAAGGCATCTATA 2335

Qy        684 TACACATATATATACGTATATATATGTATATATATA 719
                | | |||||  | | | ||  ||||  |||||||| |
Db       2336 AATATATATAAAGATGCATCTATAAATATATATAAA 2371


RESULT 99
US-10-027-632-258173
; Sequence 258173, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 258173
;   LENGTH: 3673
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-258173

  Query Match             5.7%;  Score 235.2;  DB 17;  Length 3673;
  Best Local Similarity   67.2%;  Pred. No. 1.2e-37;
  Matches  347;  Conservative    0;  Mismatches  168;  Indels    1;  Gaps    1;

Qy        204 ATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATATACACACACACAC 263
              || | |  | |    | ||      | || | | || |||| |||||||||| | |||
Db       1857 ATATAAAGACATCTATAAATATATATAAAGACATCTATAAATATATATAAAGACATCTGT 1916

Qy        264 ACATATATATATACACATATATATACACATATATACATATATACACATATATACACATAT 323
              | |||||||||| | |||| | || | | ||||| | | || || | ||||||| | | |
Db       1917 AAATATATATAAAGACATCTGTAAATATATATAAAGACATCTATAAATATATATAAAGAC 1976

Qy        324 ATATACACATACATATGTACACATATATATACACATATGTATACACATATATACACATAT 383
              || || | ||| |||| | |   || |||| | |||| || | |||| |||| | ||||
Db       1977 ATCTATAAATATATATAAAGATGCATCTATAAATATATATAAAGACATCTATAAATATAT 2036

Qy        384 ATACACACATATATACACATATATACACACATATATACACATATATACACATATATACAC 443
              ||| | |||| |||| | ||| ||| | |||| |||| | ||||||| | | || || |
Db       2037 ATAAAGACATCTATAAATATACATAAAGACATCTATAAATATATATAAAAACATCTATAA 2096

Qy        444 ATATACACATATATACACATATATACATATATACACATATATATAATATACACACATATA 503
              ||||| | | | || ||  |||| | ||||| | |||| |||| |||||| | | ||  |
Db       2097 ATATATATAAAGATGCATCTATAAATATATAAAGACATCTATA-AATATATAAAGATGCA 2155

Qy        504 TATACACATATATACACACATATATACACATATATACACATATATATACACATATATACA 563
              | || | ||||||| |||  |||| | | ||| | |   ||| |||| | | ||| | |||
Db       2156 TCTATAAATATATAGACATCTATAAATATATAAAGATGCATCTATAAATATATAAAGACA 2215

Qy        564 CATATATACATATATACACATATATATACATATATACACATATATACATATATACACATA 623
                |||| | ||||||| | | || |||| ||||||| | | | | |  |||| | | |||
Db       2216 TCTATAAATATATATAAAGACATCTATAAATATATATAAAGACACATCTATAAATATATA 2275

Qy        624 TATACATATATACACACATATATACACATACATATACACACACATAGATATACATATATA 683
              || | | || || | | ||||||| | |  ||| || | | | ||| | |  ||| ||||
Db       2276 TAAAGACATCTATAAATATATATAAAGACGCATCTATAAATATATATAAAGGCATCTATA 2335

Qy        684 TACACATATATATACGTATATATATGTATATATATA 719
                | | |||||  | | | ||  ||||  |||||||| |
Db       2336 AATATATATAAAGATGCATCTATAAATATATATAAA 2371


RESULT 100
US-10-311-455-1971/c
; Sequence 1971, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
```

```
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1971
;   LENGTH: 21537
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1971

  Query Match             5.7%;  Score 235.2;  DB 15;  Length 21537;
  Best Local Similarity   69.6%;  Pred. No. 2.5e-37;
  Matches  348;  Conservative    0;  Mismatches  148;  Indels    4;  Gaps    2;

Qy        197 AAATGTCATCTCAGAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACA 256
              ||||   | | || |||||  | |   | | |   | |       |||| ||  ||| | |
Db      16071 AAATTAAAACACACAAAAACTATATAACTTATCCAAAAACACACTATAAATCCATATATA 16012

Qy        257 CACACACACATATATATATACACATATATATACACATATATACATATATACACATATATA 316
              | | | | ||||||||||| | ||||| | | | ||| || |||||||| | |||||
Db      16011 TATATATATATATATATATATATATATACACATATATAAATCCATATATATATATATACG 15952

Qy        317 CACATATATATACACATACATATGTACACATATATATACACATATGTATACACATATATA 376
              || |||||||   | ||| ||||  ||||| |||||||||||||| || ||||||||||
Db      15951 CATATATATACGTATATATATATACACACACATATATACACATATATACACACATATATA 15892

Qy        377 CACATATATACACACATATATACACATATATACACACATATATACACATATATACACATA 436
              |||||||| ||| | ||||||| | |||||||||||||| | ||| | ||||||||||| |
Db      15891 CACATATACACATATATATATATATATATACACACACATATATATATATACACACA 15832

Qy        437 TATACACATATACACATATATACACATATATACATATATACACATATATATA---ATATA 493
              | | | ||||| | | | | |||||||||| |||||||||| | | |||   |||||
Db      15831 CACATATATATATATACACACACACATATATATATATATACACACACACATATATATATA 15772

Qy        494 CACACATATATATACACATATATACACACATATATACACATATATACACATATATATACA 553
              | | ||| | | |||||||||| |     ||  | | | ||||| | ||||| | | |
Db      15771 TATATATACACACACACATATATATAATATAATCGATATAAATATATATATATAAAAAAA 15712

Qy        554 CATATATACACATATATACATATATACACATATATATACATATATACACATATATACATA 613
              |||||   |  |||| ||         |||  || || | || ||   |   ||   |
Db      15711 AATATACCTATTTATAAACTCTCTATTCCATTAATTTATACATCTAATAAACAATTATTT 15652

Qy        614 TATACACATATATACATATATACACACATATATACACATACATATACACACACATAGATA 673
              |||| | ||||||| |||||||||||||||||||| | ||| ||||| | | ||||| |||
Db      15651 TATATATATATATATATATATACACACATATATATA-ATATATATATATATACATACATA 15593

Qy        674 TACATATATATACACATATA 693
              | | ||| ||| |||||||
Db      15592 CATACATACATATACATATA 15573


Search completed: April 22, 2005, 07:10:50
Job time : 1829.26 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 20, 2005, 22:52:20 ; Search time 8250.88 Seconds
(without alignments)
19011.657 Million cell updates/sec

Title: US-09-939-209A-3_COPY_1_4121
Perfect score: 4121
Sequence: 1 agttcaagaccagcctgagc.....tgtgagggcaaattgaagtc 4121

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_htc:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	347.8	8.4	976	9	CL830166	CL830166 OR_CBa005
	2	345.4	8.4	863	9	CL840224	CL840224 OR_CBa006
	3	342	8.3	835	9	CL800294	CL800294 OR_CBa001
	4	341.4	8.3	828	8	CC107667	CC107667 NDL.37F6.
c	5	340.4	8.3	976	9	CL830166	CL830166 OR_CBa005
c	6	338.8	8.2	835	9	CL800294	CL800294 OR_CBa001
c	7	338.4	8.2	863	9	CL840224	CL840224 OR_CBa006
c	8	338.2	8.2	828	8	CC107667	CC107667 NDL.37F6.
c	9	337.4	8.2	620	7	CF626074	CF626074 zmrws05_0
c	10	337.4	8.2	632	7	CF634146	CF634146 zmrww00_0
c	11	337.4	8.2	883	9	CW016544	CW016544 ZMMBLc000
	12	337.2	8.2	883	9	CW016544	CW016544 ZMMBLc000
	13	336.6	8.2	725	8	AQ917102	AQ917102 T233131b
	14	336.4	8.2	827	8	CC067623	CC067623 CSU-K33r.
	15	336.2	8.2	590	9	CR294134	CR294134 Medicago
c	16	335.8	8.1	827	8	CC067623	CC067623 CSU-K33r.
	17	335.2	8.1	620	7	CF626074	CF626074 zmrws05_0
	18	335.2	8.1	632	7	CF634146	CF634146 zmrww00_0
	19	335.2	8.1	717	4	BI505681	BI505681 BB170032B
c	20	334.8	8.1	590	9	CR294134	CR294134 Medicago
c	21	334.2	8.1	717	4	BI505681	BI505681 BB170032B
c	22	333.8	8.1	725	8	AQ917102	AQ917102 T233131b
c	23	333	8.1	744	9	CL623140	CL623140 OR_BBa001
c	24	332.6	8.1	512	8	BZ796807	BZ796807 PUGDF25TD
c	25	331.6	8.0	707	9	CL755178	CL755178 OR_BBa012
	26	331.2	8.0	784	9	CL934691	CL934691 OA_ABa004
	27	331	8.0	512	8	BZ796807	BZ796807 PUGDF25TD
c	28	330.4	8.0	668	9	BX145650	BX145650 Danio rer
	29	330.4	8.0	707	9	CL755178	CL755178 OR_BBa012
c	30	330.2	8.0	496	6	CA916667	CA916667 PCSC21302
c	31	329.8	8.0	751	7	CF924719	CF924719 DD005F11.
	32	329	8.0	668	9	BX145650	BX145650 Danio rer
c	33	328.4	8.0	784	9	CL934691	CL934691 OA_ABa004
c	34	328.2	8.0	587	9	CL776144	CL776144 OR_BBa009
	35	327.2	7.9	751	7	CF924719	CF924719 DD005F11.
	36	326.2	7.9	496	6	CA916667	CA916667 PCSC21302
	37	324.2	7.9	744	9	CL623140	CL623140 OR_BBa001
	38	323.8	7.9	587	9	CL776144	CL776144 OR_BBa009
c	39	322.6	7.8	491	9	CL830207	CL830207 OR_CBa005
	40	318.2	7.7	845	8	CC099198	CC099198 CSU-K34.1
c	41	317	7.7	845	8	CC099198	CC099198 CSU-K34.1
	42	315	7.6	816	9	CL855754	CL855754 OR_CBa008
	43	314.6	7.6	570	4	BI506858	BI506858 BB170024B
c	44	313.2	7.6	570	4	BI506858	BI506858 BB170024B
c	45	311.4	7.6	816	9	CL855754	CL855754 OR_CBa008

ALIGNMENTS

RESULT 1
CL830166
LOCUS CL830166 976 bp DNA linear GSS 09-AUG-2004
DEFINITION OR_CBa0052B14.f OR_CBa Oryza rufipogon genomic clone OR_CBa0052B14 5', genomic survey sequence.
ACCESSION CL830166
VERSION CL830166.1 GI:51075776
KEYWORDS GSS.
SOURCE Oryza rufipogon
ORGANISM Oryza rufipogon
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 976)
AUTHORS Kim,H., Yu,Y., Wissotski,M., Yost,D., Stum,D., Rao,K., Luo,M., Jetty,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
TITLE OMAP project
JOURNAL Unpublished (2004)
COMMENT Contact: Rod A. Wing
Arizona Genomics Institute
University of Arizona
Forbes Building Room 303, Tucson, AZ 85721-0036, USA
Tel: 520 626 9595
Fax: 520 621 1259
Email: http://genome.arizona.edu
PCR PRimers
FORWARD: TAA TAC GAC TCA CTA TAG GG
BACKWARD: CAC TCA TTA GGC ACC CCA
Plate: 0052 row: B column: 14
Seq primer: TAA TAC GAC TCA CTA TAG GG
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .976
/organism="Oryza rufipogon"
/mol_type="genomic DNA"
/db_xref="taxon:4529"
/clone="OR_CBa0052B14"
/tissue_type="young leaves"
/dev_stage="2 week old seedlings"
/lab_host="DH10B T1 phage resistant"
/clone_lib="OR_CBa"
/note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII; drk treated 36 hrs before harvest"
ORIGIN

Query Match 8.4%; Score 347.8; DB 9; Length 976;

```
  Best Local Similarity   69.5%;  Pred. No. 1.5e-48;
  Matches  486;  Conservative    0;  Mismatches  212;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db        188 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 247

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db        248 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 307

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        308 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 367

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db        368 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 427

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db        428 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 487

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db        488 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 547

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db        548 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 607

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db        608 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 667

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||||         | |||          |  |   ||.  |    |||
Db        668 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 727

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
              |   | | | |   | |    | |||    | |   | || |      |      | |
Db        728 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 787

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885
              ||       | ||| ||| || | ||   | |  ||| | |||  ||  ||   |  ||
Db        788 TATATATATATATATATATATATATATATATATATATATATAATATATATATATATATAT 847

Qy        886 CACTATATATATATAGTGAAAGCAGTGTAAACCTAAA 924
                ||||||||||||||   | |    | || |  || |
Db        848 ATATATATATATATATATATATATATATATATATATA 886


RESULT 2
CL840224
LOCUS       CL840224               863 bp    DNA     linear   GSS 09-AUG-2004
DEFINITION  OR_CBa0067C10.f OR_CBa Oryza rufipogon genomic clone OR_CBa0067C10
            5', genomic survey sequence.
ACCESSION   CL840224
VERSION     CL840224.1  GI:51085834
KEYWORDS    GSS.
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 863)
  AUTHORS   Kim,H., Yu,Y., Wissotski,M., Yost,D., Stum,D., Rao,K., Luo,M.,
            Jetty,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and
            Wing,R.
  TITLE     OMAP project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0067  row: C  column: 10
            Seq primer: TAA TAC GAC TCA CTA TAG GG
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .863
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
                     /clone="OR_CBa0067C10"
                     /tissue_type="young leaves"
                     /dev_stage="2 week old seedlings"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OR_CBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII;
                     drk treated 36 hrs before harvest"
ORIGIN

  Query Match             8.4%;  Score 345.4;  DB 9;  Length 863;
  Best Local Similarity   70.3%;  Pred. No. 3.8e-48;
  Matches  477;  Conservative    0;  Mismatches  201;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db        152 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 211

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db        212 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 271

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        272 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 331

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db        332 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 391

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db        392 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 451

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db        452 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 511

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db        512 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 571

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db        572 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 631

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||||         | |||          |  |   ||   |    |||
Db        632 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 691

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
              |   | | | |   | |    | |||    | |   | || |      |  | | |
Db        692 TATATATATATATATATATATATATATATATATATATATATATATATATATATATTTATATA 751

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885
              ||       | ||| ||| || | ||   | |  ||| | ||      |  | ||  ||
Db        752 TATATATATATATATATATATATATATATATATATATATATATATATATATATAATATAT 811
```

```
Qy        886 CACTATATATATATATAGT 904
              ||||||||||||||| |
Db        812 ATATATATATATATATAAT 830


RESULT 3
CL800294
LOCUS       CL800294                 835 bp    DNA     linear   GSS 06-AUG-2004
DEFINITION  OR_CBa0011N09.r OR_CBa Oryza rufipogon genomic clone OR_CBa0011N09
            3', genomic survey sequence.
ACCESSION   CL800294
VERSION     CL800294.1  GI:51025055
KEYWORDS    GSS.
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 835)
  AUTHORS   Kim,H., Yu,Y., Wissotski,M., Yost,D., Stum,D., Rao,K., Luo,M.,
            Jetty,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and
            Wing,R.
  TITLE     OMAP project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0011  row: N  column: 09
            Seq primer: CAC TCA TTA GGC ACC CCA
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .835
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
                     /clone="OR_CBa0011N09"
                     /tissue_type="young leaves"
                     /dev_stage="2 week old seedlings"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OR_CBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII;
                     drk treated 36 hrs before harvest"
ORIGIN

  Query Match             8.3%;  Score 342;  DB 9;  Length 835;
  Best Local Similarity   67.4%;  Pred. No. 1.4e-47;
  Matches  511;  Conservative    0;  Mismatches  245;  Indels    2;  Gaps    2;

Qy        212 AAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACATATAT 271
              ||  |   |  || | |||||||||||||||||||||||| | | | | | | ||||||
Db         67 AACAGTATTAAACAGTATATATATATATATATATATATATATATATATATATATATAT 126

Qy        272 ATATACACATATATATACACATATATACATATATACACATATATACACATATATATACAC 331
              ||||| | ||||||||| | ||||||| ||||||| | ||||||| | |||||||||  |
Db        127 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 186

Qy        332 ATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACACAC 391
              ||| |||| || | ||||||||| | |||| |||| | ||||||| | ||||||| | ||
Db        187 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 246

Qy        392 ATATATACACATATATACACACATATATACACATATATACACATATATACACATATACAC 451
              ||||||| | ||||||| | | |||||||| | ||||||| | ||||||| | ||||| |
Db        247 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 306

Qy        452 ATATATACACATATATACATATATACACATATATATA-ATATACACACATATATATACAC 510
              ||||||| | ||||||| ||||||| | ||||||||| ||||| | | |||||||||  |

Db        307 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 366

Qy        511 ATATATACACACATATATACACATATATACACATATATATACACATATATACACATATAT 570
              ||||||| | | ||||||| | ||||||| | |||||||||  | ||||||| | ||||||
Db        367 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 426

Qy        571 ACATATATACACATATATATACATATATACACATATATACATATATACACATATATACAT 630
              | ||||||| | ||||||||| ||||||| | ||||||| ||||||| | ||||||| ||
Db        427 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 486

Qy        631 ATATACACACATATATACACATACATATACACACACATAGATATACATATATATACACAT 690
              ||||| | | ||||||| | ||| ||||| | | | ||| ||||| |||||||||  | ||
Db        487 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 546

Qy        691 ATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGA 750
              |||||||  |||||||||  ||||||||||          |  |||          |  |
Db        547 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 606

Qy        751 TTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTT 810
              ||    |   ||| |  | | | |  | | |   | |||   | |  | || |
Db        607 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 666

Qy        811 CAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA-AG 869
                  |  | |||       | ||| ||| || | ||   | |  ||| | || |
Db        667 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 726

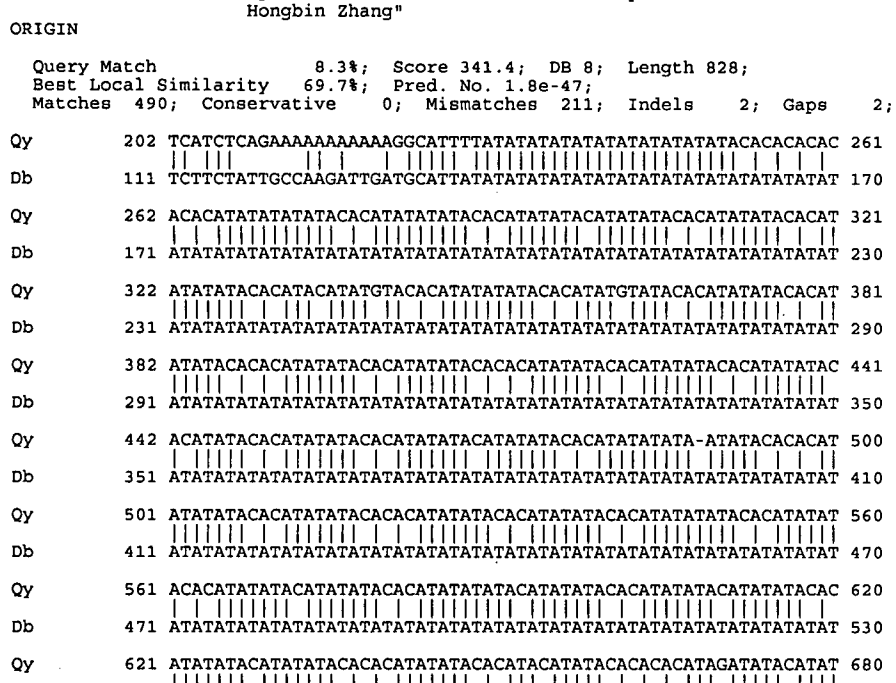
Qy        870 GTAGTTAAGAAGGTAACACTATATATATATATAGTGAAAGCAGTGTAAACCTAAAGGATG 929
              ||  ||   |  ||    |||||||||||||   | |    | || |  || |
Db        727 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 786

Qy        930 GGCCAAGGATTTAAATGTTATAGAAGAATGGCTAAGAT 967
                |  || || || | ||| |   ||    || ||
Db        787 ATATATATATATATATATAATATATATATAAAAAATAT 824


RESULT 4
CC107667
LOCUS       CC107667                 828 bp    DNA     linear   GSS 16-APR-2003
DEFINITION  NDL.37F6.T7 Notre Dame Liverpool Aedes aegypti genomic clone
            NDL.37F6, genomic survey sequence.
ACCESSION   CC107667
VERSION     CC107667.1  GI:29976722
KEYWORDS    GSS.
SOURCE      Aedes aegypti (yellow fever mosquito)
  ORGANISM  Aedes aegypti
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Aedes;
            Stegomyia.
REFERENCE   1  (bases 1 to 828)
  AUTHORS   Loftus,B., Shetty,J., Knudson,D. and Severson,D.
  TITLE     BAC end sequencing of Aedes aegypti
  JOURNAL   Unpublished (2003)
COMMENT     Other_GSSs: NDL.37F6.SP6
            Contact: Brendan Loftus
            Department of Eukaryotic Genomics
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-3543
            Fax: 301-838-0208
            Email: enta@tigr.org
            Library was provided by David Severson
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .828
                     /organism="Aedes aegypti"
                     /mol_type="genomic DNA"
                     /strain="Liverpool"
                     /db_xref="taxon:7159"
                     /clone="NDL.37F6"
                     /clone_lib="Notre Dame Liverpool"
                     /note="Vector: pECBAC1; Site_1: Hind III; The library was
                     prepared from whole body tissue of newly hatched L1 larvae
```

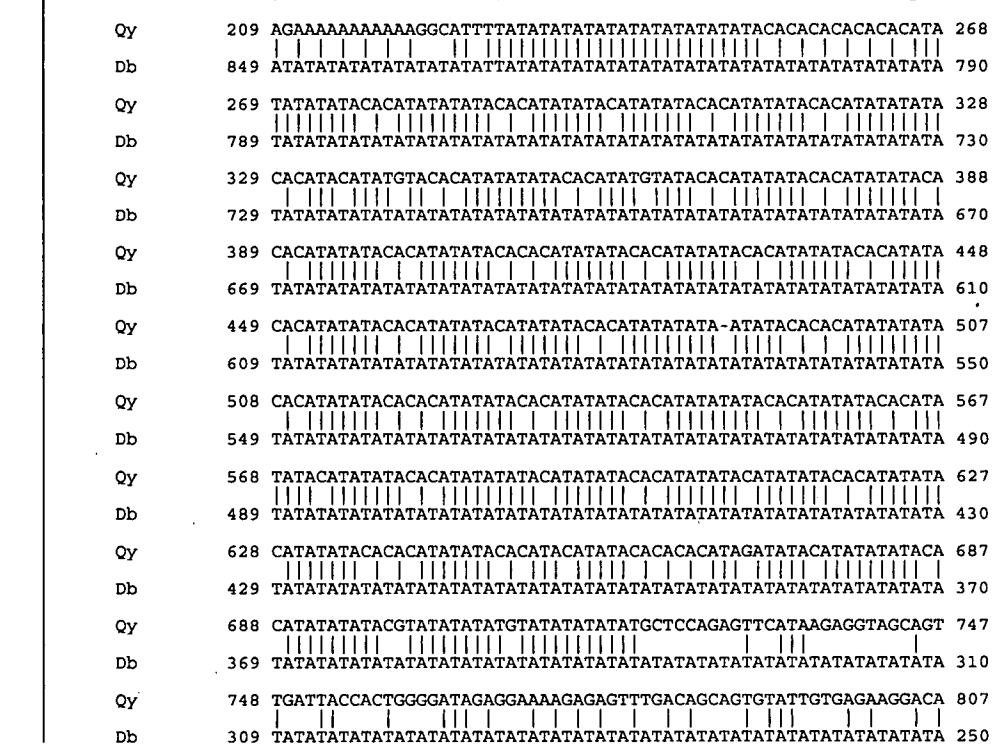
by David Severson at the University of Notre Dame and
Hongbin Zhang"
ORIGIN

Query Match 8.3%; Score 341.4; DB 8; Length 828;
Best Local Similarity 69.7%; Pred. No. 1.8e-47;
Matches 490; Conservative 0; Mismatches 211; Indels 2; Gaps 2;



RESULT 5
CL830166/c
LOCUS CL830166 976 bp DNA linear GSS 09-AUG-2004
DEFINITION OR_CBa0052B14.f OR_CBa Oryza rufipogon genomic clone OR_CBa0052B14 5', genomic survey sequence.
ACCESSION CL830166
VERSION CL830166.1 GI:51075776
KEYWORDS GSS.
SOURCE Oryza rufipogon
ORGANISM Oryza rufipogon
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 976)
AUTHORS Kim,H., Yu,Y., Wissotski,M., Yost,D., Stum,D., Rao,K., Luo,M., Jetty,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
TITLE OMAP project
JOURNAL Unpublished (2004)
COMMENT Contact: Rod A. Wing
Arizona Genomics Institute
University of Arizona
Forbes Building Room 303, Tucson, AZ 85721-0036, USA
Tel: 520 626 9595
Fax: 520 621 1259
Email: http://genome.arizona.edu
PCR PRimers
FORWARD: TAA TAC GAC TCA CTA TAG GG
BACKWARD: CAC TCA TTA GGC ACC CCA
Plate: 0052 row: B column: 14
Seq primer: TAA TAC GAC TCA CTA TAG GG
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .976
/organism="Oryza rufipogon"
/mol_type="genomic DNA"
/db_xref="taxon:4529"
/clone="OR_CBa0052B14"
/tissue_type="young leaves"
/dev_stage="2 week old seedlings"
/lab_host="DH10B T1 phage resistant"
/clone_lib="OR_CBa"
/note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII; drk treated 36 hrs before harvest"
ORIGIN

Query Match 8.3%; Score 340.4; DB 9; Length 976;
Best Local Similarity 68.9%; Pred. No. 2.6e-47;
Matches 495; Conservative 0; Mismatches 221; Indels 2; Gaps 2;



```
Qy        808 TTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA 867
              | |     |      | |||       | ||| ||| || | ||   | |  ||| | ||
Db        249 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 190

Qy        868 -AGGTAGTTAAGAAGGTAACACTATATATATATATAGTGAAAGCAGTGTAAACCTAAA 924
               |  ||  ||   |  ||    ||||||||||||||   | |    | || |  || |
Db        189 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 132


RESULT 6
CL800294/c
LOCUS       CL800294                 835 bp    DNA     linear   GSS 06-AUG-2004
DEFINITION  OR_CBa0011N09.r OR_CBa Oryza rufipogon genomic clone OR_CBa0011N09
            3', genomic survey sequence.
ACCESSION   CL800294
VERSION     CL800294.1  GI:51025055
KEYWORDS    GSS.
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 835)
  AUTHORS   Kim,H., Yu,Y., Wissotski,M., Yost,D., Stum,D., Rao,K., Luo,M.,
            Jetty,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and
            Wing,R.
  TITLE     OMAP project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0011  row: N  column: 09
            Seq primer: CAC TCA TTA GGC ACC CCA
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .835
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
                     /clone="OR_CBa0011N09"
                     /tissue_type="young leaves"
                     /dev_stage="2 week old seedlings"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OR_CBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII;
                     drk treated 36 hrs before harvest"
ORIGIN

  Query Match             8.2%;  Score 338.8;  DB 9;  Length 835;
  Best Local Similarity   69.5%;  Pred. No. 5e-47;
  Matches  488;  Conservative    0;  Mismatches  212;  Indels    2;  Gaps    2;

Qy        225 ATTTTATATATATATATATATATATATACACACACACACACATATATATATACACATATA 284
              || |||||||||||||||||||||||| | | | | | | |||||||||||| | |||||
Db        809 ATATTATATATATATATATATATATATATATATATATATATATATATATATATATATATA 750

Qy        285 TATACACATATATACATATATACACATATATACACATATATATACACATACATATGTACA 344
              ||||  | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || |
Db        749 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 690

Qy        345 CATATATATACACATATGTATACACATATATACACATATATACACACATATATACACATA 404
               |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||
Db        689 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 630

Qy        405 TATACACACATATATACACATATATACACATATATACACATATACACATATATACACATA 464
              |||| | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||
Db        629 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 570

Qy        465 TATACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACA 523
              |||| ||||||| | |||||||| ||||| | | ||||||||| | ||||||| | | |
Db        569 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 510

Qy        524 TATATACACATATATACACATATATATACACATATATACACATATATACATATATACACA 583
              |||||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |
Db        509 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 450

Qy        584 TATATATACATATATACACATATATACATATATACACATATATACATATATACACACATA 643
              |||||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||
Db        449 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 390

Qy        644 TATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATA 703
              |||| | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||
Db        389 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 330

Qy        704 TATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGA 763
              ||||| ||||||||||          | |||         | |  ||   |  |
Db        329 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 270

Qy        764 TAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCA 823
              || |  | | | |  | | |  | |||    | |   | || |    |      |
Db        269 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 210

Qy        824 AATAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA-AGGTAGTTAAGAAGG 882
               |||      | ||| ||| || | ||   | |  ||| | || |  ||  ||   |
Db        209 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 150

Qy        883 TAACACTATATATATATAGTGAAAGCAGTGTAAACCTAAA 924
              ||    |||||||||||||   | |    | || |  || |
Db        149 TATATATATATATATATATATATATATATATATATATATA 108


RESULT 7
CL840224/c
LOCUS       CL840224                 863 bp    DNA     linear   GSS 09-AUG-2004
DEFINITION  OR_CBa0067C10.f OR_CBa Oryza rufipogon genomic clone OR_CBa0067C10
            5', genomic survey sequence.
ACCESSION   CL840224
VERSION     CL840224.1  GI:51085834
KEYWORDS    GSS.
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 863)
  AUTHORS   Kim,H., Yu,Y., Wissotski,M., Yost,D., Stum,D., Rao,K., Luo,M.,
            Jetty,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and
            Wing,R.
  TITLE     OMAP project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0067  row: C  column: 10
            Seq primer: TAA TAC GAC TCA CTA TAG GG
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .863
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
```

```
                     /clone="OR_CBa0067C10"
                     /tissue_type="young leaves"
                     /dev_stage="2 week old seedlings"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OR_CBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII;
                     drk treated 36 hrs before harvest"
ORIGIN

  Query Match             8.2%;  Score 338.4;  DB 9;  Length 863;
  Best Local Similarity   69.1%;  Pred. No. 5.8e-47;
  Matches  492;  Conservative    0;  Mismatches  216;  Indels    4;  Gaps    2;

Qy        204 ATCTCAGAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACAC 263
              ||   | | | | | |    || ||||||||||||||||||||||||| | | | | |
Db        830 ATTATATATATATATATATATATATTTATATATATATATATATATATATATATATATAT 771

Qy        264 ACATATATATATACACATATATATACACATATATACATATATACACATATATACACATAT 323
              | |||||||||| | ||||||||| | ||||||| ||||||| | ||||||| | ||||
Db        770 ATATATATATATATATATATATATATAAATATATATATATATATATATATATATATATAT 711

Qy        324 ATATACACATACATATGTACACATATATATACACATATGTATACACATATATACACATAT 383
              ||||| | ||| |||| || | ||||||||| | |||| |||| | ||||||| | ||||
Db        710 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 651

Qy        384 ATACACACATATATACACATATATACACACATATATACACATATATACACATATATACAC 443
              ||| | | ||||||| | ||||||| | | ||||||| | ||||||| | ||||||| |
Db        650 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 591

Qy        444 ATATACACATATATACACATATATACATATATACACATATATATA-ATATACACACATAT 502
              ||||| | ||||||| | ||||||| ||||||| | ||||||||| ||||| | | ||||
Db        590 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 531

Qy        503 ATATACACATATATACACACATATATACACATATATACACATATATATACACATATATAC 562
              ||||| | ||||||| | | ||||||| | ||||||| | ||||||||| | |||||||
Db        530 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 471

Qy        563 ACATATATACATATATACACATATATATACATATATACACATATATACATATATACACAT 622
              | ||||||| ||||||| | ||||||||| ||||||| | ||||||| ||||||| | ||
Db        470 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 411

Qy        623 ATATACATATATACACACATATATACACATACATATACACACACATAGATATACATATAT 682
              ||||| ||||||| | | ||||||| | ||| ||||| | | | ||| ||||| ||||||
Db        410 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 351

Qy        683 ATACACATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTA 742
              ||| | |||||||||  |||||||||| ||||||||||          | |||
Db        350 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 291

Qy        743 GCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAA 802
              | |   ||    |    ||| |   | | | | | | |    | |||    | |
Db        290 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 231

Qy        803 GGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTAT 862
              | || |    |   | |||     | ||| ||| || | ||   | |  |||
Db        230 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 171

Qy        863 CTGTAAGGTAGTTAAGAAGGTAACACTATATATATATATAGTGAAAGCAGTG 914
              | ||  ||  ||  |  ||    |||||||||| | ||   | | |||
Db        170 ATATA---TATATATATATATATATATATATATGTTGTTTCACCTGTG 122


RESULT 8
CC107667/c
LOCUS       CC107667                828 bp    DNA     linear   GSS 16-APR-2003
DEFINITION  NDL.37F6.T7 Notre Dame Liverpool Aedes aegypti genomic clone
            NDL.37F6, genomic survey sequence.
ACCESSION   CC107667
VERSION     CC107667.1  GI:29976722
KEYWORDS    GSS.
SOURCE      Aedes aegypti (yellow fever mosquito)
  ORGANISM  Aedes aegypti
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Aedes;
            Stegomyia.
REFERENCE   1  (bases 1 to 828)
  AUTHORS   Loftus,B., Shetty,J., Knudson,D. and Severson,D.
  TITLE     BAC end sequencing of Aedes aegypti
  JOURNAL   Unpublished (2003)
COMMENT     Other_GSSs: NDL.37F6.SP6
            Contact: Brendan Loftus
            Department of Eukaryotic Genomics
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-3543
            Fax: 301-838-0208
            Email: enta@tigr.org
            Library was provided by David Severson
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .828
                     /organism="Aedes aegypti"
                     /mol_type="genomic DNA"
                     /strain="Liverpool"
                     /db_xref="taxon:7159"
                     /clone="NDL.37F6"
                     /clone_lib="Notre Dame Liverpool"
                     /note="Vector: pECBAC1; Site_1: Hind III; The library was
                     prepared from whole body tissue of newly hatched L1 larvae
                     by David Severson at the University of Notre Dame and
                     Hongbin Zhang"
ORIGIN

  Query Match             8.2%;  Score 338.2;  DB 8;  Length 828;
  Best Local Similarity   70.2%;  Pred. No. 6.3e-47;
  Matches  482;  Conservative    0;  Mismatches  203;  Indels    2;  Gaps    2;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | |||||||||||||||||||||||| | | | | | |||||||||| | |||||||
Db        817 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 758

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db        757 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 698

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        697 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 638

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | |||||  | ||||||| | |||||
Db        637 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 578

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db        577 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 518

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db        517 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 458

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db        457 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 398

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| |||||||||| | |||||||||  ||||||
Db        397 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 338

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||          | |||          | |   ||    |   |||
Db        337 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 278

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
              |   | | | |  | | |   | |||   | |   | || |     |     | |
```

```
Db        277 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 218

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTA 884
              ||        | ||| ||| || | ||    | |  ||| | || |  ||  ||   | ||
Db        217 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 158

Qy        885 ACACTATATATATATATAGTGAAAGCA 911
                  ||||||||||||||   || |||
Db        157 TATATATATATATATATATATAATGCA 131
```

RESULT 9
CF626074/c

LOCUS	CF626074 620 bp mRNA linear EST 02-OCT-2003
DEFINITION	zmrws05_0B10-009-d11.s3 zmrws05 Zea mays cDNA 3', mRNA sequence.
ACCESSION	CF626074
VERSION	CF626074.1 GI:37378736
KEYWORDS	EST.
SOURCE	Zea mays
ORGANISM	Zea mays Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE	1 (bases 1 to 620)
AUTHORS	Bohnert,H., Sharp,R.E., Springer,G.K., Poroyko,V., Fredricksen,M., Sharp,L.G., Spollen,W.G., Ries,J., Guillen,A., Khambati,A., Topinka,C., Davis,G.E., Schachtman,D., Wu,Y. and Nguyen,H.T.
TITLE	NSF Grant DBI-0211842: Functional Genomics of Root Growth and Root Signaling Under Drought
JOURNAL	Unpublished (2003)
COMMENT	Contact: Hans Bohnert University of Illinois, Urbana-Champaign 1201 West Gregory Drive, Urbana, IL 61801, USA Tel: 217-265-5475 Fax: 217-333-5574 Email: bohnerth@life.uiuc.edu POLYA=Yes.

FEATURES Location/Qualifiers

source 1. .620
/organism="Zea mays"
/mol_type="mRNA"
/db_xref="taxon:4577"
/clone_lib="zmrws05"
/note="Samples were collected in Robert E. Sharp's lab (University of Missouri-Columbia) to construct three normalized cDNA libraries. Dark-grown maize seedlings with primary roots 12-20 mm in length were transplanted to high (-0.03 MPa) or low water potential (-1.6 MPa) vermiculite, and harvested at 5 h and 48 h after transplanting. About 1,000 roots were used for each of the low water potential libraries (zmrws05 and zmrws48) while 500 roots were combined from each of the two time points at high water potential (zmrww00). Each root was divided into 4 segments (distances are from the junction of the root apex and root cap): segment 1, 0-3 mm plus the root cap; segment 2, 3-7 mm; segment 3, 7-12 mm; segment 4, 12-20 mm. (For detatils of conditions see (1) with nutrient modifications as in (2)). The three normalized cDNA libraries were constructed in the lab of Hans Bohnert (University of Illinois-UC). Total RNA was extracted by the 'hot Phenol' method (Plant Molecular Biology manual. D5: 1-13, 2nd ed., 1997). This method worked in eliminating carbohydrate material present in the root tips. The integrity of the RNA was verified by denaturing agarose gels and spectrophotometry (ratio A260/280). Poly(A)+mRNA was isolated twice from total RNA using the Oligotex Direct mRNA kit (Qiagen). Poly(A)+ mRNA was converted to double-stranded cDNA and tagged by using modified Oligo(dT) primers. One of 4 sequence tags corresponding to a different segment of the root was added to the 3'-end of the modified Oligo(dT) primers, including a Not1 site and used to reverse transcribe the segment-specific mRNAs into cDNAs. Each library contains all four tags. A suffix (s1, s2, s3, or s4) has been added to each sequence identifier to designate which region of the root (Root_segment_1, 2, 3, or 4) the sequence was found in based on the identification of the tag. A suffix of s0 indicates that the sequence tag, and hence the source segment, could not be identified. The double stranded cDNAs were size-selected (>450 bp). Size selected cDNAs were adaptored with EcoRI adaptors at both ends, and then digested with NotI. The cDNA was directionally cloned into EcoR1-NotI digested pBS II SK(+) phagemid vector (Stratagene) and electroporated into E.coli DH10B. The total number of white colony forming units (cfu) in the primary libraries before amplification was as follows: zmrws05: 3.37 x 106; zmrws48: 4.87 x 106; zmrww00: 3 x 106. The background of empty clones was less than 1%. Inserts ranged from ~0.5kb to >2.5 kb, as determined by PCR. Plasmid DNA from the primary libraries then was converted to single-stranded circles and used as a template for PCR amplification using the T7 and T3 priming sites that flank the cloned cDNA inserts. The purified PCR products, representing the entire cDNA population cloned in each library, were used as a driver for normalization. Hybridization between the single-stranded library and the PCR products was carried out for 44 hours at 30oC. Non-hybridized single-stranded DNA circles were separated from hybridized DNA rendered partially double-stranded and electroporated into DH10B. The total number of clones with insert was: zmrws05: 2.0x107; zmrws48: 4.2x107; zmrww00: 1.1x107. The background of empty clones was less than 2%. Insert size, determined by PCR of the entire library, ranged from 0.5kb to 2.5kb. (1) Sharp R E; Silk W K; Hsiao T C. Growth of the Maize Primary Root at Low Water Potentials I. Spatial Distribution of Expansive Growth. Plant Physiology (Rockville). 87(1). 1988. 50-57. (2) Spollen W G; LeNoble M E; Samuels T D; Bernstein N; Sharp R E. Abscisic acid accumulation maintains maize primary root elongation at low water potentials by restricting ethylene production. Plant Physiology (Rockville). 122(3). March, 2000. 967-976.
TAG_TISSUE=Root_segment_3
TAG_SEQ=TCGCA"

ORIGIN

Query Match 8.2%; Score 337.4; DB 7; Length 620;
Best Local Similarity 78.7%; Pred. No. 8.9e-47;
Matches 415; Conservative 0; Mismatches 111; Indels 1; Gaps 1;

```
Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | |||||||||||||||||||||||| | | | | | | |||||||||||| | |||||||
Db        566 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 507

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | |||||||||| | ||| |||| || | |
Db        506 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 447

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        446 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 387

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db        386 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 327

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db        326 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 267

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db        266 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 207

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db        206 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 147
```

```
Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db        146 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 87

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATT 752
              ||| ||||||||||||       ||  || ||    | ||| |||||
Db         86 TATATATATATATATATATATATATGTAAATTAGCATCACCAGCTGATT 40
```

RESULT 10
CF634146/c

```
LOCUS       CF634146                632 bp    mRNA    linear   EST 02-OCT-2003
DEFINITION  zmrww00_0A10-008-g12.s3 zmrww00 Zea mays cDNA 3', mRNA sequence.
ACCESSION   CF634146
VERSION     CF634146.1  GI:37393761
KEYWORDS    EST.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 632)
  AUTHORS   Bohnert,H., Sharp,R.E., Springer,G.K., Poroyko,V., Fredricksen,M.,
            Sharp,L.G., Spollen,W.G., Ries,J., Guillen,A., Khambati,A.,
            Topinka,C., Davis,G.E., Schachtman,D., Wu,Y. and Nguyen,H.T.
  TITLE     NSF Grant DBI-0211842: Functional Genomics of Root Growth and Root
            Signaling Under Drought
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Hans Bohnert
            University of Illinois, Urbana-Champaign
            1201 West Gregory Drive, Urbana, IL 61801, USA
            Tel: 217-265-5475
            Fax: 217-333-5574
            Email: bohnerth@life.uiuc.edu
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .632
                     /organism="Zea mays"
                     /mol_type="mRNA"
                     /db_xref="taxon:4577"
                     /clone_lib="zmrww00"
                     /note="Samples were collected in Robert E. Sharp's lab
                     (University of Missouri-Columbia) to construct three
                     normalized cDNA libraries. Dark-grown maize seedlings with
                     primary roots 12-20 mm in length were transplanted to high
                     (-0.03 MPa) or low water potential (-1.6 MPa) vermiculite,
                     and harvested at 5 h and 48 h after transplanting. About
                     1,000 roots were used for each of the low water potential
                     libraries (zmrws05 and zmrws48) while 500 roots were
                     combined from each of the two time points at high water
                     potential (zmrww00). Each root was divided into 4 segments
                     (distances are from the junction of the root apex and root
                     cap): segment 1, 0-3 mm plus the root cap; segment 2, 3-7
                     mm; segment 3, 7-12 mm; segment 4, 12-20 mm. (For detatils
                     of conditions see (1) with nutrient modifications as in
                     (2)). The three normalized cDNA libraries were constructed
                     in the lab of Hans Bohnert (University of Illinois-UC).
                     Total RNA was extracted by the 'hot Phenol' method (Plant
                     Molecular Biology manual. D5: 1-13, 2nd ed., 1997). This
                     method worked in eliminating carbohydrate material present
                     in the root tips. The integrity of the RNA was verified by
                     denaturing agarose gels and spectrophotometry (ratio
                     A260/280). Poly(A)+mRNA was isolated twice from total RNA
                     using the Oligotex Direct mRNA kit (Qiagen). Poly(A)+ mRNA
                     was converted to double-stranded cDNA and tagged by using
                     modified Oligo(dT) primers. One of 4 sequence tags
                     corresponding to a different segment of the root was added
                     to the 3'-end of the modified Oligo(dT) primers, including
                     a Not1 site and used to reverse transcribe the
                     segment-specific mRNAs into cDNAs. Each library contains
                     all four tags. A suffix (s1, s2, s3, or s4) has been added
                     to each sequence identifier to designate which region of
                     the root (Root_segment_1, 2, 3, or 4) the sequence was
                     found in based on the identification of the tag. A suffix
                     of s0 indicates that the sequence tag, and hence the
                     source segment, could not be identified. The double
                     stranded cDNAs were size-selected (>450 bp). Size selected
                     cDNAs were adaptored with EcoRI adaptors at both ends, and
                     then digested with NotI. The cDNA was directionally cloned
                     into EcoR1-NotI digested pBS II SK(+) phagemid vector
                     (Stratagene) and electroporated into E.coli DH10B. The
                     total number of white colony forming units (cfu) in the
                     primary libraries before amplification was as follows:
                     zmrws05: 3.37 x 106; zmrws48: 4.87 x 106; zmrww00: 3 x
                     106. The background of empty clones was less than 1%.
                     Inserts ranged from ~0.5kb to >2.5 kb, as determined by
                     PCR. Plasmid DNA from the primary libraries then was
                     converted to single-stranded circles and used as a
                     template for PCR amplification using the T7 and T3 priming
                     sites that flank the cloned cDNA inserts. The purified PCR
                     products, representing the entire cDNA population cloned
                     in each library, were used as a driver for normalization.
                     Hybridization between the single-stranded library and the
                     PCR products was carried out for 44 hours at 30oC.
                     Non-hybridized single-stranded DNA circles were separated
                     from hybridized DNA rendered partially double-stranded and
                     electroporated into DH10B. The total number of clones with
                     insert was: zmrws05: 2.0x107; zmrws48: 4.2x107; zmrww00:
                     1.1x107. The background of empty clones was less than 2%.
                     Insert size, determined by PCR of the entire library,
                     ranged from 0.5kb to 2.5kb. (1) Sharp R E; Silk W K; Hsiao
                     T C. Growth of the Maize Primary Root at Low Water
                     Potentials I. Spatial Distribution of Expansive Growth.
                     Plant Physiology (Rockville). 87(1). 1988. 50-57. (2)
                     Spollen W G; LeNoble M E; Samuels T D; Bernstein N; Sharp
                     R E. Abscisic acid accumulation maintains maize primary
                     root elongation at low water potentials by restricting
                     ethylene production. Plant Physiology (Rockville). 122(3).
                     March, 2000. 967-976.
                     TAG_TISSUE=Root_segment_3
                     TAG_SEQ=TCGCA"
ORIGIN
```

```
  Query Match             8.2%;  Score 337.4;  DB 7;  Length 632;
  Best Local Similarity   78.7%;  Pred. No. 8.9e-47;
  Matches  415;  Conservative    0;  Mismatches  111;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACATATATATATACACATATATA 286
              |||||||||||||||||||||||||| | | | | | |||||||||| | |||||||
Db        545 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 486

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db        485 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 426

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        425 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 366

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db        365 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 306

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db        305 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 246

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db        245 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 186

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db        185 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 126
```

```
Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db        125 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 66

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATT 752
              ||| ||||||||||||        ||  || ||    | ||| |||||
Db         65 TATATATATATATATATATATATGTAAATTAGCATCACCAGCTGATT 19


RESULT 11
CW016544/c
LOCUS       CW016544                883 bp    DNA     linear   GSS 23-SEP-2004
DEFINITION  ZMMBLc0009C11.f ZMMBLc Zea mays genomic clone ZMMBLc0009C11 5',
            genomic survey sequence.
ACCESSION   CW016544
VERSION     CW016544.1  GI:52606273
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 883)
  AUTHORS   Wing,R., Luo,M., Soderlund,C. and Haller,K.
  TITLE     ZMMBL sequences
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            Plate: 0009  row: C  column: 11
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .883
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /cultivar="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBLc0009C11"
                     /tissue_type="immature ears"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="ZMMBLc"
                     /note="Vector: pAGIBAC1; Site_1: Hpa II; Site_2: Hpa II"
ORIGIN

  Query Match             8.2%;  Score 337.4;  DB 9;  Length 883;
  Best Local Similarity   81.4%;  Pred. No. 8.5e-47;
  Matches  403;  Conservative    0;  Mismatches   91;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||||| | | ||| | | |||||||||||| |||||||||
Db        872 TATATATATATATATATATATATATATATATACATATATATATATATATATACATATATA 813

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              |||| ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db        812 TACATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 753

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        752 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 693

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db        692 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 633

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db        632 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 573

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db        572 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 513

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db        512 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 453

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db        452 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 393

Qy        706 TATGTATATATATAT 720
              ||| ||||||||||||
Db        392 TATATATATATATAT 378


RESULT 12
CW016544
LOCUS       CW016544                883 bp    DNA     linear   GSS 23-SEP-2004
DEFINITION  ZMMBLc0009C11.f ZMMBLc Zea mays genomic clone ZMMBLc0009C11 5',
            genomic survey sequence.
ACCESSION   CW016544
VERSION     CW016544.1  GI:52606273
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 883)
  AUTHORS   Wing,R., Luo,M., Soderlund,C. and Haller,K.
  TITLE     ZMMBL sequences
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            Plate: 0009  row: C  column: 11
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .883
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /cultivar="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBLc0009C11"
                     /tissue_type="immature ears"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="ZMMBLc"
                     /note="Vector: pAGIBAC1; Site_1: Hpa II; Site_2: Hpa II"
ORIGIN

  Query Match             8.2%;  Score 337.2;  DB 9;  Length 883;
  Best Local Similarity   71.3%;  Pred. No. 9.2e-47;
  Matches  458;  Conservative    0;  Mismatches  183;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | |||||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db        233 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 292

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db        293 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 352

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        353 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 412

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
```

```
Db        413 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 472

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db        473 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 532

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db        533 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 592

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db        593 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 652

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | ||||||||  ||||||
Db        653 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 712

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||||        | |||          | |   ||    |    |||
Db        713 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 772

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
              |   | | | |  | | |    | |||    | | | | || | |  |       | |
Db        773 TATATATATATATATATATATATATATATATATATATGTATATATATGTATATATATATA 832

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA 867
              ||     | | ||| ||| || | ||   | |  ||| | ||
Db        833 TATATATGTATATATATATATATATATATATATATATATA 874



RESULT 13
AQ917102
LOCUS       AQ917102                725 bp    DNA     linear   GSS 07-DEC-1999
DEFINITION  T233131b Medicago truncatula BAC library Medicago truncatula
            genomic clone 24E12, genomic survey sequence.
ACCESSION   AQ917102
VERSION     AQ917102.1  GI:6536747
KEYWORDS    GSS.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 725)
  AUTHORS   Kim,D., Limpens,E., Peng,H., Ellis,L. and Cook,D.R.
  TITLE     BAC end sequencing of Medicago truncatula
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Cook DR
            The Crop Biotechnology Center
            Texas A&M University
            Department of Plant Pathology and Microbiology, Rm 120 L.F.
            Peterson Bldg, College Station, TX 77843-2132, USA
            Tel: 409 845 8743
            Fax: 409 862 4790
            Email: dcook@ppserver.tamu.edu
            Other name: DK277-L; date: 8/12/99; Submitted to the Database of
            Genome Survey Sequences (GSS) on 12/06/99; More information is
            available at 'http://chrysie.tamu.edu/medicago'.
            Seq primer: SQ-BAC-L (AACGCCAGGGTTTTCCCAGTCACGACG)
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .725
                     /organism="Medicago truncatula"
                     /mol_type="genomic DNA"
                     /cultivar="genotype A17"
                     /db_xref="taxon:3880"
                     /clone="24E12"
                     /clone_lib="Medicago truncatula BAC library"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII; Nam. Y-W, Penmetsa, R.V., Endre, G., Kim, D., and
                     Cook, D.R.  1999. Construction of a bacterial artificial
                     chromosome  library of Medicago truncatula and
                     identification of clones containing ethylene response
                     genes. Theor Appl Genet 98:  638-646. "
ORIGIN

  Query Match             8.2%;  Score 336.6;  DB 8;  Length 725;
  Best Local Similarity   77.7%;  Pred. No. 1.2e-46;
  Matches  419;  Conservative    0;  Mismatches  119;  Indels    1;  Gaps    1;

Qy        183 AGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAGGCATTTTATATATATATATA 242
              ||  || ||   | |  |||  |   |   ||||   | | |||  |  ||||||||||
Db        158 AGCAAAGAAACCCAATTTGTGGTGGAAATGAAAAGGGATGCCATGATGGATATATATATA 217

Qy        243 TATATATATACACACACACACATATATATATACACATATATATACACATATATACATA 302
              ||||||||||| | | | | | | |||||||||||| | ||||||||| | ||||||| |||
Db        218 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 277

Qy        303 TATACACATATATACACATATATATACACATACATATGTACACATATATATACACATATG 362
              |||| | ||||||| | ||||||||| | ||| |||| || | ||||||||| | ||||
Db        278 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 337

Qy        363 TATACACATATATACACATATATACACACATATATACACATATATACACACATATATACA 422
              |||| | ||||||| | ||||||| | | ||||||| | ||||||| | | |||||||
Db        338 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 397

Qy        423 CATATATACACATATATACACATATACACATATATACACATATATACATATATACACATA 482
              ||||||| | ||||||| | ||||| | ||||||| | ||||||| ||||||| | |||
Db        398 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 457

Qy        483 TATATA-ATATACACACATATATATACACATATATACACACATATATACACATATATACA 541
              |||||| ||||| | | ||||||||| | ||||||| | | ||||||| | ||||||| |
Db        458 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 517

Qy        542 CATATATATACACATATATACACATATATACATATATACACATATATATACATATATACA 601
              ||||||||| | |||||||| | ||||||| ||||||| | | ||||||||| ||||||| |
Db        518 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 577

Qy        602 CATATATACATATATACACATATATACATATATACACACATATATACACATACATATACA 661
              ||||||| ||||||| | ||||||| ||||||| | | ||||||| | ||| ||||| |
Db        578 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 637

Qy        662 CACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATAT 720
              | | ||| ||||| |||||||| | |||||||||  |||||||| ||||||||||
Db        638 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 696



RESULT 14
CC067623
LOCUS       CC067623                827 bp    DNA     linear   GSS 16-APR-2003
DEFINITION  CSU-K33r.18M15.SP6 CSU-K33r Aedes aegypti genomic clone
            CSU-K33r.18M15, genomic survey sequence.
ACCESSION   CC067623
VERSION     CC067623.1  GI:29906129
KEYWORDS    GSS.
SOURCE      Aedes aegypti (yellow fever mosquito)
  ORGANISM  Aedes aegypti
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Aedes;
            Stegomyia.
REFERENCE   1  (bases 1 to 827)
  AUTHORS   Loftus,B., Shetty,J., Severson,D., Brown,S. and Knudson,D.
  TITLE     End sequencing of Aedes aegypti BACs
  JOURNAL   Unpublished (2003)
COMMENT     Other_GSSs: CSU-K33r.18M15.T7
            Contact: Brendan Loftus
            Department of Eukaryotic Genomics
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-3543
            Fax: 301-838-0208
            Email: enta@tigr.org
            Library was provided by Susan Brown and Dennis Knudson at Colorado
            State University.
```

```
           Seq primer: SP6
           Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .827
                     /organism="Aedes aegypti"
                     /mol_type="genomic DNA"
                     /strain="Rexville"
                     /db_xref="taxon:7159"
                     /clone="CSU-K33r.18M15"
                     /clone_lib="CSU-K33r"
                     /note="Vector: pBeloBAC11; Site_1: HindIII"
ORIGIN

  Query Match             8.2%;  Score 336.4;  DB 8;  Length 827;
  Best Local Similarity   78.7%;  Pred. No. 1.3e-46;
  Matches  414;  Conservative    0;  Mismatches  111;  Indels    1;  Gaps    1;

Qy        196 GAAATGTCATCTCAGAAAAAAAAAAAGGCATTTTATATATATATATATATATATACAC 255
              ||| |  | ||  |  ||     |||| |     ||||||||||||||||||||||| |
Db        238 GAATTTCCTTCAAACGAATGCTAAAACGTGCGATATATATATATATATATATATATAT 297

Qy        256 ACACACACACATATATATATACACATATATATACACATATATACATATATACACATATAT 315
              | | | | | |||||||||||| | ||||||||| | ||||||| ||||||| | ||||||
Db        298 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 357

Qy        316 ACACATATATATACACATACATATGTACACATATATATACACATATGTATACACATATAT 375
              | |||||||||| | ||| |||| || | ||||||||| | |||| |||| | |||||||
Db        358 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 417

Qy        376 ACACATATATACACACATATATACACATATATACACACATATATACACATATATACACAT 435
              | | ||||||| | | ||||||| | ||||||| | ||||||| | ||||||| | ||
Db        418 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 477

Qy        436 ATATACACATATACACATATATACACATATATACATATATACACATATATATA-ATATAC 494
              ||||| | ||||| | ||||||| | ||||||| ||||||| | ||||||||| |||||
Db        478 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 537

Qy        495 ACACATATATATACACATATATACACACATATATACACATATATACACATATATATACAC 554
              | | ||||||||| | ||||||| | | ||||||| | ||||||| | ||||||||| |
Db        538 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 597

Qy        555 ATATATACACATATATACATATATACACATATATATACATATATACACATATATACATAT 614
              ||||||| | ||||||| ||||||| | ||||||||| ||||||| | ||||||| ||||
Db        598 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 657

Qy        615 ATACACATATATACATATATACACACATATATACACATACATATACACACACATAGATAT 674
              |||  | ||||||| ||||||| | | ||||||| | |||  ||||| | | | ||| ||||
Db        658 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 717

Qy        675 ACATATATATACACATATATATACGTATATATATGTATATATATAT 720
              | ||||||||| | |||||||||  ||||||||| ||||||||||
Db        718 ATATATATATATATATATATATATATATATATATATATATATATAT 763


RESULT 15
CR294134
LOCUS       CR294134                590 bp    DNA     linear   GSS 28-FEB-2004
DEFINITION  Medicago truncatula BAC ends cultivar Jemalong A17 of Medicago
            truncatula, genomic survey sequence.
ACCESSION   CR294134
VERSION     CR294134.1  GI:44684558
KEYWORDS    GSS.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 590)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-FEB-2004) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
FEATURES             Location/Qualifiers
     source          1. .590
                     /organism="Medicago truncatula"
                     /mol_type="genomic DNA"
                     /cultivar="Jemalong A17"
                     /db_xref="taxon:3880"
                     /clone_lib="MTE1"
                     /note="Vector: pIndigoBAC ; Site_1: EcoRI ; Site_2: EcoRI
                     ; Debelle F. and Chalhoub B.~Genoscope sequence ID :
                     mte1-1J18FM1"
ORIGIN

  Query Match             8.2%;  Score 336.2;  DB 9;  Length 590;
  Best Local Similarity   79.7%;  Pred. No. 1.4e-46;
  Matches  409;  Conservative    0;  Mismatches  103;  Indels    1;  Gaps    1;

Qy        209 AGAAAAAAAAAAAGGCATTTTATATATATATATATATATATATACACACACACACATA 268
              || | | | | | | || |||||||||||||||||||||||| | | | | | |||
Db         15 AGGAGATATATATAGTATAGTATATATATATATATATATATATATATATATATATA 74

Qy        269 TATATATACACATATATATACACATATATACATATATACACATATATACACATATATATA 328
              |||||||| | ||||||||| | ||||||| ||||||| | ||||||| | |||||||||
Db         75 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 134

Qy        329 CACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACA 388
              | ||| |||| || | ||||||||| | |||| |||| | ||||||| | ||||||| |
Db        135 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 194

Qy        389 CACATATATACACATATATACACACATATATACACATATATACACATATATACACATATA 448
              | ||||||| | |||||||| | | ||||||| ||||||| | ||||||| | |||||
Db        195 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 254

Qy        449 CACATATATACACATATATACATATATACACATATATATA-ATATACACACATATATATA 507
              | ||||||| | ||||||| ||||||| | ||||||||| ||||| | | |||||||||
Db        255 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 314

Qy        508 CACATATATACACACATATATACACATATATACACATATATATACACATATATACACATA 567
              | ||||||| | | ||||||| | ||||||| | ||||||||| | ||||||| | |||
Db        315 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 374

Qy        568 TATACATATATACACATATATATACATATATACACATATATACATATATACACATATATA 627
              |||| ||||||| | ||||||||| ||||||| | ||||||| ||||||| | |||||||
Db        375 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 434

Qy        628 CATATATACACACATATATACACATACATATACACACACATAGATATACATATATATACA 687
              ||||||| | | ||||||| | ||| ||||| | | | ||| ||||| |||||||||
Db        435 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 494

Qy        688 CATATATATACGTATATATATGTATATATATAT 720
              |||||||||  ||||||||| ||||||||||
Db        495 TATATATATATATATATATATATATATATATAT 527


RESULT 16
CC067623/c
LOCUS       CC067623                827 bp    DNA     linear   GSS 16-APR-2003
DEFINITION  CSU-K33r.18M15.SP6 CSU-K33r Aedes aegypti genomic clone
            CSU-K33r.18M15, genomic survey sequence.
ACCESSION   CC067623
VERSION     CC067623.1  GI:29906129
KEYWORDS    GSS.
SOURCE      Aedes aegypti (yellow fever mosquito)
  ORGANISM  Aedes aegypti
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Aedes;
            Stegomyia.
REFERENCE   1  (bases 1 to 827)
  AUTHORS   Loftus,B., Shetty,J., Severson,D., Brown,S. and Knudson,D.
  TITLE     End sequencing of Aedes aegypti BACs
  JOURNAL   Unpublished (2003)
COMMENT     Other_GSSs: CSU-K33r.18M15.T7
            Contact: Brendan Loftus
```

Department of Eukaryotic Genomics
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-3543
Fax: 301-838-0208
Email: enta@tigr.org
Library was provided by Susan Brown and Dennis Knudson at Colorado State University.
Seq primer: SP6
Class: BAC ends.

```
FEATURES             Location/Qualifiers
     source          1. .827
                     /organism="Aedes aegypti"
                     /mol_type="genomic DNA"
                     /strain="Rexville"
                     /db_xref="taxon:7159"
                     /clone="CSU-K33r.18M15"
                     /clone_lib="CSU-K33r"
                     /note="Vector: pBeloBAC11; Site_1: HindIII"
ORIGIN

  Query Match             8.1%;  Score 335.8;  DB 8;  Length 827;
  Best Local Similarity   76.2%;  Pred. No. 1.6e-46;
  Matches  426;  Conservative    0;  Mismatches  132;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db        766 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 707

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db        706 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 647

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        646 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 587

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db        586 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 527

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db        526 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 467

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db        466 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 407

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db        406 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 347

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | ||||||||  ||||||
Db        346 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 287

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| |||||||||||    ||   || |  |   ||   |   | | |    |  | |  |
Db        286 TATATATATATATATATCGCACGTTTTAGCATTCGTTTGAAGGAAATTCAGGAGAAGGGA 227

Qy        766 GAGGAAAAGAGAGTTTGAC 784
              ||| |    | | ||| ||
Db        226 GAGTATTGTATAATTTTAC 208


RESULT 17
CF626074
LOCUS       CF626074                620 bp    mRNA    linear   EST 02-OCT-2003
DEFINITION  zmrws05_0B10-009-d11.s3 zmrws05 Zea mays cDNA 3', mRNA sequence.
ACCESSION   CF626074
VERSION     CF626074.1  GI:37378736
KEYWORDS    EST.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 620)
  AUTHORS   Bohnert,H., Sharp,R.E., Springer,G.K., Poroyko,V., Fredricksen,M.,
            Sharp,L.G., Spollen,W.G., Ries,J., Guillen,A., Khambati,A.,
            Topinka,C., Davis,G.E., Schachtman,D., Wu,Y. and Nguyen,H.T.
  TITLE     NSF Grant DBI-0211842: Functional Genomics of Root Growth and Root
            Signaling Under Drought
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Hans Bohnert
            University of Illinois, Urbana-Champaign
            1201 West Gregory Drive, Urbana, IL 61801, USA
            Tel: 217-265-5475
            Fax: 217-333-5574
            Email: bohnerth@life.uiuc.edu
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .620
                     /organism="Zea mays"
                     /mol_type="mRNA"
                     /db_xref="taxon:4577"
                     /clone_lib="zmrws05"
                     /note="Samples were collected in Robert E. Sharp's lab
                     (University of Missouri-Columbia) to construct three
                     normalized cDNA libraries. Dark-grown maize seedlings with
                     primary roots 12-20 mm in length were transplanted to high
                     (-0.03 MPa) or low water potential (-1.6 MPa) vermiculite,
                     and harvested at 5 h and 48 h after transplanting. About
                     1,000 roots were used for each of the low water potential
                     libraries (zmrws05 and zmrws48) while 500 roots were
                     combined from each of the two time points at high water
                     potential (zmrww00). Each root was divided into 4 segments
                     (distances are from the junction of the root apex and root
                     cap): segment 1, 0-3 mm plus the root cap; segment 2, 3-7
                     mm; segment 3, 7-12 mm; segment 4, 12-20 mm. (For detatils
                     of conditions see (1) with nutrient modifications as in
                     (2)). The three normalized cDNA libraries were constructed
                     in the lab of Hans Bohnert (University of Illinois-UC).
                     Total RNA was extracted by the 'hot Phenol' method (Plant
                     Molecular Biology manual. D5: 1-13, 2nd ed., 1997). This
                     method worked in eliminating carbohydrate material present
                     in the root tips. The integrity of the RNA was verified by
                     denaturing agarose gels and spectrophotometry (ratio
                     A260/280). Poly(A)+mRNA was isolated twice from total RNA
                     using the Oligotex Direct mRNA kit (Qiagen). Poly(A)+ mRNA
                     was converted to double-stranded cDNA and tagged by using
                     modified Oligo(dT) primers. One of 4 sequence tags
                     corresponding to a different segment of the root was added
                     to the 3'-end of the modified Oligo(dT) primers, including
                     a Not1 site and used to reverse transcribe the
                     segment-specific mRNAs into cDNAs. Each library contains
                     all four tags. A suffix (s1, s2, s3, or s4) has been added
                     to each sequence identifier to designate which region of
                     the root (Root_segment_1, 2, 3, or 4) the sequence was
                     found in based on the identification of the tag. A suffix
                     of s0 indicates that the sequence tag, and hence the
                     source segment, could not be identified. The double
                     stranded cDNAs were size-selected (>450 bp). Size selected
                     cDNAs were adaptored with EcoRI adaptors at both ends, and
                     then digested with NotI. The cDNA was directionally cloned
                     into EcoR1-NotI digested pBS II SK(+) phagemid vector
                     (Stratagene) and electroporated into E.coli DH10B. The
                     total number of white colony forming units (cfu) in the
                     primary libraries before amplification was as follows:
                     zmrws05: 3.37 x 106; zmrws48: 4.87 x 106; zmrww00: 3 x
                     106. The background of empty clones was less than 1%.
                     Inserts ranged from ~0.5kb to >2.5 kb, as determined by
                     PCR. Plasmid DNA from the primary libraries then was
                     converted to single-stranded circles and used as a
                     template for PCR amplification using the T7 and T3 priming
```

sites that flank the cloned cDNA inserts. The purified PCR products, representing the entire cDNA population cloned in each library, were used as a driver for normalization. Hybridization between the single-stranded library and the PCR products was carried out for 44 hours at 30oC. Non-hybridized single-stranded DNA circles were separated from hybridized DNA rendered partially double-stranded and electroporated into DH10B. The total number of clones with insert was: zmrws05: 2.0x107; zmrws48: 4.2x107; zmrww00: 1.1x107. The background of empty clones was less than 2%. Insert size, determined by PCR of the entire library, ranged from 0.5kb to 2.5kb. (1) Sharp R E; Silk W K; Hsiao T C. Growth of the Maize Primary Root at Low Water Potentials I. Spatial Distribution of Expansive Growth. Plant Physiology (Rockville). 87(1). 1988. 50-57. (2) Spollen W G; LeNoble M E; Samuels T D; Bernstein N; Sharp R E. Abscisic acid accumulation maintains maize primary root elongation at low water potentials by restricting ethylene production. Plant Physiology (Rockville). 122(3). March, 2000. 967-976.
TAG_TISSUE=Root_segment_3
TAG_SEQ=TCGCA"

ORIGIN

Query Match 8.1%; Score 335.2; DB 7; Length 620;
Best Local Similarity 78.4%; Pred. No. 2.1e-46;
Matches 414; Conservative 0; Mismatches 113; Indels 1; Gaps 1;

```
Qy      194 GCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATATAC 253
            || || |  ||| |    | | | | |      | ||||||||||||||||||||||||
Db       54 GCTAATTTACATATATATATATATATATATATATATATATATATATATATATATATATATAT 113

Qy      254 ACACACACACATATATATATACACATATATATACACATATATACATATATACACATAT 313
            | | | | | ||||||||||| | ||||||||| | ||||||| ||||||| | ||||
Db      114 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 173

Qy      314 ATACACATATATATACACATACATATGTACACATATATATACACATATGTATACACATAT 373
            ||| | ||||||||| | ||| |||| || | ||||||||| | |||| |||| | ||||
Db      174 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 233

Qy      374 ATACACATATATACACACATATATACACATATATACACACATATATACACATATATACAC 433
            ||| | ||||||| | | ||||||| | ||||||| | | ||||||| | ||||||| |
Db      234 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 293

Qy      434 ATATATACACATATACACATATATACACATATATACATATATACACATATATATA-ATAT 492
            ||||||| | ||||| | ||||||| | ||||||| ||||||| | ||||||||| ||||
Db      294 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 353

Qy      493 ACACACATATATATACACATATATACACACATATATACACATATATACACATATATATAC 552
            | | | ||||||||| | ||||||| | | ||||||| | ||||||| | |||||||||
Db      354 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 413

Qy      553 ACATATATACACATATATACATATATACACATATATATACATATATACACATATATACAT 612
            | ||||||| | ||||||| ||||||| | ||||||||| ||||||| | ||||||| ||
Db      414 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 473

Qy      613 ATATACACATATATACATATATACACACATATATACACATACATATACACACACATAGAT 672
            ||||| | ||||||| ||||||| | | ||||||| | ||| ||||| | | | ||| ||
Db      474 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 533

Qy      673 ATACATATATATACACATATATATACGTATATATATGTATATATATAT 720
            ||| ||||||||| | |||||||||  ||||||||| ||||||||||
Db      534 ATATATATATATATATATATATATATATATATATATATATATATATAT 581
```

RESULT 18
CF634146
LOCUS CF634146 632 bp mRNA linear EST 02-OCT-2003
DEFINITION zmrww00_0A10-008-g12.s3 zmrww00 Zea mays cDNA 3', mRNA sequence.
ACCESSION CF634146
VERSION CF634146.1 GI:37393761
KEYWORDS EST.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 632)
AUTHORS Bohnert,H., Sharp,R.E., Springer,G.K., Poroyko,V., Fredricksen,M., Sharp,L.G., Spollen,W.G., Ries,J., Guillen,A., Khambati,A., Topinka,C., Davis,G.E., Schachtman,D., Wu,Y. and Nguyen,H.T.
TITLE NSF Grant DBI-0211842: Functional Genomics of Root Growth and Root Signaling Under Drought
JOURNAL Unpublished (2003)
COMMENT Contact: Hans Bohnert
University of Illinois, Urbana-Champaign
1201 West Gregory Drive, Urbana, IL 61801, USA
Tel: 217-265-5475
Fax: 217-333-5574
Email: bohnerth@life.uiuc.edu
POLYA=Yes.
FEATURES Location/Qualifiers
source 1. .632
/organism="Zea mays"
/mol_type="mRNA"
/db_xref="taxon:4577"
/clone_lib="zmrww00"
/note="Samples were collected in Robert E. Sharp's lab (University of Missouri-Columbia) to construct three normalized cDNA libraries. Dark-grown maize seedlings with primary roots 12-20 mm in length were transplanted to high (-0.03 MPa) or low water potential (-1.6 MPa) vermiculite, and harvested at 5 h and 48 h after transplanting. About 1,000 roots were used for each of the low water potential libraries (zmrws05 and zmrws48) while 500 roots were combined from each of the two time points at high water potential (zmrww00). Each root was divided into 4 segments (distances are from the junction of the root apex and root cap): segment 1, 0-3 mm plus the root cap; segment 2, 3-7 mm; segment 3, 7-12 mm; segment 4, 12-20 mm. (For detatils of conditions see (1) with nutrient modifications as in (2)). The three normalized cDNA libraries were constructed in the lab of Hans Bohnert (University of Illinois-UC). Total RNA was extracted by the 'hot Phenol' method (Plant Molecular Biology manual. D5: 1-13, 2nd ed., 1997). This method worked in eliminating carbohydrate material present in the root tips. The integrity of the RNA was verified by denaturing agarose gels and spectrophotometry (ratio A260/280). Poly(A)+mRNA was isolated twice from total RNA using the Oligotex Direct mRNA kit (Qiagen). Poly(A)+ mRNA was converted to double-stranded cDNA and tagged by using modified Oligo(dT) primers. One of 4 sequence tags corresponding to a different segment of the root was added to the 3'-end of the modified Oligo(dT) primers, including a Not1 site and used to reverse transcribe the segment-specific mRNAs into cDNAs. Each library contains all four tags. A suffix (s1, s2, s3, or s4) has been added to each sequence identifier to designate which region of the root (Root_segment_1, 2, 3, or 4) the sequence was found in based on the identification of the tag. A suffix of s0 indicates that the sequence tag, and hence the source segment, could not be identified. The double stranded cDNAs were size-selected (>450 bp). Size selected cDNAs were adaptored with EcoRI adaptors at both ends, and then digested with NotI. The cDNA was directionally cloned into EcoR1-NotI digested pBS II SK(+) phagemid vector (Stratagene) and electroporated into E.coli DH10B. The total number of white colony forming units (cfu) in the primary libraries before amplification was as follows: zmrws05: 3.37 x 106; zmrws48: 4.87 x 106; zmrww00: 3 x 106. The background of empty clones was less than 1%. Inserts ranged from ~0.5kb to >2.5 kb, as determined by PCR. Plasmid DNA from the primary libraries then was converted to single-stranded circles and used as a template for PCR amplification using the T7 and T3 priming sites that flank the cloned cDNA inserts. The purified PCR

products, representing the entire cDNA population cloned in each library, were used as a driver for normalization. Hybridization between the single-stranded library and the PCR products was carried out for 44 hours at 30oC. Non-hybridized single-stranded DNA circles were separated from hybridized DNA rendered partially double-stranded and electroporated into DH10B. The total number of clones with insert was: zmrws05: 2.0x107; zmrws48: 4.2x107; zmrww00: 1.1x107. The background of empty clones was less than 2%. Insert size, determined by PCR of the entire library, ranged from 0.5kb to 2.5kb. (1) Sharp R E; Silk W K; Hsiao T C. Growth of the Maize Primary Root at Low Water Potentials I. Spatial Distribution of Expansive Growth. Plant Physiology (Rockville). 87(1). 1988. 50-57. (2) Spollen W G; LeNoble M E; Samuels T D; Bernstein N; Sharp R E. Abscisic acid accumulation maintains maize primary root elongation at low water potentials by restricting ethylene production. Plant Physiology (Rockville). 122(3). March, 2000. 967-976.
TAG_TISSUE=Root_segment_3
TAG_SEQ=TCGCA"

ORIGIN

Query Match 8.1%; Score 335.2; DB 7; Length 632;
Best Local Similarity 78.4%; Pred. No. 2.1e-46;
Matches 414; Conservative 0; Mismatches 113; Indels 1; Gaps 1;

```
Qy        194 GCGAAATGTCATCTCAGAAAAAAAAAAAGGCATTTTATATATATATATATATATATATAC 253
              || || |  ||| |    | | | | |      | ||||||||||||||||||||||||
Db         33 GCTAATTTACATATATATATATATATATATATATATATATATATATATATATATATATAT 92

Qy        254 ACACACACACACATATATATATACACATATATATACACATATATACATATATACACATAT 313
              | | | | | | |||||||||||| | ||||||||| | ||||||| ||||||| | ||||
Db         93 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 152

Qy        314 ATACACATATATATACACATACATATGTACACATATATATACACATATGTATACACATAT 373
              ||| | ||||||||| | ||| |||| || | ||||||||| | |||| |||| | ||||
Db        153 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 212

Qy        374 ATACACATATATACACACATATATACACATATATACACACATATATACACATATATACAC 433
              ||| | ||||||| | | ||||||| | ||||||| | | ||||||| | ||||||| |
Db        213 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 272

Qy        434 ATATATACACATATACACATATATACACATATATACATATATACACATATATATA-ATAT 492
              ||||||| | ||||| | ||||||| | ||||||| ||||||| | ||||||||| ||||
Db        273 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 332

Qy        493 ACACACATATATATACACATATATACACACATATATACACATATATACACATATATATAC 552
              | | | ||||||||| | ||||||| | | ||||||| | ||||||| | |||||||||
Db        333 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 392

Qy        553 ACATATATACACATATATACATATATACACATATATATACATATATACACATATATACAT 612
              | ||||||| | ||||||| ||||||| | ||||||||| ||||||| | ||||||| ||
Db        393 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 452

Qy        613 ATATACACATATATACATATATACACACATATATACACATACATATACACACACATAGAT 672
              ||||| | ||||||| |||||||| | | ||||||| | ||| ||||| | | | ||| ||
Db        453 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 512

Qy        673 ATACATATATATACACATATATATACGTATATATATGTATATATATAT 720
              ||| ||||||||| | |||||||||  ||||||||| ||||||||||
Db        513 ATATATATATATATATATATATATATATATATATATATATATATATAT 560
```

RESULT 19
BI505681
LOCUS BI505681 717 bp mRNA linear EST 08-APR-2002
DEFINITION BB170032B10E06.5 Bee Brain Normalized/Subtracted Library, BB17 Apis mellifera cDNA clone BB170032B10E06 5', mRNA sequence.
ACCESSION BI505681
VERSION BI505681.1 GI:15356055
KEYWORDS EST.
SOURCE Apis mellifera (honey bee)
ORGANISM Apis mellifera
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Apoidea; Apidae; Apis.
REFERENCE 1 (bases 1 to 717)
AUTHORS Whitfield,C.W., Band,M.R., Bonaldo,M.F., Kumar,C.G., Liu,L., Pardinas,J., Robertson,H.M., Soares,B. and Robinson,G.E.
TITLE Annotated expressed sequence tags and cDNA microarrays for studies of brain and behavior in the honey bee
JOURNAL Genome Res. 12 (4); 555-566 (2002)
MEDLINE 21929762
PUBMED 11932240
COMMENT Contact: Gene E. Robinson
Department of Entomology
University of Illinois
505 S. Goodwin Ave., Urbana, IL 61801, USA
Tel: 217 265 0309
Fax: 217 244 3499
Email: generobi@life.uiuc.edu
This research was funded by the University of Illinois Critical Research Initiatives Fund and a Burroughs-Wellcome Trust Innovation Award in Functional Genomics to G.E. Robinson and an NSF Postdoctoral Fellowship in Bioinformatics to C.W. Whitfield.
REPEATS IN THE SEQUENCE
Low_complexity STRAND (+) ELEMENT AT_rich LOCATION [101,132];
Simple_repeat STRAND (+) ELEMENT (TA)n LOCATION [133,717].
PCR PRImers
FORWARD: TAATACGACTCACTATAGGG
BACKWARD: ATTAACCCTCACTAAAG
Plate: BB170032B10 row: E column: 06
Seq primer: AGCGGATAACAATTTCACACAGGA
High quality sequence stop: 717.
FEATURES Location/Qualifiers
source 1. .717
/organism="Apis mellifera"
/mol_type="mRNA"
/strain="mixed strains of European bees, predominantly A.m. ligustica"
/db_xref="taxon:7460"
/clone="BB170032B10E06"
/sex="female"
/tissue_type="brain"
/dev_stage="adult worker honey bee"
/lab_host="DH10B"
/clone_lib="Bee Brain Normalized/Subtracted Library, BB17"
/note="Organ: brain; Vector: pT7T3-Pac; Site_1: EcoR1; Site_2: Not1; This BB17 cDNA library was generated by subtraction of the BB16 library with 4000 previously sequenced clones. The BB16 library was contributed by the Soares laboratory and it was constructed and normalized as described by Bonaldo, M.F., Lennon, G. and Soares, M.B. (1996), Genome Research 6(9): 791-806. RNA was prepared from dissected brains of adult worker bees of various ages and various behavioral groups. "

ORIGIN

Query Match 8.1%; Score 335.2; DB 4; Length 717;
Best Local Similarity 77.2%; Pred. No. 2.1e-46;
Matches 420; Conservative 0; Mismatches 123; Indels 1; Gaps 1;

```
Qy        178 AGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAGGCATTTTATATATAT 237
              | | ||   || | ||   | || |      | | || | | |      | |||||||||
Db         96 AACTCAATTAAAATGATTTATATATATATATATATAATATATATATATATATATATAT 155

Qy        238 ATATATATATATATACACACACACACATATATATATACACATATATATACACATATAT 297
              |||||||||||||||| | | | | | | |||||||||| | ||||||||| | ||||||
Db        156 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 215

Qy        298 ACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACAC 357
              | ||||||| | ||||||| | ||||||||| | ||| |||| || | |||||||||| |
Db        216 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 275

Qy        358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417
```

```
Db        276 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 335

Qy        418 ATACACATATATACACATATATACACATATACACATATATACACATATATACATATATAC 477

Db        336 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 395

Qy        478 ACATATATATA-ATATACACACATATATATACACATATATACACACATATATACACATAT 536

Db        396 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 455

Qy        537 ATACACATATATATACACATATATACACATATATACATATATACACATATATATACATAT 596

Db        456 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 515

Qy        597 ATACACATATATACATATATACACATATATACATATATACACACATATATACACATACAT 656

Db        516 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 575

Qy        657 ATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATAT 716

Db        576 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 635

Qy        717 ATAT 720

Db        636 ATAT 639
```

```
RESULT 20
CR294134/c
LOCUS       CR294134                590 bp    DNA     linear   GSS 28-FEB-2004
DEFINITION  Medicago truncatula BAC ends cultivar Jemalong A17 of Medicago
            truncatula, genomic survey sequence.
ACCESSION   CR294134
VERSION     CR294134.1  GI:44684558
KEYWORDS    GSS.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 590)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-FEB-2004) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
FEATURES             Location/Qualifiers
     source          1. .590
                     /organism="Medicago truncatula"
                     /mol_type="genomic DNA"
                     /cultivar="Jemalong A17"
                     /db_xref="taxon:3880"
                     /clone_lib="MTE1"
                     /note="Vector: pIndigoBAC ; Site_1: EcoRI ; Site_2: EcoRI
                     ; Debelle F. and Chalhoub B.~Genoscope sequence ID :
                     mte1-1J18FM1"
ORIGIN

  Query Match             8.1%;  Score 334.8;  DB 9;  Length 590;
  Best Local Similarity   79.8%;  Pred. No. 2.5e-46;
  Matches  407;  Conservative    0;  Mismatches  102;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACATATATATATACACATATATA 286

Db        532 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 473

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346

Db        472 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 413

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
```

```
Db        412 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 353

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466

Db        352 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 293

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525

Db        292 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 233

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585

Db        232 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 173

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645

Db        172 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 113

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705

Db        112 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 53

Qy        706 TATGTATATATATATGCTCCAGAGTTCATA 735

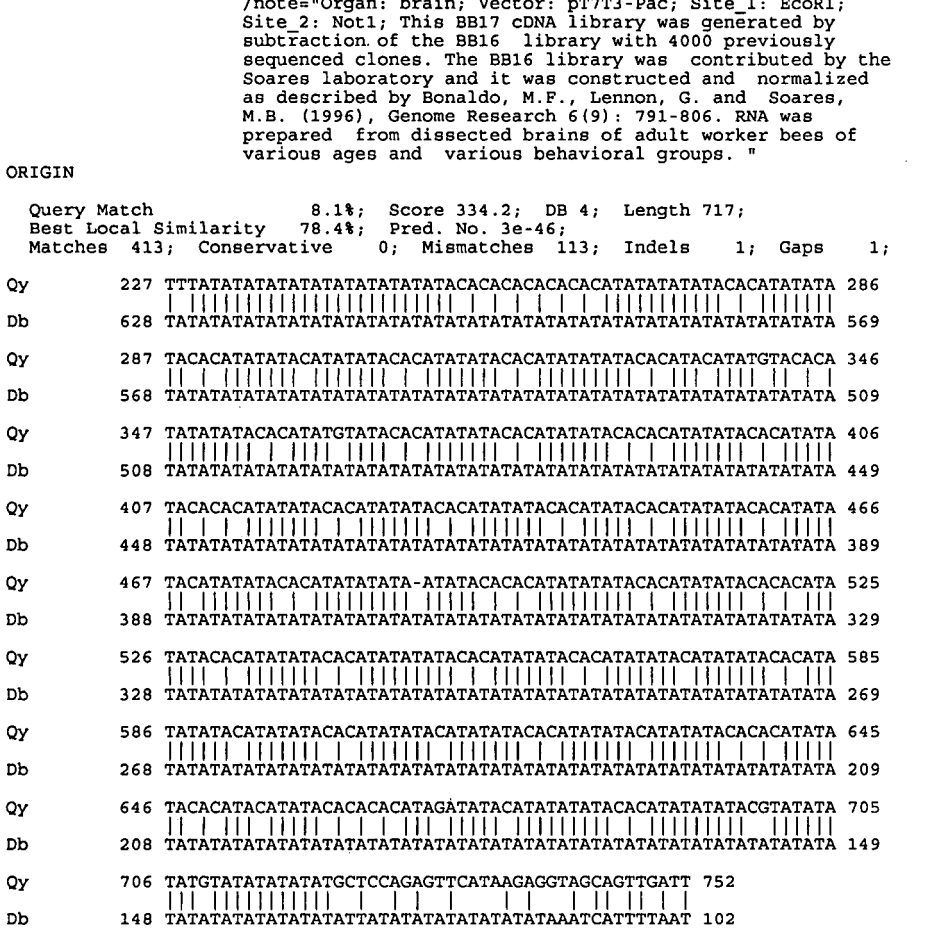
Db         52 TATATATATATATATATACTATACTATATA 23
```

```
RESULT 21
BI505681/c
LOCUS       BI505681                717 bp    mRNA    linear   EST 08-APR-2002
DEFINITION  BB170032B10E06.5 Bee Brain Normalized/Subtracted Library, BB17 Apis
            mellifera cDNA clone BB170032B10E06 5', mRNA sequence.
ACCESSION   BI505681
VERSION     BI505681.1  GI:15356055
KEYWORDS    EST.
SOURCE      Apis mellifera (honey bee)
  ORGANISM  Apis mellifera
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Apoidea;
            Apidae; Apis.
REFERENCE   1  (bases 1 to 717)
  AUTHORS   Whitfield,C.W., Band,M.R., Bonaldo,M.F., Kumar,C.G., Liu,L.,
            Pardinas,J., Robertson,H.M., Soares,B. and Robinson,G.E.
  TITLE     Annotated expressed sequence tags and cDNA microarrays for studies
            of brain and behavior in the honey bee
  JOURNAL   Genome Res. 12 (4), 555-566 (2002)
  MEDLINE   21929762
   PUBMED   11932240
COMMENT     Contact: Gene E. Robinson
            Department of Entomology
            University of Illinois
            505 S. Goodwin Ave., Urbana, IL 61801, USA
            Tel: 217 265 0309
            Fax: 217 244 3499
            Email: generobi@life.uiuc.edu
            This research was funded by the University of Illinois Critical
            Research Initiatives Fund and a Burroughs-Wellcome Trust Innovation
            Award in Functional Genomics to G.E. Robinson and an NSF
            Postdoctoral Fellowship in Bioinformatics to C.W. Whitfield.
            REPEATS IN THE SEQUENCE
            Low_complexity STRAND (+) ELEMENT AT_rich LOCATION [101,132];
            Simple_repeat STRAND (+) ELEMENT (TA)n LOCATION [133,717].
            PCR PRimers
            FORWARD: TAATACGACTCACTATAGGG
            BACKWARD: ATTAACCCTCACTAAAG
            Plate: BB170032B10  row: E  column: 06
            Seq primer: AGCGGATAACAATTTCACACAGGA
            High quality sequence stop: 717.
FEATURES             Location/Qualifiers
     source          1. .717
                     /organism="Apis mellifera"
                     /mol_type="mRNA"
```

/strain="mixed strains of European bees, predominantly A.m. ligustica"
/db_xref="taxon:7460"
/clone="BB170032B10E06"
/sex="female"
/tissue_type="brain"
/dev_stage="adult worker honey bee"
/lab_host="DH10B"
/clone_lib="Bee Brain Normalized/Subtracted Library, BB17"
/note="Organ: brain; Vector: pT7T3-Pac; Site_1: EcoR1; Site_2: Not1; This BB17 cDNA library was generated by subtraction of the BB16 library with 4000 previously sequenced clones. The BB16 library was contributed by the Soares laboratory and it was constructed and normalized as described by Bonaldo, M.F., Lennon, G. and Soares, M.B. (1996), Genome Research 6(9): 791-806. RNA was prepared from dissected brains of adult worker bees of various ages and various behavioral groups. "

ORIGIN

Query Match 8.1%; Score 334.2; DB 4; Length 717;
Best Local Similarity 78.4%; Pred. No. 3e-46;
Matches 413; Conservative 0; Mismatches 113; Indels 1; Gaps 1;



RESULT 22
AQ917102/c
LOCUS AQ917102 725 bp DNA linear GSS 07-DEC-1999
DEFINITION T233131b Medicago truncatula BAC library Medicago truncatula genomic clone 24E12, genomic survey sequence.
ACCESSION AQ917102
VERSION AQ917102.1 GI:6536747
KEYWORDS GSS.
SOURCE Medicago truncatula (barrel medic)
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.
REFERENCE 1 (bases 1 to 725)
AUTHORS Kim,D., Limpens,E., Peng,H., Ellis,L. and Cook,D.R.
TITLE BAC end sequencing of Medicago truncatula
JOURNAL Unpublished (1999)
COMMENT Contact: Cook DR
The Crop Biotechnology Center
Texas A&M University
Department of Plant Pathology and Microbiology, Rm 120 L.F. Peterson Bldg, College Station, TX 77843-2132, USA
Tel: 409 845 8743
Fax: 409 862 4790
Email: dcook@ppserver.tamu.edu
Other name: DK277-L; date: 8/12/99; Submitted to the Database of Genome Survey Sequences (GSS) on 12/06/99; More information is available at 'http://chrysie.tamu.edu/medicago'.
Seq primer: SQ-BAC-L (AACGCCAGGGTTTTCCCAGTCACGACG)
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .725
/organism="Medicago truncatula"
/mol_type="genomic DNA"
/cultivar="genotype A17"
/db_xref="taxon:3880"
/clone="24E12"
/clone_lib="Medicago truncatula BAC library"
/note="Vector: pBeloBAC11; Site_1: HindIII; Site_2: HindIII; Nam. Y-W, Penmetsa, R.V., Endre, G., Kim, D., and Cook, D.R. 1999. Construction of a bacterial artificial chromosome library of Medicago truncatula and identification of clones containing ethylene response genes. Theor Appl Genet 98: 638-646. "

ORIGIN

Query Match 8.1%; Score 333.8; DB 8; Length 725;
Best Local Similarity 80.4%; Pred. No. 3.5e-46;
Matches 403; Conservative 0; Mismatches 97; Indels 1; Gaps 1;



```
Db        221 TATATATATATATATCCATCA 201


RESULT 23
CL623140/c
LOCUS       CL623140                744 bp    DNA     linear   GSS 01-JUL-2004
DEFINITION  OR_BBa0018D11.r OR_BBa Oryza rufipogon genomic clone OR_BBa0018D11
            3', genomic survey sequence.
ACCESSION   CL623140
VERSION     CL623140.1  GI:49121028
KEYWORDS    GSS.
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 744)
  AUTHORS   Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R.,
            Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
  TITLE     OMAP Project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Insert Length: 161   Std Error: 0.00
            Plate: 0018  row: D  column: 11
            Seq primer: CAC TCA TTA GGC ACC CCA
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .744
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
                     /clone="OR_BBa0018D11"
                     /tissue_type="young leaves"
                     /lab_host="DH10B-T1 phage resistant"
                     /clone_lib="OR_BBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

  Query Match             8.1%;  Score 333;  DB 9;  Length 744;
  Best Local Similarity   79.3%;  Pred. No. 4.8e-46;
  Matches  407;  Conservative    0;  Mismatches  105;  Indels    1;  Gaps    1;

Qy        209 AGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACACATA 268
              | | | | | | |   || |||||||| |||||||||||||||| | | | | | |||
Db        702 ATATATATATATATATATATTATATATCTATATATATATATATATATATATATATATA 643

Qy        269 TATATATACACATATATATACACATATATACATATATACACATATATACACATATATATA 328
              |||||||| | ||||||||| | ||||||| ||||||| | ||||||| | |||||||||
Db        642 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 583

Qy        329 CACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACA 388
              | ||| |||| || | ||||||||| | |||| |||| | ||||||| | ||||||| |
Db        582 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 523

Qy        389 CACATATATACACATATATACACACATATATACACATATATACACATATATACACATATA 448
              | ||||||| | ||||||| | | ||||||| | ||||||| | ||||||| | |||||
Db        522 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 463

Qy        449 CACATATATACACATATATACATATATACACATATATATAATATACACACATATATATAC 508
              | |||||||| | ||||||| ||||||| | |||||||| ||||| | | |||||||||
Db        462 TATATATATATATATATATATATATATATATATATATAT-ATATATATATATATATAT 404

Qy        509 ACATATATACACACATATATACACATATATACACATATATATACACATATATACACATAT 568
              | ||||||| | | ||||||| | ||||||| | ||||||||| | ||||||| | ||||
Db        403 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 344

Qy        569 ATACATATATACACATATATATACATATATACACATATATACATATATACACATATATAC 628
              ||| ||||||| | ||||||||| ||||||| | ||||||| ||||||| | |||||||
Db        343 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 284

Qy        629 ATATATACACACATATATACACATACATATACACACACATAGATATACATATATATACAC 688
              ||||||| | | ||||||| | ||| ||||| | | | ||| ||||| |||||||||| |
Db        283 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 224

Qy        689 ATATATATACGTATATATATGTATATATATATG 721
              |||||||||  |||||||||||||| ||| ||||
Db        223 ATATATATATATATATATATGTATGTATGTATG 191


RESULT 24
BZ796807/c
LOCUS       BZ796807                512 bp    DNA     linear   GSS 17-MAR-2003
DEFINITION  PUGDF25TD ZM_0.6_1.0_KB Zea mays genomic clone ZMMBTa341F02,
            genomic survey sequence.
ACCESSION   BZ796807
VERSION     BZ796807.1  GI:28998547
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 512)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
            Bennetzen,J.
  TITLE     Maize Genomics Consortium
  JOURNAL   Unpublished (2003)
COMMENT     Other_GSSs: PUGDF25TB
            Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .512
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBTa341F02"
                     /clone_lib="ZM_0.6_1.0_KB"
                     /note="Vector: pCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high
                     CoT selected genomic DNA library"
ORIGIN

  Query Match             8.1%;  Score 332.6;  DB 8;  Length 512;
  Best Local Similarity   80.8%;  Pred. No. 5.9e-46;
  Matches  400;  Conservative    0;  Mismatches   94;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||||||| | | | | | |||||||||| | |||||||
Db        508 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 449

Qy        287 TACACATATATACACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db        448 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 389

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        388 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 329

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
```

```
Db        328 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 269

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db        268 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 209

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db        208 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 149

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db        148 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 89

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db         88 TATATATATATATATATATATATATATATATATATATATATATATATATATATAAATATATA 29

Qy        706 TATGTATATATATAT 720
              ||| ||||||||||||
Db         28 TATATATATATATAT 14


RESULT 25
CL755178/c
LOCUS       CL755178                707 bp    DNA     linear   GSS 27-JUL-2004
DEFINITION  OR_BBa0122N12.r OR_BBa Oryza rufipogon genomic clone OR_BBa0122N12
            3', genomic survey sequence.
ACCESSION   CL755178
VERSION     CL755178.1  GI:50706035
KEYWORDS    GSS.
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 707)
  AUTHORS   Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R.,
            Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
  TITLE     OMAP Project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Insert Length: 161   Std Error: 0.00
            Plate: 0122  row: N  column: 12
            Seq primer: CAC TCA TTA GGC ACC CCA
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .707
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
                     /clone="OR_BBa0122N12"
                     /tissue_type="young leaves"
                     /lab_host="DH10B-T1 phage resistant"
                     /clone_lib="OR_BBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

  Query Match             8.0%;  Score 331.6;  DB 9;  Length 707;
  Best Local Similarity   79.4%;  Pred. No. 8.3e-46;
  Matches  405;  Conservative    0;  Mismatches  104;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||| | ||||||||||||||| | | | | | |||||||||||| | |||||||

Db        563 TATATATATCTTTATATATATATATATATATATATATATATATATATATATATATATA 504

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||  |||| || | |
Db        503 TATATATATATATATATATATATATATATATATATATATATATATATCTATATTTATATA 444

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              ||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        443 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 384

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db        383 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 324

Qy        467 TACATATATACACATATATATAATATACACACATATATATACACATATATACACACATAT 526
              || ||||||| | |||||||| ||||| | | ||||||||| | ||||||| | | ||||
Db        323 TATATATATATATATATATAT-ATATATATATATATATATATATATATATATATATATAT 265

Qy        527 ATACACATATATACACATATATATACACATATATACACATATATACATATATACACATAT 586
              ||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | ||||
Db        264 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 205

Qy        587 ATATACATATATACACATATATACATATATACACATATATACATATATACACACATATAT 646
              ||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | ||||||
Db        204 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 145

Qy        647 ACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATAT 706
              | | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  |||||||
Db        144 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 85

Qy        707 ATGTATATATATATGCTCCAGAGTTCATAA 736
              || |||||||||||| | | ||  | || |
Db         84 ATATATATATATATGATGCTGATATGATTA 55


RESULT 26
CL934691
LOCUS       CL934691                784 bp    DNA     linear   GSS 14-SEP-2004
DEFINITION  OA_ABa0044J13.f OA_ABa Oryza australiensis genomic clone
            OA_ABa0044J13 5', genomic survey sequence.
ACCESSION   CL934691
VERSION     CL934691.1  GI:52063483
KEYWORDS    GSS.
SOURCE      Oryza australiensis
  ORGANISM  Oryza australiensis
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 784)
  AUTHORS   Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R.,
            Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
  TITLE     OMAP Project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0044  row: J  column: 13
            Seq primer: TAA TAC GAC TCA CTA TAG GG
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .784
                     /organism="Oryza australiensis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4532"
                     /clone="OA_ABa0044J13"
                     /tissue_type="young leaves"
```

```
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OA_ABa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

  Query Match             8.0%;  Score 331.2;  DB 9;  Length 784;
  Best Local Similarity   78.9%;  Pred. No. 9.6e-46;
  Matches  407;  Conservative    0;  Mismatches  108;  Indels    1;  Gaps    1;

Qy        205 TCTCAGAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACA 264
              ||||    | | | |       | ||||||||||||||||||||||||| | | | | | |
Db        201 TCTCTCTATATATATATATATATATATATATATATATATATATATATATATATATATA 260

Qy        265 CATATATATATACACATATATATACACATATATACATATATACACATATATACACATATA 324
              |||||||||||| | ||||||||| | ||||||| ||||||| | ||||||| | |||||
Db        261 TATATATATATATATATATATATATATATATATATATATATATATAGATATATATATATATA 320

Qy        325 TATACACATACATATGTACACATATATATACACATATGTATACACATATATACACATATA 384
              |||| | ||| |||| || | |||||||||| | |||| |||| | ||||||| | |||||
Db        321 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 380

Qy        385 TACACACATATATACACATATATACACACATATATACACATATATACACATATATACACA 444
              || | | ||||||| | ||||||| | | ||||||| | ||||||| | ||||||| | |
Db        381 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 440

Qy        445 TATACACATATATACACATATATACATATATACACATATATAATATACACACATATAT 504
              |||| | |||||||| | ||||||| ||||||| | |||||||| ||||| | | ||||||
Db        441 TATATATATATATATATATATATATATATATATATATATATAT-ATATATATATATAT 499

Qy        505 ATACACATATATACACACATATATACACATATATACACATATATATACACATATATACAC 564
              ||| | ||||| | | | ||||||| | ||||||| | ||||||||| | ||||||| |
Db        500 ATAGATATATAGATATATATATATATATATATATATATATATATATATATATATATAT 559

Qy        565 ATATATACATATATACACATATATATACATATATACACATATATACATATATACACATAT 624
              ||||||| ||||||| | ||||||||| ||||||| | ||||||| ||||||| | ||||
Db        560 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 619

Qy        625 ATACATATATACACACATATATACACATACATATACACACACATAGATATACATATATAT 684
              ||| ||||||| | | ||||||| | ||| ||||| | | | ||||||||| ||||||||
Db        620 ATATATATATATATATATATATATATATATATATATATATATATATATAGATATATATATAT 679

Qy        685 ACACATATATATACGTATATATATGTATATATATAT 720
              | | |||||||||  ||||||||  ||||||||||
Db        680 ATATATATATATATATATATATAAATATATATATAT 715


RESULT 27
BZ796807
LOCUS       BZ796807                 512 bp    DNA     linear   GSS 17-MAR-2003
DEFINITION  PUGDF25TD ZM_0.6_1.0_KB Zea mays genomic clone ZMMBTa341F02,
            genomic survey sequence.
ACCESSION   BZ796807
VERSION     BZ796807.1  GI:28998547
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 512)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
            Bennetzen,J.
  TITLE     Maize Genomics Consortium
  JOURNAL   Unpublished (2003)
COMMENT     Other_GSSs: PUGDF25TB
            Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .512
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBTa341F02"
                     /clone_lib="ZM_0.6_1.0_KB"
                     /note="Vector: pCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high
                     CoT selected genomic DNA library"
ORIGIN

  Query Match             8.0%;  Score 331;  DB 8;  Length 512;
  Best Local Similarity   80.6%;  Pred. No. 1.1e-45;
  Matches  399;  Conservative    0;  Mismatches   95;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | |||||||||||||||||||||||||| | | | |    | |||||||||| | |||||||
Db          1 TATATATATATATATATATATATATATATATATATTTATATATATATATATATATATA 60

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db         61 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 120

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        121 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 180

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db        181 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 240

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | |||||||| ||||| | | ||||||||| | ||||||| | | |||
Db        241 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 300

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db        301 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 360

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              ||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db        361 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 420

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db        421 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 480

Qy        706 TATGTATATATATAT 720
              ||| ||||||||||||
Db        481 TATATATATATATAT 495


RESULT 28
BX145650/c
LOCUS       BX145650                 668 bp    DNA     linear   GSS 28-JAN-2003
DEFINITION  Danio rerio genomic clone DKEY-109N18, genomic survey sequence.
ACCESSION   BX145650
VERSION     BX145650.1  GI:27977003
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 668)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the T7 end of BAC 109N18. 109N18
```

```
            is part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES              Location/Qualifiers
     source           1. .668
                      /organism="Danio rerio"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:7955"
                      /clone="DKEY-109N18"
                      /tissue_type="Testis"
                      /note="vector pIndigoBAC-536"
ORIGIN

  Query Match           8.0%;  Score 330.4;  DB 9;  Length 668;
  Best Local Similarity   78.5%;  Pred. No. 1.3e-45;
  Matches  408;  Conservative    0;  Mismatches  111;  Indels    1;  Gaps    1;

Qy        229 TATATATATATATATATATATATACACACACACACACATATATATATACACATATATATA 288
              ||||||||||||||||||||||| | | | | | |||||||||||| | |||||||||
Db        668 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 609

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              | ||||||| ||||||| | |||||||| | ||||||||| | ||| |||| || | |||
Db        608 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 549

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              |||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||||
Db        548 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 489

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||||
Db        488 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 429

Qy        469 CATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATATA 527
              ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||||
Db        428 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 369

Qy        528 TACACATATATACACATATATATACACATATATACACATATATACATATATACACACATATA 587
              || | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||||
Db        368 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 309

Qy        588 TATACATATATACACATATATACATATATACACATATATACATATATACACACATATATA 647
              |||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||||
Db        308 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 249

Qy        648 CACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATA 707
              | ||| ||||| | | | ||| ||||| |||||||||| | |||||||||  ||||||||
Db        248 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 189

Qy        708 TGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGT 747
              | |||| | |  |  ||   ||  || || |   | |||
Db        188 TATATAAACACTTCATCTCTAGACCAAAACACCCATGAGT 149


RESULT 29
CL755178
LOCUS       CL755178                707 bp    DNA     linear   GSS 27-JUL-2004
DEFINITION  OR_BBa0122N12.r OR_BBa Oryza rufipogon genomic clone OR_BBa0122N12
            3', genomic survey sequence.
ACCESSION   CL755178
VERSION     CL755178.1  GI:50706035
KEYWORDS    GSS.
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 707)
  AUTHORS   Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R.,
            Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
  TITLE     OMAP Project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Insert Length: 161   Std Error: 0.00
            Plate: 0122  row: N  column: 12
            Seq primer: CAC TCA TTA GGC ACC CCA
            Class: BAC ends.
FEATURES              Location/Qualifiers
     source           1. .707
                      /organism="Oryza rufipogon"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:4529"
                      /clone="OR_BBa0122N12"
                      /tissue_type="young leaves"
                      /lab_host="DH10B-T1 phage resistant"
                      /clone_lib="OR_BBa"
                      /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

  Query Match           8.0%;  Score 330.4;  DB 9;  Length 707;
  Best Local Similarity   70.8%;  Pred. No. 1.3e-45;
  Matches  453;  Conservative    0;  Mismatches  186;  Indels    1;  Gaps    1;

Qy        228 TTATATATATATATATATATATATACACACACACACATATATATATACACATATATAT 287
              | ||||||||||||||||||||||| | | | | | |||||||||||| | ||||||||
Db         69 TCATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 128

Qy        288 ACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACAT 347
              |_| ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | ||
Db        129 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 188

Qy        348 ATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATAT 407
              ||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | ||||||
Db        189 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 248

Qy        408 ACACACATATATACACATATATACACATATATACACATATACACATATATACACATATAT 467
              | | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | ||||||
Db        249 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 308

Qy        468 ACATATATACACATATATATAATATACACACATATATATACACATATATACACACATATA 527
              | ||||||| | |||||||| ||||| | | ||||||||| | ||||||| | | |||||
Db        309 ATATATATATATATATATAT-ATATATATATATATATATATATATATATATATATATATA 367

Qy        528 TACACATATATACACATATATATACACATATATACACATATATACATATATACACATATA 587
              || | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||||
Db        368 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 427

Qy        588 TATACATATATACACATATATACATATATACACATATATACATATATACACACATATATA 647
              |||| ||||||| | ||||||| ||||| | | ||||||| ||||||| | | |||||||
Db        428 TATATATATATATATATATATATAAATATAGATATATATATATATATATATATATATATA 487

Qy        648 CACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATA 707
              | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||||
Db        488 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 547

Qy        708 TGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGA 767
              | || | ||||||          | |||          |  |   ||  | |    ||| |
Db        548 TATAAAGATATATATAGATATATATATATATATATATATATATATATATAAATATATATA 607

Qy        768 GGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATA 827
              | | | |  | | | | | |||  | |  | || |      |       | |||
Db        608 TATATATATATATATATATATAAATATATATATATATATATATATATATATATATATA 667

Qy        828 GTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA 867
                |  ||||| | | || | ||   | |  ||| | ||
Db        668 TATAGACAAATAGAGATATATATATATATATATATATATA 707
```

```
RESULT 30
CA916667/c
LOCUS       CA916667                496 bp    mRNA    linear   EST 27-DEC-2002
DEFINITION  PCSC21302 Scarlet Runner Bean Suspensor Region TriplEx2 Phaseolus
            coccineus cDNA 5' similar to No significant hits, mRNA sequence.
ACCESSION   CA916667
VERSION     CA916667.1  GI:27403659
KEYWORDS    EST.
SOURCE      Phaseolus coccineus
  ORGANISM  Phaseolus coccineus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Phaseolus.
REFERENCE   1  (bases 1 to 496)
  AUTHORS   Bui,A.Q., Le,B.H., Weterings,K., Bi,Y.-P., Choi,J.-S.,
            McElroy,K.E., Choi,P.S., Harada,J.J., Fischer,R.L. and
            Goldberg,R.B.
  TITLE     Gene Activity in Different Regions of a Post-Fertilization Plant
            Embryo by EST Analysis
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Goldberg, R.B.
            Department of Molecular, Cell, & Developmental Biology
            University of California, Los Angeles
            621 Charles E. Young Drive South, Los Angeles, CA 90095-1606, USA
            Tel: 310 825 3270
            Fax: 310 825 8201
            Email: bobg@ucla.edu
            Seq primer: 5' TriplEx
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .496
                     /organism="Phaseolus coccineus"
                     /mol_type="mRNA"
                     /cultivar="Hammond's Dwarf Scarlet"
                     /db_xref="taxon:3886"
                     /dev_stage="6-days post-pollination"
                     /clone_lib="Scarlet Runner Bean Suspensor Region TriplEx2"
                     /note="Organ: Suspensor Region of Globular-Stage Embryos;
                     Vector: TriplEx2; Site_1: SfiIA; Site_2: SfiIB; Suspensor
                     regions were micro-dissected from globular-stage embryos
                     six days after pollination from greenhouse-grown plants
                     [Weterings et al., Plant Cell 13, 2409-2425 (2001)].
                     Double-stranded cDNA was synthesized from suspensor mRNA
                     using the SMART cDNA Library Construction Kit according to
                     the manufacturer (Clontech). The suspensor cDNA fragments
                     were directionally ligated into the SfiI restriction site
                     of the lambda TriplEx2 vector (Clontech), and the
                     recombinant cDNAs were transformed into E. coli XL1-Blue
                     cells (Clontech). Suspensor cDNA plasmids used for
                     directional sequencing were obtained by in vivo excision
                     from the lambda TriplEx2 recombinants in E. coli BM25.8
                     cells (Clontech)."
ORIGIN

  Query Match             8.0%;  Score 330.2;  DB 6;  Length 496;
  Best Local Similarity   80.9%;  Pred. No. 1.5e-45;
  Matches  397;  Conservative    0;  Mismatches   93;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db        491 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 432

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db        431 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 372

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        371 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 312

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db        311 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 252

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db        251 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 192

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db        191 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 132

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db        131 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 72

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||||||||||||||| |||||||||  |||| |
Db         71 TATATATATATATATATATATATATATATACATATATATACATATATATATATATACA 12

Qy        706 TATGTATATAT 716
              ||| |||||||
Db         11 TATCTATATAT 1


RESULT 31
CF924719/c
LOCUS       CF924719                751 bp    mRNA    linear   EST 10-NOV-2003
DEFINITION  DD005F11.ab1 Codonopsis lanceolata cDNA library Codonopsis
            lanceolata cDNA, mRNA sequence.
ACCESSION   CF924719
VERSION     CF924719.1  GI:38243245
KEYWORDS    EST.
SOURCE      Codonopsis lanceolata
  ORGANISM  Codonopsis lanceolata
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            asterids; campanulids; Asterales; Campanulaceae; Codonopsis.
REFERENCE   1  (bases 1 to 751)
  AUTHORS   Yang,D.C., In,J.G. and Lee,B.S.
  TITLE     Codonopsis lanceolata
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Deok Chun Yang
            BioPia
            1 Seochonri, Kihung, Yongin, Kyungki-do, 449-701, Korea
            Tel: 82 31 201 2688
            Fax: 82 31 205 6323
            Email: dcyang@ibiopia.com.
FEATURES             Location/Qualifiers
     source          1. .751
                     /organism="Codonopsis lanceolata"
                     /mol_type="mRNA"
                     /db_xref="taxon:103999"
                     /clone_lib="Codonopsis lanceolata cDNA library"
                     /note="Organ: root; Vector: pTriplEx2; Site_1: Sfi IA;
                     Site_2: Sfi IB"
ORIGIN

  Query Match             8.0%;  Score 329.8;  DB 7;  Length 751;
  Best Local Similarity   80.3%;  Pred. No. 1.7e-45;
  Matches  399;  Conservative    0;  Mismatches   97;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db        749 TATATATATATATATATATATATATAGATATATATATATATATATATATATATATATATA 690

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db        689 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 630

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        629 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 570
```

```
Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db        569 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 510

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || |||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db        509 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 450

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db        449 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 390

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db        389 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 330

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | ||| |||||  ||||||
Db        329 TATATATATATATATATATATATATATATATATATATATATATATATAGATATATATATATA 270

Qy        706 TATGTATATATATATGC 722
              ||  ||| |||||||||
Db        269 TAGATATGTATATATGC 253
```

RESULT 32
BX145650
LOCUS BX145650 668 bp DNA linear GSS 28-JAN-2003
DEFINITION Danio rerio genomic clone DKEY-109N18, genomic survey sequence.
ACCESSION BX145650
VERSION BX145650.1 GI:27977003
KEYWORDS GSS.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 668)
AUTHORS Humphray,S.J., Huckle,E. and Durham,J.L.
TITLE Direct Submission
JOURNAL Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Unpublished
COMMENT This sequence was generated from the T7 end of BAC 109N18. 109N18 is part of the Daniokey BAC Library created by R. Plasterk and N.V. Keygene. Further details:
http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES Location/Qualifiers
source 1. .668
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="DKEY-109N18"
/tissue_type="Testis"
/note="vector pIndigoBAC-536"
ORIGIN

Query Match 8.0%; Score 329; DB 9; Length 668;
Best Local Similarity 79.8%; Pred. No. 2.3e-45;
Matches 400; Conservative 0; Mismatches 100; Indels 1; Gaps 1;

```
Qy        217 AAAAAGGCATTTTATATATATATATATATATATACACACACACACACATATATATATA 276
              | | | | | | ||||||||||||||||||||||| | | | | | | ||||||||||||
Db        169 AGAGATGAAGTGTTTATATATATATATATATATATATATATATATATATATATATATATA 228

Qy        277 CACATATATATACACATATATACATATATACACATATATACACATATATATACACATACA 336
                | ||||||||| | ||||||| ||||||| | ||||||| | ||||||||| | ||| |
Db        229 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 288

Qy        337 TATGTACACATATATATACACATATGTATACACATATATACACATATATACACACATATA 396
              ||| || | ||||||||| | |||| |||| | ||||||| | ||||||| | | |||||
Db        289 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 348
```

```
Qy        397 TACACATATATACACACATATATACACATATATACACATATATACACATATACACATATA 456
              || | ||||||| | | ||||||| | ||||||| | ||||||| | ||||| | |||||
Db        349 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 408

Qy        457 TACACATATATACATATATACACATATATATAATATACACACATATATATACACATATAT 516
              || | ||||||| ||||||| | |||||||| ||||| | | ||||||||| | ||||||
Db        409 TATATATATATATATATATATATATATATAT-ATATATATATATATATATATATATATAT 467

Qy        517 ACACACATATATACACATATATACACATATATATACACATATATACACATATATACATAT 576
              | | | ||||||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||
Db        468 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 527

Qy        577 ATACACATATATATACATATATACACATATATACATATATACACATATATACATATATAC 636
              ||| | ||||||||| ||||||| | ||||||| ||||||| | ||||||| |||||||
Db        528 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 587

Qy        637 ACACATATATACACATACATATACACACACATAGATATACATATATATACACATATATAT 696
              | | ||||||| | ||| ||||| | | | ||| ||||| ||||||||| | ||||||||
Db        588 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 647

Qy        697 ACGTATATATATGTATATATA 717
              |  ||||||||| ||||||||
Db        648 ATATATATATATATATATATA 668
```

RESULT 33
CL934691/c
LOCUS CL934691 784 bp DNA linear GSS 14-SEP-2004
DEFINITION OA_ABa0044J13.f OA_ABa Oryza australiensis genomic clone OA_ABa0044J13 5', genomic survey sequence.
ACCESSION CL934691
VERSION CL934691.1 GI:52063483
KEYWORDS GSS.
SOURCE Oryza australiensis
ORGANISM Oryza australiensis
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 784)
AUTHORS Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
TITLE OMAP Project
JOURNAL Unpublished (2004)
COMMENT Contact: Rod A. Wing
Arizona Genomics Institute
University of Arizona
Forbes Building Room 303, Tucson, AZ 85721-0036, USA
Tel: 520 626 9595
Fax: 520 621 1259
Email: http://genome.arizona.edu
PCR PRimers
FORWARD: TAA TAC GAC TCA CTA TAG GG
BACKWARD: CAC TCA TTA GGC ACC CCA
Plate: 0044 row: J column: 13
Seq primer: TAA TAC GAC TCA CTA TAG GG
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .784
/organism="Oryza australiensis"
/mol_type="genomic DNA"
/db_xref="taxon:4532"
/clone="OA_ABa0044J13"
/tissue_type="young leaves"
/lab_host="DH10B T1 phage resistant"
/clone_lib="OA_ABa"
/note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

Query Match 8.0%; Score 328.4; DB 9; Length 784;
Best Local Similarity 79.0%; Pred. No. 2.8e-45;
Matches 403; Conservative 0; Mismatches 106; Indels 1; Gaps 1;

```
Qy        212 AAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACATATAT 271
```

```
Db        719 ATAAATATATATATATTTATATATATATATATATATATATATATATATATATATCTATAT 660

Qy        272 ATATACACATATATATACACATATATACATATATACACATATATACACATATATATACAC 331

Db        659 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 600

Qy        332 ATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACACAC 391

Db        599 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 540

Qy        392 ATATATACACATATATACACACATATATACACATATATACACATATATACACATATACAC 451

Db        539 ATATATATATATATATATATATATATATCTATATATCTATATATATATATATATATATAT 480

Qy        452 ATATATACACATATATACATATATACACATATATATA-ATATACACACATATATATACAC 510

Db        479 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 420

Qy        511 ATATATACACACATATATACACATATATACACATATATATACACATATATACACATATAT 570

Db        419 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 360

Qy        571 ACATATATACACATATATATACATATATACACATATATACATATATACACATATATACAT 630

Db        359 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATCTAT 300

Qy        631 ATATACACACATATATACACATACATATACACACACATAGATATACATATATATACACAT 690

Db        299 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 240

Qy        691 ATATATACGTATATATATGTATATATATAT 720

Db        239 ATATATATATATATATATATATATATAT 210
```

RESULT 34
CL776144/c

```
LOCUS       CL776144                587 bp    DNA     linear   GSS 30-JUL-2004
DEFINITION  OR_BBa0090B12.r OR_BBa Oryza rufipogon genomic clone OR_BBa0090B12
            3', genomic survey sequence.
ACCESSION   CL776144
VERSION     CL776144.1  GI:50850530
KEYWORDS    GSS.
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 587)
  AUTHORS   Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R.,
            Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
  TITLE     OMAP Project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Insert Length: 161   Std Error: 0.00
            Plate: 0090  row: B  column: 12
            Seq primer: CAC TCA TTA GGC ACC CCA
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .587
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
                     /clone="OR_BBa0090B12"
                     /tissue_type="young leaves"
                     /lab_host="DH10B-T1 phage resistant"
                     /clone_lib="OR_BBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN
```

```
  Query Match             8.0%;  Score 328.2;  DB 9;  Length 587;
  Best Local Similarity   77.5%;  Pred. No. 3.2e-45;
  Matches  410;  Conservative    0;  Mismatches  118;  Indels    1;  Gaps    1;

Qy        209 AGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACATA 268

Db        537 ATATATATATATATATATATAATATATATATATATATATATATATATATATATATATA 478

Qy        269 TATATATACACATATATATACACATATATACATATATACACATATATACACATATATATA 328

Db        477 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 418

Qy        329 CACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACA 388

Db        417 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 358

Qy        389 CACATATATACACATATATACACACATATATACACATATATACACATATATACACATATA 448

Db        357 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 298

Qy        449 CACATATATACACATATATACATATATACACATATATATA-ATATACACACATATATATA 507

Db        297 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 238

Qy        508 CACATATATACACACATATATACACATATATACACATATATATACACATATATACACATA 567

Db        237 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 178

Qy        568 TATACATATATACACATATATATACATATATACACATATATACATATATACACATATATA 627

Db        177 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 118

Qy        628 CATATATACACACATATATACACATACATATACACACACATAGATATACATATATATACA 687

Db        117 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 58

Qy        688 CATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAA 736

Db         57 TATATATATATATATATATGCATATGCGCCTGTGTATAAGAAATCATAA 9
```

RESULT 35
CF924719

```
LOCUS       CF924719                751 bp    mRNA    linear   EST 10-NOV-2003
DEFINITION  DD005F11.ab1 Codonopsis lanceolata cDNA library Codonopsis
            lanceolata cDNA, mRNA sequence.
ACCESSION   CF924719
VERSION     CF924719.1  GI:38243245
KEYWORDS    EST.
SOURCE      Codonopsis lanceolata
  ORGANISM  Codonopsis lanceolata
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            asterids; campanulids; Asterales; Campanulaceae; Codonopsis.
REFERENCE   1  (bases 1 to 751)
  AUTHORS   Yang,D.C., In,J.G. and Lee,B.S.
  TITLE     Codonopsis lanceolata
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Deok Chun Yang
            BioPia
            1 Seochonri, Kihung, Yongin, Kyungki-do, 449-701, Korea
            Tel: 82 31 201 2688
            Fax: 82 31 205 6323
            Email: dcyang@ibiopia.com.
FEATURES             Location/Qualifiers
     source          1. .751
```

```
                     /organism="Codonopsis lanceolata"
                     /mol_type="mRNA"
                     /db_xref="taxon:103999"
                     /clone_lib="Codonopsis lanceolata cDNA library"
                     /note="Organ: root; Vector: pTriplEx2; Site_1: Sfi IA;
                     Site_2: Sfi IB"
ORIGIN

  Query Match             7.9%;  Score 327.2;  DB 7;  Length 751;
  Best Local Similarity   80.0%;  Pred. No. 4.5e-45;
  Matches  397;  Conservative    0;  Mismatches   98;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | |||| |||| |||||||||||||||    | | | |  |||||||||| | |||||||
Db        256 TATATACATATCTATATATATATATATCTATATATATATATATATATATATATATATATA 315

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db        316 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 375

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        376 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 435

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db        436 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 495

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db        496 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 555

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db        556 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 615

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db        616 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 675

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||| ||||  ||||||
Db        676 TATATATATATATATATATATATATATATATATATATATATATATATATATCTATATATATA 735

Qy        706 TATGTATATATATATG 721
              ||| ||||||||||||
Db        736 TATATATATATATATG 751


RESULT 36
CA916667
LOCUS       CA916667               496 bp    mRNA    linear   EST 27-DEC-2002
DEFINITION  PCSC21302 Scarlet Runner Bean Suspensor Region TriplEx2 Phaseolus
            coccineus cDNA 5' similar to No significant hits, mRNA sequence.
ACCESSION   CA916667
VERSION     CA916667.1  GI:27403659
KEYWORDS    EST.
SOURCE      Phaseolus coccineus
  ORGANISM  Phaseolus coccineus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Phaseolus.
REFERENCE   1  (bases 1 to 496)
  AUTHORS   Bui,A.Q., Le,B.H., Weterings,K., Bi,Y.-P., Choi,J.-S.,
            McElroy,K.E., Choi,P.S., Harada,J.J., Fischer,R.L. and
            Goldberg,R.B.
  TITLE     Gene Activity in Different Regions of a Post-Fertilization Plant
            Embryo by EST Analysis
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Goldberg, R.B.
            Department of Molecular, Cell, & Developmental Biology
            University of California, Los Angeles
            621 Charles E. Young Drive South, Los Angeles, CA 90095-1606, USA
            Tel: 310 825 3270
            Fax: 310 825 8201
            Email: bobg@ucla.edu
            Seq primer: 5' TriplEx
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .496
                     /organism="Phaseolus coccineus"
                     /mol_type="mRNA"
                     /cultivar="Hammond's Dwarf Scarlet"
                     /db_xref="taxon:3886"
                     /dev_stage="6-days post-pollination"
                     /clone_lib="Scarlet Runner Bean Suspensor Region TriplEx2"
                     /note="Organ: Suspensor Region of Globular-Stage Embryos;
                     Vector: TriplEx2; Site_1: SfiIA; Site_2: SfiIB; Suspensor
                     regions were micro-dissected from globular-stage embryos
                     six days after pollination from greenhouse-grown plants
                     [Weterings et al., Plant Cell 13, 2409-2425 (2001)].
                     Double-stranded cDNA was synthesized from suspensor mRNA
                     using the SMART cDNA Library Construction Kit according to
                     the manufacturer (Clontech). The suspensor cDNA fragments
                     were directionally ligated into the SfiI restriction site
                     of the lambda TriplEx2 vector (Clontech), and the
                     recombinant cDNAs were transformed into E. coli XL1-Blue
                     cells (Clontech). Suspensor cDNA plasmids used for
                     directional sequencing were obtained by in vivo excision
                     from the lambda TriplEx2 recombinants in E. coli BM25.8
                     cells (Clontech)."
ORIGIN

  Query Match             7.9%;  Score 326.2;  DB 6;  Length 496;
  Best Local Similarity   80.0%;  Pred. No. 7.1e-45;
  Matches  396;  Conservative    0;  Mismatches   98;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | |||| |||| |||||||||||||||    | | | |  |||||||||| | |||||||
Db          2 TATATAGATATGTATATATATATATATATGTATATATATGTATATATATATATATATATA 61

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db         62 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 121

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        122 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 181

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db        182 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 241

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db        242 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 301

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db        302 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 361

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db        362 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 421

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||| ||||  ||||||
Db        422 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 481

Qy        706 TATGTATATATATAT 720
              ||| ||||||||||||
Db        482 TATATATATATATAT 496


RESULT 37
```

```
CL623140
LOCUS       CL623140                744 bp    DNA     linear   GSS 01-JUL-2004
DEFINITION  OR_BBa0018D11.r OR_BBa Oryza rufipogon genomic clone OR_BBa0018D11
            3', genomic survey sequence.
ACCESSION   CL623140
VERSION     CL623140.1  GI:49121028
KEYWORDS    GSS.
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 744)
  AUTHORS   Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R.,
            Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
  TITLE     OMAP Project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Insert Length: 161   Std Error: 0.00
            Plate: 0018  row: D  column: 11
            Seq primer: CAC TCA TTA GGC ACC CCA
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .744
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
                     /clone="OR_BBa0018D11"
                     /tissue_type="young leaves"
                     /lab_host="DH10B-T1 phage resistant"
                     /clone_lib="OR_BBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

  Query Match             7.9%;  Score 324.2;  DB 9;  Length 744;
  Best Local Similarity   79.2%;  Pred. No. 1.5e-44;
  Matches  397;  Conservative    0;  Mismatches  103;  Indels    1;  Gaps    1;

Qy        217 AAAAAGGCATTTTATATATATATATATATATATACACACACACACACATATATATATA 276
              |||||      |  ||| ||| ||||||||||||||| | | | | | ||||||||||||
Db        183 AAAAAATACATACATACATACATATATATATATATATATATATATATATATATATATA 242

Qy        277 CACATATATATACACATATATACATATATACACATATATACACATATATATACACATACA 336
              | ||||||||| | ||||||| ||||||| | ||||||| | ||||||||| | ||| |
Db        243 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 302

Qy        337 TATGTACACATATATATACACATATGTATACACATATATACACATATATACACACATATA 396
              ||| || | ||||||||| | |||| |||| | ||||||| | ||||||| | | |||||
Db        303 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 362

Qy        397 TACACATATATACACACATATATACACATATATACACATATATACACATATACACATATA 456
              || | ||||||| | | ||||||| | ||||||| | ||||||| | ||||| | |||||
Db        363 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 422

Qy        457 TACACATATATACATATATACACATATATATAATATACACACATATATATACACATATAT 516
              || | ||||||| ||||||| | |||||||| ||||| | | ||||||||| | ||||||
Db        423 TATATATATATATATATATATATATATATAT-ATATATATATATATATATATATATAT 481

Qy        517 ACACACATATATACACATATATACACATATATATACACATATATACACATATATACATAT 576
              | | | ||||||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||
Db        482 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 541

Qy        577 ATACACATATATATACATATATACACATATATACATATATACACATATATACATATATAC 636
              ||| | ||||||||| ||||||| | ||||||| ||||||| | ||||||| |||||||
Db        542 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 601

Qy        637 ACACATATATACACATACATATACACACATAGATATACATATATATACACATATATAT 696
              | | ||||||| | ||| ||||| | | | ||| ||||| ||||||||| | ||||||||
Db        602 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 661

Qy        697 ACGTATATATATGTATATATA 717
              |  |||||||||  |||||||
Db        662 ATATATATATATAGATATATA 682


RESULT 38
CL776144
LOCUS       CL776144                587 bp    DNA     linear   GSS 30-JUL-2004
DEFINITION  OR_BBa0090B12.r OR_BBa Oryza rufipogon genomic clone OR_BBa0090B12
            3', genomic survey sequence.
ACCESSION   CL776144
VERSION     CL776144.1  GI:50850530
KEYWORDS    GSS.
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 587)
  AUTHORS   Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R.,
            Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
  TITLE     OMAP Project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Insert Length: 161   Std Error: 0.00
            Plate: 0090  row: B  column: 12
            Seq primer: CAC TCA TTA GGC ACC CCA
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .587
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
                     /clone="OR_BBa0090B12"
                     /tissue_type="young leaves"
                     /lab_host="DH10B-T1 phage resistant"
                     /clone_lib="OR_BBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

  Query Match             7.9%;  Score 323.8;  DB 9;  Length 587;
  Best Local Similarity   77.8%;  Pred. No. 1.8e-44;
  Matches  416;  Conservative    0;  Mismatches  117;  Indels    2;  Gaps    2;

Qy        188 ACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATAT 247
              |||   ||| |     ||| |     | | | | |       | ||||||||||||||||||
Db         23 ACACAGGCGCATATGCATATATATATATATATATATATATATATATATATATATATAT 82

Qy        248 ATATACACACACACACACATATATATACACATATATATACACATATATACATATATAC 307
              ||||| | | | | | | |||||||||| | ||||||||| | ||||||| |||||||
Db         83 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 142

Qy        308 ACATATATACACATATATATACACATACATATGTACACATATATATACACATATGTATAC 367
              | ||||||| | ||||||||| | ||| |||| || | ||||||||| | |||| ||||
Db        143 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 202

Qy        368 ACATATATACACATATATACACACATATATACACATATATACACACATATATACACATAT 427
              | ||||||| | ||||||| | | ||||||| | ||||||| | | ||||||| | ||||
```

```
Db        203 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 262

Qy        428 ATACACATATATACACATATACACATATATACACATATATACATATATACACATATATAT 487
              ||| | ||||||| | ||||| | ||||||| | ||||||| ||||||| | ||||||||
Db        263 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 322

Qy        488 A-ATATACACACATATATATACACATATATACACACATATATACACATATATACACATAT 546
              | ||||| | | ||||||||| | ||||||| | | ||||||| | ||||||| | ||||
Db        323 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 382

Qy        547 ATATACACATATATACACATATATACATATATACACATATATATACATATATACACATAT 606
              ||||| | ||||||| | ||||||| ||||||| | ||||||||| ||||||| | ||||
Db        383 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 442

Qy        607 ATACATATATACACATATATACATATATACACACATATATACACATACATATACACACAC 666
              ||| ||||||| | ||||||| ||||||| | | ||||||| | ||| ||||| | | |
Db        443 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 502

Qy        667 ATAGATATACATA-TATATACACATATATATACGTATATATATGTATATATATAT 720
              ||| ||||| ||| |||||| | ||||||||| ||||||||| ||||||||||
Db        503 ATATATATATATATTATATATATATATATATATATATATATATATATATATATATAT 557


RESULT 39
CL830207/c
LOCUS       CL830207                491 bp    DNA     linear   GSS 09-AUG-2004
DEFINITION  OR_CBa0052C06.f OR_CBa Oryza rufipogon genomic clone OR_CBa0052C06
            5', genomic survey sequence.
ACCESSION   CL830207
VERSION     CL830207.1  GI:51075808
KEYWORDS    GSS.
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 491)
  AUTHORS   Kim,H., Yu,Y., Wissotski,M., Yost,D., Stum,D., Rao,K., Luo,M.,
            Jetty,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and
            Wing,R.
  TITLE     OMAP project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0052  row: C  column: 06
            Seq primer: TAA TAC GAC TCA CTA TAG GG
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .491
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
                     /clone="OR_CBa0052C06"
                     /tissue_type="young leaves"
                     /dev_stage="2 week old seedlings"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OR_CBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII;
                     drk treated 36 hrs before harvest"
ORIGIN

  Query Match             7.8%;  Score 322.6;  DB 9;  Length 491;
  Best Local Similarity   80.2%;  Pred. No. 2.9e-44;
  Matches  392;  Conservative    0;  Mismatches   94;  Indels    3;  Gaps    1;

Qy        232 ATATATATATATATATATATACACACACACACATATATATATACACATATATATACAC 291
              ||||| |||||||||||||||| | | | | | |||||||||||| |  ||| |||||||
Db        491 ATATACATATATATATATATATATATATATATATATATATATATAGATCTATGTATACAC 432

Qy        292 ATATATACATATATACACATATATACACATATATATACACATACATATGTACACATATAT 351
              ||||||| ||||||| | ||| | | | |||| |||  | ||| ||| ||| ||||||||
Db        431 ATATATATATATATATATATACAGATAGATATGTATGTATATATATAGGTATACATATAT 372

Qy        352 ATACACATATGTATACACATATATACACATATATACACACATATATACACATATATACAC 411
              ||| |||||| |||| | ||||||| | ||||||  | | ||||||| | ||||||| |
Db        371 ATATACATATATATATATATATATATATATATCTATATATATATATAGATATATATAT 312

Qy        412 ACATATATACACATATATACACATATATACACATATACACATATATACACATATATACAT 471
              | ||||| | | ||||||| | ||||||| | ||||| | ||||||| | ||||||| ||
Db        311 ATATATAGAGATATATATATATATATATATATATATATATATATATATATATATATAT 252

Qy        472 ATATACACATATATATAATATACACACATATATATACACATATATACACACATATATACA 531
              ||||| | ||||||    ||||| | | ||||||||| | ||||||| | | ||||||| |
Db        251 ATATATATATATAT---ATATATATATATATATATATATATATATAGATATATATATATA 195

Qy        532 CATATATACACATATATATACACATATATACACATATATACATATATACACATATATATA 591
              ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||||||||
Db        194 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 135

Qy        592 CATATATACACATATATACATATATACACATATATACATATATACACACATATATACACA 651
              ||||||| | ||||||| ||||||| |||||||||||||||||||||| ||||||| | |
Db        134 TATATATATATATATATATATATATATACATATATACATATATACACATATATATATATA 75

Qy        652 TACATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTA 711
              || ||||| | | | ||| ||||| ||||||||| | ||||||||| |||||||||| ||
Db         74 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 15

Qy        712 TATATATAT 720
              |||||||||
Db         14 TATATATAT 6


RESULT 40
CC099198
LOCUS       CC099198                845 bp    DNA     linear   GSS 16-APR-2003
DEFINITION  CSU-K34.116B12.T7 CSU-K34 Aedes aegypti genomic clone
            CSU-K34.116B12, genomic survey sequence.
ACCESSION   CC099198
VERSION     CC099198.1  GI:29965572
KEYWORDS    GSS.
SOURCE      Aedes aegypti (yellow fever mosquito)
  ORGANISM  Aedes aegypti
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Aedes;
            Stegomyia.
REFERENCE   1  (bases 1 to 845)
  AUTHORS   Loftus,B., Shetty,J., Severson,D., Brown,S. and Knudson,D.
  TITLE     End sequencing of Aedes aegypti BACs
  JOURNAL   Unpublished (2003)
COMMENT     Other_GSSs: CSU-K34.116B12.SP6
            Contact: Brendan Loftus
            Department of Eukaryotic Genomics
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-3543
            Fax: 301-838-0208
            Email: enta@tigr.org
            Library was provided by Susan Brown and Dennis Knudson at Colorado
            State University.
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .845
                     /organism="Aedes aegypti"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7159"
                     /clone="CSU-K34.116B12"
                     /clone_lib="CSU-K34"
```

/note="Vector: pBACe3.6; Site_1: EcoRI; Source DNA: Aedes aegypti; strain unknown [derived from freshly hatched larvae at the Virus Research Centre, Poona, India. Reference: SINGH, K. R. P., 1967 Cell cultures derived from larvae of Aedes albopictus (Skuse) and Aedes aegypti (L.). Current Science 36: 506-508]; ATC-10 cell line ATCC CCL-125"

ORIGIN

Query Match 7.7%; Score 318.2; DB 8; Length 845;
Best Local Similarity 79.7%; Pred. No. 1.5e-43;
Matches 388; Conservative 0; Mismatches 98; Indels 1; Gaps 1;

```
Qy        211 AAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACATATA 270
              |||  || |||   | | |||||||||||||||||||||||| | | | | | | |||||
Db        360 AAACTAACAAATTAAATATATATATATATATATATATATATATATATATATATATATA 419

Qy        271 TATATACACATATATATACACATATATACATATATACACATATATACACATATATATACA 330
              |||||| | ||||||||| | ||||||| ||||||| | ||||||| | |||||||||| |
Db        420 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 479

Qy        331 CATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACACA 390
              ||| |||| || | ||||||||| | |||| |||| | ||||||| | ||||||| | |
Db        480 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 539

Qy        391 CATATATACACATATATACACACATATATACACATATATACACATATATACACATATACA 450
              ||||||| | ||||||| | | ||||||| | ||||||| | ||||||| | ||||| |
Db        540 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 599

Qy        451 CATATATACACATATATACATATATACACATATATATAATATACACACATATATATACAC 510
              ||||||| | ||||||| ||||||| | |||||||| |||| | | |||||||||| |
Db        600 TATATATATATATATATATATATATATATATATATAT-ATATATATATATATATATATAT 658

Qy        511 ATATATACACACATATATACACATATATACACATATATATACACATATATACACATATAT 570
              ||||||| | | ||||||| | ||||||| | ||||||||| | ||||||| | ||||||
Db        659 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 718

Qy        571 ACATATATACACATATATATACATATATACACATATATACATATATACACATATATACAT 630
              | ||||||| | ||||||||| ||||||| | ||||||| ||||||| | ||||||| ||
Db        719 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 778

Qy        631 ATATACACACATATATACACATACATATACACACACATAGATATACATATATATACACAT 690
              ||||| | | ||||||| | ||| ||||| | | | ||| ||||| ||||||||| | ||
Db        779 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 838

Qy        691 ATATATA 697
              |||||||
Db        839 ATATATA 845
```

RESULT 41
CC099198/c
LOCUS CC099198 845 bp DNA linear GSS 16-APR-2003
DEFINITION CSU-K34.116B12.T7 CSU-K34 Aedes aegypti genomic clone CSU-K34.116B12, genomic survey sequence.
ACCESSION CC099198
VERSION CC099198.1 GI:29965572
KEYWORDS GSS.
SOURCE Aedes aegypti (yellow fever mosquito)
ORGANISM Aedes aegypti
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Aedes; Stegomyia.
REFERENCE 1 (bases 1 to 845)
AUTHORS Loftus,B., Shetty,J., Severson,D., Brown,S. and Knudson,D.
TITLE End sequencing of Aedes aegypti BACs
JOURNAL Unpublished (2003)
COMMENT Other_GSSs: CSU-K34.116B12.SP6
Contact: Brendan Loftus
Department of Eukaryotic Genomics
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-3543
Fax: 301-838-0208
Email: enta@tigr.org
Library was provided by Susan Brown and Dennis Knudson at Colorado State University.
Seq primer: T7
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .845
/organism="Aedes aegypti"
/mol_type="genomic DNA"
/db_xref="taxon:7159"
/clone="CSU-K34.116B12"
/clone_lib="CSU-K34"
/note="Vector: pBACe3.6; Site_1: EcoRI; Source DNA: Aedes aegypti; strain unknown [derived from freshly hatched larvae at the Virus Research Centre, Poona, India. Reference: SINGH, K. R. P., 1967 Cell cultures derived from larvae of Aedes albopictus (Skuse) and Aedes aegypti (L.). Current Science 36: 506-508]; ATC-10 cell line ATCC CCL-125"

ORIGIN

Query Match 7.7%; Score 317; DB 8; Length 845;
Best Local Similarity 78.0%; Pred. No. 2.3e-43;
Matches 394; Conservative 0; Mismatches 110; Indels 1; Gaps 1;

```
Qy        229 TATATATATATATATATATATACACACACACACATATATATATACACATATATATA 288
              |||||||||||||||||||||||| | | | | | |||||||||| | |||||||||
Db        845 TATATATATATATATATATATATATATATATATATATATATATATATATATATATA 786

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |||
Db        785 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 726

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              |||||| | |||| |||| | ||||||| | ||||||| | | |||||||| | |||||||
Db        725 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 666

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              | | ||||||| | ||||||| | ||||||| | ||||| | |||||||| | |||||||
Db        665 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 606

Qy        469 CATATATACACATATATATAATATACACACATATATATACACATATATACACACATATAT 528
              ||||||| | |||||||| ||||| | | ||||||||| | ||||||| | | ||||||
Db        605 TATATATATATATATATAT-ATATATATATATATATATATATATATATATATATATAT 547

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACATATAT 588
              | | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | ||||||
Db        546 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 487

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
              ||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||||
Db        486 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 427

Qy        649 ACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATAT 708
              | ||| ||||| | | | ||| ||||| ||||||||| | ||||||||| | || |
Db        426 ATATATATATATATATATATATATATATATATATATATATATATATATATATTTAATTTG 367

Qy        709 GTATATATATATGCTCCAGAGTTCA 733
              || | | || | | || ||
Db        366 TTAGTTTGAAATCATTGACTGTGCA 342
```

RESULT 42
CL855754
LOCUS CL855754 816 bp DNA linear GSS 16-AUG-2004
DEFINITION OR_CBa0085L12.r OR_CBa Oryza rufipogon genomic clone OR_CBa0085L12 3', genomic survey sequence.
ACCESSION CL855754
VERSION CL855754.1 GI:51264405
KEYWORDS GSS.
SOURCE Oryza rufipogon
ORGANISM Oryza rufipogon

```
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 816)
  AUTHORS   Kim,H., Yu,Y., Wissotski,M., Yost,D., Stum,D., Rao,K., Luo,M.,
            Jetty,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and
            Wing,R.
  TITLE     OMAP project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0085  row: L  column: 12
            Seq primer: CAC TCA TTA GGC ACC CCA
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .816
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
                     /clone="OR_CBa0085L12"
                     /tissue_type="young leaves"
                     /dev_stage="2 week old seedlings"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OR_CBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII;
                     drk treated 36 hrs before harvest"
ORIGIN

  Query Match             7.6%;  Score 315;  DB 9;  Length 816;
  Best Local Similarity   79.2%;  Pred. No. 5.1e-43;
  Matches  411;  Conservative    0;  Mismatches  105;  Indels    3;  Gaps    3;

Qy        204 ATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATATACACACACACAC 263
              || | |  | | || | |    || | ||||||||||||||||||||||| | | | | |
Db        250 ATATAATCATATAATATATATATATCATATATATATATATATATATATATATATATATAT 309

Qy        264 ACATATATATATACACATATATATACACATATATACATATATACACATATATACACATAT 323
              | |||||||||||| | ||||||||| | ||||||| ||||||| | ||||||| | ||||
Db        310 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 369

Qy        324 ATATACACATACATATGTACACATATATATACACATATGTATACACATATATACACATAT 383
              ||||| | ||| |||| || | ||||||||| | |||| |||| | |||||||| | ||||
Db        370 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 429

Qy        384 ATACACACATATATACACATATATACACA-CATATATACACATATATACACATATATACA 442
              ||| | | |||||||| | |||||||| | | |||||||| | ||||||| | ||||||| |
Db        430 ATATATATATATATATATATATATATATATCATATATATATATATATATATATATATATA 489

Qy        443 CATATACACATATATACACATATATACATATATACACATATATATA-ATATACACACATA 501
              |||||| | ||||||| | ||||||| ||||||| | ||||||||| ||||| | | |||
Db        490 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 549

Qy        502 TATATACACATATATACACACATATATACACATATATACACATATATATACACATATATA 561
              |||||| | ||||||| | | ||||||| | ||||||| | ||||||||| | |||||||
Db        550 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 609

Qy        562 CACATATATACATATATACACATATATATACATATATACACATATATACATATATACACA 621
              | ||||||| ||||||| | ||||||||| ||||||| | ||||||| ||||||| | |
Db        610 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 669

Qy        622 TATATACATATATACACACATATATACACATACATATACACACACATAGATATACATATA 681
              |||||| ||||||| | | ||||||| | ||| ||||| | | | ||| ||||| |||||
Db        670 TATATATATATATATA-ATATATATATATATATATATATATATATATATATATATATATA 728

Qy        682 TATACACATATATATACGTATATATATGTATATATATAT 720
```

```
              |||| | |||||||||  ||||||||| |||||||| ||
Db        729 TATATATATATATATATATATATATATATATATATATAAAT 767


RESULT 43
BI506858
LOCUS       BI506858                570 bp    mRNA    linear   EST 08-APR-2002
DEFINITION  BB170024B20C02.5 Bee Brain Normalized/Subtracted Library, BB17 Apis
            mellifera cDNA clone BB170024B20C02 5', mRNA sequence.
ACCESSION   BI506858
VERSION     BI506858.1  GI:15357232
KEYWORDS    EST.
SOURCE      Apis mellifera (honey bee)
  ORGANISM  Apis mellifera
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Apoidea;
            Apidae; Apis.
REFERENCE   1  (bases 1 to 570)
  AUTHORS   Whitfield,C.W., Band,M.R., Bonaldo,M.F., Kumar,C.G., Liu,L.,
            Pardinas,J., Robertson,H.M., Soares,B. and Robinson,G.E.
  TITLE     Annotated expressed sequence tags and cDNA microarrays for studies
            of brain and behavior in the honey bee
  JOURNAL   Genome Res. 12 (4), 555-566 (2002)
  MEDLINE   21929762
   PUBMED   11932240
COMMENT     Contact: Gene E. Robinson
            Department of Entomology
            University of Illinois
            505 S. Goodwin Ave., Urbana, IL 61801, USA
            Tel: 217 265 0309
            Fax: 217 244 3499
            Email: generobi@life.uiuc.edu
            This research was funded by the University of Illinois Critical
            Research Initiatives Fund and a Burroughs-Wellcome Trust Innovation
            Award in Functional Genomics to G.E. Robinson and an NSF
            Postdoctoral Fellowship in Bioinformatics to C.W. Whitfield.
            REPEAT IN THE SEQUENCE
            Simple_repeat STRAND (+) ELEMENT (TA)n LOCATION [90,570].
            PCR PRimers
            FORWARD: TAATACGACTCACTATAGGG
            BACKWARD: ATTAACCCTCACTAAAG
            Plate: BB170024B20  row: C  column: 02
            Seq primer: AGCGGATAACAATTTCACACAGGA
            High quality sequence stop: 570.
FEATURES             Location/Qualifiers
     source          1. .570
                     /organism="Apis mellifera"
                     /mol_type="mRNA"
                     /strain="mixed strains of European bees, predominantly
                     A.m. ligustica"
                     /db_xref="taxon:7460"
                     /clone="BB170024B20C02"
                     /sex="female"
                     /tissue_type="brain"
                     /dev_stage="adult worker honey bee"
                     /lab_host="DH10B"
                     /clone_lib="Bee Brain Normalized/Subtracted Library, BB17"
                     /note="Organ: brain; Vector: pT7T3-Pac; Site_1: EcoR1;
                     Site_2: Not1; This BB17 cDNA library was generated by
                     subtraction of the BB16  library with 4000 previously
                     sequenced clones. The BB16 library was  contributed by the
                     Soares laboratory and it was constructed and  normalized
                     as described by Bonaldo, M.F., Lennon, G. and  Soares,
                     M.B. (1996), Genome Research 6(9): 791-806. RNA was
                     prepared  from dissected brains of adult worker bees of
                     various ages and  various behavioral groups. "
ORIGIN

  Query Match             7.6%;  Score 314.6;  DB 4;  Length 570;
  Best Local Similarity   77.7%;  Pred. No. 6.3e-43;
  Matches  405;  Conservative    0;  Mismatches  114;  Indels    2;  Gaps    2;

Qy        201 GTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACA 260
```

```
Db         48 GTCTCCACAGATAGAGAGGGGTGTAACTCAATTAAAATGATTTATATATATATATATATA 107

Qy        261 CACACATATATATATACACATATATATACACATATATACATATATACACATATATACACA 320
              | | |||||||||||| | ||||||||| | ||||||| ||||||| | ||||||| | |
Db        108 -ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 166

Qy        321 TATATATACACATACATATGTACACATATATATACACATATGTATACACATATATACACA 380
              |||||||| | ||| |||| || | ||||||||| | |||| |||| | ||||||| | |
Db        167 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 226

Qy        381 TATATACACACATATATACACATATATACACACATATATACACATATATACACATATATA 440
              |||||| | | ||||||| | ||||||| | | ||||||| | ||||||| | |||||||
Db        227 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 286

Qy        441 CACATATACACATATATACACATATATACATATATACACATATATATA-ATATACACACA 499
              | ||||| | ||||||| | ||||||| ||||||| | ||||||||| ||||| | | |
Db        287 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 346

Qy        500 TATATATACACATATATACACACATATATACACATATATACACATATATATACACATATA 559
              |||||||| | ||||||| | | ||||||| | ||||||| | ||||||||| | |||||
Db        347 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 406

Qy        560 TACACATATATACATATATACACATATATATACATATATACACATATATACATATATACA 619
              || | ||||||| ||||||| | ||||||||| ||||||| | ||||||| ||||||| |
Db        407 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 466

Qy        620 CATATATACATATATACACACATATATACACATACATATACACACACATAGATATACATA 679
              |||||| ||||||| | | ||||||| | ||| ||||| | | | ||| ||||| |||
Db        467 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 526

Qy        680 TATATACACATATATATACGTATATATATGTATATATATAT 720
              |||||| | |||||||||  ||||||||| ||||||||||
Db        527 TATATATATATATATATATATATATATATATATATATATAT 567


RESULT 44
BI506858/c
LOCUS       BI506858                570 bp    mRNA    linear   EST 08-APR-2002
DEFINITION  BB170024B20C02.5 Bee Brain Normalized/Subtracted Library, BB17 Apis
            mellifera cDNA clone BB170024B20C02 5', mRNA sequence.
ACCESSION   BI506858
VERSION     BI506858.1  GI:15357232
KEYWORDS    EST.
SOURCE      Apis mellifera (honey bee)
  ORGANISM  Apis mellifera
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Apoidea;
            Apidae; Apis.
REFERENCE   1  (bases 1 to 570)
  AUTHORS   Whitfield,C.W., Band,M.R., Bonaldo,M.F., Kumar,C.G., Liu,L.,
            Pardinas,J., Robertson,H.M., Soares,B. and Robinson,G.E.
  TITLE     Annotated expressed sequence tags and cDNA microarrays for studies
            of brain and behavior in the honey bee
  JOURNAL   Genome Res. 12 (4), 555-566 (2002)
  MEDLINE   21929762
   PUBMED   11932240
COMMENT     Contact: Gene E. Robinson
            Department of Entomology
            University of Illinois
            505 S. Goodwin Ave., Urbana, IL 61801, USA
            Tel: 217 265 0309
            Fax: 217 244 3499
            Email: generobi@life.uiuc.edu
            This research was funded by the University of Illinois Critical
            Research Initiatives Fund and a Burroughs-Wellcome Trust Innovation
            Award in Functional Genomics to G.E. Robinson and an NSF
            Postdoctoral Fellowship in Bioinformatics to C.W. Whitfield.
            REPEAT IN THE SEQUENCE
            Simple_repeat STRAND (+) ELEMENT (TA)n LOCATION [90,570].
            PCR PRimers
            FORWARD: TAATACGACTCACTATAGGG
            BACKWARD: ATTAACCCTCACTAAAG
            Plate: BB170024B20  row: C  column: 02
            Seq primer: AGCGGATAACAATTTCACACAGGA
            High quality sequence stop: 570.
FEATURES              Location/Qualifiers
     source           1. .570
                      /organism="Apis mellifera"
                      /mol_type="mRNA"
                      /strain="mixed strains of European bees, predominantly
                      A.m. ligustica"
                      /db_xref="taxon:7460"
                      /clone="BB170024B20C02"
                      /sex="female"
                      /tissue_type="brain"
                      /dev_stage="adult worker honey bee"
                      /lab_host="DH10B"
                      /clone_lib="Bee Brain Normalized/Subtracted Library, BB17"
                      /note="Organ: brain; Vector: pT7T3-Pac; Site_1: EcoR1;
                      Site_2: Not1; This BB17 cDNA library was generated by
                      subtraction of the BB16  library with 4000 previously
                      sequenced clones. The BB16 library was  contributed by the
                      Soares laboratory and it was constructed and  normalized
                      as described by Bonaldo, M.F., Lennon, G. and  Soares,
                      M.B. (1996), Genome Research 6(9): 791-806. RNA was
                      prepared  from dissected brains of adult worker bees of
                      various ages and  various behavioral groups. "
ORIGIN

  Query Match             7.6%;  Score 313.2;  DB 4;  Length 570;
  Best Local Similarity   80.5%;  Pred. No. 1.1e-42;
  Matches  391;  Conservative    0;  Mismatches   93;  Indels    2;  Gaps    2;

Qy        235 TATATATATATATATATACACACACACACATATATATATACACATATATATACACATA 294
              |||||||||||||||||| | | | | | |||||||||||| | ||||||||| | |||
Db        570 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 511

Qy        295 TATACATATATACACATATATACACATATATATACACATACATATGTACACATATATATA 354
              |||| ||||||| | ||||||| | |||||||||| | ||| |||| || | |||||||||
Db        510 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 451

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              | |||| |||| | ||||||| | ||||||| | | ||||||| | ||||||| | | |
Db        450 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 391

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              |||||| | ||||||| | ||||||| | ||||| | ||||||| | ||||||| |||||
Db        390 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 331

Qy        475 TACACATATATATAATATACACACATATATATACACATATATACACACATATATACACAT 534
              || | |||||||| ||||| | | ||||||||| | ||||||| | | ||||||| | ||
Db        330 TATATATATATAT-ATATATATATATATATATATATATATATATATATATATATATAT 272

Qy        535 ATATACACATATATATACACATATATACACATATATACATATATACACATATATATACAT 594
              ||||| | ||||||||| | ||||||| | ||||||| ||||||| | ||||||||| ||
Db        271 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 212

Qy        595 ATATACACATATATACATATATACACATATATACATATATACACACATATATACACATAC 654
              ||||| | ||||||| ||||||| | ||||||| ||||||| | | ||||||| | |||
Db        211 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 152

Qy        655 ATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATAT 714
              ||||| | | | ||| ||||| |||||||||| | |||||||||  |||||||||| |||||
Db        151 ATATATATATATATATATATATATATATATATATATATATATA-TTATATATATATATAT 93

Qy        715 ATATAT 720
              ||| ||
Db         92 ATAAAT 87


RESULT 45
CL855754/c
LOCUS       CL855754                816 bp    DNA     linear   GSS 16-AUG-2004
DEFINITION  OR_CBa0085L12.r OR_CBa Oryza rufipogon genomic clone OR_CBa0085L12
            3', genomic survey sequence.
```

```
ACCESSION   CL855754
VERSION     CL855754.1  GI:51264405
KEYWORDS    GSS.
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 816)
  AUTHORS   Kim,H., Yu,Y., Wissotski,M., Yost,D., Stum,D., Rao,K., Luo,M.,
            Jetty,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and
            Wing,R.
  TITLE     OMAP project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0085  row: L  column: 12
            Seq primer: CAC TCA TTA GGC ACC CCA
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .816
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
                     /clone="OR_CBa0085L12"
                     /tissue_type="young leaves"
                     /dev_stage="2 week old seedlings"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OR_CBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII;
                     drk treated 36 hrs before harvest"
ORIGIN

  Query Match             7.6%;  Score 311.4;  DB 9;  Length 816;
  Best Local Similarity   77.6%;  Pred. No. 2.1e-42;
  Matches  413;  Conservative    0;  Mismatches  116;  Indels    3;  Gaps    3;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||| |||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db        770 TATATTTATATATATATATATATATATATATATATATATATATATATATATATATATATA 711

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | || |||| | ||||||||| | ||| |||| || | |
Db        710 TATATATATATATATATATATATAT-TATATATATATATATATATATATATATATATATA 652

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        651 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 592

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db        591 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 532

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db        531 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 472

Qy        526 TATACACATAT-ATACACATATATATACACATATATACACATATATACATATATACACAT 584
              |||| | |||| ||| | ||||||||| | ||||||| | ||||||| ||||||| | ||
Db        471 TATATATATATGATATATATATATATATATATATATATATATATATATATATATATATAT 412

Qy        585 ATATATACATATATACACATATATACATATATACACATATATACATATATACACACATAT 644
              ||||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | ||||
Db        411 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 352

Qy        645 ATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATAT 704
              ||| | ||| ||||| | | | ||| ||||| |||||||||| | |||||||||  |||||
Db        351 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 292

Qy        705 ATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCA 756
              |||| ||||||||||||| | | | | || | | | | || ||
Db        291 ATATATATATATATATGATATATATATATTATATGATTATATATATATATCA 240


RESULT 46
CL858220/c
LOCUS       CL858220                546 bp    DNA     linear   GSS 16-AUG-2004
DEFINITION  OR_CBa0089A17.f OR_CBa Oryza rufipogon genomic clone OR_CBa0089A17
            5', genomic survey sequence.
ACCESSION   CL858220
VERSION     CL858220.1  GI:51269336
KEYWORDS    GSS.
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 546)
  AUTHORS   Kim,H., Yu,Y., Wissotski,M., Yost,D., Stum,D., Rao,K., Luo,M.,
            Jetty,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and
            Wing,R.
  TITLE     OMAP project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0089  row: A  column: 17
            Seq primer: TAA TAC GAC TCA CTA TAG GG
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .546
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
                     /clone="OR_CBa0089A17"
                     /tissue_type="young leaves"
                     /dev_stage="2 week old seedlings"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OR_CBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII;
                     drk treated 36 hrs before harvest"
ORIGIN

  Query Match             7.5%;  Score 310.6;  DB 9;  Length 546;
  Best Local Similarity   76.0%;  Pred. No. 3e-42;
  Matches  396;  Conservative    0;  Mismatches  124;  Indels    1;  Gaps    1;

Qy        242 ATATATATATACACACACACACACATATATATATACACATATATATACACATATATACAT 301
              |||||||||| | | | |   | |||||          |||||| |||||||| ||
Db        546 ATATATATATATATATATATCTATATATAATATATATATCATATATATACATATATATAT 487

Qy        302 ATATACACATATATACACATATATATACACATACATATGTACACATATATATACACATAT 361
              ||||| | ||||||| | |||||||||| | |||  ||| || | |||||||||| | ||||
Db        486 ATATATATATATATATATATATATATATATATATCTATATATATATATATATATATAT 427

Qy        362 GTATACACATATATACACATATATACACACATATATACACATATATACACACATATATAC 421
              |||| | ||||||| | |||||||||| | ||||| | | ||||||  | | |||||||
Db        426 ATATATATATATATATATATATATACATATATATACATATATATATCTATATATATAT 367

Qy        422 ACATATATACACATATATACACATATACACATATATACACATATATACATATATACACAT 481
              | ||||||| | ||||||| | ||||| | ||||||| | ||||||| ||||||| | ||
```

```
Db        366 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 307

Qy        482 ATATATAATATACACACATATATATACACATATATACACACATATATACACATATATACA 541
              |||||| ||||| | | |||||||||| | ||||||| | | ||||||| | |||||||
Db        306 ATATAT-ATATATATATATATATATATATATATATATATATATATATATATATATATATA 248

Qy        542 CATATATATACACATATATACACATATATACATATATACACATATATATACATATATACA 601
              ||||||||| | ||||||| | ||||||| ||||||| | |||||||| ||||||| |
Db        247 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 188

Qy        602 CATATATACATATATACACATATATACATATATACACACATATATACACATACATATACA 661
              ||||||| ||||||| | ||||||| ||||||| | | ||||||| | ||| ||||| |
Db        187 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 128

Qy        662 CACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATATG 721
              | | ||| ||||| ||||||||| | |||||||||   ||||||||| ||||||||||||
Db        127 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATG 68

Qy        722 CTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGG 762
                  |  ||  || |||  |  | |   |||||| ||||
Db         67 AGGTGAAACTCTAAAAAGGATGACGATCCATACCACAGGGG 27


RESULT 47
BX145980
LOCUS       BX145980                709 bp    DNA     linear   GSS 28-JAN-2003
DEFINITION  Danio rerio genomic clone DKEY-104E13, genomic survey sequence.
ACCESSION   BX145980
VERSION     BX145980.1  GI:27977335
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 709)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the T7 end of BAC 104E13. 104E13
            is part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .709
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-104E13"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             7.5%;  Score 308;  DB 9;  Length 709;
  Best Local Similarity   79.7%;  Pred. No. 7.9e-42;
  Matches  376;  Conservative    0;  Mismatches   95;  Indels    1;  Gaps    1;

Qy        250 ATACACACACACACACATATATATATACACATATATATACACATATATACATATATACAC 309
              | |  || |   | |||||||||||| | ||||||||| | ||||||| |||||||
Db        238 AAAGTCAAATTTATGCATATATATATATATATATATATATATATATATATATATATATAT 297

Qy        310 ATATATACACATATATATACACATACATATGTACACATATATATACACATATGTATACAC 369
              ||||||| | ||| | ||| | ||| |||| |||||||||||||| | |||| |||||||
Db        298 ATATATATATATACACATATATATATATATACACATATATATATATATATATATACAC 357

Qy        370 ATATATACACATATATACACACATATATACACATATATACACACATATATACACATATAT 429
              | ||||| | ||||||| | | ||||||| | ||||||| | | ||||||| | ||||||
Db        358 ACATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 417

Qy        430 ACACATATATACACATATACACATATATACACATATATACATATATACACATATATATA- 488
              | | ||||| | | ||||| | ||||||| | ||| ||| ||| ||| | |||||||||
```

```
Db        418 ATATATATACATATATATATATATATATATATATACATATATACATATATATATATAT 477

Qy        489 ATATACACACATATATATACACATATATACACACATATATACACATATATACACATATAT 548
              ||||||| |||||||||| | ||||||| | | ||||||| | ||||||| | ||||||
Db        478 ATATACATACATATATATATATATATATATATATATATATATATATATATATATATAT 537

Qy        549 ATACACATATATACACATATATACATATATACACATATATATACATATATACACATATAT 608
              ||| | ||||||| | ||||||| ||||||| ||| ||||||| ||||||| ||||||||
Db        538 ATATATATATATATATATATATATATATATATACACATATATATATATATATACATATAT 597

Qy        609 ACATATATACACATATATACATATATACACACATATATACACATACATATACACACACAT 668
              | ||||||| | ||||||| ||||||| | | ||||||| | ||| ||||| | | |  |
Db        598 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATCT 657

Qy        669 AGATATACATATATATACACATATATATACGTATATATATGTATATATATAT 720
              | ||||| |||||||||| | |||||||||  |||||||| ||||||||||
Db        658 ATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 709


RESULT 48
CR177142
LOCUS       CR177142                648 bp    DNA     linear   GSS 06-JUL-2004
DEFINITION  Reverse strand read from insert in 5'HPRT insertion targeting and
            chromosome engineering clone MHPN105k12, genomic survey sequence.
ACCESSION   CR177142
VERSION     CR177142.1  GI:49955991
KEYWORDS    GSS; genome survey sequence; MICER.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 648)
  AUTHORS   Adams,D.J., Biggs,P.J., Cox,A.V., Davies,R.M., van der Weyden,L.,
            Jonkers,J., Smith,J., Plumb,R.W., Taylor,R.G., Nishijima,I., Yu,Y.,
            Rogers,J. and Bradley,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-FEB-2004) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. http://www.sanger.ac.uk/MICER
FEATURES             Location/Qualifiers
     source          1. .648
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="MHPN105k12"
                     /clone_lib="MHPN"
ORIGIN

  Query Match             7.5%;  Score 307.2;  DB 9;  Length 648;
  Best Local Similarity   75.4%;  Pred. No. 1.1e-41;
  Matches  395;  Conservative    0;  Mismatches  128;  Indels    1;  Gaps    1;

Qy        197 AAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACA 256
              |||   ||   | |||||  |||   |||| |||||||||||| | ||||||| | | |
Db        107 AAACTACAAGCAATAAAAATCAAATAACATTATATATATATATACACATATATACATATA 166

Qy        257 CACACACACATATATATATACACATATATATACACATATATACATATATACACATATATA 316
              |||  | | | ||||| | | | | | ||||| ||||||| ||||||| ||| ||||| |
Db        167 TACATATATACATATACACATATACACATATATACATATACACATATACACATATATACA 226

Qy        317 CACATATATATACACATACATATGTACACATATATATACACATATGTATACACATATATA 376
              | | | |||||||||| | |  || ||| ||||| ||| |||   ||| ||| |||||
Db        227 TATACACATATACACATATACACATATACACATATACACATATACACATATACACATATA 286

Qy        377 CACATATATACACACATATATACACATATATACACACATATATACACATATATACACATA 436
              |||||||| ||| | | | ||| ||| ||||||||| ||| | | | | | ||| ||| |
Db        287 CACATATACACATATACACATATACACATATACACATATACACATATACACATATACACA 346

Qy        437 TATACACATATACACATATATACACATATATACATATATACACATATATATAATATACAC 496
              |||||||||||||||||||| ||| ||| | | ||| | | | | | |||| | | | |
Db        347 TATACACATATACACATATACACATATACACATATACACATATACACATAT-ACACATAT 405

Qy        497 ACATATATATACACATATATACACACATATATACACATATATACACATATATATACACAT 556
              ||| ||||| ||| ||| | | | ||| |||||||||||| ||| ||||| ||| | |
```

```
Db        406 ACACATATACACATATACACATATACACATATACACATATACACATATATACATATATAC 465

Qy        557 ATATACACATATATACATATATACACATATATATACATATATACACATATATACATATAT 616
              ||||| |||||||||||||||||||| | | | ||| ||||| | ||| ||| ||| ||
Db        466 ATATATACATATATACATATATACACACACACACACACATATATATATACATATATACAT 525

Qy        617 ACACATATATACATATATACACACATATATACACATACATATACACACACATAGATATAC 676
              | | ||||||||| | | |||||||||||||| | | | ||||||| | | ||| ||||||
Db        526 ATATATATATACACACACACACACATATATATATACATATATACATATATATATATATAC 585

Qy        677 ATATATATACACATATATATACGTATATATATGTATATATATAT 720
              | | | | |||||||||||||   ||||||||| ||||||||||
Db        586 ACACACACACACATATATATATATATATATATATATATATAT 629


RESULT 49
CL626353/c
LOCUS       CL626353                581 bp    DNA     linear   GSS 01-JUL-2004
DEFINITION  OR_BBa0022K14.f OR_BBa Oryza rufipogon genomic clone OR_BBa0022K14
            5', genomic survey sequence.
ACCESSION   CL626353
VERSION     CL626353.1  GI:49128388
KEYWORDS    GSS.
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 581)
  AUTHORS   Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R.,
            Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
  TITLE     OMAP Project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Insert Length: 161   Std Error: 0.00
            Plate: 0022  row: K  column: 14
            Seq primer: TAA TAC GAC TCA CTA TAG GG
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .581
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
                     /clone="OR_BBa0022K14"
                     /tissue_type="young leaves"
                     /lab_host="DH10B-T1 phage resistant"
                     /clone_lib="OR_BBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

  Query Match             7.4%;  Score 306.4;  DB 9;  Length 581;
  Best Local Similarity   77.4%;  Pred. No. 1.5e-41;
  Matches  384;  Conservative    0;  Mismatches  111;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | || |||||||||||| ||| | ||| |||   || ||| ||||| ||| ||| ||| |||
Db        574 TATACATATATATATACATACACATATACATTCATACATATATACATACACATATACATA 515

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              | |||||||||||||| ||||| | ||| | |||||||||||| |||||||||  |||| |
Db        514 CATACATATATACATACATACATACATACATACATATATATATACATACATACATACATA 455

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              ||||||| | | ||  ||||| | ||||| | ||||||||| ||||| ||||| |||||
Db        454 CATATATATATACATACATACATACATATATATATATATACATACATACATACATATATA 395

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | | ||| | ||||| ||| | ||||||| ||||| | ||||| ||| | |||||
Db        394 TATATATACATACATACATACATATATATATATATACATACATACATACATATATATATA 335

Qy        467 TACATATATACACATATATATAATATACACACATATATATACACATATATACACACATAT 526
              |||||| ||||| | |||||| ||||| | ||||| ||| | |||||||||| | | ||||
Db        334 TACATACATACATACATATAT-ATATATATACATACATACATACATATATATATATATAT 276

Qy        527 ATACACATATATACACATATATATACACATATATACACATATATACATATATACACATAT 586
              | | ||||| ||| | ||||||||| ||||| ||||| ||||||| ||||||| | || |
Db        275 ACATACATACATATATATATATATACATACATACATATATATATATATATATATGT 216

Qy        587 ATATACATATATACACATATATACATATATACACATATATACATATATACACACATATAT 646
              ||||| ||||||  | ||||| | | ||| | | || |||| ||||||| |   ||||||
Db        215 ATATATATATATGTATATATACATACATACATATATGTATATATATATATATGTATATAT 156

Qy        647 ACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATAT 706
              | | ||| ||||| | | ||| ||||| |||| |||| | ||||||||| |||||||
Db        155 ATATATATATATATGTATATATATATATATATATGTATATATATATATATATATATAT 96

Qy        707 ATGTATATATATATGC 722
              || |||||||  | ||
Db         95 ATATATATATGGACGC 80


RESULT 50
CA255084
LOCUS       CA255084                613 bp    mRNA    linear   EST 26-SEP-2003
DEFINITION  SCEPFL4175C11.g FL4 Saccharum officinarum cDNA clone SCEPFL4175C11
            5', mRNA sequence.
ACCESSION   CA255084
VERSION     CA255084.1  GI:35939612
KEYWORDS    EST.
SOURCE      Saccharum officinarum
  ORGANISM  Saccharum officinarum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Saccharum; Saccharum officinarum
            complex.
REFERENCE   1  (bases 1 to 613)
  AUTHORS   Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.
  TITLE     The libraries that made SUCEST
  JOURNAL   Genet. Mol. Biol. 24 (1-4), 1-7 (2001)
COMMENT     Contact: Arruda P
            Centro de Biologia Molecular e Engenharia Genetica
            Universidade Estadual de Campinas
            Caixa Postal 6010, 13083-970, Campinas SP, Brazil
            Tel: 55 19 3788 1137
            Fax: 55 19 3788 1089
            Email: parruda@unicamp.br
            Clone distribution: clone distribution information can be found
            through the Brazilian Clone Collection Center (BCCC) at
            http://www.bcccenter.fcav.unesp.br
            Plate: 175  row: C  column: 11
            Seq primer: T7 Promoter Primer.
FEATURES             Location/Qualifiers
     source          1. .613
                     /organism="Saccharum officinarum"
                     /mol_type="mRNA"
                     /db_xref="taxon:4547"
                     /clone="SCEPFL4175C11"
                     /lab_host="DH10B"
                     /clone_lib="FL4"
                     /note="Organ: Developed inflorescence and rachis
                     (20cm-long); Vector: pSport1; Site_1: SalI; Site_2: NotI;
                     An unidirectional cDNA library generated from [Developed
                     inflorescence  and rachis (20cm-long)]. cDNA was prepared
                     from polyA+ mRNA using  SuperScript Plasmid System Kit
                     (Invitrogen). The double-strand cDNAs were  fractionated
                     in a sepharose CL-2B 40cm-columns and fragments sizing
                     between 0.8 and 1.5 Kb were directionally cloned into the
```

vector. Details of each source of RNA and library construction can be obtained at http://sucest.lad.ic.unicamp.br/public"

ORIGIN

Query Match 7.4%; Score 306; DB 6; Length 613;
Best Local Similarity 76.9%; Pred. No. 1.7e-41;
Matches 386; Conservative 0; Mismatches 115; Indels 1; Gaps 1;

```
Qy        220 AAGGCATTTTATATATATATATATATATATACACACACACACACATATATATATACAC 279
              | |||  | | |||  |||| | ||| ||| |  |      | | | ||||||| ||  |
Db        109 ATGGCTCTCTGTATGCATATGTCTATCTATGTGGAGCGCTAGATAGATATATACATCTAT 168

Qy        280 ATATATATACACATATATACATATATACACATATATACACATATATATACACATACATAT 339
              ||||||||| | ||||||  ||||||| | ||||||| | ||||||||| | ||| || |
Db        169 ATATATATATATATATATCGATATATATATATATATATATATATATATATATATATGT 228

Qy        340 GTACACATATATATACACATATGTATACACATATATACACATATATACACACATATATAC 399
              || | |||||||||| | |||| |||| | ||||||| | ||||||| | | |||||||
Db        229 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 288

Qy        400 ACATATATACACACATATATACACATATATACACATATATACACATATACACATATATAC 459
              | ||||||| | | ||||||| | ||||||| | ||||||| | ||||| | |||||||
Db        289 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 348

Qy        460 ACATATATACATATATACACATATATATAATATACACACATATATATACACATATATACA 519
              | ||||||| ||||||| | |||||||| ||||| | | ||||||||| | ||||||| |
Db        349 ATATATATATATATATATATATATATAT-ATATATATATATATATATATATATATATATA 407

Qy        520 CACATATATACACATATATACACATATATATACACATATATACACATATATACATATATA 579
              | ||||||| | ||||||| | ||||||| | ||||||| | ||||||| |||||||
Db        408 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 467

Qy        580 CACATATATATACATATATACACATATATACATATATACACATATATACATATATACACA 639
              | ||||||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | |
Db        468 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 527

Qy        640 CATATATACACATACATATACACACACATAGATATACATATATATACACATATATATACG 699
              ||||||| | ||| ||||| | | | ||| ||||| |||||||||| | ||||| |||
Db        528 TATATATATATATATATATATATATATATATATATATATATATATATATATATATAGATATA 587

Qy        700 TATATATATGTATATATATATG 721
              ||||| ||| ||||||| |||
Db        588 TATATCTATATATATATCGATG 609
```

RESULT 51
CA255084/c

```
LOCUS       CA255084                613 bp    mRNA    linear   EST 26-SEP-2003
DEFINITION  SCEPFL4175C11.g FL4 Saccharum officinarum cDNA clone SCEPFL4175C11
            5', mRNA sequence.
ACCESSION   CA255084
VERSION     CA255084.1  GI:35939612
KEYWORDS    EST.
SOURCE      Saccharum officinarum
  ORGANISM  Saccharum officinarum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Saccharum; Saccharum officinarum
            complex.
REFERENCE   1  (bases 1 to 613)
  AUTHORS   Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.
  TITLE     The libraries that made SUCEST
  JOURNAL   Genet. Mol. Biol. 24 (1-4), 1-7 (2001)
COMMENT     Contact: Arruda P
            Centro de Biologia Molecular e Engenharia Genetica
            Universidade Estadual de Campinas
            Caixa Postal 6010, 13083-970, Campinas SP, Brazil
            Tel: 55 19 3788 1137
            Fax: 55 19 3788 1089
            Email: parruda@unicamp.br
            Clone distribution: clone distribution information can be found
            through the Brazilian Clone Collection Center (BCCC) at
            http://www.bcccenter.fcav.unesp.br
            Plate: 175  row: C  column: 11
            Seq primer: T7 Promoter Primer.
FEATURES             Location/Qualifiers
     source          1. .613
                     /organism="Saccharum officinarum"
                     /mol_type="mRNA"
                     /db_xref="taxon:4547"
                     /clone="SCEPFL4175C11"
                     /lab_host="DH10B"
                     /clone_lib="FL4"
                     /note="Organ: Developed inflorescence and rachis
                     (20cm-long); Vector: pSport1; Site_1: SalI; Site_2: NotI;
                     An unidirectional cDNA library generated from [Developed
                     inflorescence and rachis (20cm-long)]. cDNA was prepared
                     from polyA+ mRNA using SuperScript Plasmid System Kit
                     (Invitrogen). The double-strand cDNAs were fractionated
                     in a sepharose CL-2B 40cm-columns and fragments sizing
                     between 0.8 and 1.5 Kb were directionally cloned into the
                     vector. Details of each source of RNA and library
                     construction can be obtained at
                     http://sucest.lad.ic.unicamp.br/public"
ORIGIN
```

Query Match 7.4%; Score 304.4; DB 6; Length 613;
Best Local Similarity 77.3%; Pred. No. 3.2e-41;
Matches 382; Conservative 0; Mismatches 111; Indels 1; Gaps 1;

```
Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | |||||| |||||||||| |||||| | | | | | ||||||||||| | |||||||
Db        601 TATATATAGATATATATATCTATATATATATATATATATATATATATATATATATATA 542

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db        541 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 482

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        481 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 422

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db        421 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 362

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db        361 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 302

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db        301 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 242

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||||| ||||||| ||||||| | ||||||| ||||||| |    |||||
Db        241 TATATATATATATACATATATATATATATATATATATATATATATATATATCGATATA 182

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | |    ||| || || ||    | |||||| ||| || |||  |
Db        181 TATATATATATATATAGATGTATATATCTATCTAGCGCTCCACATAGATAGACATATGCA 122

Qy        706 TATGTATATATATA 719
              ||   | |   |||
Db        121 TACAGAGAGCCATA 108
```

RESULT 52
CL830207

```
LOCUS       CL830207                491 bp    DNA     linear   GSS 09-AUG-2004
DEFINITION  OR_CBa0052C06.f OR_CBa Oryza rufipogon genomic clone OR_CBa0052C06
            5', genomic survey sequence.
ACCESSION   CL830207
VERSION     CL830207.1  GI:51075808
KEYWORDS    GSS.
```

```
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 491)
  AUTHORS   Kim,H., Yu,Y., Wissotski,M., Yost,D., Stum,D., Rao,K., Luo,M.,
            Jetty,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and
            Wing,R.
  TITLE     OMAP project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0052  row: C  column: 06
            Seq primer: TAA TAC GAC TCA CTA TAG GG
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .491
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
                     /clone="OR_CBa0052C06"
                     /tissue_type="young leaves"
                     /dev_stage="2 week old seedlings"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OR_CBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII;
                     drk treated 36 hrs before harvest"
ORIGIN

  Query Match             7.3%;  Score 302;  DB 9;  Length 491;
  Best Local Similarity   77.3%;  Pred. No. 8.5e-41;
  Matches  379;  Conservative    0;  Mismatches  110;  Indels    1;  Gaps    1;

Qy        230 ATATATATATATATATATATATACACACACACACATATATATATACACATATATATAC 289
              ||||||||||||||||||||||| | | | | | |||||||||| | |||||||||
Db          2 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 61

Qy        290 ACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATAT 349
              | ||||||| ||||||| | ||||||    |||||| || | ||  |||| || | ||||
Db         62 ATATATATATATATATATATATATATATGTGTATATATGTATATATGTATATATATATATAT 121

Qy        350 ATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATAC 409
              ||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||||
Db        122 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 181

Qy        410 ACACATATATACACATATATACACATATATACACATATACACATATATACACATATATAC 469
              | | ||||||| | ||||||| | || |||| | ||||| | ||||||| | |||||||
Db        182 ATATATATATATATATATATATATATATCTATATATATATATATATATATATATATATAT 241

Qy        470 ATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATATAT 528
              ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | ||||||
Db        242 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 301

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACACATATAT 588
                  ||||||| | |||||| || | ||||||| | ||||||| ||||||| | ||||||
Db        302 CTCTATATATATATATATATCTATATATATATATAGATATATATATATATATATATATAT 361

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
              |||  |||||| | || ||||| |||||| | | ||| | || |||     | |||||||
Db        362 ATATGTATATATATATGTATACCTATATATATACATACATATCTATCTGTATATATATAT 421

Qy        649 ACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATAT 708
              | ||| ||||    | |||||||| || ||||||||| | |||||||||  |||||||||
Db        422 ATATATATATGTGTATACATAGATCTATATATATATATATATATATATATATATATATAT 481

Qy        709 GTATATATAT 718
              ||| |||||
Db        482 ATATGTATAT 491


RESULT 53
CL858220
LOCUS       CL858220                546 bp    DNA     linear   GSS 16-AUG-2004
DEFINITION  OR_CBa0089A17.f OR_CBa Oryza rufipogon genomic clone OR_CBa0089A17
            5', genomic survey sequence.
ACCESSION   CL858220
VERSION     CL858220.1  GI:51269336
KEYWORDS    GSS.
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 546)
  AUTHORS   Kim,H., Yu,Y., Wissotski,M., Yost,D., Stum,D., Rao,K., Luo,M.,
            Jetty,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and
            Wing,R.
  TITLE     OMAP project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0089  row: A  column: 17
            Seq primer: TAA TAC GAC TCA CTA TAG GG
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .546
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
                     /clone="OR_CBa0089A17"
                     /tissue_type="young leaves"
                     /dev_stage="2 week old seedlings"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OR_CBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII;
                     drk treated 36 hrs before harvest"
ORIGIN

  Query Match             7.3%;  Score 300.2;  DB 9;  Length 546;
  Best Local Similarity   77.5%;  Pred. No. 1.7e-40;
  Matches  376;  Conservative    0;  Mismatches  108;  Indels    1;  Gaps    1;

Qy        224 CATTTTATATATATATATATATATATACACACACACACATATATATATACACATAT 283
              ||  | |||||||||||||||||||||| | | | | | |||||||||| | ||||
Db         63 CACCTCATATATATATATATATATATATATATATATATATATATATATATATATATAT 122

Qy        284 ATATACACATATATACATATATACACATATATACACATATATATACACATACATATGTAC 343
              ||||| | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| ||
Db        123 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 182

Qy        344 ACATATATATACACATATGTATACACATATATACACATATATACACACATATATACACAT 403
              | ||||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | ||
Db        183 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 242

Qy        404 ATATACACACATATATACACATATATACACATATATACACATATACACATATATACACAT 463
              ||||| | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | ||
Db        243 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 302

Qy        464 ATATACATATATACACATATATATAATATACACACATATATATACACATATATACACACA 523
```

```
Db        303 ATATATATATATATATATATATAT-ATATATATATATATATATATATATATATATATA 361

Qy        524 TATATACACATATATACACATATATATACACATATATACACATATATACATATATACACA 583
              |||||| | ||||||| | |||||||||   |||||||   ||||||| ||||||| | |
Db        362 TATATATATATATATATATAGATATATATATGTATATATATGTATATATATATATATATATA 421

Qy        584 TATATATACATATATACACATATATACATATATACACATATATACATATATACACACATA 643
              |||||||| ||||||| | ||||||| ||| ||| | ||||||| ||||||| | | |||
Db        422 TATATATATATATATATATATATATATATAGATATATATATATATATATATATATATATA 481

Qy        644 TATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATA 703
              |||| | ||| |||| | | | ||            |||||| | |||||||||  ||||
Db        482 TATATATATATATATGTATATATATGATATATATATTTATATATAGATATATATATATATA 541

Qy        704 TATAT 708
              |||||
Db        542 TATAT 546


RESULT 54
BX180548/c
LOCUS       BX180548                592 bp    DNA     linear   GSS 28-JAN-2003
DEFINITION  Danio rerio genomic clone DKEY-180C22, genomic survey sequence.
ACCESSION   BX180548
VERSION     BX180548.1  GI:28012355
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 592)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the T7 end of BAC 180C22. 180C22
            is part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .592
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-180C22"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             7.2%;  Score 298.6;  DB 9;  Length 592;
  Best Local Similarity   78.0%;  Pred. No. 3.1e-40;
  Matches  372;  Conservative    0;  Mismatches  104;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | |||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db        590 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 531

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db        530 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 471

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | |||||||| | ||||||| | | ||||||| | |||||
Db        470 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 411

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db        410 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 351

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | |||||. | | ||||||| | ||||||| |
Db        350 TATATATATATATATATATATATATATATATATACATACACACACACATACACACACACA 291

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              | ||||||| | ||||| | | | |||||||||||| ||||| ||||||| ||||| |||
Db        290 CACACACATACACACACACACACACACACATATATATACATACATACATACATACATATA 231

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db        230 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 171

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTAT 702
              || | ||| ||||| | | | ||| |||||| | | |        || | | |  |||
Db        170 TATATATATATATATATATATATATATATACTTTTTTTCTGTGCTAAACAAATTTAT 114


RESULT 55
BX134913
LOCUS       BX134913                696 bp    DNA     linear   GSS 28-JAN-2003
DEFINITION  Danio rerio genomic clone DKEY-89F23, genomic survey sequence.
ACCESSION   BX134913
VERSION     BX134913.1  GI:27966206
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 696)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the T7 end of BAC 89F23. 89F23 is
            part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .696
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-89F23"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             7.2%;  Score 297.4;  DB 9;  Length 696;
  Best Local Similarity   79.8%;  Pred. No. 4.8e-40;
  Matches  363;  Conservative    0;  Mismatches   91;  Indels    1;  Gaps    1;

Qy        231 TATATATATATATATATATATACACACACACACATATATATATACACATATATATACA 290
              ||||||| |||||||||||||| | | | | | |||||||||||| | |||||||| |
Db        243 TATATATTTATATATATATATATATATATATATATATATATATATATATATATATA 302

Qy        291 CATATATACATATATACACATATATACACATATATATACACATACATATGTACACATATA 350
              ||||||| |||||||| | |||||||     |||||| | | ||| |||| || | |||||
Db        303 TATATATATATATATATATATATATAGTAGTATATAGATAGATATATATATATATATATA 362

Qy        351 TATACACATATGTATACACATATATACACATATATACACACATATATACACATATATACA 410
              |||| | |||||||| | ||||||| | ||||||| | | ||||||| | ||||||| |
Db        363 TATATATATATGTATATATATATATATATATATATATATATATATATATATATATATA 422

Qy        411 CACATATATACACATATATACACATATATACACATATACACATATATACACATATATACA 470
              | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | ||||||| |
Db        423 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 482

Qy        471 TATATACACATATATATAATATACACACATATATATACACATATATACACACATATATAC 530
              |||||| | |||||||| ||||| | | |||||||||  | ||||||| | | |||||||
Db        483 TATATATATATATATAT-ATATATATATATATATATATATATATATATATATATAT 541

Qy        531 ACATATATACACATATATATACACATATATACACATATATACATATATACACATATATAT 590
```

```
Db        542 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 601

Qy        591 ACATATATACACATATATACATATATACACATATATACATATATACACACATATATACAC 650
              | ||||||| | ||||||| | ||||||| | ||||||| ||||||| | ||||||||
Db        602 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 661

Qy        651 ATACATATACACACACATAGATATACATATATATA 685
              ||| ||||| | | | ||| ||||| |||||||||
Db        662 ATATATATATATATATATATATATATATATATATA 696


RESULT 56
BX163144
LOCUS       BX163144                700 bp    DNA     linear   GSS 13-MAR-2003
DEFINITION  Danio rerio genomic clone DKEY-154G8, genomic survey sequence.
ACCESSION   BX163144
VERSION     BX163144.1  GI:27994657
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 700)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAR-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the T7 end of BAC 154G8. 154G8 is
            part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .700
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-154G8"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             7.2%;  Score 297.4;  DB 9;  Length 700;
  Best Local Similarity   79.0%;  Pred. No. 4.8e-40;
  Matches  391;  Conservative    0;  Mismatches  101;  Indels    3;  Gaps    3;

Qy        227 TTTATATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||||||||||||||||||||||| |||||||| | |||||||
Db        198 TATATATATATATATATATATATATATACACACACACACAT-TATATATATATATATATA 256

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | | | | ||||| ||| ||||||||||  |||| |
Db        257 TATATATATATATATATATATATATACACACATACATACATACACACATACATACATACATA 316

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              ||| | ||||| |   | ||||| | ||||| | ||| | ||| | ||| | ||||| |
Db        317 CATACACACACACACACACACACACACATACATACATACATACATACATACATACA 376

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              |||| ||||| ||||| | ||| | ||||| ||||| | ||||| | ||| | ||||| |
Db        377 TACATACATACATACATACATACATACATACATACATACATACATACATACATACA 436

Qy        467 TACATATATACACATATATATAATATACACACATATATATACACATATATACACACATAT 526
              |||||| ||||| | ||| || | ||||| ||||| ||||| | ||||||| | | ||||
Db        437 TACATACATACATACATACAT-ACATACATACATACATATATATATATATATATATATAT 495

Qy        527 ATACACATATATACACATATATATACACATATATACACATATATACATATATACACATAT 586
              |||      ||||| | | ||||||||| | ||| | ||||||||| ||||||| | ||||
Db        496 ATAATATCATATATATACATATATACATACATACATACATATATATATATATATATATAT 555

Qy        587 ATATACATATATACACATATATACATATATACACATATATACATATATACACACATATAT 646
```

```
              ||||| ||||||| | ||||||| ||||| ||||| | | ||||| ||||| ||||| ||
Db        556 ATATATATATATATATATATATATATATATACACACACACACACATACATACATACATACAT 615

Qy        647 ACACATACATATACACAC-ACATAGATATACATATATATACACATATATATACGTATATA 705
              ||||| ||| | |||||| || ||||| |||||||||| | ||||||||  |||||||
Db        616 ACACACACACACACACGTGTTATATATATATATATATATATATATATATATATATATA 675

Qy        706 TATGTATATATATAT 720
              ||| ||||||||||||
Db        676 TATATATATATATAT 690


RESULT 57
BX180768/c
LOCUS       BX180768                701 bp    DNA     linear   GSS 28-JAN-2003
DEFINITION  Danio rerio genomic clone DKEY-176A10, genomic survey sequence.
ACCESSION   BX180768
VERSION     BX180768.1  GI:28012577
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 701)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the SP6 end of BAC 176A10. 176A10
            is part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .701
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-176A10"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             7.2%;  Score 297;  DB 9;  Length 701;
  Best Local Similarity   70.1%;  Pred. No. 5.6e-40;
  Matches  413;  Conservative    0;  Mismatches  175;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||||||||| | | | | | | |||||||||| | |||||||
Db        701 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 642

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db        641 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 582

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        581 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 522

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | |||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db        521 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 462

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db        461 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 402

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db        401 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 342

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
```

```
                |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db        341 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 282

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| |||||  | |  | |  | |    | |     || |  || ||      |
Db        281 TATATATATATATATTCTCGAAAACCTCTGATCAAACTCTTACCTTCATTTAAAACATTT 222

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              |  |  |   |||  |  |  |    | | | |  | || |   | | | |   ||| |
Db        221 TCATTTAAGCCATACACAGCTTAACATAAATGCGCCATCATTAAGTCATCGCCAAGGAAA 162

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGG 814
              |   | ||     | | |  |||     || |   ||    | ||||||
Db        161 GTAAATAAATACATATCAGCGCATACCTTTCTAGAAAACTGAGTTCAGG 113


RESULT 58
CL763005
LOCUS       CL763005                617 bp    DNA     linear   GSS 27-JUL-2004
DEFINITION  OR_BBa0133F16.f OR_BBa Oryza rufipogon genomic clone OR_BBa0133F16
            5', genomic survey sequence.
ACCESSION   CL763005
VERSION     CL763005.1  GI:50720952
KEYWORDS    GSS.
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 617)
  AUTHORS   Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R.,
            Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
  TITLE     OMAP Project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Insert Length: 161   Std Error: 0.00
            Plate: 0133  row: F  column: 16
            Seq primer: TAA TAC GAC TCA CTA TAG GG
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .617
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
                     /clone="OR_BBa0133F16"
                     /tissue_type="young leaves"
                     /lab_host="DH10B-T1 phage resistant"
                     /clone_lib="OR_BBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

  Query Match             7.2%;  Score 295.4;  DB 9;  Length 617;
  Best Local Similarity   76.3%;  Pred. No. 1.1e-39;
  Matches  376;  Conservative    0;  Mismatches  116;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | |||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db         85 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 144

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db        145 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 204

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
```

```
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        205 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 264

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db        265 TATATATATATATATATATATATATATATATATATATATATATATATATATATAAATATATA 324

Qy        467 TACATATATACACATATATATAATATACACACATATATATACACATATATACACACATAT 526
              || |||||||  | |||||||||  |||| | | ||||||||      ||| |      ||||
Db        325 TAAATATATATATATATATATTATATA-ATATATATATTATTTTATTTTTTTTAATAT 383

Qy        527 ATACACATATATACACATATATATACACATATATACACATATATACATATATACACATAT 586
              ||| | |||||||   |||||||| | ||||||| | ||||||| | | ||| |  | |
Db        384 ATATATATATATATTTATATATATATATATATATATATATATATAAAAAAATATATTTTT 443

Qy        587 ATATACATATATACACATATATACATATATACACATATATACATATATACACACATATAT 646
              ||||| ||| ||||| ||||||| | ||||| | | | ||| ||||| | | | | | |
Db        444 ATATATATAGATACATATATATAAAAATATAGAGAGAGATATATATAAATATAAAAAAAG 503

Qy        647 ACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATAT 706
              | | ||| ||||| | |  || | | ||| ||||||||||| ||| |||||  |||| |
Db        504 ATATATATATATATATATGCAGATACATAGATATATATACATATAGATATATATATAGTT 563

Qy        707 ATGTATATATATA 719
              |||||||||||||
Db        564 ATGTATATATATA 576


RESULT 59
BX134913/c
LOCUS       BX134913                696 bp    DNA     linear   GSS 28-JAN-2003
DEFINITION  Danio rerio genomic clone DKEY-89F23, genomic survey sequence.
ACCESSION   BX134913
VERSION     BX134913.1  GI:27966206
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 696)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the T7 end of BAC 89F23. 89F23 is
            part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .696
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-89F23"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             7.1%;  Score 294.4;  DB 9;  Length 696;
  Best Local Similarity   79.6%;  Pred. No. 1.5e-39;
  Matches  360;  Conservative    0;  Mismatches   91;  Indels    1;  Gaps    1;

Qy        269 TATATATACACATATATATACACATATATACATATATACACATATATACACATATATATA 328
              |||||||| | |||||||| | ||||||| ||||||| | ||||||| | |||||||||
Db        696 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 637

Qy        329 CACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACA 388
              | ||| |||| || | |||||||||| | |||| |||| | ||||||| | |||||||  |
Db        636 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 577

Qy        389 CACATATATACACATATATACACACATATATACACATATATACACATATATACACATATA 448
```

```
Db        576 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 517

Qy        449 CACATATATACACATATATACATATATACACATATATATAATATACACACATATATATAC 508
              | ||||||| | ||||||| ||||||| | |||||||| ||||| | | |||||||||
Db        516 TATATATATATATATATATATATATATATATATATATAT-ATATATATATATATATATAT 458

Qy        509 ACATATATACACACATATATACACATATATACACATATATATACACATATATACACATAT 568
              | ||||||| | | ||||||| | ||||||| | |||||||||| | ||||||| | ||||
Db        457 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 398

Qy        569 ATACATATATACACATATATATACATATATACACATATATACATATATACACATATATAC 628
              ||| ||||||| | |||||||||||||||| | ||||||| ||||||| | || |||
Db        397 ATATATATATATATATATATATATACATATATATATATATATATATATATATATATCTATCT 338

Qy        629 ATATATATACACACATATATACACATACATATACACACACATAGATATACATATATATACAC 688
              |||||      | ||||||| | ||| ||||| | | | ||| ||||| |||||||||
Db        337 ATATACTACTATATATATATATATATATATATATATATATATATATATATATATATATATAT 278

Qy        689 ATATATATACGTATATATATGTATATATATAT 720
              |||||||||  |||||||||| |||||| ||||
Db        277 ATATATATATATATATATATATATATAAATAT 246


RESULT 60
CF924712/c
LOCUS       CF924712                760 bp    mRNA    linear   EST 10-NOV-2003
DEFINITION  DD005E11.ab1 Codonopsis lanceolata cDNA library Codonopsis
            lanceolata cDNA, mRNA sequence.
ACCESSION   CF924712
VERSION     CF924712.1  GI:38243228
KEYWORDS    EST.
SOURCE      Codonopsis lanceolata
  ORGANISM  Codonopsis lanceolata
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            asterids; campanulids; Asterales; Campanulaceae; Codonopsis.
REFERENCE   1  (bases 1 to 760)
  AUTHORS   Yang,D.C., In,J.G. and Lee,B.S.
  TITLE     Codonopsis lanceolata
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Deok Chun Yang
            BioPia
            1 Seochonri, Kihung, Yongin, Kyungki-do, 449-701, Korea
            Tel: 82 31 201 2688
            Fax: 82 31 205 6323
            Email: dcyang@ibiopia.com.
FEATURES             Location/Qualifiers
     source          1. .760
                     /organism="Codonopsis lanceolata"
                     /mol_type="mRNA"
                     /db_xref="taxon:103999"
                     /clone_lib="Codonopsis lanceolata cDNA library"
                     /note="Organ: root; Vector: pTriplEx2; Site_1: Sfi IA;
                     Site_2: Sfi IB"
ORIGIN

  Query Match             7.1%;  Score 294.2;  DB 7;  Length 760;
  Best Local Similarity   79.8%;  Pred. No. 1.6e-39;
  Matches  347;  Conservative    0;  Mismatches   88;  Indels    0;  Gaps    0;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | | |||||||| |||||||||||||| | | | | | |||||||||||| | |||||||
Db        748 TATGTATATATAGATATATATATATATAGATATATATATATATATATATATATATATA 689

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | |||| || | ||||||||| | ||| |||| || | |
Db        688 TATATATATATATATATATATATATATATCTATATATATATATATATATATATATATATATA 629

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | ||||||
Db        628 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 569

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db        568 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 509

Qy        467 TACATATATACACATATATATAATATACACACATATATATACACATATATACACACATAT 526
              || ||||||| | |||||||| ||||| | | ||||||||| | ||||||| | | ||||
Db        508 TATATATATATATATATATATCATATATATATATATATATATATATATATATATATATAT 449

Qy        527 ATACACATATATACACATATATATACACATATATACACATATATACACATATATACACATAT 586
              ||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | ||||
Db        448 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 389

Qy        587 ATATACATATATACACATATATACATATATACACATATATACATATATACACACATATAT 646
              ||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| |  |||||||
Db        388 ATATATATATATATATATATATATATATATATATATATATATATATATATATATGCATATAT 329

Qy        647 ACACATACATATACA 661
              | | ||| ||||| |
Db        328 ATATATATATATATA 314


RESULT 61
BX197183
LOCUS       BX197183                771 bp    DNA     linear   GSS 28-JAN-2003
DEFINITION  Danio rerio genomic clone DKEY-207A10, genomic survey sequence.
ACCESSION   BX197183
VERSION     BX197183.1  GI:28029069
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 771)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the SP6 end of BAC 207A10. 207A10
            is part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .771
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-207A10"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             7.1%;  Score 293.6;  DB 9;  Length 771;
  Best Local Similarity   77.1%;  Pred. No. 2.1e-39;
  Matches  370;  Conservative    0;  Mismatches  109;  Indels    1;  Gaps    1;

Qy        168 ATTGCACTCCAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATT 227
              ||| | || |  ||  |  || | | ||  |  | | |  |||||||||     |
Db        292 ATTCAATTCAAAAGCATAAATAAATTGGGGAAAAAAAAATAAATAAAAAAAAAAAAATAT 351

Qy        228 TTATATATATATATATATATATACACACACACACATATATATATACACATATATAT 287
              |||||||||||||||||||||||| | | | | | |||||||||||| | ||||||||
Db        352 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 411

Qy        288 ACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACAT 347
              | | ||||||| ||||||| | ||||||| ||||||||| | ||| |||| || | ||
Db        412 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 471

Qy        348 ATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATAT 407
              ||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | ||||||
Db        472 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 531
```

```
Qy        408 ACACACATATATACACATATATACACATATATACACATATACACATATATACACATATAT 467
              | | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | ||||||
Db        532 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 591

Qy        468 ACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATAT 526
              | ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | ||||
Db        592 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 651

Qy        527 ATACACATATATACACATATATATACACATATATACACATATATACATATATACACACAT 586
              ||| | ||||||| | ||||||||| | |||||||| | ||||||| ||||||| | ||||
Db        652 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 711

Qy        587 ATATACATATATACACATATATACATATATACACATATATACATATATACACACATATAT 646
              ||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | ||||||||
Db        712 ATATATATATATATATATATATATATATATATATATATATATATATATATATATACATATAT 771
```

RESULT 62
BZ215061

```
LOCUS       BZ215061                643 bp    DNA     linear   GSS 11-OCT-2002
DEFINITION  CH230-280D4.TV CHORI-230 Segment 2 Rattus norvegicus genomic clone
            CH230-280D4, genomic survey sequence.
ACCESSION   BZ215061
VERSION     BZ215061.1  GI:23873314
KEYWORDS    GSS.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 643)
  AUTHORS   Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K.,
            Shvartsbeyn,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D.,
            Riggs,F., de Jong,P. and Fraser,C.M.
  TITLE     Rat BAC End Sequences from Library CHORI-230 MboI segment
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: CH230-280D4.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the rat BAC library CHORI-230
            (http://www.chori.org/bacpac/rat230.htm). For BAC library
            availability, please contact Pieter de Jong (pdejong@mail.cho.org).
            Clones may be purchased from BACPAC Resources
            (http://www.chori.org/bacpac/or ering_information.htm). BAC end
            page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
            Plate: 280  row: D  column: 4
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .643
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /strain="BN/SsNHsd/MCW"
                     /db_xref="taxon:10116"
                     /clone="CH230-280D4"
                     /sex="Female"
                     /cell_type="Brain"
                     /clone_lib="CHORI-230 Segment 2"
                     /note="Vector: pTARBAC1.3; Site_1: MboI; Site_2: MboI;
                     CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
                     Pieter de Jong"
ORIGIN
```

```
  Query Match             7.1%;  Score 293;  DB 8;  Length 643;
  Best Local Similarity   75.7%;  Pred. No. 2.7e-39;
  Matches  376;  Conservative    0;  Mismatches  120;  Indels    1;  Gaps    1;

Qy        229 TATATATATATATATATATATATACACACACACACATATATATATACACATATATATA 288
              || | | | | ||| | | |||||||||| |||||||||| ||||||||| ||| | |
Db        118 TACAAACACACATACACACATATACACACATACACACATATACATATACACACATACACA 177

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              |||||||| ||||| ||||||| ||||| ||| | |||||||||||||| || ||||   |
Db        178 CACATATACACATACATACACACATATAAACACACATATACACATACACATATACATGCA 237

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              | ||||||| || || ||||| ||| ||    |||||||||||||||| ||| | ||| |
Db        238 CACATACACACATATACACACACATACACTTCCATATACACACATATACACAGACATACA 297

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              |||| ||| | | ||||| | ||| ||| | | ||||| |||||||||| ||||| ||| |
Db        298 CACATATACACAGACATACACACATATACACAGACATACACACATATAGACACACATACA 357

Qy        469 CATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATATA 527
              | ||  || ||| | ||||| | | ||| ||| | |||||||||| ||||||||||||||||
Db        358 CTTACTTATACACACATATACACAGACATACACACATATACACAGATATACACACATATA 417

Qy        528 TACACATATATACACATATATATACACATATATACACATATATACATATATACACATATA 587
              ||| | ||| |||||||||| | ||||||| |    ||||||| ||| |||||||||| | |
Db        418 CACAGACATACACACATATAGACACACATACACTTACATATACACACATATACACAGACA 477

Qy        588 TATACATATATACACATATATACATATATACACATATATACATATATACACACATATATA 647
              || ||| |||||||||| ||||||| ||| ||||| ||| ||| |||||||||| | | | |
Db        478 TACACACATATACACACATATACACATACACACACATACACACATATACACATACACACA 537

Qy        648 CACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATA 707
              || | ||| ||| ||||| | | ||| ||||| | | | ||| | ||| ||  ||| | |
Db        538 CATACACACATATACACATACATATACACATACACACAAACACACATACACACATACACA 597

Qy        708 TGTATATATATATGCTC 724
              ||   | | |  | |
Db        598 CATACGAACACACACAC 614
```

RESULT 63
BM906116

```
LOCUS       BM906116                1156 bp    mRNA    linear   EST 12-MAR-2002
DEFINITION  AGENCOURT_6621231 NIH_MGC_125 Homo sapiens cDNA clone IMAGE:5589966
            5', mRNA sequence.
ACCESSION   BM906116
VERSION     BM906116.1  GI:19356495
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1156)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Invitrogen
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12363  row: i  column: 07
            High quality sequence start: 323
            High quality sequence stop: 613.
FEATURES             Location/Qualifiers
     source          1. .1156
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5589966"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_125"
                     /note="Organ: ovary (pool of 3); Vector: pCMV-SPORT6;
```

Site_1: EcoRV (destroyed); Site_2: NotI; RNA source pool of three ovaries, from females ranging in age from 38 to 49 yo. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 2.1 kb, insert size range 1-3.5 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 036."

ORIGIN

Query Match 7.1%; Score 292.6; DB 5; Length 1156;
Best Local Similarity 76.5%; Pred. No. 2.9e-39;
Matches 385; Conservative 0; Mismatches 114; Indels 4; Gaps 2;

```
Qy        210 GAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACACATAT 269
              || | | |   |||    |  | || | |||||| |||||||| | | | ||| | ||||
Db        223 GACATAGATGGAGGAGGATCCTCTAGACATATATGTATATATAAATAGATACATATATAT 282

Qy        270 ATATATACACATATATATACACATATATACATATATACACATATATACACATATATATAC 329
              | |||||   | ||  ||| | ||||||||||||||| | ||||||||||||||||||
Db        283 AGATATATCTACATGCATATATATATATACATATATAGATATATACACATATATATAT 342

Qy        330 ACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACAC 389
              | ||| |||| || | ||||||||| | |||| |||| | ||||||| | ||||||| |
Db        343 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 402

Qy        390 ACATATATACACATATATACACACATATATACACATATATACACATATATACACATATAC 449
              | ||||||| | ||||||| | | ||||||| | ||||||| | ||||||| | |||||
Db        403 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 462

Qy        450 ACATATATACACATATATACATATATACACATATATATA-ATATACACACATATATATAC 508
              | ||||||| | ||||||| ||||||| | ||||||||| ||||| | | |||||||||
Db        463 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 522

Qy        509 ACATATATACACACATATATACACATATATACACATATATATACACATATATACACATAT 568
              | ||||||| | | ||||||| | ||||||| | ||||||||| | ||||||| | ||||
Db        523 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 582

Qy        569 ATACATATATACACATATATATACATATATACACATATATACATATATACACATATATAC 628
              ||| ||||||| | ||||||||| ||||||| | ||||||| ||||||| ||||||||||
Db        583 ATATATATATATATATATATATATATATATATATATATATATATATATATACATATATAC 642

Qy        629 ATATATACACACATATATACACATACATATACACACACATAGATATACATATATATACAC 688
              ||||||| | | ||||||| | ||| ||||| | | | ||| |||| |    | ||||
Db        643 ATATATATATATATATATATATATATATATATATATATATACATATGGA---CTGTACAT 699

Qy        689 ATATATATACGTATATATATGTA 711
              ||||||| ||  | ||   ||||
Db        700 ATATATACACCCACATGACTGTA 722
```

RESULT 64
CL907900/c
LOCUS CL907900 714 bp DNA linear GSS 14-SEP-2004
DEFINITION OA_ABa0007A21.r OA_ABa Oryza australiensis genomic clone OA_ABa0007A21 3', genomic survey sequence.
ACCESSION CL907900
VERSION CL907900.1 GI:52016779
KEYWORDS GSS.
SOURCE Oryza australiensis
ORGANISM Oryza australiensis
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 714)
AUTHORS Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
TITLE OMAP Project
JOURNAL Unpublished (2004)
COMMENT Contact: Rod A. Wing
Arizona Genomics Institute
University of Arizona
Forbes Building Room 303, Tucson, AZ 85721-0036, USA
Tel: 520 626 9595
Fax: 520 621 1259
Email: http://genome.arizona.edu
PCR PRimers
FORWARD: TAA TAC GAC TCA CTA TAG GG
BACKWARD: CAC TCA TTA GGC ACC CCA
Plate: 0007 row: A column: 21
Seq primer: CAC TCA TTA GGC ACC CCA
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .714
/organism="Oryza australiensis"
/mol_type="genomic DNA"
/db_xref="taxon:4532"
/clone="OA_ABa0007A21"
/tissue_type="young leaves"
/lab_host="DH10B T1 phage resistant"
/clone_lib="OA_ABa"
/note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

Query Match 7.0%; Score 290.2; DB 9; Length 714;
Best Local Similarity 77.7%; Pred. No. 7.8e-39;
Matches 363; Conservative 0; Mismatches 103; Indels 1; Gaps 1;

```
Qy        286 ATACACATATATACATATATACACATATATACACATATATATACACATACATATGTACAC 345
              ||| | ||||||| ||||||| | ||||||  | |||| |||| | ||| |||| || |
Db        714 ATATATATATATATATATATATATATATTTATATTTATATATATATATATATATAT 655

Qy        346 ATATATATACACATATGTATACACATATATACACATATATACACACATATATACACATAT 405
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | ||||
Db        654 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 595

Qy        406 ATACACACATATATACACATATATACACATATATACACATATACACATATATACACATAT 465
              ||| | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | ||||
Db        594 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 535

Qy        466 ATACATATATACACATATATATAATATACACACATATATATACACATATATACACACATA 525
              ||| |||||||| | |||||||| ||||| | | ||||||||| | ||||||| | | |||
Db        534 ATATATATATATATATATATAT-ATATATATATATATATATATATATATATATATATA 476

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | |||||||| | ||||||| | ||||||| ||||||| | |||
Db        475 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 416

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db        415 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 356

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||||||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db        355 TATATATACATATATATATATATATATATATATATATATATATATATATATATATATA 296

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATT 752
              ||| |||||||| | || |  | || |  | |   | ||  |  |||
Db        295 TATATATATATACACGCGCGCGTGTGCGCACGTTATTGCGTTGAATT 249
```

RESULT 65
BX180768
LOCUS BX180768 701 bp DNA linear GSS 28-JAN-2003
DEFINITION Danio rerio genomic clone DKEY-176A10, genomic survey sequence.
ACCESSION BX180768
VERSION BX180768.1 GI:28012577
KEYWORDS GSS.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 701)
AUTHORS Humphray,S.J., Huckle,E. and Durham,J.L.
TITLE Direct Submission

```
  JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the SP6 end of BAC 176A10. 176A10
            is part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .701
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-176A10"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             7.0%;  Score 290;  DB 9;  Length 701;
  Best Local Similarity   76.1%;  Pred. No. 8.5e-39;
  Matches  370;  Conservative    0;  Mismatches  115;  Indels    1;  Gaps    1;

Qy        213 AAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACACATATATA 272
              ||| |||| |  ||  ||  |  ||    | |    | |      |    | |||||||
Db        216 AAATGAAAATGTTTTAAATGAAGGTAAGAGTTTGATCAGAGGTTTTCGAGAATATATATA 275

Qy        273 TATACACATATATATACACATATATACATATATACACATATATACACATATATATACACA 332
              |||| | |||||||| | ||||||| ||||||| | ||||||| | |||||||| | |
Db        276 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 335

Qy        333 TACATATGTACACATATATATACACATATGTATACACATATATACACATATATACACACA 392
              || |||| || | |||||||| | |||| |||| | ||||||| | |||||||| | | |
Db        336 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 395

Qy        393 TATATACACATATATACACACATATATACACATATATACACATATATACACATATACACA 452
              |||||| | ||||||| | | ||||||| | ||||||| | ||||||| | ||||| | |
Db        396 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 455

Qy        453 TATATACACATATATACATATATACACATATATATA-ATATACACACATATATATACACA 511
              |||||| | ||||||| ||||||| | ||||||||| ||||| | | |||||||| | |
Db        456 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 515

Qy        512 TATATACACACATATATACACATATATACACATATATATACACATATATACACATATATA 571
              |||||| | | ||||||| | ||||||| | ||||||||| | ||||||| | |||||||
Db        516 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 575

Qy        572 CATATATACACATATATATACATATATACACATATATACATATATACACATATATACATA 631
               ||||||| | ||||||||| ||||||| | ||||||| ||||||| | ||||||| |||
Db        576 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 635

Qy        632 TATACACACATATATACACATACATATACACACACATAGATATACATATATATACACATA 691
              |||| | | ||||||| | ||| ||||| | | | ||| ||||| |||||||||| | |||
Db        636 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 695

Qy        692 TATATA 697
              ||||||
Db        696 TATATA 701


RESULT 66
CL907900
LOCUS       CL907900                 714 bp    DNA     linear   GSS 14-SEP-2004
DEFINITION  OA_ABa0007A21.r OA_ABa Oryza australiensis genomic clone
            OA_ABa0007A21 3', genomic survey sequence.
ACCESSION   CL907900
VERSION     CL907900.1  GI:52016779
KEYWORDS    GSS.
SOURCE      Oryza australiensis
  ORGANISM  Oryza australiensis
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 714)
  AUTHORS   Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R.,
            Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
  TITLE     OMAP Project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0007  row: A  column: 21
            Seq primer: CAC TCA TTA GGC ACC CCA
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .714
                     /organism="Oryza australiensis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4532"
                     /clone="OA_ABa0007A21"
                     /tissue_type="young leaves"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OA_ABa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

  Query Match             7.0%;  Score 289.4;  DB 9;  Length 714;
  Best Local Similarity   79.4%;  Pred. No. 1.1e-38;
  Matches  355;  Conservative    0;  Mismatches   91;  Indels    1;  Gaps    1;

Qy        251 TACACACACACACACATATATATATACACATATATATACACATATATACATATATACACA 310
              | | ||||| | | | | |||||||| | ||||||||| | ||||||| ||||||| | |
Db        266 TGCGCACACGCGCGCGTGTATATATATATATATATATATATATATATATATATATATATA 325

Qy        311 TATATACACATATATATACACATACATATGTACACATATATATACACATATGTATACACA 370
              |||||| | ||||||||| |  || |||| || | ||||||||| | |||| |||| | |
Db        326 TATATATATATATATATATATGTATATATATATATATATATATATATATATATATATATA 385

Qy        371 TATATACACATATATACACACATATATACACATATATACACACATATATACACATATATA 430
              |||||| | ||||||| | | ||||||| | ||||||| | | ||||||| | |||||||
Db        386 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 445

Qy        431 CACATATATACACATATACACATATATACACATATATACATATATACACATATATATAAT 490
              | ||||||| | ||||| | ||||||| | ||||||| ||||||| | |||||||| ||
Db        446 TATATATATATATATATATATATATATATATATATATATATATATATATATATATAT-AT 504

Qy        491 ATACACACATATATATACACATATATACACACATATATACACATATATACACATATATAT 550
              ||| | | ||||||||| | ||||||| | | ||||||| | ||||||| | ||||||||
Db        505 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 564

Qy        551 ACACATATATACACATATATACATATATACACATATATATACATATATACACATATATAC 610
              | | ||||||| | ||||||| ||||||| | ||||||||| ||||||| | |||||||
Db        565 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 624

Qy        611 ATATATACACATATATACATATATACACACATATATACACATACATATACACACACATAG 670
              ||||||| | ||||||| ||||||| | | ||||||| | ||| ||||| | | | |||
Db        625 ATATATATATATATATATATATATATATATATATATATATATATATATATATATAAATATATAA 684

Qy        671 ATATACATATATATACACATATATATA 697
              ||||| ||||||||| | |||||||||
Db        685 ATATATATATATATATATATATATATA 711


RESULT 67
DR42I22S
LOCUS       DR42I22S                 673 bp    DNA     linear   GSS 22-NOV-2002
DEFINITION  Danio rerio genomic clone DKEY-42I22, genomic survey sequence.
ACCESSION   AL983929
VERSION     AL983929.1  GI:25183962
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
```

```
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 673)
  AUTHORS   Humphray,S.J., Huckle,E. and Hunt,S.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-NOV-2002) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail contact:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the SP6 end of BAC 42I22. 42I22 is
            part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene.
            Further details: http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .673
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-42I22"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             7.0%;  Score 288.8;  DB 9;  Length 673;
  Best Local Similarity   80.0%;  Pred. No. 1.4e-38;
  Matches  352;  Conservative    0;  Mismatches  87;  Indels    1;  Gaps    1;

Qy        281 TATATATACACATATATACATATATACACATATATACACATATATATACACATACATATG 340
              | | | || | ||||||| ||||||| | ||||||| | ||||||||| | ||| ||||
Db        234 TGTGTGTATATATATATATATATATATATATATATATATATATATATATATATATATA 293

Qy        341 TACACATATATATACACATATGTATACACATATATACACATATATACACACATATATACA 400
              || | ||||||||| | |||| ||||||| | | ||||| | | ||||| ||| | |||
Db        294 TATATATATATATATATATATATATATACACACACACACACACACACACATATACACACA 353

Qy        401 CATATATACACACATATATACACATATATACACATATATACACATATACACATATATACA 460
              |||||||| | | ||||||| | ||||||| | ||||||| | ||||| | |||||||| |
Db        354 CATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 413

Qy        461 CATATATACATATATACACATATATATA-ATATACACACATATATATACACATATATACA 519
              ||||||| ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| |
Db        414 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 473

Qy        520 CACATATATACACATATATACACATATATATACACATATATACACATATATACATATATA 579
              | ||||||| | ||||||| | ||||||||| | ||||||| | ||||||| |||||||
Db        474 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 533

Qy        580 CACATATATATACATATATACACATATATACATATATACACATATATACATATATACACA 639
              | ||||||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | |
Db        534 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 593

Qy        640 CATATATACACATACATATACACACACATAGATATACATATATATACACATATATATACG 699
              |||||| | ||| ||||| | | | ||| ||||| |||||||||| | ||||||||||
Db        594 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 653

Qy        700 TATATATATGTATATATATA 719
              ||||||||| ||||||||||
Db        654 TATATATATATATATATA 673


RESULT 68
CW007564
LOCUS       CW007564               569 bp    DNA     linear   GSS 23-SEP-2004
DEFINITION  ZMMBLa0009D17.r ZMMBLa Zea mays genomic clone ZMMBLa0009D17 3',
            genomic survey sequence.
ACCESSION   CW007564
VERSION     CW007564.1  GI:52588327
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 569)
  AUTHORS   Wing,R., Luo,M., Soderlund,C. and Haller,K.
  TITLE     ZMMBL sequences
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            Plate: 0009  row: D  column: 17
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .569
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /cultivar="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBLa0009D17"
                     /tissue_type="immature ears"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="ZMMBLa"
                     /note="Vector: pAGIBAC1; Site_1: SalI; Site_2: SalI"
ORIGIN

  Query Match             7.0%;  Score 288.4;  DB 9;  Length 569;
  Best Local Similarity   74.5%;  Pred. No. 1.6e-38;
  Matches  389;  Conservative    0;  Mismatches  131;  Indels    2;  Gaps    2;

Qy        200 TGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACAC 259
              ||| |  || ||||| ||  | |||  ||   |||||  |    |  |    ||
Db         48 TGTAAAGTCTGAAAACAACCACGGCCATTGTAATATAATTTCCCAACTGCTCACGATGCT 107

Qy        260 ACA-CACATATATATATACACATATATATACACATATATACATATATACACATATATACA 318
              ||| | || |  || | |  |   |  |||| ||||||| ||||||| | ||||||| |
Db        108 ACATCTCAAAACTACAAAACCTCGTTGCTACAAATATATATATATATATATATATATATA 167

Qy        319 CATATATATACACATACATATGTACACATATATATACACATATGTATACACATATATACA 378
              |||||||| | ||| |||| || | |||||||| | |||| |||| | ||||||| |
Db        168 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 227

Qy        379 CATATATACACACATATATACACATATATACACACATATATACACATATATACACATATA 438
              ||||||| | | ||||||| ||||||| | | ||||||| | ||||||| | |||||
Db        228 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 287

Qy        439 TACACATATACACATATATACACATATATACATATATACACATATATATA-ATATACACA 497
              || | ||||| | ||||||| ||||||| ||||||| ||||||||| ||||| | |
Db        288 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 347

Qy        498 CATATATATACACATATATACACACATATATACACATATATACACATATATATACACATA 557
              ||||||||| | ||||||| | | ||||||| | ||||||| | ||||||||| | |||
Db        348 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 407

Qy        558 TATACACATATATACATATATACACATATATATACATATATACACATATATACATATATA 617
              |||| | ||||||| ||||||| | ||||||||| ||||||| | ||||||| |||||||
Db        408 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 467

Qy        618 CACATATATACATATATACACACATATATACACATACATATACACACACATAGATATACA 677
              | ||||||| ||||||| | | ||||||| | ||| ||||| | | | ||| ||||| |
Db        468 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 527

Qy        678 TATATATACACATATATATACGTATATATATGTATATATATA 719
              |||||||| | |||||||||  |||| |||| |||| | |||
Db        528 TATATATATATATATATATATATATAAATATATATAGAAATA 569


RESULT 69
DR42I22S/c
LOCUS       DR42I22S               673 bp    DNA     linear   GSS 22-NOV-2002
DEFINITION  Danio rerio genomic clone DKEY-42I22, genomic survey sequence.
ACCESSION   AL983929
```

```
VERSION     AL983929.1  GI:25183962
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 673)
  AUTHORS   Humphray,S.J., Huckle,E. and Hunt,S.E.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-NOV-2002) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail contact:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the SP6 end of BAC 42I22. 42I22 is
            part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene.
            Further details: http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .673
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-42I22"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             7.0%;  Score 288;  DB 9;  Length 673;
  Best Local Similarity   75.4%;  Pred. No. 1.9e-38;
  Matches  371;  Conservative    0;  Mismatches  120;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||||| | | | | | | |||||||||||| | |||||||
Db        673 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 614

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db        613 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 554

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db        553 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 494

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db        493 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 434

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db        433 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 374

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | || | | |  | || | |     | | |  | | |  | |||
Db        373 TATATATATATATATATATGTGTGTGTATATGTGTGTGTGTGTGTGTGTGTGTATATA 314

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db        313 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 254

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| |||||||||||||| | | | ||| | | | |||| | | | || | | |
Db        253 TATATATATATATACACACACACACACACACACACACACACACACACACACACACACACA 194

Qy        706 TATGTATATATA 717
                |  || ||||
Db        193 GTTTCATTTATA 182


RESULT 70
CL917654
LOCUS       CL917654                720 bp    DNA     linear   GSS 14-SEP-2004
DEFINITION  OA_ABa0020N13.r OA_ABa Oryza australiensis genomic clone
            OA_ABa0020N13 3', genomic survey sequence.
ACCESSION   CL917654
VERSION     CL917654.1  GI:52029479
KEYWORDS    GSS.
SOURCE      Oryza australiensis
  ORGANISM  Oryza australiensis
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 720)
  AUTHORS   Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R.,
            Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
  TITLE     OMAP Project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0020  row: N  column: 13
            Seq primer: CAC TCA TTA GGC ACC CCA
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .720
                     /organism="Oryza australiensis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4532"
                     /clone="OA_ABa0020N13"
                     /tissue_type="young leaves"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OA_ABa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

  Query Match             7.0%;  Score 288;  DB 9;  Length 720;
  Best Local Similarity   75.2%;  Pred. No. 1.8e-38;
  Matches  385;  Conservative    0;  Mismatches  125;  Indels    2;  Gaps    2;

Qy        196 GAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACAC 255
              || || ||||   || |||||   ||    | | ||   || |   ||    ||| |
Db        200 GACATATCATAGAAGGAAAAAGTTAAATTTTGTAATTAGTAAAACAATTGGAATACATTT 259

Qy        256 ACACACACACATATATATATACACATATATATACACATATATACATATATACACATATAT 315
              ||         |||||||||| | ||||||||| | ||||||| ||||||| | ||||||
Db        260 ACTGTGTGGAATATATATATATATATATATATATATATATATATATATATATATATATAT 319

Qy        316 ACACATATATATACACATACATATGTACACATATATATACACATATGTATACACATATAT 375
              | | |||||||||| | ||| |||| || | ||||||||| | |||| |||| | ||||||
Db        320 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 379

Qy        376 ACACATATATACACACATATATACACATATATACACACATATATACACATATATACACAT 435
              | | ||||||| | | ||||||| | ||||||| | | ||||||| | ||||||| | ||
Db        380 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 439

Qy        436 ATATACACATATACACATATATACACATATATACATATATACACATATATATAATATACA 495
              ||||| | ||||| | ||||||| | ||||||| ||||||| | ||||||||| ||||| |
Db        440 ATATATATATATATATATATATATATATATATATATATATATATATATATATATAT-ATATATA 498

Qy        496 CACATATATATACACATATATACACACATATATACACATATATACACATATATATACACA 555
              | ||||||||| | ||||||| | | ||||||| | ||||||| | ||||||||| | |
Db        499 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 558

Qy        556 TATATACACATATATACATATATACACATATATATACATATATACACATATATACATATA 615
              |||||| | ||||||| ||||||| | ||||||||| ||||||| | ||||||| |||||
Db        559 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 618

Qy        616 TACACATATATACATATATACACACATATATACACATACATATACACACACATAGATATA 675
              || | ||||||| |||||||    | ||||||| | ||| ||||| | |   ||| |||||
Db        619 TATATATATATATATATATATTTATATATATATATATATATATATATATTTATATATATA 678
```

```
Qy        676 CATATATATACACATATATATACGTATATATA 707
              |||| ||||   |||||||| |||||||||
Db        679 TATATGTATA-GTATATATATATGTATATATA 709


RESULT 71
BX157670
LOCUS       BX157670                668 bp    DNA     linear   GSS 28-JAN-2003
DEFINITION  Danio rerio genomic clone DKEY-125M16, genomic survey sequence.
ACCESSION   BX157670
VERSION     BX157670.1  GI:27989269
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 668)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the T7 end of BAC 125M16. 125M16
            is part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .668
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-125M16"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             7.0%;  Score 287.4;  DB 9;  Length 668;
  Best Local Similarity   77.5%;  Pred. No. 2.3e-38;
  Matches  373;  Conservative    0;  Mismatches  106;  Indels    2;  Gaps    2;

Qy        242 ATATATATATACACACACACACACATATATATATACACATATATATACACATATATACAT 301
              | | | ||  ||| | | | | | ||| ||| ||| |  ||||| || |  |||||| ||
Db        185 AAAAAAATGAACATAAATAAAGAAATAAATAAATAAATTTATATCTATATCTATATATAT 244

Qy        302 ATATACACATATATACACATATATATACACATACATATGTACACA-TATATATACACATA 360
              ||||| | |||||||| | ||||||||| | ||| |||| || ||| |||||| | ||| |
Db        245 ATATATATATATATATATATATATATATATATATATATATACAGTATATACATACACA 304

Qy        361 TGTATACACATATATACACATATATACACACATATATACACATATATACACACATATATA 420
              || | ||||||||| ||||| ||| | | ||||||| ||||| ||||| ||||| |||
Db        305 CATACATACATATATATACATACATATATATATATATATACATACATACATACATACATA 364

Qy        421 CACATATATACACATATATACACATATACACATATATACACATATATACATATATACACA 480
              || | ||| | ||||| ||||| | ||||| | ||| | ||||| ||||||| ||||| |
Db        365 CATACATACATACATACATACATACATACATACATACATACATACATACATACATACATA 424

Qy        481 TATATATA-ATATACACACATATATATACACATATATACACACATATATACACATATATA 539
              |||||||| |||||  | | |||||||||  | |||||||  | |  |||||||  |||||||
Db        425 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 484

Qy        540 CACATATATATACACATATATACACATATATACATATATACACATATATATACATATATA 599
              | |||||||||  | |||||||  |||||||  |||||||  | |  |||||||  |||||||
Db        485 TATATATATATATATATATATATATATATATATATATATATATATAGATATATATATATATA 544

Qy        600 CACATATATACATATATACACATATATACATATATACACACATATATACACATACATATA 659
              | |||||||  |||||||  | |||||||  |||||||  | |  |||||||  | |||  |||||
Db        545 TATATATATATATATATATAGATATATATATATATATATATATATATATATATATATATA 604

Qy        660 CACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATA 719
              | | | |||  |||||  |||||||||  | |||||||||   |||||  ||  ||||||||||
Db        605 TATATATATATATATATATATATATATATATATATATATATATATATATGTAGATATATATATA 664
```

```
Qy        720 T 720
              |
Db        665 T 665


RESULT 72
BX149919
LOCUS       BX149919                737 bp    DNA     linear   GSS 13-MAR-2003
DEFINITION  Danio rerio genomic clone DKEY-112P5, genomic survey sequence.
ACCESSION   BX149919
VERSION     BX149919.1  GI:27981385
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 737)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAR-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the T7 end of BAC 112P5. 112P5 is
            part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .737
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-112P5"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             7.0%;  Score 287.4;  DB 9;  Length 737;
  Best Local Similarity   76.5%;  Pred. No. 2.3e-38;
  Matches  365;  Conservative    0;  Mismatches  111;  Indels    1;  Gaps    1;

Qy        242 ATATATATATACACACACACACACATATATATATACACATATATATACACATATATACAT 301
              | || | ||||||       ||  | | | | | |      | | | |    | |  || ||
Db        262 AGATTTCTATACAGTTTGGCAACCTTTTCTGTTTTTTGGTTTTTTTTTTAAAAGCTATAT 321

Qy        302 ATATACACATATATACACATATATATACACATACATATGTACACATATATATACACATAT 361
              ||||| | |||||||| | ||||||||| | ||| |||| || | ||||||||| | ||||
Db        322 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 381

Qy        362 GTATACACATATATACACATATATACACACATATATACACATATATACACACATATATAC 421
              ||||  | |||||||  | |||||||  | |  |||||||  | |||||||  | |  |||||||
Db        382 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 441

Qy        422 ACATATATACACATATATACACATATACACATATATACACATATATACATATATACACAT 481
              | |||||||  | |||||||  | |||||  | |||||||  | |||||||  |||||||  | ||
Db        442 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 501

Qy        482 ATATATAATATACACACATATATATACACATATATACACACATATATACACATATATACA 541
              ||||||  |||||||| |||||||||  | |||||||  | |  |||||||  |||||||  |
Db        502 ATATAT-ATATACACATATATATATATATATATATATATATATATATATATATATATATA 560

Qy        542 CATATATATACACATATATACACATATATACATATATACACATATATATACATATATACA 601
              |||||||||  | |||||||  |||||||  |||||||  |||||||||  |||||||  |
Db        561 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 620

Qy        602 CATATATACATATATACACATATATACATATATACACACATATATACACATACATATACA 661
              |||||||  |||||||  | |||||||  |||||||  | |  |||||||  | |||  |||||  |
Db        621 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 680

Qy        662 CACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATAT 718
              | | |||  |||||||  |||||||  | |||||||||  |||||||||  |||||||||
Db        681 TATATATATATATACACATATATATATATATATATATATATATATATATATATATATAT 737
```

```
RESULT 73
BX197183/c
LOCUS       BX197183                771 bp    DNA     linear   GSS 28-JAN-2003
DEFINITION  Danio rerio genomic clone DKEY-207A10, genomic survey sequence.
ACCESSION   BX197183
VERSION     BX197183.1  GI:28029069
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 771)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the SP6 end of BAC 207A10. 207A10
            is part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .771
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-207A10"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             7.0%;  Score 286.6;  DB 9;  Length 771;
  Best Local Similarity   78.1%;  Pred. No. 3.1e-38;
  Matches  357;  Conservative    0;  Mismatches   99;  Indels    1;  Gaps    1;

Qy        270 ATATATACACATATATATACACATATATACATATATACACATATATACACATATATATAC 329
              ||||||  | |||||||||  | |||||||  |||||||  | |||||||  | |||||||||
Db        771 ATATATGTATATATATATATATATATATATATATATATATATATATATATATATATATAT 712

Qy        330 ACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACAC 389
              | |||  ||||  ||  |  |||||||||  |  ||||  ||||  |  |||||||  |  |||||||  |
Db        711 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 652

Qy        390 ACATATATACACATATATACACACATATATACACATATATACACATATATACACATATAC 449
              | |||||||  |  |||||||  |  |  |||||||  |  |||||||  |  |||||||  |  |||||
Db        651 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 592

Qy        450 ACATATATACACATATATACATATATACACATATATATAATATACACACATATATATACA 509
              | |||||||  |  |||||||  |||||||  |  |||||||  |||||  |  |  |||||||||  |
Db        591 ATATATATATATATATATATATATATATATATATATAT-ATATATATATATATATATATA 533

Qy        510 CATATATACACACATATATACACATATATACACATATATATACACATATATACACATATA 569
                |||||||  |  |  |||||||  |  |||||||  |  |||||||||  |  |||||||  |  |||||
Db        532 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 473

Qy        570 TACATATATACACATATATATACATATATACACATATATACATATATACACATATATACA 629
              ||  |||||||  |  |||||||||  |||||||  |  |||||||  |||||||  |  |||||||  |
Db        472 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 413

Qy        630 TATATACACACATATATACACATACATATACACACACATAGATATACATATATATACACA 689
              ||||||  |  |||||||  |  |||  |||||  |  |  |||  |||||  |||||||||  |  |
Db        412 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 353

Qy        690 TATATATACGTATATATATGTATATATATATGCTCCA 726
              |||||  |   |  |  |||  |||  |  |  |  |  |  |
Db        352 TATATTTTTTTTTTTTTATTTATTTTTTTTTCCCCAA 316


RESULT 74
CW007564/c
LOCUS       CW007564                569 bp    DNA     linear   GSS 23-SEP-2004
DEFINITION  ZMMBLa0009D17.r ZMMBLa Zea mays genomic clone ZMMBLa0009D17 3',
            genomic survey sequence.
ACCESSION   CW007564
VERSION     CW007564.1  GI:52588327
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 569)
  AUTHORS   Wing,R., Luo,M., Soderlund,C. and Haller,K.
  TITLE     ZMMBL sequences
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            Plate: 0009  row: D  column: 17
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .569
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /cultivar="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBLa0009D17"
                     /tissue_type="immature ears"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="ZMMBLa"
                     /note="Vector: pAGIBAC1; Site_1: SalI; Site_2: SalI"
ORIGIN

  Query Match             6.9%;  Score 286;  DB 9;  Length 569;
  Best Local Similarity   77.3%;  Pred. No. 4.1e-38;
  Matches  360;  Conservative    0;  Mismatches  105;  Indels    1;  Gaps    1;

Qy        281 TATATATACACATATATACATATATACACATATATACACATATATATACACATACATATG 340
              |||  |  ||  |  ||||  ||  |||||||  |  |||||||  |  |||||||||  |  |||  ||||
Db        569 TATTTCTATATATATTTATATATATATATATATATATATATATATATATATATATATATA 510

Qy        341 TACACATATATATACACATATGTATACACATATATACACATATATACACACATATATACA 400
              ||  |  |||||||||  |  ||||  ||||  |  |||||||  |  |||||||  |  |  |||||||  |
Db        509 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 450

Qy        401 CATATATACACACATATATACACATATATACACATATATACACATATACACATATATACA 460
              |||||||  |  |  |||||||  |  |||||||  |  |||||||  |  |||||  |  |||||||  |
Db        449 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 390

Qy        461 CATATATACACATATATACACATATATATA-ATATACACACATATATATACACATATATACA 519
              |||||||  |||||||  |  |||||||||  |||||  |  |  |||||||||  |  |||||||  |
Db        389 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 330

Qy        520 CACATATATACACATATATACACATATATATACACATATATACACATATATACATATATA 579
              |  |||||||  |  |||||||  |  |||||||||  |  |||||||  |  |||||||  |||||||
Db        329 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 270

Qy        580 CACATATATATACATATATACACATATATACATATATACACATATATACATATATACACA 639
              |  |||||||||  |||||||  |  |||||||  |||||||  |  |||||||  |||||||  |  |
Db        269 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 210

Qy        640 CATATATACACATACATATACACACACATAGATATACATATATATACACATATATATACG 699
              |||||||  |  |||  |||||  |  |  |||  |||||  |||||||||  |  |||||||||
Db        209 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 150

Qy        700 TATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCA 745
              |||||||||  |  ||    |  |  |    |  |  |||  ||||||
Db        149 TATATATATTTGTAGCAACGAGGTTTTGTAGTTTTGAGATGTAGCA 104
```

```
RESULT 75
CL917654/c
LOCUS       CL917654                720 bp    DNA     linear   GSS 14-SEP-2004
DEFINITION  OA_ABa0020N13.r OA_ABa Oryza australiensis genomic clone
            OA_ABa0020N13 3', genomic survey sequence.
ACCESSION   CL917654
VERSION     CL917654.1  GI:52029479
KEYWORDS    GSS.
SOURCE      Oryza australiensis
  ORGANISM  Oryza australiensis
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 720)
  AUTHORS   Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R.,
            Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
  TITLE     OMAP Project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0020  row: N  column: 13
            Seq primer: CAC TCA TTA GGC ACC CCA
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .720
                     /organism="Oryza australiensis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4532"
                     /clone="OA_ABa0020N13"
                     /tissue_type="young leaves"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OA_ABa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

  Query Match             6.9%;  Score 286;  DB 9;  Length 720;
  Best Local Similarity   79.7%;  Pred. No. 4e-38;
  Matches  362;  Conservative    0;  Mismatches   90;  Indels    2;  Gaps    2;

Qy        283 TATATACACATATATACATA-TATACACATATATACACATATATATACACATACATATGT 341
              |||||| |||||||| ||| |||||| ||||||| | ||| ||||| | ||| |||| |
Db        709 TATATATACATATATATATACTATACATATATATATATATAAATATATATATATATATAT 650

Qy        342 ACACATATATATACACATATGTATACACATATATACACATATATACACACATATATACAC 401
              | | ||||| ||| | |||| |||| | ||||||| | ||||||| | | ||||||| |
Db        649 ATATATATAAATATATATATATATATATATATATATATATATATATATATATATATAT 590

Qy        402 ATATATACACACATATATACACATATATACACATATATACACATATACACATATATACAC 461
              ||||||| | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| |
Db        589 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 530

Qy        462 ATATATACATATATACACATATATATA-ATATACACACATATATATACACATATATACAC 520
              ||||||| ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| |
Db        529 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 470

Qy        521 ACATATATACACATATATACACATATATATACACATATATACACATATATACATATATAC 580
              | ||||||| | ||||||| | ||||||||| | ||||||| | ||||||| |||||||
Db        469 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 410

Qy        581 ACATATATATACATATATACACATATATACATATATACACATATATACATATATACACAC 640
              | ||||||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | |
Db        409 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 350

Qy        641 ATATATACACATACATATACACACACATAGATATACATATATATACACATATATATACGT 700
              ||||||| | ||| ||||| | | | ||| ||||| |||||||| | |||||||||  |
Db        349 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 290

Qy        701 ATATATATGTATATATATATGCTCCAGAGTTCAT 734
              |||||||| |||||||||||  || |||  ||
Db        289 ATATATATATATATATATATTCCACACAGTAAAT 256


RESULT 76
CF924712
LOCUS       CF924712                760 bp    mRNA    linear   EST 10-NOV-2003
DEFINITION  DD005E11.ab1 Codonopsis lanceolata cDNA library Codonopsis
            lanceolata cDNA, mRNA sequence.
ACCESSION   CF924712
VERSION     CF924712.1  GI:38243228
KEYWORDS    EST.
SOURCE      Codonopsis lanceolata
  ORGANISM  Codonopsis lanceolata
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            asterids; campanulids; Asterales; Campanulaceae; Codonopsis.
REFERENCE   1  (bases 1 to 760)
  AUTHORS   Yang,D.C., In,J.G. and Lee,B.S.
  TITLE     Codonopsis lanceolata
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Deok Chun Yang
            BioPia
            1 Seochonri, Kihung, Yongin, Kyungki-do, 449-701, Korea
            Tel: 82 31 201 2688
            Fax: 82 31 205 6323
            Email: dcyang@ibiopia.com.
FEATURES             Location/Qualifiers
     source          1. .760
                     /organism="Codonopsis lanceolata"
                     /mol_type="mRNA"
                     /db_xref="taxon:103999"
                     /clone_lib="Codonopsis lanceolata cDNA library"
                     /note="Organ: root; Vector: pTriplEx2; Site_1: Sfi IA;
                     Site_2: Sfi IB"
ORIGIN

  Query Match             6.9%;  Score 285.8;  DB 7;  Length 760;
  Best Local Similarity   80.0%;  Pred. No. 4.3e-38;
  Matches  360;  Conservative    0;  Mismatches   88;  Indels    2;  Gaps    2;

Qy        268 ATATATATACACATATATATACACATATATACATATATACACATATATACACATATATAT 327
              ||||||||| | |||||||||  |||||||| |||||||| | |||||||| | |||||||||
Db        313 ATATATATATATATATATATATGCATATATATATATATATATATATATATATATATAT 372

Qy        328 ACACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATAC 387
              | | ||| |||| || | ||||||||| | |||| |||| | |||||||| | ||||||||
Db        373 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 432

Qy        388 ACACATATATACACATATATACACACATATATACACATATATACACATATATACACATAT 447
              | | ||||||| | |||||||| | | ||||||| | |||||||| | |||||||    ||||
Db        433 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATA-TGATAT 491

Qy        448 ACACATATATACACATATATACATATATACACATATATATAATATACACACATATATATA 507
              | | ||||||| | |||||||| |||||||| | |||||||| ||||| | | |||||||||
Db        492 ATATATATATATATATATATATATATATATATATATATATAT-ATATATATATATATATA 550

Qy        508 CACATATATACACACATATATACACATATATACACATATATATACACATATATACACATA 567
              | ||||||| | | ||||||| | ||||||| | ||||||||| | ||||||| | |||
Db        551 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 610

Qy        568 TATACATATATACACATATATATACATATATACACATATATACATATATACACATATATA 627
              |||| ||||||| | ||||||||| ||||||| | ||||||| ||||||| | |||||||
Db        611 TATATATATATATATATATATATATATATATATATATATATATATATATATAGATATATATA 670

Qy        628 CATATATACACACATATATACACATACATATACACACACATAGATATACATATATATACA 687
              ||||||| | | ||||||| | ||| ||||| | | | ||| ||||  ||||||||| |
Db        671 TATATATATATATATATATATATATATATATATATATATATATATATATATCTATATATATA 730
```

```
Qy        688 CATATATATACGTATATATATGTATATATA 717
              |||| ||||  || ||||||||||||||||
Db        731 TATATCTATATATACATATATGTATATATA 760


RESULT 77
BX145980/c
LOCUS       BX145980               709 bp    DNA     linear   GSS 28-JAN-2003
DEFINITION  Danio rerio genomic clone DKEY-104E13, genomic survey sequence.
ACCESSION   BX145980
VERSION     BX145980.1  GI:27977335
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 709)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the T7 end of BAC 104E13. 104E13
            is part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .709
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-104E13"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             6.9%;  Score 285.4;  DB 9;  Length 709;
  Best Local Similarity   74.0%;  Pred. No. 5.1e-38;
  Matches  375;  Conservative    0;  Mismatches  131;  Indels    1;  Gaps    1;

Qy        230 ATATATATATATATATATATATACACACACACACACATATATATATACACATATATATAC 289
              |||||||||||||||||||||||| | | | | | |||||||||||| | ||| |||||
Db        709 ATATATATATATATATATATATATATATATATATATATATATATATATATATAGATATAT 650

Qy        290 ACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATAT 349
              | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | || |
Db        649 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATGT 590

Qy        350 ATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATAC 409
              ||||| | |||| |||    ||||||| | ||||||| | | ||||||| | |||||||
Db        589 ATATATATATATATATGTGTATATATATATATATATATATATATATATATATATATATAT 530

Qy        410 ACACATATATACACATATATACACATATATACACATATACACATATATACACATATATAC 469
              | | ||||||| | ||||||| | ||||||| | ||||| |  ||| || | |||||||
Db        529 ATATATATATATATATATATATATATATATATATATATATATGTATGTATATATATATAT 470

Qy        470 ATATATACACATATATATAATATACACACATATATATACACATATATACACACATATATA 529
              |||||||   |||||| | ||||| | | ||||||||| | || |||| | | |||||||
Db        469 ATATATATGTATATATGT-ATATATATATATATATATATATGTATATATATATATATA 411

Qy        530 CACATATATACACATATATATACACATATATACACATATATACATATATACACATATATA 589
              | ||||||| | ||||||||| | ||||||| | ||||||| |||||||     | ||||
Db        410 TATATATATATATATATATATATATATATATATATATATATATATATATATGTGTGTATA 351

Qy        590 TACATATATACACATATATACATATATACACATATATACATATATACACACATATATACA 649
              || ||||||| | |||| |  ||||||| | ||||||   |||||| | | ||||||| |
Db        350 TATATATATATATATATGTGTATATATATATATATGTGTATATATATATATATATATA 291

Qy        650 CATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATATG 709
              ||| ||||| | | | ||| ||||| |||||||||  |||| || |   | |
Db        290 TATATATATATATATATATATATATATATATATATGCATAAATTTGACTTTTGCATGC 231
```

```
Qy        710 TATATATATATGCTCCAGAGTTCATAA 736
              || |  ||  || | |||  || | |
Db        230 TAAAATTACTTGATTAAGATTTTGTCA 204


RESULT 78
BU507480/c
LOCUS       BU507480               1836 bp    mRNA    linear   EST 12-SEP-2002
DEFINITION  AGENCOURT_10309824 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:6501420
            5', mRNA sequence.
ACCESSION   BU507480
VERSION     BU507480.1  GI:22813713
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1836)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM14055  row: n  column: 13
            High quality sequence start: 74
            High quality sequence stop: 447.
FEATURES             Location/Qualifiers
     source          1. .1836
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6501420"
                     /tissue_type="leiomyosarcoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_71"
                     /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 2.1 kb.  "
ORIGIN

  Query Match             6.9%;  Score 284.6;  DB 5;  Length 1836;
  Best Local Similarity   75.5%;  Pred. No. 6.1e-38;
  Matches  361;  Conservative    0;  Mismatches  116;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | | | | | | | | | | | |||||||||||||||| |       | | | | |
Db        548 TCTNTTTCTCTNTTTCTTTTTTTGTACACACACACACACACTCTCTCTCTCTTTTTTC 489

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | | | | ||| | | ||| |||||||| | |||||||||||| ||| |||| || | |
Db        488 TATATACACACACACACACATATATATANATATATATATACATATATATATATATATA 429

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | |||||||| | |||||||  | | |||||||| | |||||
Db        428 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 369

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | |||||||| | |||||||| | |||||||| | |||||| | |||||||| | |||||
Db        368 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 309

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | |||||||  | | |||
Db        308 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 249

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | |||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
```

```
Db        248 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 189

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| |||||| | |||||| ||||||| | ||||||| ||||||| | | |||||
Db        188 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 129

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATA 703
              || | ||| ||||| | | | ||| ||||| ||||||||| | ||| || | ||   |
Db        128 TATATATATATATATATATATATATATATATATATATATATATATANATNTCCGNGNA 71


RESULT 79
CA364719/c
LOCUS       CA364719                683 bp    mRNA    linear   EST 06-NOV-2002
DEFINITION  639681 NCCCWA 1RT Oncorhynchus mykiss cDNA clone 1RT70B09_B_A05 5',
            mRNA sequence.
ACCESSION   CA364719
VERSION     CA364719.1  GI:24674188
KEYWORDS    EST.
SOURCE      Oncorhynchus mykiss (rainbow trout)
  ORGANISM  Oncorhynchus mykiss
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei;
            Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
REFERENCE   1  (bases 1 to 683)
  AUTHORS   Rexroad,C.E. 3rd, Lee,Y., Keele,J.W., Karamycheva,S., Brown,G.,
            Koop,B., Gahr,S.A., Palti,Y. and Quackenbush,J.
  TITLE     Sequence analysis of a rainbow trout cDNA library and creation of a
            gene index
  JOURNAL   Cytogenet. Genome Res. 102 (1-4), 347-354 (2003)
COMMENT     Contact: Rexroad CE
            USDA, ARS, National Center for Cool and Cold Water Aquaculture
            11876 Leetown Road, Kearneysville, WV 25430, USA
            Tel: 304 724 8340 x2129
            Fax: 304 725 0351
            Email: crexroad@ncccwa.ars.usda.gov
            Single pass sequencing. Bases called with phred v0.020425.c and
            trimmed with the aid of the trim_alt option. Vector identified by
            cross_match v0.990329.
            Seq primer: AGCGGATAACAATTTCACACAGGA.
FEATURES             Location/Qualifiers
     source          1. .683
                     /organism="Oncorhynchus mykiss"
                     /mol_type="mRNA"
                     /db_xref="taxon:8022"
                     /clone="1RT70B09_B_A05"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /clone_lib="NCCCWA 1RT"
                     /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
                     Library made from pooled tissue from brain, gill, liver,
                     spleen, muscle, and kidney."
ORIGIN

  Query Match             6.8%;  Score 282.2;  DB 6;  Length 683;
  Best Local Similarity   74.5%;  Pred. No. 1.8e-37;
  Matches  369;  Conservative    0;  Mismatches  123;  Indels    3;  Gaps    1;

Qy        229 TATATATATATATATATATATATACACACACACACATATATATATACACATATATATA 288
              |||||||||||||| |||||||||| ||| ||| | |||||||| ||     | ||||||
Db        674 TATATATATATATACATATATATACATACATACATATATATATATGTATGTGTGTATATA 615

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              || | || |||| ||  | |||||| ||| || ||||||||| || |    | |||||
Db        614 TGTATGTGTATATATTTATGCGTATATATACACATGTATACACATGCACACACATACATA 555

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              | |||  | ||||||||| |||||||| | | | |||    |  |||||||| |||||||
Db        554 TGTATGTATATATGTATATACATATATATATACACATATGTATGTATATACATATATATA 495

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              ||||||  |||  | ||| ||| | ||||||| ||| | |||||  |||| | |||||||
Db        494 CACACACGTATGTATATACATATATATATATACACACACACGTATATATATATATA 435

Qy        469 CATATATACACATATATATA---ATATACACACATATATATACACATATATACACACATA 525
              ||| |   ||||||| |||   ||||| | | ||| | |||||||  |||| | | |||
Db        434 TATACACGTACATATACATACACATATATATATATACACATACACACGTATATATATATA 375

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||||||||| ||| | ||||||||||    | |||     ||| ||| ||||||| ||| |
Db        374 TATACACATACATATATATATATATACGTGTGTATGTGTATACATATATATATATACACA 315

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              || | || |||| | | |||||||| ||||| ||  |||||||  ||||| | | | |||||
Db        314 TACACACGTATACATATATATATATATATACACGTATATATATATATACATATATATATA 255

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              |||||||||| |  |||||||||| ||||| |||||||||    || | ||||  ||||||
Db        254 TACACATACACACGCACACACATATATATATATATATATATGTATGTGTATATATATATA 195

Qy        706 TATGTATATATATAT 720
              |||  ||||   |||
Db        194 TATACATATGCGTAT 180


RESULT 80
CW518121
LOCUS       CW518121                663 bp    DNA     linear   GSS 08-OCT-2004
DEFINITION  OP__Ba0025C13.f OP__Ba Oryza punctata genomic clone OP__Ba0025C13
            5', genomic survey sequence.
ACCESSION   CW518121
VERSION     CW518121.1  GI:53992343
KEYWORDS    GSS.
SOURCE      Oryza punctata
  ORGANISM  Oryza punctata
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 663)
  AUTHORS   SanMiguel,P., Westerman,R., Kim,H., Yu,Y., Wissotski,M., Yost,D.,
            Stum,D., Rao,K., Luo,M., Jetty,R., Kudrna,D., Muller,C.,
            Hatfield,J., Soderlund,C., Wing,R. and Jackson,S.A.
  TITLE     OMAP Project - Purdue University
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Scott A. Jackson
            Jackson Laboratory
            Purdue University
            915 W. State St., West Lafayette, IN 47907, USA
            Tel: 7654963621
            Fax: 7654967255
            Email: sjackson@purdue.edu
            Basecalling by phred version 0.020425.c. This sequence was derived
            from the raw sequence read by clipping with lucy version 1.19s.
            Bases 50-712 of the raw sequence (length 1364) were retained after
            clipping.
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Insert Length: 161000   Std Error: 0.00
            Plate: 0025  row: C  column: 13
            Seq primer: TAA TAC GAC TCA CTA TAG GG
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .663
                     /organism="Oryza punctata"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4537"
                     /clone="OP__Ba0025C13"
                     /tissue_type="young leaves"
                     /lab_host="DH10B-T1 phage resistant"
                     /clone_lib="OP__Ba"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

  Query Match             6.8%;  Score 280.4;  DB 9;  Length 663;
  Best Local Similarity   77.2%;  Pred. No. 3.5e-37;
```

```
  Matches  366;  Conservative    0;  Mismatches  106;  Indels    2;  Gaps    2;

Qy        208 CAGAAAAAAAAAAAGGCATTTTATATATATATATATATATACACACACACACACAT 267
              || ||||||||||    |||    || |  ||  | | | ||||| | | | | | ||
Db        191 CAAAAAAAAAAAATCCAATTCACCATGTCAATGCAAAAAGATATATATATATATATAT 250

Qy        268 ATATATATACACATATATATACACATATATACATATATACACATATATACACATATATAT 327
              ||||||||| | |||||||||| | ||||||| ||||||| | ||||||| | ||||||||
Db        251 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 310

Qy        328 ACACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATAC 387
              | | ||| |||| || | ||||||||| | |||| |||| | ||||||| | |||||||
Db        311 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 370

Qy        388 ACACATATATACACATATATACACACATATATACACATATATACACATATATACACATAT 447
              | | ||||||| | ||||||| | | ||||||| | ||||||| | ||||||| | ||||
Db        371 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 430

Qy        448 ACACATATATACACATATATACATATATACACATATATATA-ATATACACACATATATAT 506
              | | ||||||| | ||||||| ||||||| | ||||||||| ||||| | | ||||||||
Db        431 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 490

Qy        507 ACACATATATACACACATATATACACATATATACACATATATATACACATATATACACAT 566
              | | ||||||| | | |||||||||||||||| | | ||||| ||| | ||||| | | |
Db        491 ATATATATATATATATATATATATACACATATAAATAAATATAGATAGATATATA-ATATAA 549

Qy        567 ATATACATATATACACATATATATACATATATACACATATATACATATATACACATATAT 626
              |||||  ||||||| | ||||||||| ||||||| | ||||||| ||||||| | ||||||
Db        550 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 609

Qy        627 ACATATATACACACATATATACACATACATATACACACACATAGATATACATAT 680
              | ||||||| | | ||||||| | ||| ||||| | | | ||| ||||| ||||
Db        610 ATATATATATATATATATATATATATATATATATATATATATATATATATAAATAT 663


RESULT 81
BX149919/c
LOCUS       BX149919                737 bp    DNA     linear   GSS 13-MAR-2003
DEFINITION  Danio rerio genomic clone DKEY-112P5, genomic survey sequence.
ACCESSION   BX149919
VERSION     BX149919.1  GI:27981385
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 737)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAR-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the T7 end of BAC 112P5. 112P5 is
            part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .737
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-112P5"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             6.8%;  Score 280;  DB 9;  Length 737;
  Best Local Similarity   75.6%;  Pred. No. 4.1e-37;
  Matches  360;  Conservative    0;  Mismatches  115;  Indels    1;  Gaps    1;

Qy        230 ATATATATATATATATATATATACACACACACACACATATATATATACACATATATATAC 289
              |||||||||||||||||||||||| | | | | | |||| | |||| | |||||||||

Db        737 ATATATATATATATATATATATATATATATATATATATATATATGTGTATATATATATATATAT 678

Qy        290 ACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATAT 349
              | ||||||| ||||||| | ||||||| | |||||||||| | ||| |||| || | ||||
Db        677 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 618

Qy        350 ATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATAC 409
              ||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||||
Db        617 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 558

Qy        410 ACACATATATACACATATATACACATATATACACATATACACATATATACACATATATAC 469
              | | ||||||| | ||||||| | ||||||| | ||||| |  | |||| | |||||||
Db        557 ATATATATATATATATATATATATATATATATATATATATATATATGTGTATATATATATAT 498

Qy        470 ATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATATAT 528
              ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | ||||||
Db        497 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 438

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACACATATAT 588
              | | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | ||||||
Db        437 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 378

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
              ||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||||
Db        377 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 318

Qy        649 ACATACATATACACACACATAGATATACATATATATACACATATATATACGTATAT 704
              | |   | | | | ||  | | | | |   | |  | | ||||  || |
Db        317 AGCTTTTAAAAAAAAAACCAAAAAACAGAAAAGGTTGCCAAACTGTATAGAAATCT 262


RESULT 82
CL044834
LOCUS       CL044834                1031 bp    DNA     linear   GSS 31-DEC-2003
DEFINITION  CH216-61O18_RM1.1 CH216 Xenopus tropicalis genomic clone
            CH216-61O18, genomic survey sequence.
ACCESSION   CL044834
VERSION     CL044834.1  GI:40500747
KEYWORDS    GSS.
SOURCE      Xenopus tropicalis (western clawed frog)
  ORGANISM  Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 1031)
  AUTHORS   Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
            Mardis,E. and Wilson,R.
  TITLE     A physical map of the xenopus tropicalis genome
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 175000   Std Error: 0.00
            Seq primer: RM1 TACGACTCACTATAGGGAGA
            Class: BAC ends
            High quality sequence start: 47
            High quality sequence stop: 682.
FEATURES             Location/Qualifiers
     source          1. .1031
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /strain="Nigerian frog"
                     /db_xref="taxon:8364"
                     /clone="CH216-61O18"
                     /sex="male"
                     /cell_line="Stock 248 F7A2, inbred N7"
                     /clone_lib="CH216"
                     /note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis
                     BAC library"
ORIGIN

  Query Match             6.8%;  Score 280;  DB 9;  Length 1031;
```

```
  Best Local Similarity   73.4%;  Pred. No. 3.9e-37;
  Matches  373;  Conservative    0;  Mismatches  130;  Indels    5;  Gaps    1;

Qy        215 AAAAAAAGGCATTTTATATATATATATATATATACACACACACACACATATATATA 274
              ||||  |       |  ||| | | | | ||| ||| |||||||||| ||| ||||| ||| |
Db        203 AAAATACATACATACATACACACACACATACATACATACACACATACATACATACATACA 262

Qy        275 TACACATATATATACACATATATACATATATACACATATATACACATATATATACACATA 334
              |||| | ||| ||||| | ||| | | ||| ||||| ||| ||||||| ||| | ||| |
Db        263 TACATACATACATACATACATAAAGACATACACACACATACACACATACATACATACACA 322

Qy        335 CATATGTACACATATATATACACATATGTATACACATATATACACATATATACACACATA 394
              ||||  |||||||| ||| | |||||  || | ||| | ||||||| ||| | ||| | |
Db        323 CATACATACACATACATACATACATACATACACACATACACACATACATACATACA 382

Qy        395 TATACACATATATACACACATATATACACATATATACACATATATACACATATACACATA 454
              | | ||||| ||||| ||||| ||||| | | | | ||||| ||||||| ||||| | |
Db        383 CACATACATACATACAGACATACATACATACACACATACATACATACACACATACATACA 442

Qy        455 TATACACATATATACATATATACACATATATATAATATACACACATATATATACACATAT 514
              || ||||||| ||||||| | ||||| ||      | | ||| ||||| ||| |||||||
Db        443 TACACACATACATACATACACACACACAT-----ACACACATACATACATACACACATAC 497

Qy        515 ATACACACATATATACACATATATACACATATATATACACATATATACACATATATACAT 574
              ||||||||| | ||||| | ||| | ||||| ||| | ||||| ||||||| ||| | |
Db        498 ATACACACACACATACATACATACATACATACATACATACATACACACATACATAC 557

Qy        575 ATATACACATATATATACATATATACACATATATACATATATACACATATATACATATAT 634
              ||| |||||||| ||| | | ||| ||||||| ||||||| | ||| | | ||| | |||
Db        558 ATACACACATACATACATACATACACACATACATACATACAGACATACATACATACATAC 617

Qy        635 ACACACATATATATACACATACATATACACACACATAGATATACATATATATACACATATAT 694
              | ||||| ||| |||||||||||| | ||| | | | ||| ||| ||| ||||||| |||
Db        618 ATACACACATACACACATACATACATACATAAAGACATACACACATACATACACACATAC 677

Qy        695 ATACGTATATATATGTATATATATATGC 722
              |||| || ||  ||  ||| | | || |
Db        678 ATACATACATTCATACATACACACATAC 705


RESULT 83
AQ506817/c
LOCUS       AQ506817                537 bp    DNA     linear   GSS 29-APR-1999
DEFINITION  RPCI-11-281J17.TV RPCI-11 Homo sapiens genomic clone
            RPCI-11-281J17, genomic survey sequence.
ACCESSION   AQ506817
VERSION     AQ506817.1  GI:4711564
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 537)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and
            Venter,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: RPCI-11-281J17.TJ
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genet cs (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: T7
            Class: BAC ends.
FEATURES            Location/Qualifiers
     source         1. .537
                    /organism="Homo sapiens"
                    /mol_type="genomic DNA"
                    /db_xref="GDB:7607752"
                    /db_xref="taxon:9606"
                    /clone="RPCI-11-281J17"
                    /sex="Male"
                    /cell_type="Lymphocytes"
                    /clone_lib="RPCI-11"
                    /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                    RPCI11 Human Male BAC Library"
ORIGIN

  Query Match             6.8%;  Score 278.8;  DB 8;  Length 537;
  Best Local Similarity   76.0%;  Pred. No. 6.8e-37;
  Matches  357;  Conservative    0;  Mismatches  112;  Indels    1;  Gaps    1;

Qy        230 ATATATATATATATATATATATACACACACACACATATATATATACACATATATATAC 289
              |||||||||||||||||||| ||| | | ||| | | | ||||||||| | ||||| |||
Db        530 ATATATATATATATATATACATATATACATATATACATATATATAAATATATACATAT 471

Qy        290 ACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATAT 349
              | ||||| | ||||||| | ||||| | | ||||| ||| | |||||||| || | ||||
Db        470 ATATATAAATATATATATATATATAAATATATATAAATATATATACATATATAAATATAT 411

Qy        350 ATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATAC 409
              ||||| | |||| || | | ||| ||| | ||||||| | | | ||||| | |||||||
Db        410 ATATAAATATATATAAATATATACATATATATATATATATATAAATATATAAATATATAT 351

Qy        410 ACACATATATACACATATATACACATATATACACATATACACATATATACACATATATAC 469
              | | ||||||||| |||||| || ||||||| | ||||||| ||||||| | |||||||
Db        350 ATAAATATATACATATATATCCATATATATAAATATATACATATATATAAATATATATAA 291

Qy        470 ATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATATAT 528
              ||||||| | ||||||| | ||||| ||| ||||  ||| | ||  ||| | | | ||||
Db        290 ATATATAAATATATATAAATATATATACATATATCAATATATATCAATATATATAAATAT 231

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACATATAT 588
              |||   ||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||||||
Db        230 ATATCAATATATAAATATATATAAATATATATAAATATATATAAATATATAAATATATAT 171

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
              | | ||||||| | ||||||| ||||||| | ||||||| | ||||| | | |||| |
Db        170 ACATATATATAAATATATATAAATATATAAATATATATATAAATATAAATATATATCAAT 111

Qy        649 ACATACATATACACACACATAGATATACATATATATACACATATATATAC 698
              | ||||||||| | | | ||| ||||| |||     | ||  ||||| | |
Db        110 ATATACATATATATATAAATATATATATATACTGTTTCATTTATATCTGC 61


RESULT 84
BX159146
LOCUS       BX159146                789 bp    DNA     linear   GSS 28-JAN-2003
DEFINITION  Danio rerio genomic clone DKEY-152P14, genomic survey sequence.
ACCESSION   BX159146
VERSION     BX159146.1  GI:27990759
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 789)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the SP6 end of BAC 152P14. 152P14
            is part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
```

```
          http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .789
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-152P14"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             6.8%;  Score 278.8;  DB 9;  Length 789;
  Best Local Similarity   75.5%;  Pred. No. 6.4e-37;
  Matches  373;  Conservative    0;  Mismatches  117;  Indels    4;  Gaps    2;

Qy        213 AAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACACATATATA 272
              ||| || |   |||||||||||||||| | | | ||| |   |||  |||||||| |||||
Db        300 AAATAAGATACCATTTTATATATATACACACACATACA-GTACATGCACACACACATATA 358

Qy        273 TATACACATATATATACACATATATACATATATACACATATATACACATATATATACACA 332
              |||||||| | | | |||||||||||| |||| |||  ||||| | | |||||| |
Db        359 TATACACACACACAAACACATATATATATTTACGTATATACATATATATATGTGTGTG 418

Qy        333 TACATATGTACACATATATATACACATATGTATACACATATATACACATATATACACACA 392
              || || ||||    ||| |||| | |||| |||| | |||||||| | ||||| |||||||
Db        419 TATATGTGTATGTGTATGTATATATATATATATATATATATATATATATATACACACACA 478

Qy        393 TATATACACATATATACACACATATATACACATATATACACATATATACACATATACACA 452
                | ||| | ||||||| | | ||||||| | ||||||| | ||||||| ||| | |||||
Db        479 CACATATATATATATATATATATATATATATATATATATATATATATATACACACACACA 538

Qy        453 TATATACACATATATACATATATACACATATATATAATATACACACATATATATACACAT 512
              |||||| | ||||||| |||| || | ||||||   ||||| | | |||||||||| | ||
Db        539 TATATATATATATATATATATGTATATATATAT---ATATATATATATATATATATATAT 595

Qy        513 ATATACACACATATATACACATATATACACATATATATACACATATATACACATATATAC 572
              ||||| | | |||||||||| | | ||| ||||||||| | ||||||| | |||||||
Db        596 ATATATATATATATATATACACACACACATATATATATATATATATATATATATATAT 655

Qy        573 ATATATACACATATATATACATATATACACATATATACATATATACACATATATACATAT 632
              ||||||| | ||||||||| ||||| ||||| | | ||| | | ||||| | | ||| |
Db        656 ATATATATATATATATATATATATATACACACACACACACACACACACACACACACAC 715

Qy        633 ATACACACATATATACACATACATATACACACACATAGATATACATATATATACACATAT 692
              ||| | | ||||||| | ||| ||||| | | |   | |||||||||| ||| | ||||
Db        716 ATATATATATATATATATATATATATATATATATATGTGTATATACATATACATATATATAT 775

Qy        693 ATATACGTATATAT 706
              |||||  |||||||
Db        776 ATATATATATATAT 789


RESULT 85
BX240321
LOCUS       BX240321                446 bp    DNA     linear   GSS 29-JAN-2003
DEFINITION  Danio rerio genomic clone DKEY-276D22, genomic survey sequence.
ACCESSION   BX240321
VERSION     BX240321.1  GI:28162655
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 446)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the SP6 end of BAC 276D22. 276D22
            is part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
```

```
          http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .446
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-276D22"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             6.7%;  Score 277.6;  DB 9;  Length 446;
  Best Local Similarity   79.5%;  Pred. No. 1.1e-36;
  Matches  353;  Conservative    0;  Mismatches   89;  Indels    2;  Gaps    2;

Qy        268 ATATATATACACATATATATACACATATATACATATATACACATATATACACATATATAT 327
              ||||||||| | |||||||||| | ||||||| ||||||| | ||||||| | ||||||||
Db          4 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 63

Qy        328 ACACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATAC 387
              | | ||| | || |||| | ||| ||| | |||  |||| | ||||||| |||||||||
Db         64 ATATATATAAATATACATA-ATAAATAGAAATAGTTATATATATATATAGACATATATAT 122

Qy        388 ACACATATATACACATATATACACACATATATACACATATATACACATATATACACATAT 447
              | | ||||||| | | ||||| | | ||||||| | ||||| | | | ||||| | |||
Db        123 ATATATATATATATAGATATATATATATATATATATATATAGATAGACATATATATATAG 182

Qy        448 ACACATATATACACATATATACATATATACACATATATATA-ATATACACACATATATAT 506
              | | ||||||| | ||||||| ||||||| | |||||||||  ||||| | | ||||||||
Db        183 ATATATATATATATATATATATATATATATATATATATATAGATATATATATATATATAT 242

Qy        507 ACACATATATACACACATATATACACATATATACACATATATATACACATATATACACAT 566
              | | ||||||| | | ||||||| | ||||||| | ||||||||| | ||||||| | ||
Db        243 ATATATATATATATATATATATATATATATATATAGATATATATATATATATATATATAT 302

Qy        567 ATATACATATATACACATATATATACATATATACACATATATACATATATACACATATAT 626
              ||||| ||||||| | ||||||||| ||||||||| ||||||| ||||||| | ||||||
Db        303 ATATATATATATATATATATATATATATATATATACATATATATATATATATATATATAT 362

Qy        627 ACATATATACACACATATATACACATACATATACACACACATAGATATACATATATATAC 686
              | ||||||| | | ||||||| | ||| ||||| | | ||||| ||||| |||||||||
Db        363 ATATATATATATATATATATATATATATATATATATATACATATATATATATATATAT 422

Qy        687 ACATATATATACGTATATATATGT 710
              | |||||||||  |||||||| |
Db        423 ATATATATATATATATATATAT 446


RESULT 86
BX241290
LOCUS       BX241290                722 bp    DNA     linear   GSS 29-JAN-2003
DEFINITION  Danio rerio genomic clone DKEY-282A22, genomic survey sequence.
ACCESSION   BX241290
VERSION     BX241290.1  GI:28163624
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 722)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the T7 end of BAC 282A22. 282A22
            is part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .722
                     /organism="Danio rerio"
```

```
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-282A22"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             6.7%;  Score 276.6;  DB 9;  Length 722;
  Best Local Similarity   77.6%;  Pred. No. 1.5e-36;
  Matches  350;  Conservative    0;  Mismatches   94;  Indels    7;  Gaps    1;

Qy        250 ATACACACACACACACATATATATATACACATATATATACACATATATACATATATACAC 309
              ||||| |     |   |||| |||||| | ||||||||| | ||||||| ||||||| |
Db         13 ATACATATTATTATTTATATTTATATATATATATATATATATATATATATATATATATAT 72

Qy        310 ATATATACACATATATATACACATACATATGTACACATATATATACACATATGTATACAC 369
              ||||||| ||||||| ||||||| | | || |||| ||||||| | ||| || |||||||
Db         73 ATATATATACATATACATACACACATACATATACATATATATACATACACATATATACAC 132

Qy        370 ATATATACACATATATACACACATATATACACATATATACACACATATATACACATATAT 429
              ||| ||| ||||| ||| | | | ||| | ||||| ||||| ||||| ||||| | ||
Db        133 ATACATATACATACATATATATACATACATACATACATACATACATACATACATACATGC 192

Qy        430 ACACATATATACACATATACACATATATACACATATATACATATATACACATATATATA- 488
              | ||||  ||||| ||||| | |||||||||||||| | ||||| ||| ||||||||||
Db        193 ATACATGCATACATATATATATATATATACACATACACACATACATATACATATATATAC 252

Qy        489 ------ATATACACACATATATATACACATATATACACACATATATACACATATATACAC 542
                    ||||| ||||||| ||||||| | ||||| | | ||||| | ||||| |||||
Db        253 ATACACATATATACACATACATATACATACATATATATATATATACATACATACATACAT 312

Qy        543 ATATATATACACATATATACACATATATACATATATACACATATATATACATATATACAC 602
              | ||| |||||  | ||| | ||||| ||||||| ||||| | ||| ||||||| ||| |
Db        313 ACATACATACATACATACATACATACATACATACATACATACATACATACATACATATAT 372

Qy        603 ATATATACATATATACACATATATACATATATACACACATATATACACATACATATACAC 662
              |||||||||||||||||||| ||||||||| | | | ||||||||||||||||| | ||
Db        373 ATATATACATATATACACATACATACATATACATATATATATATACACATACATACATAC 432

Qy        663 ACACATAGATATACATATATATACACATATA 693
              | ||||| ||| | | ||| ||||| | |||
Db        433 ATACATACATACATACATACATACATACATA 463


RESULT 87
CL763005/c
LOCUS       CL763005                617 bp    DNA     linear   GSS 27-JUL-2004
DEFINITION  OR_BBa0133F16.f OR_BBa Oryza rufipogon genomic clone OR_BBa0133F16
            5', genomic survey sequence.
ACCESSION   CL763005
VERSION     CL763005.1  GI:50720952
KEYWORDS    GSS.
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 617)
  AUTHORS   Kim,H., Yu,Y., Stum,D., Yost,D., Rao,K., Luo,M., Jetty,R.,
            Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and Wing,R.
  TITLE     OMAP Project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Insert Length: 161   Std Error: 0.00
            Plate: 0133  row: F  column: 16
            Seq primer: TAA TAC GAC TCA CTA TAG GG
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .617
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
                     /clone="OR_BBa0133F16"
                     /tissue_type="young leaves"
                     /lab_host="DH10B-T1 phage resistant"
                     /clone_lib="OR_BBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

  Query Match             6.7%;  Score 276;  DB 9;  Length 617;
  Best Local Similarity   73.6%;  Pred. No. 2e-36;
  Matches  365;  Conservative    0;  Mismatches  130;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | |||| |||  |||||||||||| || |   | | | |  ||| ||| | || |||||||
Db        572 TATATACATAACTATATATATATCTATATGTATATATATCTATGTATCTGCATATATATA 513

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | |||| |   | ||||   |||||||    | | ||||    || |||| | | |
Db        512 TATATATATCTTTTTTTATATTTATATATATCTCTCTCTATATTTTTATATATATGTATC 453

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | | || |||      |||||| | ||||||| | | ||||||| | |||||
Db        452 TATATATATAAAAATATATTTTTTTATATATATATATATATATATATATATAAATATA 393

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||  | | | ||| |   ||||| | ||       ||||||| | |||||
Db        392 TATATATATATATTAAAAAAAATAAAATAATATATATATTATATAATATATATATATATA 333

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || || |||| | || |||||| ||||| | | |||||||| | ||||||| | | |||
Db        332 TATATTTATATATATTTATATATATATATATATATATATATATATATATATATATATA 273

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | |||||||| | |||||||| | ||||||| | ||||||| ||||||| | |||
Db        272 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 213

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db        212 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 153

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db        152 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 93

Qy        706 TATGTATATATATATG 721
              ||| ||||||||||||
Db         92 TATATATATATATATG 77


RESULT 88
CL800151
LOCUS       CL800151                715 bp    DNA     linear   GSS 06-AUG-2004
DEFINITION  OR_CBa0011K05.r OR_CBa Oryza rufipogon genomic clone OR_CBa0011K05
            3', genomic survey sequence.
ACCESSION   CL800151
VERSION     CL800151.1  GI:51024771
KEYWORDS    GSS.
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 715)
  AUTHORS   Kim,H., Yu,Y., Wissotski,M., Yost,D., Stum,D., Rao,K., Luo,M.,
            Jetty,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and
            Wing,R.
```

```
  TITLE     OMAP project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0011  row: K  column: 05
            Seq primer: CAC TCA TTA GGC ACC CCA
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .715
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
                     /clone="OR_CBa0011K05"
                     /tissue_type="young leaves"
                     /dev_stage="2 week old seedlings"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OR_CBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII;
                     drk treated 36 hrs before harvest"
ORIGIN

  Query Match             6.7%;  Score 275.4;  DB 9;  Length 715;
  Best Local Similarity   80.3%;  Pred. No. 2.4e-36;
  Matches  335;  Conservative    0;  Mismatches   81;  Indels    1;  Gaps    1;

Qy        303 TATACACATATATACACATATATATACACATACATATGTACACATATATATACACATATG 362
              ||||  |||||||| | ||||||||| | ||| |||| || | ||||||||| | ||||
Db        300 TATAATCATATATATATATATATATATATATATATATATATATATATATATATATATATA 359

Qy        363 TATACACATATATACACATATATACACACATATATACACATATATACACACATATATACA 422
              |||| | ||||||| | ||||||| | | ||||||| | ||||||| | | ||||||| |
Db        360 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 419

Qy        423 CATATATACACATATATACACATATACACATATATACACATATATACATATATACACATA 482
               ||||||| | ||||||| | ||||| | ||||||| | ||||||| ||||||| | |||
Db        420 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 479

Qy        483 TATATAATATACACACATATATATACACATATATACACACATATATACACATATATACAC 542
              ||||| ||||| | | ||||||||| | ||||||| | | ||||||| | ||||||| |
Db        480 TATAT-ATATATATATATATATATATATATATATATATATATATATATATATATATATAT 538

Qy        543 ATATATATACACATATATACACATATATACATATATACACATATATATACATATATACAC 602
              ||||||||| | ||||||| | ||||||| ||||||| | ||||||||| ||||||| |
Db        539 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 598

Qy        603 ATATATACATATATACACATATATACATATATACACACATATATACACATACATATACAC 662
              ||||||| ||||||| | ||||||| ||||||| | | ||||||| | ||| ||||| |
Db        599 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 658

Qy        663 ACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATA 719
              | | ||| ||||| ||||||||| | |||||||||  ||||||||| ||||||||||
Db        659 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 715


RESULT 89
CL800151/c
LOCUS       CL800151                715 bp    DNA     linear   GSS 06-AUG-2004
DEFINITION  OR_CBa0011K05.r OR_CBa Oryza rufipogon genomic clone OR_CBa0011K05
            3', genomic survey sequence.
ACCESSION   CL800151
VERSION     CL800151.1  GI:51024771
KEYWORDS    GSS.
SOURCE      Oryza rufipogon
  ORGANISM  Oryza rufipogon
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 715)
  AUTHORS   Kim,H., Yu,Y., Wissotski,M., Yost,D., Stum,D., Rao,K., Luo,M.,
            Jetty,R., Kudrna,D., Muller,C., Hatfield,J., Soderlund,C. and
            Wing,R.
  TITLE     OMAP project
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Plate: 0011  row: K  column: 05
            Seq primer: CAC TCA TTA GGC ACC CCA
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .715
                     /organism="Oryza rufipogon"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4529"
                     /clone="OR_CBa0011K05"
                     /tissue_type="young leaves"
                     /dev_stage="2 week old seedlings"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="OR_CBa"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII;
                     drk treated 36 hrs before harvest"
ORIGIN

  Query Match             6.7%;  Score 275.4;  DB 9;  Length 715;
  Best Local Similarity   80.3%;  Pred. No. 2.4e-36;
  Matches  335;  Conservative    0;  Mismatches   81;  Indels    1;  Gaps    1;

Qy        229 TATATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 288
              ||||||||||||||||||||||||| | | | | | ||||||||||| | |||||||||
Db        715 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 656

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |||
Db        655 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 596

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              |||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||||
Db        595 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 536

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||||
Db        535 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 476

Qy        469 CATATATACACATATATATAATATACACACATATATATACACATATATACACACATATAT 528
              ||||||| | |||||||| ||||| | | ||||||||| | ||||||| | | ||||||
Db        475 TATATATATATATATATAT-ATATATATATATATATATATATATATATATATATATAT 417

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACATATAT 588
              | | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | ||||||
Db        416 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 357

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              ||| ||||||| | ||||||| ||||||| | ||||||| ||||||| |    ||||
Db        356 ATATATATATATATATATATATATATATATATATATATATATATATATATGATTATA 300


RESULT 90
BX158149
LOCUS       BX158149                705 bp    DNA     linear   GSS 28-JAN-2003
DEFINITION  Danio rerio genomic clone DKEY-151O11, genomic survey sequence.
ACCESSION   BX158149
```

```
VERSION     BX158149.1  GI:27989752
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 705)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the SP6 end of BAC 151O11. 151O11
            is part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .705
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-151O11"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             6.7%;  Score 275;  DB 9;  Length 705;
  Best Local Similarity   78.9%;  Pred. No. 2.9e-36;
  Matches  343;  Conservative    0;  Mismatches   85;  Indels    7;  Gaps    1;

Qy        266 ATATATATATACACATATATATACACATATATACATATATACACATATATACACATATAT 325
              |||| ||||||| | |||||||||| | ||||||| ||||||| | ||||||| |||||||
Db         12 ATATTTATATATATATATATATATATATATATATATATATATATATATATACATATAC 71

Qy        326 ATACACATACATATGTACACATATATATACACATATGTATACACATATATACACATATAT 385
              ||||||| | | || |||| ||||||| | ||| || |||||||||| ||| ||||| ||
Db         72 ATACACACATACATATACATATATATACACACATATATACACATACATACATATACACAT 131

Qy        386 ACACACATATATACACATATATACACACATATATACACATATATACACATATATACACAT 445
              | | | | ||| | ||||| ||||| ||||| ||||| | ||  | ||||  ||||| ||
Db        132 ATATATACATACATACATACATACATACATACATACATGCATACATGCATACATAT 191

Qy        446 ATACACATATATACACATATATACATATATACACATATATATA-------ATATACACAC 498
              ||| | |||||||||||||| | ||||| ||| |||||||||||       ||||| ||||
Db        192 ATATATATATATACACATACACACATACATATACATATATATACATACACATATATACAC 251

Qy        499 ATATATATACACATATATACACACATATATACACATATATACACATATATATACACATAT 558
              ||| ||||||| | ||||| | | ||||| | ||||| ||||| | ||| ||||| | ||
Db        252 ATACATATACATACATATATATATATATACATACATACATACATACATACATACATACAT 311

Qy        559 ATACACATATATACATATATACACATATATATACATATATACACATATATACATATATAC 618
              | | ||||| ||||||| ||||| | ||| ||||||| ||| | ||||||||||||||||
Db        312 ACATACATACATACATACATACATACATACATACATACATATATATATATACATATATAC 371

Qy        619 ACATATATACATATATACACACATATATACACATACATATACACACACATAGATATACAT 678
              ||||| |||||||||| | | | |||||||||||||||||| | ||| ||||| ||| | |
Db        372 ACATACATACATATACATATATATATATACACATACATACATACATACATACATACATAC 431

Qy        679 ATATATACACATATA 693
              ||| ||||| | |||
Db        432 ATACATACATACATA 446


RESULT 91
BM906116/c
LOCUS       BM906116               1156 bp    mRNA    linear   EST 12-MAR-2002
DEFINITION  AGENCOURT_6621231 NIH_MGC_125 Homo sapiens cDNA clone IMAGE:5589966
            5', mRNA sequence.
ACCESSION   BM906116
VERSION     BM906116.1  GI:19356495
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1156)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Invitrogen
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12363  row: i  column: 07
            High quality sequence start: 323
            High quality sequence stop: 613.
FEATURES             Location/Qualifiers
     source          1. .1156
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5589966"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_125"
                     /note="Organ: ovary (pool of 3); Vector: pCMV-SPORT6;
                     Site_1: EcoRV (destroyed); Site_2: NotI; RNA source pool
                     of three ovaries, from females ranging in age from 38 to
                     49 yo. Library is oligo-dT primed and directionally cloned
                     (EcoRV site is destroyed upon cloning). Average insert
                     size 2.1 kb, insert size range 1-3.5 kb. Library is
                     normalized and enriched for full-length clones and was
                     constructed by C. Gruber (Invitrogen). Research Genetics
                     tracking code 036."
ORIGIN

  Query Match             6.7%;  Score 274.2;  DB 5;  Length 1156;
  Best Local Similarity   76.3%;  Pred. No. 3.6e-36;
  Matches  350;  Conservative    0;  Mismatches  108;  Indels    1;  Gaps    1;

Qy        268 ATATATATACACATATATATACACATATATACATATATACACATATATACACATATATAT 327
              |||| |||| | ||||||||| | ||||||| ||||||| | |||| || | || |||||
Db        688 ATATGTATATATATATATATATATATATATATATATATATATATATGTATATATGTATAT 629

Qy        328 ACACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATAC 387
              | | ||| |||| || | ||||||||| | |||| |||| | ||||||| | |||||||
Db        628 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 569

Qy        388 ACACATATATACACATATATACACACATATATACACATATATACACATATATACACATAT 447
              | | ||||||| | ||||||| | | ||||||| | ||||||| | ||||||| | ||||
Db        568 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 509

Qy        448 ACACATATATACACATATATACATATATACACATATATATAATATACACACATATATATA 507
              | | ||||||| | ||||||| ||||||| | |||||||| ||||| | | |||||||||
Db        508 ATATATATATATATATATATATATATATATATATATATAT-ATATATATATATATATATA 450

Qy        508 CACATATATACACACATATATACACATATATACACATATATATACACATATATACACATA 567
              | ||||||| | | ||||||| | ||||||| | ||||||||| | ||||||| | |||
Db        449 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 390

Qy        568 TATACATATATACACATATATATACATATATACACATATATACATATATACACATATATA 627
              |||| ||||||| | ||||||||| ||||||| | ||||||| ||||||| | |||| |
Db        389 TATATATATATATATATATATATATATATATATATATATATATATATATATATATGTG 330

Qy        628 CATATATACACACATATATACACATACATATACACACACATAGATATACATATATATACA 687
              ||||||  | | |||| || | ||| |||| ||    | ||| || || |||||| ||
Db        329 TATATATCTATATATATGTATATATATATATGCATGTAGATATATCTATATATATGTATC 270

Qy        688 CATATATATACGTATATATATGTATATATATATGCTCCA 726
              || ||||||   |||||| | || |       | |||||
Db        269 TATTTATATATACATATATGTCTAGAGGATCCTCCTCCA 231
```

```
RESULT 92
CC075205/c
LOCUS       CC075205                556 bp    DNA     linear   GSS 16-APR-2003
DEFINITION  CSU-K33r.31C19.T7 CSU-K33r Aedes aegypti genomic clone
            CSU-K33r.31C19, genomic survey sequence.
ACCESSION   CC075205
VERSION     CC075205.1  GI:29916666
KEYWORDS    GSS.
SOURCE      Aedes aegypti (yellow fever mosquito)
  ORGANISM  Aedes aegypti
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Aedes;
            Stegomyia.
REFERENCE   1  (bases 1 to 556)
  AUTHORS   Loftus,B., Shetty,J., Severson,D., Brown,S. and Knudson,D.
  TITLE     End sequencing of Aedes aegypti BACs
  JOURNAL   Unpublished (2003)
COMMENT     Other_GSSs: CSU-K33r.31C19.SP6
            Contact: Brendan Loftus
            Department of Eukaryotic Genomics
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-3543
            Fax: 301-838-0208
            Email: enta@tigr.org
            Library was provided by Susan Brown and Dennis Knudson at Colorado
            State University.
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .556
                     /organism="Aedes aegypti"
                     /mol_type="genomic DNA"
                     /strain="Rexville"
                     /db_xref="taxon:7159"
                     /clone="CSU-K33r.31C19"
                     /clone_lib="CSU-K33r"
                     /note="Vector: pBeloBAC11; Site_1: HindIII"
ORIGIN

  Query Match             6.6%;  Score 273.4;  DB 8;  Length 556;
  Best Local Similarity   71.7%;  Pred. No. 5.5e-36;
  Matches  372;  Conservative    0;  Mismatches  146;  Indels    1;  Gaps    1;

Qy        230 ATATATATATATATATATATATACACACACACACACATATATATATACACATATATATAC 289
              ||||||||||||||||||||||| | | | | | |  |||||.||| | |||||||||
Db        556 ATATATATATATATATATATATATATATATATATATCTATATATATATATATATATATAT 497

Qy        290 ACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATAT 349
              | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | ||||
Db        496 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 437

Qy        350 ATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATAC 409
              |||||.| |||| |||| | ||||||| | ||||||| | | ||||||| | |||||||
Db        436 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 377

Qy        410 ACACATATATACACATATATACACATATATACACATATACACATATATACACATATATAC 469
              | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||||
Db        376 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 317

Qy        470 ATATATACACATATATATAATATACACACATATATATACACATATATACACACATATATA 529
              ||||||| | |||||||| ||||| | | ||||||||| | ||||||| | | |||||||
Db        316 ATATATATATATATATAT-ATATATATATATATATATATATATATATATATATATATATA 258

Qy        530 CACATATATACACATATATATACACATATATACACATATATACATATATACACATATATA 589
              | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||||||
Db        257 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 198

Qy        590 TACATATATACACATATATACATATATACACATATATACATATATACACACATATATACA 649
              || ||||||| | ||||||| |||||   | |||||  |   ||  ||  | | |  |
Db        197 TATATATATATATATATATATATATAAGTATATATAAATTAATCAACCTAGAAAAGTTTC 138

Qy        650 CATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATATG 709
              | |  | |          | |   | | |  || |||   |||  || ||  ||   ||
Db        137 AACAGCTGTCGTGGTGAAAAATAAAAAAAGTAGATAATAATAACTACACTAATCCCTAAT 78

Qy        710 TATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTT 748
              ||| ||  ||||  |          |||| | | |||||
Db         77 TATCTACCTATGTAGCTTGTCCTTTAAGGTGAAACAGTT 39


RESULT 93
BX146163/c
LOCUS       BX146163                648 bp    DNA     linear   GSS 28-JAN-2003
DEFINITION  Danio rerio genomic clone DKEY-104K20, genomic survey sequence.
ACCESSION   BX146163
VERSION     BX146163.1  GI:27977521
KEYWORDS    GSS.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 648)
  AUTHORS   Humphray,S.J., Huckle,E. and Durham,J.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-2003) The Sanger Institute, Wellcome Trust Genome
            Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Unpublished
COMMENT     This sequence was generated from the SP6 end of BAC 104K20. 104K20
            is part of the Daniokey BAC Library created by R. Plasterk and N.V.
            Keygene. Further details:
            http://www.sanger.ac.uk/Projects/D_rerio/.
FEATURES             Location/Qualifiers
     source          1. .648
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-104K20"
                     /tissue_type="Testis"
                     /note="vector pIndigoBAC-536"
ORIGIN

  Query Match             6.6%;  Score 272.2;  DB 9;  Length 648;
  Best Local Similarity   70.4%;  Pred. No. 8.6e-36;
  Matches  392;  Conservative    0;  Mismatches  163;  Indels    2;  Gaps    2;

Qy        238 ATATATATATATATACACACACACACATATATATATACACATATATATACACATATATAT 297
              |||||||||||||| | | | | | | |||||||||||| | ||||||||| | ||||||
Db        648 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 589

Qy        298 ACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACAC 357
              | ||||||| | ||||||| | ||||||||| | ||| |||| || | |||||||||  |
Db        588 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 529

Qy        358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417
              |||| | || | || | ||   ||||||| | | ||||||| | ||||||| | | ||||
Db        528 ATATATGTATATATGTGTATGTATATATATATATATATATATATATATATATATATAT 469

Qy        418 ATACACATATATACACATATATACACATATACACATATATACACATATATACATATATAC 477
              ||| | ||||||| | ||||||| | ||||| | ||||||| | ||||||| ||| ||||
Db        468 ATATATATATATATATATATATATATATATATATATATATATATATATATATAAATACATAC 409

Qy        478 ACATATATATAATATACACACATATATATACACATATATACACACATATATACACATATA 537
              | |||||||| |||||||||| | | | |||| | | ||||||| | | | |||| | |
Db        408 ATATATATAT-ATATACACACACACACACACACACACACACACACACACACACACACACA 350

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATACATATA 597
              |||||.| ||||||| ||||||| | |||||||||||||||| | |||||||| |||||
Db        349 CACACACACATATACATATATATATATATATATACATATATATATATATATATATATA 290

Qy        598 TACACATATATACATATATACACATATATACATATATACACACATATATACACATACATA 657
              || | ||||||| ||||||| | ||||||| ||||||||| |||||| ||  |||   ||
Db        289 TATATATATATATATATATATATATATATATATATATATACATACATATTTAATCATTTTTA 230
```

```
Qy        658 TACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATA 717
              |    |  || | || |       ||||  |  |    |  |  || |||| | | ||| ||
Db        229 T-TTTATTCAAATATTTTATATAATATTTATTTTATTTCATTTAAATATTTTTTTATTTA 171

Qy        718 TATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGA 777
               ||  |  |  |  |||   | | | || ||   |  |   ||| |   | | | |
Db        170 AATATTTAATTTTGTTTAAATATTTGTATTTCATATGCTAGGTAAATATATATATATATA 111

Qy        778 GTTTGACAGCAGTGTAT 794
                |  | |    | |||
Db        110 TATATATATATATATAT 94


RESULT 94
CW020046/c
LOCUS       CW020046               892 bp    DNA     linear   GSS 23-SEP-2004
DEFINITION  ZMMBLc0013O08.f ZMMBLc Zea mays genomic clone ZMMBLc0013O08 5',
            genomic survey sequence.
ACCESSION   CW020046
VERSION     CW020046.1  GI:52613214
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 892)
  AUTHORS   Wing,R., Luo,M., Soderlund,C. and Haller,K.
  TITLE     ZMMBL sequences
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            Plate: 0013  row: O  column: 08
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .892
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /cultivar="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBLc0013O08"
                     /tissue_type="immature ears"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="ZMMBLc"
                     /note="Vector: pAGIBAC1; Site_1: Hpa II; Site_2: Hpa II"
ORIGIN

  Query Match             6.6%;  Score 271;  DB 9;  Length 892;
  Best Local Similarity   72.7%;  Pred. No. 1.3e-35;
  Matches  363;  Conservative    0;  Mismatches  135;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||| | | | | | | |||||||||| | |||||||
Db        790 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 731

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | |||||||||                   |
Db        730 TATATATATATATATATATATATATATATATATATATATATAATATATATATATATTA 671

Qy        347 TATATATACACATATGTATACACATATATACACATATA-TACACACATATATACACATAT 405
              |||||||| | |||| |||| | ||||||| | ||||| || | | ||||||| | ||||
Db        670 TATATATATATATATATATATATATATATATATATATACTATATATATATATATATATAT 611

Qy        406 ATACACACATATATACACATATATACACATATATACACATATACACATATATACACATAT 465
              ||| | | ||||||| | ||||||| | ||||||| | ||||  | ||||||| | ||||
Db        610 ATATATATATATATATATATATATATATATATATATATATATTTATATATATATATAT 551
```

```
Qy        466 ATACATATATACACATATATATAATATACACACATATATATACACATATATACACACATA 525
              ||| ||||||| | |||||||||         | |||    || |    |||    | ||
Db        550 ATATATATATATATATATATATATATATATAATATAATATTATAATATTATTATTATATT 491

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
               ||    |||    |    |||||    ||||||| | ||||||| ||||||| | |||
Db        490 ATTAATATTATTATTATTATTATATTATTATATATATATATATATATATATATATATA 431

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db        430 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 371

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db        370 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 311

Qy        706 TATGTATATATATATGCTC 724
              ||| ||||||||||||| |
Db        310 TATATATATATATATGCAC 292


RESULT 95
AZ318983
LOCUS       AZ318983               647 bp    DNA     linear   GSS 29-SEP-2000
DEFINITION  1M0038D20F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0038D20 F, genomic survey sequence.
ACCESSION   AZ318983
VERSION     AZ318983.1  GI:10369306
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 647)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,
            Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von
            Niederhausern,A. and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0038  row: D  column: 20
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 647.
FEATURES             Location/Qualifiers
     source          1. .647
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0038D20"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
```

adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to adaptored vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

ORIGIN

Query Match 6.6%; Score 270.6; DB 8; Length 647;
Best Local Similarity 77.6%; Pred. No. 1.6e-35;
Matches 353; Conservative 0; Mismatches 99; Indels 3; Gaps 2;

```
Qy      236 ATATATATATATATATACACACACACACATATATATATACACATATATATACACATAT 295
            ||| ||||||| ||| | ||| | | ||| | ||||||| | ||||| ||||| |||||
Db        1 ATACATATATACATACATACATATATACATACATATATACATACATACATATATACATAC 60

Qy      296 ATACATATATACACATATATACACATATATATACACATACATATGTACACATATATATAC 355
            ||| ||| ||||| | ||||| ||||| ||  | | |||||||  |||| ||||| ||||
Db       61 ATATATACATACATACATATATACATACATGCATATATACATACATACATATATACATAC 120

Qy      356 ACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACAT 415
            | | ||  ||| | | |||||||||||||| ||| ||||||| ||| | |||||||| |
Db      121 ATACATACATATATACATATACACATATACACATACATATACACATACATATACACATAC 180

Qy      416 ATATACACATATATACACATATATACACATATACACATATATACACATATATACATATAT 475
            |||||||||||| ||| ||| |||||||||||||||||||||| | | ||| ||||||| ||
Db      181 ATATACACATACATATACACATATACACATATACACATATACATATATACATACATACAT 240

Qy      476 ACACATATATATAATATACACACATATATATACACATATATACACACATATATACACATA 535
            | | | ||| || ||||||| ||||| ||||| ||||| ||| | ||||| ||  |||||
Db      241 ATATACATACAT-ATATACATACATACATATATACATACATATATACATACAT--ACATA 297

Qy      536 TATACACATATATATACACATATATACACATATATACATATATACACATATATATACATA 595
            |||||| | ||| ||||| ||||| | |||||||||||||| ||||| | ||||||||||
Db      298 TATACATACATACATACATATATACATACATATATACATACATACATACATATATACATA 357

Qy      596 TATACACATATATACATATATACACATATATACATATATACACACATATATACACATACA 655
            |||||| | ||| | | ||||| |||||||||||||| ||||| | ||| ||| | | | |
Db      358 TATACATACATACATACATATATACATATATACATACATACATATATACATATATACATA 417

Qy      656 TATACACACACATAGATATACATATATATACACAT 690
            |||||| |    |     ||   | ||| ||||||
Db      418 TATACATAACACTTCTGTTAAGCAAATAGACACAT 452
```

RESULT 96
BU507480
LOCUS BU507480 1836 bp mRNA linear EST 12-SEP-2002
DEFINITION AGENCOURT_10309824 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:6501420 5', mRNA sequence.
ACCESSION BU507480
VERSION BU507480.1 GI:22813713
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1836)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM14055 row: n column: 13
High quality sequence start: 74
High quality sequence stop: 447.
FEATURES Location/Qualifiers
source 1. .1836
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6501420"
/tissue_type="leiomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_71"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2.1 kb. "
ORIGIN

Query Match 6.6%; Score 270.6; DB 5; Length 1836;
Best Local Similarity 72.5%; Pred. No. 1.4e-35;
Matches 359; Conservative 0; Mismatches 135; Indels 1; Gaps 1;

```
Qy      308 ACATATATACACATATATATACACATACATATGTACACATATATATACACATATGTATAC 367
            | || |||| | |||||||||| | ||| |||| || | |||||||||| | |||| ||||
Db       78 ANATNTATATATATATATATATATATATATATATATATATATATATATATATATATAT 137

Qy      368 ACATATATACACATATATACACACATATATACACATATATACACACATATATACACATAT 427
            | ||||||| | ||||||| | ||||||| | ||||||| | | ||||||| | ||||
Db      138 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 197

Qy      428 ATACACATATATACACATATACACATATATACACATATATACATATATACACATATATAT 487
            ||| | ||||||| | ||||| | ||||||| | ||||||| ||||||| | ||||||||
Db      198 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 257

Qy      488 A-ATATACACACATATATATACACATATATACACACATATATACACATATATACACATAT 546
            | ||||| | | |||||||||| | ||||||| | | ||||||| | ||||||| | ||||
Db      258 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 317

Qy      547 ATATACACATATATACACATATATACATATATACACATATATATACATATATACACATAT 606
            ||||| | ||||||| | ||||||| ||||||| | ||||||||| ||||||| | ||||
Db      318 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 377

Qy      607 ATACATATATACACATATATACATATATACACACATATATACACATACATATACACACAC 666
            ||| ||||||| | ||||||| ||||||| | | ||||||| | ||| ||||| | | |
Db      378 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 437

Qy      667 ATAGATATACATATATATACACATATATATACGTATATATATGTATATATATATGCTCCA 726
            ||| ||||  |||||||||   |||||||  || | | | ||| |||||| |      |
Db      438 ATATATATGTATATATATATNTATATATATGTGTGTGTGTGTGTGTATATAGAAAAAAGA 497

Qy      727 GAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAG 786
            |||        | |   |    ||  ||| |     || | ||  ||| |||||   || ||
Db      498 GAGAGAGAGTGTGTGTGTGTGTGTACAAAAAAAGAAAAGAGAAAAGAGAGAAAGAGAG 557

Qy      787 CAGTGTATTGTGAGA 801
                | | ||||
Db      558 ACAGAGAGAGAGAGA 572
```

RESULT 97
BH329576/c
LOCUS BH329576 458 bp DNA linear GSS 03-DEC-2001
DEFINITION CH230-106A4.TV CHORI-230 Segment 1 Rattus norvegicus genomic clone CH230-106A4, genomic survey sequence.
ACCESSION BH329576
VERSION BH329576.1 GI:17260290
KEYWORDS GSS.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 458)

```
  AUTHORS   Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K.,
            Shvartsbeyn,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D.,
            Riggs,F., de Jong,P. and Fraser,C.M.
  TITLE     Rat BAC End Sequences from Library CHORI-230 EcoRI segment
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: CH230-106A4.TJ
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the rat BAC library CHORI-230
            (http://www.chori.org/bacpac/rat230.htm). For BAC library
            availability, please contact Pieter de Jong (pdejong@mail.cho.org).
            Clones may be purchased from BACPAC Resources
            (http://www.chori.org/bacpac/or ering_information.htm). BAC end
            page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
            Plate: 106  row: A  column: 4
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .458
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /strain="BN/SsNHsd/MCW"
                     /db_xref="taxon:10116"
                     /clone="CH230-106A4"
                     /sex="Female"
                     /cell_type="Brain"
                     /clone_lib="CHORI-230 Segment 1"
                     /note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;
                     CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
                     Pieter de Jong"
ORIGIN

  Query Match             6.5%;  Score 269.8;  DB 8;  Length 458;
  Best Local Similarity   77.2%;  Pred. No. 2.3e-35;
  Matches  353;  Conservative    0;  Mismatches  102;  Indels    2;  Gaps    2;

Qy        230 ATATATATATATATATATATATACACACACACACACATATATATATACACATATATATAC 289
              ||||||||||||||||||||||| | | | |   |||||||||||| | |||||||||
Db        458 ATATATATATATATATATATATATATATATAGGGATATATATATATATATATATATATAT 399

Qy        290 ACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATAT 349
              | ||||||| ||||||| | ||||||| | |||||||||| | ||| |||| || | ||||
Db        398 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 339

Qy        350 ATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATAC 409
              ||||| | |||| |||| | ||||||| | ||||||| | | |||||||| | |||||||
Db        338 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 279

Qy        410 ACACATATATACACATATATACACATATATACACATATACACATATATACACATATATAC 469
              | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||||
Db        278 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 219

Qy        470 ATATATACACATATATATAATATACACACATATATATACACATATATACACACATATATA 529
              ||||||| | |||||||| ||||| | | ||||||||| | ||||||| | | |||||||
Db        218 ATATATATATATATATAT-ATATATATATATATATATATATATATATATATATATATATA 160

Qy        530 CACATATATACACATATATATACACATATATACACATATATACATATATACACATATATA 589
                | |||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||||||
Db        159 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 100

Qy        590 TACATATATACACATATATACATATATACACATATATACATATAT-ACACACATATATAC 648
              || ||||||| | ||||||| ||||||| | |||||||  | | | ||||   | | ||
Db         99 TATATATATATATATATATATATATATATATATATATATATTTCTCTCACACTATCACAGAC 40

Qy        649 ACATACATATACACACACATAGATATACATATATATA 685
              ||| | | | | | || |||  | | | | | | |
Db         39 ACAGAGAGACAGAGAGACGTAGTGACAGACACAGAGA 3
```

```
RESULT 98
CW518121/c
LOCUS       CW518121               663 bp    DNA     linear   GSS 08-OCT-2004
DEFINITION  OP__Ba0025C13.f OP__Ba Oryza punctata genomic clone OP__Ba0025C13
            5', genomic survey sequence.
ACCESSION   CW518121
VERSION     CW518121.1  GI:53992343
KEYWORDS    GSS.
SOURCE      Oryza punctata
  ORGANISM  Oryza punctata
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 663)
  AUTHORS   SanMiguel,P., Westerman,R., Kim,H., Yu,Y., Wissotski,M., Yost,D.,
            Stum,D., Rao,K., Luo,M., Jetty,R., Kudrna,D., Muller,C.,
            Hatfield,J., Soderlund,C., Wing,R. and Jackson,S.A.
  TITLE     OMAP Project - Purdue University
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Scott A. Jackson
            Jackson Laboratory
            Purdue University
            915 W. State St., West Lafayette, IN 47907, USA
            Tel: 7654963621
            Fax: 7654967255
            Email: sjackson@purdue.edu
            Basecalling by phred version 0.020425.c. This sequence was derived
            from the raw sequence read by clipping with lucy version 1.19s.
            Bases 50-712 of the raw sequence (length 1364) were retained after
            clipping.
            PCR PRimers
            FORWARD: TAA TAC GAC TCA CTA TAG GG
            BACKWARD: CAC TCA TTA GGC ACC CCA
            Insert Length: 161000   Std Error: 0.00
            Plate: 0025  row: C  column: 13
            Seq primer: TAA TAC GAC TCA CTA TAG GG
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .663
                     /organism="Oryza punctata"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4537"
                     /clone="OP__Ba0025C13"
                     /tissue_type="young leaves"
                     /lab_host="DH10B-T1 phage resistant"
                     /clone_lib="OP__Ba"
                     /note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"
ORIGIN

  Query Match             6.5%;  Score 268.8;  DB 9;  Length 663;
  Best Local Similarity   77.9%;  Pred. No. 3.2e-35;
  Matches  349;  Conservative    0;  Mismatches   97;  Indels    2;  Gaps    2;

Qy        227 TTTATATATATATATATATATATATACACACACACACATATATATATACACATATATA 286
              |||||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db        660 TTTATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 601

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | |||||||||| | ||| || | ||   |
Db        600 TATATATATATATATATATATATATATATATATATATATATATATATATATTTATA-TTA 542

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              ||||| ||   |||| |||  | || | || | ||||||| | | ||||||| | |||||
Db        541 TATATCTATCTATATTTATTTATATGTGTATATATATATATATATATATATATATATATA 482

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db        481 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 422

Qy        467 TACATATATACACATATATATAATATACACACATATATATACACATATATACACACATAT 526
              || ||||||| | |||||||| ||||| | | ||||||||| | ||||||| | | ||||
Db        421 TATATATATATATATATATAT-ATATATATATATATATATATATATATATATATATATAT 363
```

```
Qy        527 ATACACATATATACACATATATATACACATATATACACATATATACATATATACACATAT 586
              ||| | ||||||| | |||||||||| | ||||||| | ||||||| ||||||| | ||||
Db        362 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 303

Qy        587 ATATACATATATACACATATATACATATATACACATATATACATATATACACACATATAT 646
              ||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | ||||||
Db        302 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 243

Qy        647 ACACATACATATACACACACATAGATAT 674
              | | ||| |||        ||| || ||
Db        242 ATATATATATATCTTTTTGCATTGACAT 215


RESULT 99
CW019737
LOCUS       CW019737                600 bp    DNA     linear   GSS 23-SEP-2004
DEFINITION  ZMMBLc0013H15.r ZMMBLc Zea mays genomic clone ZMMBLc0013H15 3',
            genomic survey sequence.
ACCESSION   CW019737
VERSION     CW019737.1  GI:52612602
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 600)
  AUTHORS   Wing,R., Luo,M., Soderlund,C. and Haller,K.
  TITLE     ZMMBL sequences
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Rod A. Wing
            Arizona Genomics Institute
            University of Arizona
            Forbes Building Room 303, Tucson, AZ 85721-0036, USA
            Tel: 520 626 9595
            Fax: 520 621 1259
            Email: http://genome.arizona.edu
            Plate: 0013  row: H  column: 15
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .600
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /cultivar="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBLc0013H15"
                     /tissue_type="immature ears"
                     /lab_host="DH10B T1 phage resistant"
                     /clone_lib="ZMMBLc"
                     /note="Vector: pAGIBAC1; Site_1: Hpa II; Site_2: Hpa II"
ORIGIN

  Query Match             6.5%;  Score 267.2;  DB 9;  Length 600;
  Best Local Similarity   78.8%;  Pred. No. 6e-35;
  Matches  331;  Conservative    0;  Mismatches   88;  Indels    1;  Gaps    1;

Qy        228 TTATATATATATATATATATATATACACACACACACACATATATATATACACATATATAT 287
              || | ||  | | | |     | ||||||||||||||||||||||||| | ||||||||
Db        182 TTCTCTAATTTTTTTTGCTCGAACACACACACACACACATATATATATATATATATATAT 241

Qy        288 ACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACAT 347
              | | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | ||
Db        242 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 301

Qy        348 ATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATAT 407
              ||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||||
Db        302 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 361

Qy        408 ACACACATATATACACATATATACACATATATACACATATACACATATATACACATATAT 467
              | | | ||||||| | ||||||| | ||||||| | ||||| | |||||||   |||| |
Db        362 ATATATATATATATATATATATATATATATATATATATATATATATATATATGTATATGT 421
```

```
Qy        468 ACATATATACACATATATATAATATACACACATATATATACACATATATACACACATATA 527
              | ||||||| | |||||||| ||||| | | ||||||||| | ||||||| | | |||||
Db        422 ATATATATATATATATATAT-ATATATATATATATATATATATATATATATATATATA 480

Qy        528 TACACATATATACACATATATATACACATATATACACATATATACATATATACACATATA 587
              || | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||||
Db        481 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 540

Qy        588 TATACATATATACACATATATACATATATACACATATATACATATATACACACATATATA 647
              |||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||||
Db        541 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 600


RESULT 100
CC075205
LOCUS       CC075205                556 bp    DNA     linear   GSS 16-APR-2003
DEFINITION  CSU-K33r.31C19.T7 CSU-K33r Aedes aegypti genomic clone
            CSU-K33r.31C19, genomic survey sequence.
ACCESSION   CC075205
VERSION     CC075205.1  GI:29916666
KEYWORDS    GSS.
SOURCE      Aedes aegypti (yellow fever mosquito)
  ORGANISM  Aedes aegypti
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Aedes;
            Stegomyia.
REFERENCE   1  (bases 1 to 556)
  AUTHORS   Loftus,B., Shetty,J., Severson,D., Brown,S. and Knudson,D.
  TITLE     End sequencing of Aedes aegypti BACs
  JOURNAL   Unpublished (2003)
COMMENT     Other_GSSs: CSU-K33r.31C19.SP6
            Contact: Brendan Loftus
            Department of Eukaryotic Genomics
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-3543
            Fax: 301-838-0208
            Email: enta@tigr.org
            Library was provided by Susan Brown and Dennis Knudson at Colorado
            State University.
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .556
                     /organism="Aedes aegypti"
                     /mol_type="genomic DNA"
                     /strain="Rexville"
                     /db_xref="taxon:7159"
                     /clone="CSU-K33r.31C19"
                     /clone_lib="CSU-K33r"
                     /note="Vector: pBeloBAC11; Site_1: HindIII"
ORIGIN

  Query Match             6.5%;  Score 267;  DB 8;  Length 556;
  Best Local Similarity   71.4%;  Pred. No. 6.6e-35;
  Matches  365;  Conservative    0;  Mismatches  145;  Indels    1;  Gaps    1;

Qy        186 CAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATAT 245
              || |  |  || | |  |  | | | | || | |  | | ||  |||   ||| |
Db         47 CACCTTAAAGGACAAGCTACATAGGTAGATAATTAGGGATTAGTGTAGTTATTATTATCT 106

Qy        246 ATATATACACACACACACATATATATATACACATATATATACACATATATACATATAT 305
              | | |          |||    | |  | |  | |  | | ||     ||  |  |||||
Db        107 ACTTTTTTTATTTTTCACCACGACAGCTGTTGAAACTTTTCTAGGTTGATTAATTTATAT 166

Qy        306 ACACATATATACACATATATATACACATACATATGTACACATATATATACACATATGTAT 365
              | || |||||| | ||||||||| | ||| |||| || | ||||||||| | |||| |||
Db        167 ATACTTATATATATATATATATATATATATATATATATATATATATATATATATATAT 226

Qy        366 ACACATATATACACATATATACACACATATATACACATATATACACACATATATACACAT 425
              | | ||||||| | ||||||| | | ||||||| | ||||||| | | ||||||| | ||
Db        227 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 286
```

```
Qy        426 ATATACACATATATACACATATACACATATATACACATATATACATATATACACATATAT 485
              ||||| | ||||||| | ||||| | ||||||| | ||||||| ||||||| | ||||||
Db        287 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 346

Qy        486 ATAATATACACACATATATATACACATATATACACACATATATACACATATATACACATA 545
              || ||||| | | ||||||||| | ||||||| | | ||||||| | ||||||| | |||
Db        347 AT-ATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 405

Qy        546 TATATACACATATATACACATATATACATATATACACATATATATACATATATACACATA 605
              |||||| | ||||||| | ||||||| ||||||| | ||||||||| ||||||| | |||
Db        406 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 465

Qy        606 TATACATATATACACATATATACATATATACACACATATATACACATACATATACACACA 665
              |||| ||||||| | ||||||| ||||||| | | ||||||| | ||| ||||| | | |
Db        466 TATATATATATATATATATATATATATATATATATATATATATATATATATATAGATATA 525

Qy        666 CATAGATATACATATATATACACATATATAT 696
               ||| ||||| ||||||||| | ||||||||
Db        526 TATATATATATATATATATATATATATATAT 556
```

Search completed: April 22, 2005, 04:38:45
Job time : 8269.88 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 20, 2005, 22:52:45 ; Search time 474.843 Seconds
(without alignments)
13936.065 Million cell updates/sec

Title: US-09-939-209A-3_COPY_11056_12145
Perfect score: 1090
Sequence: 1 tgctttccgtgtggtggtgc.....taactctaggatctcaggtg 1090

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5633728 seqs, 3035525691 residues

Total number of hits satisfying chosen parameters: 11267456

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq:*
19: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
20: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:*
21: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
22: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1090	100.0	20300	10	US-09-939-209A-3	Sequence 3, Appli
c 2	54.8	5.0	6118	15	US-10-311-455-1004	Sequence 1004, Ap
3	54.4	5.0	461	10	US-09-814-353-17724	Sequence 17724, A
c 4	52.8	4.8	6759	15	US-10-311-455-711	Sequence 711, App
c 5	52.8	4.8	198522	13	US-10-087-192-244	Sequence 244, App
6	52.4	4.8	6191	15	US-10-311-455-1189	Sequence 1189, Ap
c 7	52.2	4.8	499	18	US-10-021-323-10189	Sequence 10189, A
c 8	51.8	4.8	8056	18	US-10-473-126-240	Sequence 240, App
9	51.4	4.7	15373	15	US-10-311-455-440	Sequence 440, App
c 10	51.4	4.7	394468	19	US-10-741-600-17952	Sequence 17952, A
11	51.2	4.7	523	17	US-10-424-599-69112	Sequence 69112, A
c 12	51.2	4.7	12393	15	US-10-311-455-1236	Sequence 1236, Ap
13	51	4.7	8056	18	US-10-473-126-386	Sequence 386, App
14	50.8	4.7	8056	18	US-10-473-126-240	Sequence 240, App
c 15	50.6	4.6	8222	19	US-10-486-319A-43	Sequence 43, Appl
c 16	50.6	4.6	8222	19	US-10-486-319A-65	Sequence 65, Appl
c 17	50.4	4.6	436	10	US-09-814-353-5048	Sequence 5048, Ap
c 18	50.4	4.6	436	10	US-09-814-353-11340	Sequence 11340, A
c 19	50.4	4.6	34688	18	US-10-433-793-90	Sequence 90, Appl
c 20	50.2	4.6	8056	18	US-10-473-126-386	Sequence 386, App
21	50.2	4.6	50000	19	US-10-706-635-24	Sequence 24, Appl
c 22	50	4.6	18988	15	US-10-311-455-674	Sequence 674, App
c 23	50	4.6	18988	15	US-10-240-485-62	Sequence 62, Appl
c 24	50	4.6	18988	17	US-10-221-714A-64	Sequence 64, Appl
c 25	49.6	4.6	5195	15	US-10-311-455-1570	Sequence 1570, Ap
26	49.6	4.6	8169	17	US-10-221-714A-10	Sequence 10, Appl
c 27	49.6	4.6	20486	15	US-10-240-485-164	Sequence 164, App
c 28	49.6	4.6	1691139	14	US-10-067-514-1	Sequence 1, Appli
c 29	49.6	4.6	1691139	17	US-10-419-723-1	Sequence 1, Appli
c 30	49.4	4.5	2000	9	US-09-938-842A-4474	Sequence 4474, Ap
c 31	49.4	4.5	2000	11	US-09-938-842A-4474	Sequence 4474, Ap
32	49.2	4.5	6102	15	US-10-311-455-1472	Sequence 1472, Ap
c 33	49.2	4.5	6244	15	US-10-311-455-457	Sequence 457, App
c 34	49.2	4.5	34688	18	US-10-433-793-89	Sequence 89, Appl
c 35	48.8	4.5	6265	15	US-10-311-455-1389	Sequence 1389, Ap
c 36	48.8	4.5	7455	15	US-10-311-455-1731	Sequence 1731, Ap
37	48.8	4.5	19659	15	US-10-311-455-739	Sequence 739, App
c 38	48.6	4.5	93544	18	US-10-322-281-676	Sequence 676, App
39	48.6	4.5	142318	13	US-10-087-192-1504	Sequence 1504, Ap
40	48.4	4.4	5641	15	US-10-311-455-1369	Sequence 1369, Ap
c 41	48.4	4.4	5689	14	US-10-239-676-90	Sequence 90, Appl
c 42	48.4	4.4	5689	15	US-10-240-453-100	Sequence 100, App
c 43	48.4	4.4	5689	17	US-10-221-714A-148	Sequence 148, App
c 44	48.4	4.4	6106	15	US-10-311-455-1445	Sequence 1445, Ap
c 45	48.4	4.4	6106	17	US-10-257-166-113	Sequence 113, App

ALIGNMENTS

RESULT 1
US-09-939-209A-3
; Sequence 3, Application US/09939209A
; Publication No. US20030113721A1
; GENERAL INFORMATION:
; APPLICANT: LEVITT, PAT R.
; APPLICANT: MIRNICS, KAROLY
; APPLICANT: KODAVALI, MENKATA CHOWDARI
; APPLICANT: NIMGAONKAR, VISHWAJIT L.
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR FACILITATING THE DIAGNOSIS AND TREATMENT
; TITLE OF INVENTION: SCHIZOPHRENIA
; FILE REFERENCE: 00-539-US
; CURRENT APPLICATION NUMBER: US/09/939,209A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: 60/228,021
; PRIOR FILING DATE: 2000-08-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 20300
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: A genomic sequence containing RGS4 nucleic acid sequence and sequ
; OTHER INFORMATION: ences upstream and downstream to the RGS4 nucleic acid sequence
US-09-939-209A-3

Query Match 100.0%; Score 1090; DB 10; Length 20300;
Best Local Similarity 100.0%; Pred. No. 6.8e-242;
Matches 1090; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGCTTTCCGTGTGGTGGTGCAGTAACTTTCACTCTCTATGTTCTTATAAGCAAATGTTAC 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

```
Db      11056 TGCTTTCCGTGTGGTGGTGCAGTAACTTTCACTCTCTATGTTCTTATAAGCAAATGTTAC 11115

Qy         61 AATGAGATATGAGTTTTAAAGCCAGATCTTCCTTATCTCTCTGCCCCATCTCTAGTTCTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11116 AATGAGATATGAGTTTTAAAGCCAGATCTTCCTTATCTCTCTGCCCCATCTCTAGTTCTT 11175

Qy        121 GAAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11176 GAAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTT 11235

Qy        181 TGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11236 TGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAA 11295

Qy        241 AGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11296 AGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGA 11355

Qy        301 AATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11356 AATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCA 11415

Qy        361 TATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11416 TATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTG 11475

Qy        421 CTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11476 CTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGA 11535

Qy        481 AAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGTCATTCAGGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11536 AAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGTCATTCAGGA 11595

Qy        541 GTATGGAAGTATCGATGGCTTCTTTAAAATGAAGCAGGAGGGTCTGGCAGAGAGTATCTA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11596 GTATGGAAGTATCGATGGCTTCTTTAAAATGAAGCAGGAGGGTCTGGCAGAGAGTATCTA 11655

Qy        601 TGAAATAAGTTCCTCTGACCTTCACGCTTAATTTTCTGAATGGAGTGGAGCAAATTACTT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11656 TGAAATAAGTTCCTCTGACCTTCACGCTTAATTTTCTGAATGGAGTGGAGCAAATTACTT 11715

Qy        661 CAAGCTTCACTTAACTTGCATATGAAATGAACCGTACAAAAAATACAAGAGTGTCAGGAGA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11716 CAAGCTTCACTTAACTTGCATATGAAATGAACCGTACAAAAATACAAGAGTGTCAGGAGA 11775

Qy        721 AAGTTATGCTCTGGTAAATATTTTGCAAAACAGATAAAAGATAATACTAGAGCTCTGTCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11776 AAGTTATGCTCTGGTAAATATTTTGCAAAACAGATAAAAGATAATACTAGAGCTCTGTCC 11835

Qy        781 TCAAAGAGTTAAGCAGCTAATCTAAGGAGGTAAACTCTATGTCAGCAGGATGAACTGCTC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11836 TCAAAGAGTTAAGCAGCTAATCTAAGGAGGTAAACTCTATGTCAGCAGGATGAACTGCTC 11895

Qy        841 TTCCCTTTCCTCCTCAATAAATTGCAAATCATCTAGTCCAACATCTTTACCACCAGTGCC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11896 TTCCCTTTCCTCCTCAATAAATTGCAAATCATCTAGTCCAACATCTTTACCACCAGTGCC 11955

Qy        901 TGAGGCTCCAGAGGAGCCATTGCCTTCTCAAGGTCACATAGGTGGTGGGTGAGTTAGGAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11956 TGAGGCTCCAGAGGAGCCATTGCCTTCTCAAGGTCACATAGGTGGTGGGTGAGTTAGGAC 12015

Qy        961 CAAATCTAGAATTCCTGACTCCAGTAACTTCTGAAGTCATTTTGTTTTTTATTTTTATGG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12016 CAAATCTAGAATTCCTGACTCCAGTAACTTCTGAAGTCATTTTGTTTTTTATTTTTATGG 12075

Qy       1021 TTTTATTATAAGAATACTTGCTAAGCACACTTACCCCCTGCATTGATTAATAACTCTAGG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12076 TTTTATTATAAGAATACTTGCTAAGCACACTTACCCCCTGCATTGATTAATAACTCTAGG 12135

Qy       1081 ATCTCAGGTG 1090
              ||||||||||
Db      12136 ATCTCAGGTG 12145
```

```
RESULT 2
US-10-311-455-1004/c
; Sequence 1004, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1004
;   LENGTH: 6118
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1004

  Query Match             5.0%;  Score 54.8;  DB 15;  Length 6118;
  Best Local Similarity   48.4%;  Pred. No. 0.062;
  Matches  210;  Conservative    0;  Mismatches  222;  Indels    2;  Gaps    2;

Qy         94 TATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATAT 153
              ||   || |  ||    || | || || || |  |   |||    ||  |  |  || |
Db       2830 TAAACCTATAACCTTAATCAAATTATTTAAATTCCCTTTATCTCCTTTATAAACTAAAAA 2771

Qy        154 TGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTG 213
              | || || || |  |    |||  | ||  | |  | || |  | | | |||     |
Db       2770 TAATAATAAC-ACCCTCCAAATTATATTAAAAATTACCTAAATTAATATACACAAAATCC 2712

Qy        214 TTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTT 273
              |||  ||    ||||  | ||||  ||| ||  || | | |       ||| |  |   |
Db       2711 TTATAATTACACCTTTTTATTTTATAAACACTATATAAATACTTCAATATACTATATATA 2652

Qy        274 TTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAG 333
              |        |  | |       ||| |   |  | | |   | | |||| | | ||
Db       2651 TATAAAAATCTATACCCAAAATATATTTAAAAAATAAATAATAAATCAAAAAATATTTTA 2592

Qy        334 TTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAA 393
              ||   | || |    || |  | |  | |||   |||  | || ||    | ||    |
Db       2591 TTATTACTTCATACAAATAACATTTTAAATTCTTTTAAATTATTTTCTCTTTAATTCAAT 2532

Qy        394 AAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCAC 453
              ||     |   ||||   ||||||  | |     ||||  |    |||||  |||||   |
Db       2531 AACAATATTATAAACAAATACTAAATCCC-AAATATCCTAAAATAAATAAAAATTTTAAA 2473

Qy        454 AAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTT 513
               ||||   |||   ||||||| ||| |||| | ||| |  ||   | |  || |||
Db       2472 CAATTAATTTTCATAAATAATATCAAAAAATAAAATATCATAAACATTTCTATAACTCCA 2413

Qy        514 ACTTTCCAAATATT 527
              | |   | ||||||
Db       2412 ATTACACTAATATT 2399


RESULT 3
US-09-814-353-17724
; Sequence 17724, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
;  APPLICANT: Lee, John
```

```
;  APPLICANT:  Thompson, Pamela
;  APPLICANT:  Lillie, James
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
;  TITLE OF INVENTION:  IDENTIFICATION, ASSESSMENT, PREVENTION, AND
;  TITLE OF INVENTION:  THERAPY OF OVARIAN CANCER
;  FILE REFERENCE: MRI-006B
;  CURRENT APPLICATION NUMBER: US/09/814,353
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: US 60/191,031
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: US 60/207,124
;  PRIOR FILING DATE: 2000-05-25
;  PRIOR APPLICATION NUMBER: US 60/211,940
;  PRIOR FILING DATE: 2000-06-15
;  PRIOR APPLICATION NUMBER: US 60/216,820
;  PRIOR FILING DATE: 2000-07-07
;  PRIOR APPLICATION NUMBER: US 60/220,661
;  PRIOR FILING DATE: 2000-07-25
;  PRIOR APPLICATION NUMBER: US 60/257,672
;  PRIOR FILING DATE: 2000-12-21
;  NUMBER OF SEQ ID NOS: 22037
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17724
;   LENGTH: 461
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 195, 243, 244, 245, 397, 398, 399
;   OTHER INFORMATION: n = A,T,C or G
US-09-814-353-17724

  Query Match             5.0%;  Score 54.4;  DB 10;  Length 461;
  Best Local Similarity   47.1%;  Pred. No. 0.022;
  Matches  154;  Conservative    0;  Mismatches  173;  Indels    0;  Gaps    0;

Qy        199 AAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACA 258
              |||   |   |||| || | |          |||||   ||||      | | | | | |
Db        110 AAAATAAAAATTTTTTTTAAAATAATAAAAAATTTTAAAAAAATTAAAAAAAAAAAAAAA 169

Qy        259 GCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGA 318
              |||  |  | |||   ||| || ||||||   | ||| |||   |   | | ||| |
Db        170 AATATAAAAATAATTTATTTTTTTTNAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 229

Qy        319 TCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCT 378
              |||  ||          ||||  ||| | ||||  |    | |  | |  | ||||
Db        230 AATTATTTTTTTTNNNAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATTTAAAAA 289

Qy        379 TGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCA 438
              |   |||   |||||   ||   |||   | | |       | || |   |    |
Db        290 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATTTTTAAAAAAAAA 349

Qy        439 AATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGT 498
              || | |||||        || | ||||  ||  ||    | ||||     ||||| | |
Db        350 AAAAAAAATTTTTTTTTTTTATTTTTTAAATAAAAAAAAAAAAATTNNNTTTTTTTTT 409

Qy        499 ACTATTAAAACATTTACTTTCCAAATA 525
              | ||      |||  ||   ||| |
Db        410 TTTTTTTTTTTTTTTTTTAAAAAAAA 436


RESULT 4
US-10-311-455-711/c
; Sequence 711, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 711
;   LENGTH: 6759
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-711

  Query Match             4.8%;  Score 52.8;  DB 15;  Length 6759;
  Best Local Similarity   46.7%;  Pred. No. 0.19;
  Matches  168;  Conservative    0;  Mismatches  192;  Indels    0;  Gaps    0;

Qy        157 TCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTA 216
              | |  | |||  |  || || ||    | |  || | |  || || |   | | || | |
Db       2866 TAAAAAAAAAATATATACTATTTAATCAAATAATTTTACTTTTAAAAATATATCTTATAA 2807

Qy        217 ATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTG 276
              | ||||     | |   |  |  || |  ||| | | | |  ||||    | || ||
Db       2806 AAATACAAATATAAACATACTTTAACAAATTATATAAAAAAATTTATTTCAATATTATTC 2747

Qy        277 GTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTG 336
                  |||||||    |||||  ||    ||   |  || |  ||| || |   | |
Db       2746 AAAATATCAAAAATAAAATAAAATAAATACTTTTTATTAAAAAAATAAATAAATTATTCT 2687

Qy        337 AAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAA 396
              | || | || || ||    |   | ||  | | |||   ||          |||   |||||
Db       2686 ATAAATTAAAACTAACAAATAACTTAAAAAAATACCCCATTAAAATACTAAAAAAAAAAA 2627

Qy        397 GTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAA 456
              | |  |    | | ||| |    | | |||    |   || |  || ||| |
Db       2626 AAACAACATACAAAAAAATACTTATAATATAATCCTTATTTCAAAAATAAATACTAATTT 2567

Qy        457 TTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACT 516
              ||  |  |     |  ||| | |  | | | |    |  |  ||||  || ||  | || |
Db       2566 TTAAAAATCTATATAAATATAAACACAAACACACATAAACATACTCATATAATCTATAAT 2507


RESULT 5
US-10-087-192-244/c
; Sequence 244, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
;  APPLICANT: Morris, David W.
;  APPLICANT:  Engelhard, Eric K.
;  TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
;  TITLE OF INVENTION:  CANCER
;  FILE REFERENCE: 529452000122
;  CURRENT APPLICATION NUMBER: US/10/087,192
;  CURRENT FILING DATE:  2002-03-01
;  PRIOR APPLICATION NUMBER: US 09/747,377
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: US 09/798,586
;  PRIOR FILING DATE: 2001-03-02
;  NUMBER OF SEQ ID NOS: 2059
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 244
;   LENGTH: 198522
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(198522)
;   OTHER INFORMATION: n = A,T,C or G
US-10-087-192-244
```

```
  Query Match             4.8%;  Score 52.8;  DB 13;  Length 198522;
  Best Local Similarity   46.9%;  Pred. No. 0.98;
  Matches  238;  Conservative    0;  Mismatches  262;  Indels    8;  Gaps    2;

Qy         33 TCTCTATGTTCTTATAAGCAAATGTTACAATGAGATATGAGTTTTAAAGCCAGATCTTCC 92
              | | |||  | | |  |  |||   || | | | |||| |  | | |  |  | |   |
Db      84632 TATATATAATGTAAATATAAAAATATATATTTATATATAATATATTTATATATAATATAT 84573

Qy         93 TTATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATA 152
              |||| |         || | ||  | |  |   | |   |  | |||  |  |   |||
Db      84572 TTATATTATATTTATATTTATAAATTTATATAAATTTATAAATAATTTTATAAAATTATA 84513

Qy        153 TTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTT 212
              |  |  ||| |||| | || |||    | || |  |  | |   |||| | | |  | |
Db      84512 ATTTTATTTATAAATAAATTTATATAAATTTATATAAATTTA---TAAATATAATTATAT 84456

Qy        213 GTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTT 272
              | |||||| | | |   | | | | ||  |   ||   | ||| |     | || |  |
Db      84455 ATAAATATATTGCATGTTTATGTATTAATAAATATATTAATATATAATATATTATATACT 84396

Qy        273 TTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCA 332
              |      | |||||  ||| |  | |  ||  |   |  |  ||||||||   |
Db      84395 ATATAAAATATATAAAAATATATTATATACTATATAAAATATATAAAAATATATTATATA 84336

Qy        333 GTTGAAAA-----TTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAA 387
              |  | ||      |||| | | |   | |    || | ||  |||   |  | | |  |
Db      84335 CTATATAAAATATATAAAAATATATTATATACTATATAAAATATATAAAAATATATTATA 84276

Qy        388 AGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAAT 447
              || | |    |  |||| |||| ||  | || | | |      |  |||||
Db      84275 TACTATATAAAATATATAAAAATATATTATATACTATATAAAATATATAAAAATATATTA 84216

Qy        448 TACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAA 507
              ||    | ||  || |   |||  || | | | |  |||  ||  ||| |  |||  |
Db      84215 TATACTATATAAAATATATAAAAATATATAAAACATTATATATTAATATATTATATATAG 84156

Qy        508 ACATTTACTTTCCAAATATTCTGTCATT 535
              || |||| |  | |   | | |  ||
Db      84155 TTATATACTATATATAGTATATATATTT 84128


RESULT 6
US-10-311-455-1189
; Sequence 1189, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1189
;   LENGTH: 6191
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1189

  Query Match             4.8%;  Score 52.4;  DB 15;  Length 6191;
  Best Local Similarity   47.6%;  Pred. No. 0.22;
  Matches  219;  Conservative    0;  Mismatches  236;  Indels    5;  Gaps    2;

Qy        108 ATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAAT 167
              || | || ||  | |  | | |   || | ||| |||          |  | | ||   ||
Db      2650 ATTTGTAATTTGTAATTTTTTTTTAATTAATTTTGTTTTTTTTATGTTTTAATAGTTAT 2709

Qy        168 CAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCT 227
              |  |   | | | | |  | |   |   ||  || || | | || | | | |   |
Db      2710 TTTTTTTTTTGTGTATTGTAATTTTTTGTTGGTTGATTTTTTTGTTTTTAAAAATTTATT 2769

Qy        228 TGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAA 287
              |  ||||| |    |  | | | |  |   ||   || |||||| ||| |||
Db      2770 TTTTTTTTATTTTTTATTGTTAATATTTTGTTATAAAGGATTTTTTTTGTTTTATTCGTT 2829

Qy        288 AAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACA 347
              || |  | ||     |  |  |||  || |  ||| ||| |  | |||| ||| ||
Db      2830 TTAGTTTAATAGTGTAGTTTGTAGATGAATAATATAGATTGG--ATTTGAGAATAAATAG 2887

Qy        348 CAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAA 407
              ||||    | ||| || ||   | ||       |   ||  ||  || || ||    ||
Db      2888 GTAAATATAGGTATAATAGATAGGTAAA---AAAAAAAAATTGTTGAATGTAAAAAGGAA 2944

Qy        408 GGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGG 467
              | |||     |   || |||  | | || |||||    ||      ||||||| |||
Db      2945 TGAATAAATGGTAAGTTGTTAAGGAAGTGTAAATATTTTTTTAGGAAAATTTTTTTTGAC 3004

Qy        468 AAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATT 527
              |||  | | |   | || |||| |    ||||    |  ||| ||   |  | ||
Db      3005 GTTTAAAGTAAAATTAGATGGTGTTTTTTTGTGTATTTGTAATTTTTTTTGTAATTTTTT 3064

Qy        528 CTGTCATTCAGGAGTATGGAAGTATCGATGGCTTCTTTAA 567
              |     | | | ||||   | |   | | | || ||| |
Db      3065 TTAATTGTTATGTGTATTTTAATTATGGTAGTTTTTTTTA 3104


RESULT 7
US-10-021-323-10189/c
; Sequence 10189, Application US/10021323
; Publication No. US20040123340A1
; GENERAL INFORMATION:
;  APPLICANT: Deikman, Jill
;  APPLICANT:  Feng, Paul C.C.
;  APPLICANT:  Fincher, Karen L.
;  APPLICANT:  Ziegler, Todd E.
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(52274)B
;  CURRENT APPLICATION NUMBER: US/10/021,323
;  CURRENT FILING DATE:  2001-12-12
;  PRIOR APPLICATION NUMBER: US 60/255, 619
;  PRIOR FILING DATE: 2000-12-14
;  NUMBER OF SEQ ID NOS: 17880
; SEQ ID NO 10189
;   LENGTH: 499
;   TYPE: DNA
;   ORGANISM: Gossypium hirsutum
;   FEATURE:
;   OTHER INFORMATION: Clone ID: LIB3828-023-Q6-N6-C10
US-10-021-323-10189

  Query Match             4.8%;  Score 52.2;  DB 18;  Length 499;
  Best Local Similarity   47.7%;  Pred. No. 0.073;
  Matches  153;  Conservative    0;  Mismatches  168;  Indels    0;  Gaps    0;

Qy         87 TCTTCCTTATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGA 146
              | ||  || | | | | |||  | | |    || |  | || | | |   |||  ||
Db       359 TTTTTTTTTTTTTTTTTCCCCTTTTTTTTTAAATTTATTTTTCCCTTTTTTTTTTTTTTT 300

Qy        147 GAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACAT 206
              |  ||  || ||     |   ||  |  |||| |      | |  | ||     | |
Db       299 TTTTTTTTTTCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCCCCCCTT 240

Qy        207 TGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCA 266
              | |||| ||  | |  |   ||  |||||||          ||| | | | | |  | ||
Db       239 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAAAAAAAAAAATTTTTTT 180
```

```
Qy        267 TTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATAT 326
              ||| |||||  |||    ||||||   | ||| |||   |   | | | | |  ||| ||
Db        179 TTTTTTTTTTTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAATAAAT 120

Qy        327 TAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCA 386
              |   |    ||||  ||| | ||||  |    | | | | |   ||       |   |
Db        119 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 60

Qy        387 AAGGTAAAAAGTCAACCTAAA 407
              ||   |||||   ||   |||
Db         59 AAAAAAAAAAAAAAAAAAAA 39


RESULT 8
US-10-473-126-240/c
; Sequence 240, Application US/10473126
; Publication No. US20040234973A1
; GENERAL INFORMATION:
;  APPLICANT: Epigenomics AG
;  TITLE OF INVENTION: Methods and nucleic acids for the analysis of  hematopoietic cell
;  TITLE OF INVENTION:  proliferative disorders
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/10/473,126
;  CURRENT FILING DATE:  2003-09-26
;  NUMBER OF SEQ ID NOS: 1258
; SEQ ID NO 240
;   LENGTH: 8056
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-473-126-240

  Query Match             4.8%;  Score 51.8;  DB 18;  Length 8056;
  Best Local Similarity   44.5%;  Pred. No. 0.35;
  Matches  206;  Conservative    0;  Mismatches  257;  Indels    0;  Gaps    0;

Qy        111 TCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAG 170
              | || || || || |   |  | |     ||  ||   ||| |   |  || |  ||| |
Db        473 TTTATTTTTTAAAATTAATTTTTTTTTTATTATTTTTAAAAAAAATAAAATAATTAATAAA 414

Qy        171 TTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGA 230
              |  ||  ||   |  ||  |   ||  ||| | | || ||| |||  | |  |   || |
Db        413 TATATTATTAAATTAAATTTTAAATTATAAATATTTTATTTCGTTTTTTTTTTTTTTTAA 354

Qy        231 TTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAA 290
              || || | |      ||         |  ||||  |||        || |  ||||||
Db        353 TTATTATTATTTTTATTTATTTTTATTAATTATTTTTTTAAAAAATAATTTTAAAAAAAA 294

Qy        291 GGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAA 350
              | |       |  ||| | | |||  | | |  ||    | ||  |||||    | ||
Db        293 TAAAAATTTTTTTTTATTAAAATAAATTTAAAAATTTTAAAATTTTTAAATTTTTAAAAA 234

Qy        351 AATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGT 410
              | |    || ||  |  | |||   |   | | |||  ||| || | ||   ||   |
Db        233 TTTTAAAATATTTTTTAAAAATTATAATAATAATTAAAATTAATAATTAAAAAAAATAAT 174

Qy        411 ATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAA 470
              | |  |         ||  |    ||  | |||     | ||| || |   |   ||
Db        173 AAAAAATAAATAAAAAAATAAAAAATAAATAAAAAAAAATATAAAAAATTAAAAATAATAA 114

Qy        471 TAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTG 530
              | |  |    ||| | |   ||  |  | | ||| |   | |    ||| | | |
Db        113 TTAAATATATAAAAAAATATAAAAATTAAAAAATTAAAAAAAAAAATAAAAAAAAAATTTT 54

Qy        531 TCATTCAGGAGTATGGAAGTATCGATGGCTTCTTTAAAATGAA 573
              | |   | |    | |||  ||   || ||||   | ||
Db         53 TATTATTTTAATTAATTAATATTAATTTTTTATTTATTTTAAA 11


RESULT 9
US-10-311-455-440
; Sequence 440, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determi
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 440
;   LENGTH: 15373
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-440

  Query Match             4.7%;  Score 51.4;  DB 15;  Length 15373;
  Best Local Similarity   45.4%;  Pred. No. 0.59;
  Matches  223;  Conservative    0;  Mismatches  266;  Indels    2;  Gaps    1;

Qy         28 TTCACTCTCTATGTTCTTATAAGCAAATGTTACAATGAGATATGAGTTTTAAAGCCAGAT 87
              || | | |   | ||.|||  |  ||| |||   |  ||    |  | |      |
Db       4150 TTTATTTTTGTTATTTTTAGGAAAAAAAGTTGATAAATTATTATTTGTAGATATTTTTTT 4209

Qy         88 CTTCCTTATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAG 147
              ||   ||| |    |    || | |  ||  |    |     || |   ||  |||||
Db       4210 GTTTTATATTTTATAGATATATGTTTTTTTAATTTTTTTAGAAATGTTTTTTAAGTTGAT 4269

Qy        148 AAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATT 207
              ||| ||    | |  || | | ||  |||  ||||  || |  | |  ||       |||
Db       4270 TAATTTTTTAGAATTTAATTAATTTTTTAGAATTGTTTAAAAGATTAATTTTTTAGAATT 4329

Qy        208 GTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCAT 267
              ||||      ||  | |   | || || |||||||||  |||      || |  | ||| |
Db       4330 GTTTAAAAGATTAATTTTTTAGAATTGTTTAAAAGA--TTAATTTTTTAGAATTGTTTAA 4387

Qy        268 TTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATT 327
              | ||    |||    |  ||  | |  |||||||  ||    ||      | |||| ||
Db       4388 AAGATTAATTTTTTAGAATTTAATTTTTTTAGAAATGTTTTTTAAGTTGATTAATATTTT 4447

Qy        328 AGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAA 387
              |    ||| ||||| |||   |  |         | |  |  |||  || | |      | |
Db       4448 ATTTTGTTAAAAATAAAAATAATTAGATAATTAGAAAATTTTAAGTATTATATATTTAAT 4507

Qy        388 AGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAAT 447
              ||  || | |  || |  | | |   |   |  |  | | |          | || |||
Db       4508 TATTAGTAATTGTATATATAATAAAATTTTTTTAGTGTTAAATTATAAAAATATTGAAGA 4567

Qy        448 TACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTTATGTACTATTAAA 507
              |  | |  || | | ||  |    || |   |||  ||  |   || | || | ||| |
Db       4568 GAATATATTTTATTTGTTTTATTATATTTGGTAAATAATTTTAAATTGTATAGTTTTATA 4627

Qy        508 ACATTTACTTT 518
              |    | |||
Db       4628 AATAAAATTTT 4638


RESULT 10
US-10-741-600-17952/c
; Sequence 17952, Application US/10741600
; Publication No. US20050026169A1
; GENERAL INFORMATION:
;  APPLICANT: CARGILL, Michele et al.
```

```
;  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
;  TITLE OF INVENTION:  MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001499
;  CURRENT APPLICATION NUMBER: US/10/741,600
;  CURRENT FILING DATE:  2003-12-22
;  NUMBER OF SEQ ID NOS: 73997
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17952
;   LENGTH: 394468
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(394468)
;   OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-741-600-17952

  Query Match             4.7%;  Score 51.4;  DB 19;  Length 394468;
  Best Local Similarity   43.5%;  Pred. No. 2.9;
  Matches  226;  Conservative    1;  Mismatches  292;  Indels    0;  Gaps    0;

Qy         33 TCTCTATGTTCTTATAAGCAAATGTTACAATGAGATATGAGTTTTAAAGCCAGATCTTCC 92
              | | ||| |   |||||| | || |  | || | ||||   | ||| ||  | || |
Db     341132 TATATATTTATATATAAGTATATTTATCTATAATATATTTATATTATAGATAAATATAAA 341073

Qy         93 TTATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATA 152
              |    | | |        |   ||| |  |        ||||   | |:  |   | ||
Db     341072 TATATTATATATAAATATTTATGTTATATATAATATAAATATATATATKTATATTATATT 341013

Qy        153 TTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTT 212
              ||  |  |||   ||  |||    | |    | || |  |  || | |||   || |
Db     341012 TTATTTTTTAAGTATATATTACACATACTTATATAAATATATTTATATATACAAATTTAT 340953

Qy        213 GTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTT 272
              |||  |    |   |  | | |  | |    |||| | | ||||   |||  || | |
Db     340952 ATTACTTATATAAATACTTATATAAATATATACTTATATAAATACTTATATAAATATATA 340893

Qy        273 TTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCA 332
              ||   |   | | ||| | ||| |   || ||| | | | | ||       |    |
Db     340892 CTTATATAAATATATAAATGCATATATTTATACATAAATATATAAATGCATATATTTATA 340833

Qy        333 GTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTA 392
              | || ||  ||   | ||   |  | || |   ||   |  | |  | | | |  ||
Db     340832 CATAAATATATAAATGCATATATTTATACATAAATATATAAATGCATATATTTATACATA 340773

Qy        393 AAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCA 452
              || |   ||    |      |  ||| |   |   ||      |  | || | ||  |
Db     340772 AATATATAAATGCATATATTTATACATAAATATATAAATGCATATATTTATACATAAATA 340713

Qy        453 CAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATT 512
              | |  |   | |    |||     | | | |     | |  |     || || | ||
Db     340712 TATAAATGCATATATTTATACATAAATATATAAATGCATATATTTATACATAAATATATA 340653

Qy        513 TACTTTCCAAATATTCTGTCATTCAGGAGTATGGAAGTA 551
              ||| ||  | | | | | ||  |     ||| ||  ||
Db     340652 TACATTTTATAAAATATATCTATATTTCATATAGATATA 340614


RESULT 11
US-10-424-599-69112
; Sequence 69112, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 69112
;   LENGTH: 523
;   TYPE: DNA
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_33420C.1
US-10-424-599-69112

  Query Match             4.7%;  Score 51.2;  DB 17;  Length 523;
  Best Local Similarity   51.3%;  Pred. No. 0.13;
  Matches  119;  Conservative    0;  Mismatches  113;  Indels    0;  Gaps    0;

Qy        149 AATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTG 208
              |||||    |  | | ||||    |||| | | || | || ||   ||   |||   | |
Db         44 AATATAATTTTTAAAAAATATAATAATTGATGTGAAAAAAATTAGATTAAAAAATTAATT 103

Qy        209 TTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATT 268
              |||  ||| ||||       |  ||  | |||   |  |          ||     | | ||
Db        104 TTTAATTATTATAAAAGTATTTTTCATCTAATTCATTTATTCATTTCACCCTCTTCCTTT 163

Qy        269 TGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTA 328
              | ||||| |   |||  |  |  |||||||   | | ||  ||      ||     ||||
Db        164 TATTTTTAGAAAGTTATATTACAGTATAAATTTAGGAATATAGTTTTTTATTTATAATTA 223

Qy        329 GCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTG 380
                 || | |||||   | ||||||   |  ||||  |||| || || |||
Db        224 TTATTTTTATAATTATTTATAAAATGTCAGATTATTTTATTATTTTATTTTG 275


RESULT 12
US-10-311-455-1236/c
; Sequence 1236, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1236
;   LENGTH: 12393
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: 9300, 9362
;   OTHER INFORMATION: n is a or g or c or t
US-10-311-455-1236

  Query Match             4.7%;  Score 51.2;  DB 15;  Length 12393;
  Best Local Similarity   49.4%;  Pred. No. 0.59;
  Matches  160;  Conservative    0;  Mismatches  163;  Indels    1;  Gaps    1;

Qy        209 TTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATT 268
              |||  ||||||   |   ||     | |||| | |  ||  | |   | |||  | |
Db       1820 TTTATTTAATAACATTATTTTTCAATATTAATACATATTCTATAACAAAAAAACTTCGTA 1761

Qy        269 TGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTA 328
              | ||||   |||  || |||  ||   | ||| |||   | | | |  || | || | |  |
Db       1760 TTTTTTATTTTTTTTTAAATAAAAAAAAAATAAACAAAATAAAAATAATAACAAAAAAAA 1701
```

```
Qy        329 GCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAA 388
              |    ||||    | | | ||    |  |||| | || ||  || |       |  |
Db       1700 AAAAAAATAAAAAACCAAAAATAAAACTTATATATAAAATAACAAAAACAACATCTTTTA 1641

Qy        389 GGTAAAAAGTCAACCTAAAGG-TATACTACCTGCTTTCTTATCGCACTGCAAATAGAAAT 447
              | |||| |     ||||   || | ||  | |   |||        |  ||| |  |
Db       1640 AATCAAAACTTTTTTTAAACCATAAATTATTTACAAACTTTAAAATATTTAAAAAATACA 1581

Qy        448 TACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAA 507
              ||  ||||||||||        | |||   | || | | ||  |      | ||||| ||
Db       1580 TAATACAAATTTTAAAAATTTTAAAATACTAAAATATAAAAATAAAATTTTACTATAAAT 1521

Qy        508 ACATTTACTTTCCAAATATTCTGT 531
              | ||| | || |   | | | | |
Db       1520 AAATTAATTTCCTTTAAAATATAT 1497
```

RESULT 13
US-10-473-126-386
; Sequence 386, Application US/10473126
; Publication No. US20040234973A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Methods and nucleic acids for the analysis of hematopoietic cell
; TITLE OF INVENTION: proliferative disorders
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/473,126
; CURRENT FILING DATE: 2003-09-26
; NUMBER OF SEQ ID NOS: 1258
; SEQ ID NO 386
; LENGTH: 8056
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-473-126-386

Query Match 4.7%; Score 51; DB 18; Length 8056;
Best Local Similarity 45.1%; Pred. No. 0.54;
Matches 242; Conservative 0; Mismatches 285; Indels 10; Gaps 1;

```
Qy         31 ACTCTCTATGTTCTTATAAGCAAATGTTACAATGAGATATGAGTTTTAAAGCCAGATCTT 90
              | | | | ||  |||     ||    |  ||  | | | |||| ||    |  | ||
Db       2241 AATTTATTTTTTAATATTTAAAATAAATTAAAAAAATTTTAATTTTTTAATTTATTTATT 2300

Qy         91 CCTTATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAA 150
              | |   |    ||  | |   | |  |    | | || | | |||      | ||
Db       2301 TTTAAAATAATTTATTAATTTTTGTATTTAAATAAATTTTATTTAATTTTTAAATATAAT 2360

Qy        151 TATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTT 210
              | || ||  ||| |||| | ||| ||  |||    |   | | |   | || | |
Db       2361 TTTTTATTTTTATAAATAATTTATTAAATTTTATTATTTTTTGAAAATTAAAAAA----- 2415

Qy        211 TTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTG 270
                 || |||  |   || |||||| |||  |  ||      ||     | |  || |
Db       2416 -----AAAATAATTTATTTATTTTTAAAAATTAATTTTTTTTTATTTTTATTTATATATT 2470

Qy        271 TTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGC 330
              ||||   || |    |||  || || | ||  ||  | |  || | |  | |||
Db       2471 TTTTATTATTTTGTTTTAAAATAATTAAAATATTAATGAAAAATAAAAAAAAATAAATTA 2530

Qy        331 CAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGG 390
              | |  ||||| ||  | |||||      || | |    | ||  |     |   |||
Db       2531 AATTATAAAATAAATTAAAAAATTTAAAAATTTAAATAAAATTTTTAAAAAATAAAAAAT 2590

Qy        391 TAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTAC 450
              ||  |   ||   |    |  || |  |||  |     |    ||| | ||||
Db       2591 AAATGAAAAAAAAAATTATAAATATAAATAATTTAAAAAAAAAAAAATAAAAAAAATGTT 2650

Qy        451 CACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACA 510
              | |  |  |||||  |||||||        | ||  ||||  || || ||||
Db       2651 AAAATTTAAAATTTTAAAAATAATAAATTTTTAAATTTTTTTAAATTTATTATTTTTTTT 2710
```

```
Qy        511 TTTACTTTCCAAATATTCTGTCATTCAGGAGTATGGAAGTATCGATGGCTTCTTTAA 567
              |||   |   ||  |||   | | | |    |    ||  ||  |     | |||||
Db       2711 TTTTAATAAAAATAATTTGTTAAATAAATTATTAATAAAAATAAAAAAAATATTTAA 2767
```

RESULT 14
US-10-473-126-240
; Sequence 240, Application US/10473126
; Publication No. US20040234973A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Methods and nucleic acids for the analysis of hematopoietic cell
; TITLE OF INVENTION: proliferative disorders
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/473,126
; CURRENT FILING DATE: 2003-09-26
; NUMBER OF SEQ ID NOS: 1258
; SEQ ID NO 240
; LENGTH: 8056
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-473-126-240

Query Match 4.7%; Score 50.8; DB 18; Length 8056;
Best Local Similarity 51.3%; Pred. No. 0.6;
Matches 118; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

```
Qy        138 TTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCT 197
              |||  |  | |   ||  || || | |||      | ||| ||||  | || |  | || |
Db       1290 TTTTTTAAAAATTAATAAATTATAAAAAAATTTATTATAATTTTTAATAAAAAATAATTT 1349

Qy        198 TAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATAC 257
               ||    | | |||| |||| |   |   ||  |||| | |||| |   |  |    |
Db       1350 AAATTTTAATTTTTTATTAAAAATTTATTTTTTTTTTATAAAAATAATATCGATTTTTTT 1409

Qy        258 AGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAG 317
                  ||  ||  |||||  |||    ||||||  || |||  | |   || | | ||
Db       1410 TTTATTTTTTTATTTTTTTTTTTTAAAAAAAAAAATAAAAATTATTTTTTTAATAAAATT 1469

Qy        318 ATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACA 367
              ||  || | ||   | ||   || | || |  | || | |  | | || |
Db       1470 ATATTAAAATAATTAATTATTAAATTAATATTATATAAATATAAAATATA 1519
```

RESULT 15
US-10-486-319A-43/c
; Sequence 43, Application US/10486319A
; Publication No. US20050064410A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Method and nucleic acids for the analysis of colon cancer
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/486,319A
; CURRENT FILING DATE: 2004-02-09
; NUMBER OF SEQ ID NOS: 527
; SEQ ID NO 43
; LENGTH: 8222
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-486-319A-43

Query Match 4.6%; Score 50.6; DB 19; Length 8222;
Best Local Similarity 46.9%; Pred. No. 0.67;
Matches 158; Conservative 0; Mismatches 179; Indels 0; Gaps 0;

```
Qy        146 AGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACA 205
              | |||||   |  |||||||| | | || |  |   || || |    ||    || | |
Db       7970 ATAAATAAATAAAATTACAAATAAATAAAAAAATAAATAAAAAAATAAATAAATAATAAA 7911
```

```
Qy        206 TTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTC 265
              |    |   | |||| |         | |   | |||| |   || | | ||| |   |
Db       7910 TAAATAAATAAATAAAAAAAAAAATAATAAATAAAAAAATAATAAATAAATAAATAAAAA 7851

Qy        266 ATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATA 325
              ||   |   |   |   | ||||||   | ||| ||||  ||  | | ||| || |   |
Db       7850 ATAAATAAATAAATAAATAAAAAAATAAAAAAACAAATAAATAAATAAAAAAATAAATAA 7791

Qy        326 TTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTC 385
               ||   | |  | || |||| | ||||  |    |||    |  |  ||||     | |
Db       7790 ATAATAAATAAATAAATAAATAAAAAAAAAAAATAATAAATAAAAAAATAATAAATAAATA 7731

Qy        386 AAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAA 445
              ||    ||||      |  ||||   |||  |       |   ||    | |  |   | ||
Db       7730 AATAAAAAAATAAATAAATAAATAAATAAAAAAATAATAAATAAATAAATAAATTAAAAA 7671

Qy        446 ATTACCACAAATTTTATTTTGGAAATAATCTCAGAAA 482
              || |  | | |      |    ||||||    | |||
Db       7670 ATAAATAAATAAACACATAAATAAATAAATAAATAAA 7634


RESULT 16
US-10-486-319A-65/c
; Sequence 65, Application US/10486319A
; Publication No. US20050064410A1
; GENERAL INFORMATION:
;  APPLICANT: Epigenomics AG
;  TITLE OF INVENTION: Method and nucleic acids for the analysis of colon cancer
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/10/486,319A
;  CURRENT FILING DATE:  2004-02-09
;  NUMBER OF SEQ ID NOS: 527
; SEQ ID NO 65
;   LENGTH: 8222
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-486-319A-65

  Query Match             4.6%;  Score 50.6;  DB 19;  Length 8222;
  Best Local Similarity   46.9%;  Pred. No. 0.67;
  Matches  158;  Conservative    0;  Mismatches  179;  Indels    0;  Gaps    0;

Qy        146 AGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACA 205
              | |||||    |  |||||||||| | | || |  |   || || |     ||     || | |
Db       7970 ATAAATAAATAAAATTACAAATAAATAAAAAAAATAAATAAAAAAAATAAATAAATAATAAA 7911

Qy        206 TTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTC 265
              |    |   | |||| |         | |   | |||| |   || | | ||| |   |
Db       7910 TAAATAAATAAATAAAAAAAAAAATAATAAATAAAAAAATAATAAATAAATAAATAAAAA 7851

Qy        266 ATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATA 325
              ||   |   |   |   | ||||||   | ||| ||||  ||  | | ||| || |   |
Db       7850 ATAAATAAATAAATAAATAAAAAAATAAAAAAACAAATAAATAAATAAAAAAATAAATAA 7791

Qy        326 TTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTC 385
               ||   | |  | || |||| | ||||  |    |||    |  |  ||||     | |
Db       7790 ATAATAAATAAATAAATAAATAAAAAAAAAAAATAATAAATAAAAAAATAATAAATAAATA 7731

Qy        386 AAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAA 445
              ||    ||||      |  ||||   |||  |       |   ||    | |  |   | ||
Db       7730 AATAAAAAAATAAATAAATAAATAAATAAAAAAATAATAAATAAATAAATAAATTAAAAA 7671

Qy        446 ATTACCACAAATTTTATTTTGGAAATAATCTCAGAAA 482
              || |  | | |      |    ||||||    | |||
Db       7670 ATAAATAAATAAACACATAAATAAATAAATAAATAAA 7634


RESULT 17
US-09-814-353-5048/c
; Sequence 5048, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
;  APPLICANT: Lee, John
;  APPLICANT:  Thompson, Pamela
;  APPLICANT:  Lillie, James
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
;  TITLE OF INVENTION:  IDENTIFICATION, ASSESSMENT, PREVENTION, AND
;  TITLE OF INVENTION:  THERAPY OF OVARIAN CANCER
;  FILE REFERENCE: MRI-006B
;  CURRENT APPLICATION NUMBER: US/09/814,353
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: US 60/191,031
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: US 60/207,124
;  PRIOR FILING DATE: 2000-05-25
;  PRIOR APPLICATION NUMBER: US 60/211,940
;  PRIOR FILING DATE: 2000-06-15
;  PRIOR APPLICATION NUMBER: US 60/216,820
;  PRIOR FILING DATE: 2000-07-07
;  PRIOR APPLICATION NUMBER: US 60/220,661
;  PRIOR FILING DATE: 2000-07-25
;  PRIOR APPLICATION NUMBER: US 60/257,672
;  PRIOR FILING DATE: 2000-12-21
;  NUMBER OF SEQ ID NOS: 22037
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5048
;   LENGTH: 436
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 3, 175, 176, 177, 178, 179, 180, 181, 182, 185, 187, 191,
;   LOCATION: 208, 209, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230,
;   LOCATION: 238, 239, 240, 242, 244, 246, 253, 254, 255, 256, 257, 258,
;   LOCATION: 260, 268, 270, 271, 272, 275, 276, 277, 278, 279, 283
;   OTHER INFORMATION: n = A,T,C or G
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 289, 290, 291, 297, 298, 299, 300, 301, 310, 311, 312, 313,
;   LOCATION: 314, 315, 316, 317, 318, 324, 337, 339, 340, 341, 342, 343,
;   LOCATION: 344, 345, 346, 348, 350, 352, 353, 354, 355, 356, 357, 358,
;   LOCATION: 359, 360, 361, 362, 363, 364, 365, 366, 367, 368, 369
;   OTHER INFORMATION: n = A,T,C or G
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 370, 371, 372, 373, 374, 375, 376, 377, 378, 379, 380, 381,
;   LOCATION: 382, 383, 384, 385, 386, 387, 403, 404, 405, 406, 412, 413,
;   LOCATION: 414, 415, 416, 422, 423, 424, 425, 429
;   OTHER INFORMATION: n = A,T,C or G
US-09-814-353-5048

  Query Match             4.6%;  Score 50.4;  DB 10;  Length 436;
  Best Local Similarity   34.4%;  Pred. No. 0.18;
  Matches  114;  Conservative    0;  Mismatches  217;  Indels    0;  Gaps    0;

Qy         78 AAAGCCAGATCTTCCTTATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAG 137
              |||                                                      | |
Db       390 AAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNANANANNNNNNNNTNTTTTTT 331

Qy        138 TTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCT 197
              |||  |  | ||          |  | |||             || |  ||       |
Db       330 TTTTTTNAAAAANNNNNNNNNAAAAAAAAANNNNNGGGGGNNNTTTTTNAAANNNNNTTNN 271

Qy        198 TAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATAC 257
               | | | | |       ||  |  |   |  |||||||          | | | | |
Db       270 NANAAAAAATNTNNNNNNAAATTGNTNTNTNNNTTTTTTTNNNNNNNNNAAAAAAAAAA 211

Qy        258 AGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAG 317
              | | ||  ||| |||||    |   |          | ||| |||   |   | | |||
Db       210 ANNTTTTTTTTTTTTTTTANAATNCNCCNNNNNNNNAAAAAAAAAAAAAAAAAAAAAAAA 151

Qy        318 ATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATC 377
              | | | |  |    |    |||| ||| | |||| |    | |  | |  |   ||
Db       150 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 91
```

```
Qy        378 TTGCAGTCAAAGGTAAAAAGTCAACCTAAAG 408
              |   |||   |||||   ||   ||||
Db         90 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAG 60


RESULT 18
US-09-814-353-11340/c
; Sequence 11340, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
;  APPLICANT: Lee, John
;  APPLICANT:  Thompson, Pamela
;  APPLICANT:  Lillie, James
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
;  TITLE OF INVENTION:  IDENTIFICATION, ASSESSMENT, PREVENTION, AND
;  TITLE OF INVENTION:  THERAPY OF OVARIAN CANCER
;  FILE REFERENCE: MRI-006B
;  CURRENT APPLICATION NUMBER: US/09/814,353
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: US 60/191,031
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: US 60/207,124
;  PRIOR FILING DATE: 2000-05-25
;  PRIOR APPLICATION NUMBER: US 60/211,940
;  PRIOR FILING DATE: 2000-06-15
;  PRIOR APPLICATION NUMBER: US 60/216,820
;  PRIOR FILING DATE: 2000-07-07
;  PRIOR APPLICATION NUMBER: US 60/220,661
;  PRIOR FILING DATE: 2000-07-25
;  PRIOR APPLICATION NUMBER: US 60/257,672
;  PRIOR FILING DATE: 2000-12-21
;  NUMBER OF SEQ ID NOS: 22037
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11340
;   LENGTH: 436
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 3, 175, 176, 177, 178, 179, 180, 181, 182, 185, 187, 191,
;   LOCATION: 208, 209, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230,
;   LOCATION: 238, 239, 240, 242, 244, 246, 253, 254, 255, 256, 257, 258,
;   LOCATION: 260, 268, 270, 271, 272, 275, 276, 277, 278, 279, 283
;   OTHER INFORMATION: n = A,T,C or G
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 289, 290, 291, 297, 298, 299, 300, 301, 310, 311, 312, 313,
;   LOCATION: 314, 315, 316, 317, 318, 324, 337, 339, 340, 341, 342, 343,
;   LOCATION: 344, 345, 346, 348, 350, 352, 353, 354, 355, 356, 357, 358,
;   LOCATION: 359, 360, 361, 362, 363, 364, 365, 366, 367, 368, 369
;   OTHER INFORMATION: n = A,T,C or G
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 370, 371, 372, 373, 374, 375, 376, 377, 378, 379, 380, 381,
;   LOCATION: 382, 383, 384, 385, 386, 387, 403, 404, 405, 406, 412, 413,
;   LOCATION: 414, 415, 416, 422, 423, 424, 425, 429
;   OTHER INFORMATION: n = A,T,C or G
US-09-814-353-11340

  Query Match             4.6%;  Score 50.4;  DB 10;  Length 436;
  Best Local Similarity   34.4%;  Pred. No. 0.18;
  Matches  114;  Conservative    0;  Mismatches  217;  Indels    0;  Gaps    0;

Qy         78 AAAGCCAGATCTTCCTTATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAG 137
              |||                                                      | |
Db        390 AAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNANANANNNNNNNNTNTTTTTT 331

Qy        138 TTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCT 197
              |||  |  | ||         |  | |||            || |  ||       |
Db        330 TTTTTTNAAAAANNNNNNNNAAAAAAAANNNNNGGGGGNNNTTTTTNAAANNNNNTTNN 271

Qy        198 TAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATAC 257
```

```
              | | | | | |      ||  |  |   |  |||||||          | | | | |
Db        270 NANAAAAAATNTNNNNNNAAATTGNTNTNTNNNTTTTTTTNNNNNNNNNAAAAAAAAAA 211

Qy        258 AGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAG 317
              | | ||  ||| |||||     |   |       | ||| |||   |   | | |||
Db        210 ANNTTTTTTTTTTTTTTTANAATNCNCCNNNNNNNNAAAAAAAAAAAAAAAAAAAAAAAA 151

Qy        318 ATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATC 377
              |  | | |  |   |    |||| ||| | ||||  |    | |  | |  |   ||
Db        150 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 91

Qy        378 TTGCAGTCAAAGGTAAAAAGTCAACCTAAAG 408
              |   |||   |||||   ||   ||||
Db         90 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAG 60


RESULT 19
US-10-433-793-90/c
; Sequence 90, Application US/10433793
; Publication No. US20040142334A1
; GENERAL INFORMATION:
;  APPLICANT: Epigenomics AG
;  TITLE OF INVENTION: Diagnose von mit Angiogenese assoziierten Krankheiten
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/10/433,793
;  CURRENT FILING DATE:  2003-06-06
;  NUMBER OF SEQ ID NOS: 212
; SEQ ID NO 90
;   LENGTH: 34688
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-433-793-90

  Query Match             4.6%;  Score 50.4;  DB 18;  Length 34688;
  Best Local Similarity   45.0%;  Pred. No. 1.5;
  Matches  189;  Conservative    0;  Mismatches  231;  Indels    0;  Gaps    0;

Qy        151 TATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTT 210
              | || || ||||||    |   ||||  |  | |   |     | |  |  |   ||| |
Db      26204 TTTTAATAATTACACTCCCAATAATTTTAATATTCCAACGTACTTAATTTACTTATTTTC 26145

Qy        211 TTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTG 270
              ||||||||  ||  || |||    ||||| | | |  | |    ||   |    |||
Db      26144 ACGTTAATATTTTCTATTTATTAAAATAAAAAAACATTAAAATTATCATTAAAAACTTTA 26085

Qy        271 TTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGC 330
              | |||     | ||  || |       |  |  ||    |  | | |||  | |  | |
Db      26084 TCTTTAACATATTATAATATTTATTTCAATAACTAACCTATAAAAAATTAAATTCCTTTA 26025

Qy        331 CAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGG 390
              |   ||   | | | |  | |  | | | | |||||  |  |  ||
Db      26024 TAAAAATAATAATTAAAATATTTTTATCTAATAAAAAAAAATTAATAATAAAACCATTTC 25965

Qy        391 TAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTAC 450
              |   ||   ||   |||   ||| | |   ||         |   ||||  || |
Db      25964 TCCGAAAAAAATTAAAAACAATAATTCAATTTTAAAATATTATCGCAAAATTAAATTATT 25905

Qy        451 CACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACA 510
              |  ||| ||||||    |||| || |    ||   |||| |   || || || ||||
Db      25904 CTAAAAATTTATTAATAAAATTATATACTTAATATTAATCTAAAATATAATACTAAATTT 25845

Qy        511 TTTACTTTCCAAATATTCTGTCATTCAGGAGTATGGAAGTATCGATGGCTTCTTTAAAAT 570
              |||  |  |||  |||   |  |   | |||  ||  |       |||||  | | ||
Db      25844 AAAACTATAAAAAAATTAAATATATTTAAAATATATAAAAACGTTAAACTTCTATTATAT 25785


RESULT 20
US-10-473-126-386/c
; Sequence 386, Application US/10473126
; Publication No. US20040234973A1
; GENERAL INFORMATION:
```

```
;  APPLICANT: Epigenomics AG
;  TITLE OF INVENTION: Methods and nucleic acids for the analysis of  hematopoietic cell
;  TITLE OF INVENTION:  proliferative disorders
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/10/473,126
;  CURRENT FILING DATE:  2003-09-26
;  NUMBER OF SEQ ID NOS: 1258
; SEQ ID NO 386
;   LENGTH: 8056
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-473-126-386

  Query Match             4.6%;  Score 50.2;  DB 18;  Length 8056;
  Best Local Similarity   44.3%;  Pred. No. 0.82;
  Matches  205;  Conservative    0;  Mismatches  258;  Indels    0;  Gaps    0;

Qy        111 TCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAG 170
              | || || || || |   |  | |    ||  ||   ||| |   |  || |  ||| |
Db        473 TTTATTTTTTAAAATTAATTTTTTTTTATTATTTTTAAAAAAAATAAAATAATTAATAAA 414

Qy        171 TTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGA 230
              |  ||  ||   |  ||  |   ||  |||| |  || |||  ||  | | |   || |
Db        413 TATATTATTAAATTAAATTTTAAATTATAAATATTTTATTTCATTTTTTTTTTTTTTTAA 354

Qy        231 TTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAA 290
              || || | |      ||          |  ||||  |||         || |  ||||||
Db        353 TTATTATTATTTTTATTTATTTTTATTAATTATTTTTTTAAAAAATAATTTTAAAAAAAA 294

Qy        291 GGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAA 350
                | |       |  ||| | | |||  | | |  ||    | ||  |||||    | ||
Db        293 TAAAAATTTTTTTTTATTAAAATAAATTTAAAAATTTAAAATTTTTAAATTTTTAAAAA 234

Qy        351 AATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGT 410
              | |    || ||  |  | |||   |   | | |||  ||| || | ||   ||   |
Db        233 TTTTAAAATATTTTTTAAAAATTATAATAATAATTAAAATTAATAATTAAAAAAAATAAT 174

Qy        411 ATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAA 470
              | | |           ||   |    ||  | |||     | ||| || | |   |   ||
Db        173 AAAAAATAAATAAAAAAATAAAAAATAAATAAAAAAAAATATAAAAAATTAAAAATAATAA 114

Qy        471 TAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTG 530
              | | |    ||| | |  |    ||  |  | | ||| |   | |    ||| | | |
Db        113 TTAAATATATAAAAAAATATAAAAATTAAAAAATTAAAAAAAAAAATAAAAAAAAAATTTT 54

Qy        531 TCATTCAGGAGTATGGAAGTATCGATGGCTTCTTTAAAATGAA 573
              | |   | |   | |||  ||   || ||||   | ||
Db         53 TATTATTTTAATTAATTAATATTAATTTTTTATTTATTTTAAA 11


RESULT 21
US-10-706-635-24
; Sequence 24, Application US/10706635
; Publication No. US20050014263A1
; GENERAL INFORMATION:
;  APPLICANT: Moyer, Richard W.
;  APPLICANT:  Li, Yi
;  TITLE OF INVENTION: Materials and Methods for Delivery and Expression of Heterologous
;  TITLE OF INVENTION:  Vertebrate Cells
;  FILE REFERENCE: UF-221C1XCZ1
;  CURRENT APPLICATION NUMBER: US/10/706,635
;  CURRENT FILING DATE:  2003-11-12
;  PRIOR APPLICATION NUMBER: 09/086,651
;  PRIOR FILING DATE: 1998-05-29
;  PRIOR APPLICATION NUMBER: 09/662,254
;  PRIOR FILING DATE: 2000-09-14
;  PRIOR APPLICATION NUMBER: 60/224,479
;  PRIOR FILING DATE: 2000-08-10
;  NUMBER OF SEQ ID NOS: 80
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 24
;   LENGTH: 50000
;   TYPE: DNA
;   ORGANISM: Amsacta moorei entomopoxvirus
US-10-706-635-24

  Query Match             4.6%;  Score 50.2;  DB 19;  Length 50000;
  Best Local Similarity   47.9%;  Pred. No. 2;
  Matches  145;  Conservative    0;  Mismatches  158;  Indels    0;  Gaps    0;

Qy        228 TGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAA 287
              | | |||| ||        |   |   |  ||  |||| |||| | | |  ||| ||  |
Db      39496 TTAATTTTATAGTTTTTTTATTATTAAGCCAATTATTTATTTTTCTATAATTTTTTATAC 39555

Qy        288 AAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACA 347
              | |  || ||| | || |              |     | |  | | |||| || ||   |
Db      39556 ACACATACGAAGCATTGAAAAATCCAGTTTTTTTATCAATTACTAATTGATAACTATTTA 39615

Qy        348 CAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAA 407
              |  |  ||   | | |  |||  ||| | || ||    |   || | |    | | ||
Db      39616 TATGAGCAGAATAAAATGTATTTTTTATTTTTTCATCATATATTATAGAAGACATTTTAA 39675

Qy        408 GGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGG 467
                 || || ||  |||  | |  ||     | |||    ||||    |||  | ||||
Db      39676 CTATTATTATCTTATTTTATTTTTCAAAAAACATATTTTTTACATTTAATACTGTTTTTA 39735

Qy        468 AAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATT 527
              ||||   |  || |    | |||| |   || |||| ||| ||||   |      ||  |
Db      39736 TAATATGTTTTGATATTGAATTTTTATCAATATTATTTAAATATTTTAATAAATTATTGT 39795

Qy        528 CTG 530
              |||
Db      39796 CTG 39798


RESULT 22
US-10-311-455-674/c
; Sequence 674, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 674
;   LENGTH: 18988
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-674

  Query Match             4.6%;  Score 50;  DB 15;  Length 18988;
  Best Local Similarity   47.4%;  Pred. No. 1.4;
  Matches  183;  Conservative    0;  Mismatches  200;  Indels    3;  Gaps    1;

Qy        146 AGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACA 205
              | ||| | | | ||||  ||| | | | || | |   | || |||   |      |
Db       8192 AAAAAAAATAACCATTATTAATTATTAACTACTCTACCAAAATATCATTTAAATTACTTC 8133

Qy        206 TTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTC 265
              |  || |   || | |  |  | ||    | |       | ||| | |  || | || |
Db       8132 TATTTATAAAAAAAAATACATTAAATAAAATAATCTCCATTCACAAAAAATCAACAATCC 8073
```

```
Qy       266 ATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATA 325
Db      8072 ATTAAACTTCTCCTTAAACAAAATTTATCCTAAAATATAAATAAAAAAAAAATAATTAAA 8013

Qy       326 TTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTC 385
Db      8012 AAATAAAAAATATATAAAAAAAAAAAAAAAAAAAAAATAACACAAAAATACCTCACCCTACC 7953

Qy       386 AAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAA 445
Db      7952 AAAAATAAACACACAACCGCTACCTACCTATCCTA---TCTTATCGTTCCCAAAAACACA 7896

Qy       446 ATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTA 505
Db      7895 AAACAACTAAAACCCGTCATCTTCGTAATACCAACAAACATATTTTATAAATTCATATAA 7836

Qy       506 AAACATTTACTTTCCAAATATTCTGT 531
Db      7835 ATAACTAAATTTTATATAAAAACTAT 7810
```

RESULT 23
US-10-240-485-62/c
; Sequence 62, Application US/10240485
; Publication No. US20030148327A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with
; TITLE OF INVENTION: Metastasis
; FILE REFERENCE: 5013.1007
; CURRENT APPLICATION NUMBER: US/10/240,485
; CURRENT FILING DATE: 2002-10-02
; PRIOR APPLICATION NUMBER: PCT/EP01/03970
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 202
; SEQ ID NO 62
; LENGTH: 18988
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-485-62

Query Match 4.6%; Score 50; DB 15; Length 18988;
Best Local Similarity 47.4%; Pred. No. 1.4;
Matches 183; Conservative 0; Mismatches 200; Indels 3; Gaps 1;

```
Qy       146 AGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACA 205
Db      8192 AAAAAAAATAACCATTATTAATTATTAACTACTCTACCAAAATATCATTTAAATTACTTC 8133

Qy       206 TTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTC 265
Db      8132 TATTTATAAAAAAAAATACATTAAATAAAATAATCTCCATTCACAAAAAATCAACAATCC 8073

Qy       266 ATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATA 325
Db      8072 ATTAAACTTCTCCTTAAACAAAATTTATCCTAAAATATAAATAAAAAAAAAATAATTAAA 8013

Qy       326 TTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTC 385
Db      8012 AAATAAAAAATATATAAAAAAAAAAAAAAAAAAAAAATAACACAAAAATACCTCACCCTACC 7953

Qy       386 AAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAA 445
Db      7952 AAAAATAAACACACAACCGCTACCTACCTATCCTA---TCTTATCGTTCCCAAAAACACA 7896

Qy       446 ATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTA 505
Db      7895 AAACAACTAAAACCCGTCATCTTCGTAATACCAACAAACATATTTTATAAATTCATATAA 7836

Qy       506 AAACATTTACTTTCCAAATATTCTGT 531
Db      7835 ATAACTAAATTTTATATAAAAACTAT 7810
```

RESULT 24
US-10-221-714A-64/c
; Sequence 64, Application US/10221714A
; Publication No. US20040048254A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with
; TITLE OF INVENTION: tumor suppressor genes and oncogenes
; FILE REFERENCE: 5013.1005
; CURRENT APPLICATION NUMBER: US/10/221,714A
; CURRENT FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: PCT/EP01/02955
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: DE 10013847.0
; PRIOR FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 540
; SEQ ID NO 64
; LENGTH: 18988
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-714A-64

Query Match 4.6%; Score 50; DB 17; Length 18988;
Best Local Similarity 47.4%; Pred. No. 1.4;
Matches 183; Conservative 0; Mismatches 200; Indels 3; Gaps 1;

```
Qy       146 AGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACA 205
Db      8192 AAAAAAAATAACCATTATTAATTATTAACTACTCTACCAAAATATCATTTAAATTACTTC 8133

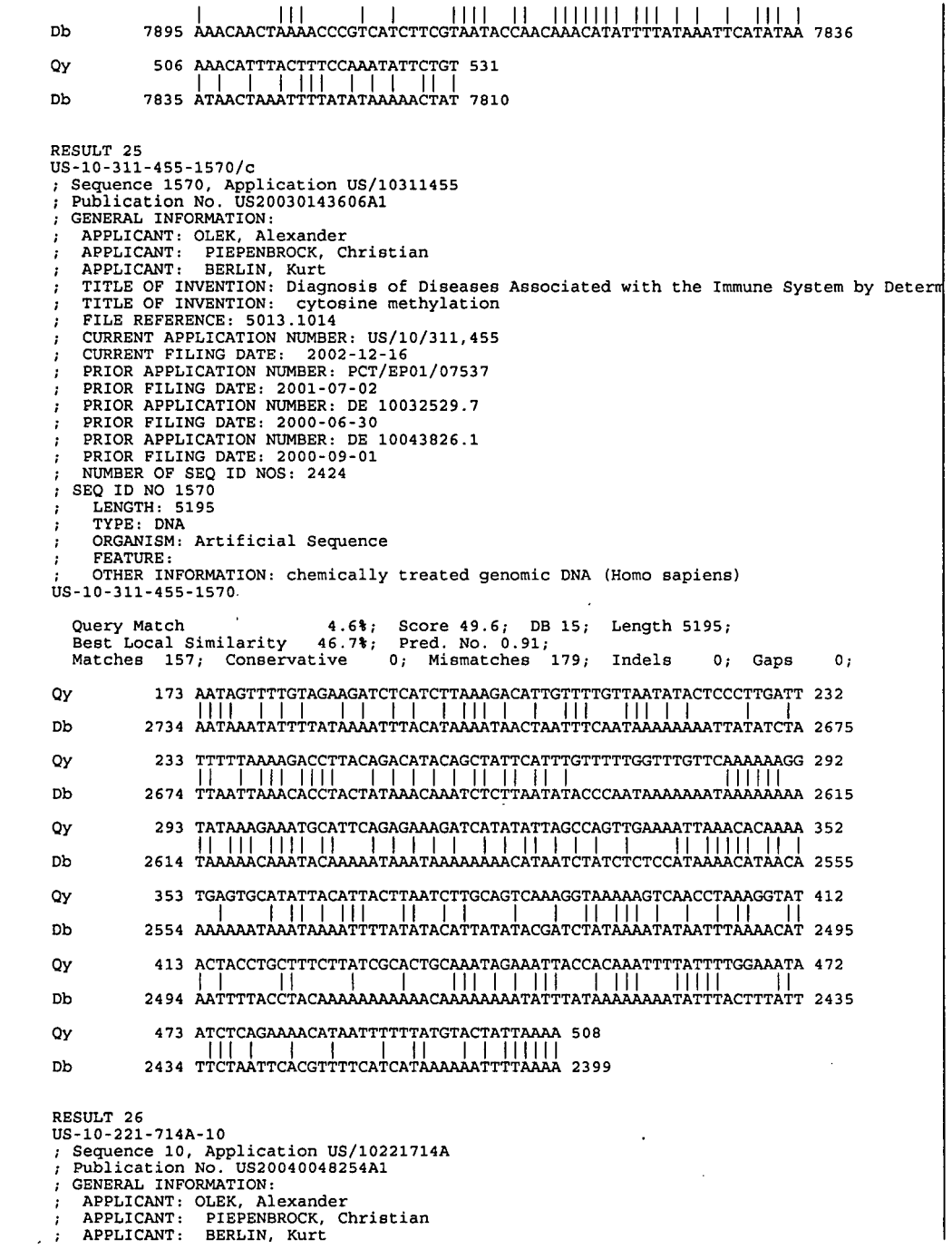
Qy       206 TTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTC 265
Db      8132 TATTTATAAAAAAAAATACATTAAATAAAATAATCTCCATTCACAAAAAATCAACAATCC 8073

Qy       266 ATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATA 325
Db      8072 ATTAAACTTCTCCTTAAACAAAATTTATCCTAAAATATAAATAAAAAAAAAATAATTAAA 8013

Qy       326 TTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTC 385
Db      8012 AAATAAAAAATATATAAAAAAAAAAAAAAAAAAAAAATAACACAAAAATACCTCACCCTACC 7953

Qy       386 AAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAA 445
Db      7952 AAAAATAAACACACAACCGCTACCTACCTATCCTA---TCTTATCGTTCCCAAAAACACA 7896

Qy       446 ATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTA 505
```



```
Db       7895 AAACAACTAAAACCCGTCATCTTCGTAATACCAACAAACATATTTTATAAATTCATATAA 7836

Qy        506 AAACATTTACTTTCCAAATATTCTGT 531
              | |  |  | |||  | | |  || |
Db       7835 ATAACTAAATTTTATATAAAAACTAT 7810


RESULT 25
US-10-311-455-1570/c
; Sequence 1570, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1570
;   LENGTH: 5195
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1570

  Query Match             4.6%;  Score 49.6;  DB 15;  Length 5195;
  Best Local Similarity   46.7%;  Pred. No. 0.91;
  Matches  157;  Conservative    0;  Mismatches  179;  Indels    0;  Gaps    0;

Qy        173 AATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATT 232
              ||||  | | |   | |  | |  | ||| |  |  |||   ||| | |     |   |
Db       2734 AATAAATATTTTATAAAATTTACATAAAATAACTAATTTCAATAAAAAAAATTATATCTA 2675

Qy        233 TTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGG 292
              ||  | ||| ||||   | | | | | || || || |                 ||||||
Db       2674 TTAATTAAACACCTACTATAAACAAATCTCTTAATATACCCAATAAAAAAATAAAAAAAA 2615

Qy        293 TATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAA 352
              || ||| |||| ||   | | | | |  | | || | | |   |    || ||||| || |
Db       2614 TAAAAACAAATACAAAAATAAATAAAAAAAACATAATCTATCTCTCCATAAAACATAACA 2555

Qy        353 TGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTAT 412
                |    | || | |||   ||  | |     |   | || ||| | |  | ||  ||
Db       2554 AAAAAATAAATAAAATTTTATATACATTATATACGATCTATAAAATATAATTTAAAACAT 2495

Qy        413 ACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATA 472
              | |    ||    |   |   ||| | | |||   | |||   |||||      ||
Db       2494 AATTTTACCTACAAAAAAAAAAACAAAAAAAATATTTATAAAAAAAATATTTACTTTATT 2435

Qy        473 ATCTCAGAAAACATAATTTTTTATGTACTATTAAAA 508
              |||  |   |   |    |  ||   | | ||||||
Db       2434 TTCTAATTCACGTTTTCATCATAAAAAATTTTAAAA 2399


RESULT 26
US-10-221-714A-10
; Sequence 10, Application US/10221714A
; Publication No. US20040048254A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with
;  TITLE OF INVENTION:  tumor suppressor genes and oncogenes
;  FILE REFERENCE: 5013.1005
;  CURRENT APPLICATION NUMBER: US/10/221,714A
;  CURRENT FILING DATE:  2003-01-21
;  PRIOR APPLICATION NUMBER: PCT/EP01/02955
;  PRIOR FILING DATE: 2001-03-15
;  PRIOR APPLICATION NUMBER: DE 10013847.0
;  PRIOR FILING DATE: 2000-03-15
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 540
; SEQ ID NO 10
;   LENGTH: 8169
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: 659, 665, 667..668, 678, 689, 6034, 6123
;   OTHER INFORMATION: n is a or g or c or t
US-10-221-714A-10

  Query Match             4.6%;  Score 49.6;  DB 17;  Length 8169;
  Best Local Similarity   45.7%;  Pred. No. 1.1;
  Matches  172;  Conservative    0;  Mismatches  204;  Indels    0;  Gaps    0;

Qy         35 TCTATGTTCTTATAAGCAAATGTTACAATGAGATATGAGTTTTAAAGCCAGATCTTCCTT 94
              | | ||   ||| |    || |||| | |||| ||||| | |   |       | |   ||
Db       6987 TTTTTGAGTTTAGAGTTTAAGGTTAGAGTGAGTTATGATTATATTATTGTATTTTAGTTT 7046

Qy         95 ATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATATT 154
                 |   |   |  | | ||| || | |      |||  ||||   ||     |  |
Db       7047 GGGTAATAGAGTGAGTTTTTGTTTTTAAGGAAAAAAATAAAAGTTGTATTTTTTGAGTTA 7106

Qy        155 GATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGT 214
                 |  | |  |  ||   || |||| | |    | |     ||||  |  || | ||
Db       7107 TTTATAGATATTTATGTATTTAATTTTATGGTTTTTATGTGGGTAAATATTTTATGTTAA 7166

Qy        215 TAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTT 274
              ||   ||     ||  ||  ||||||| |  ||  |   ||       || ||||  | |
Db       7167 TAGGTTATATAATTATTTAATTTAAAAAAATTTGAAATAATGATTAGTTTTATTTAATAT 7226

Qy        275 TGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGT 334
              |    | ||   || |  ||||   | |    || | | | | || |  | | |    |
Db       7227 TTTAATTTTTGTAAGAAATATAGGTATAATTGTTAAAAAATAAATTAATTGTAAATGAAA 7286

Qy        335 TGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAA 394
              | ||  || |   |  |||   || |  ||| | |   | ||  ||   || |||    |
Db       7287 TAAATGTTTATAAATGAATTTATGTAATTTAAAGTTTTAAAATTTTATGTTAAATTATAT 7346

Qy        395 AAGTCAACCTAAAGGT 410
              |  | ||   || | |
Db       7347 ATATAAAATAAATGTT 7362


RESULT 27
US-10-240-485-164/c
; Sequence 164, Application US/10240485
; Publication No. US20030148327A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with
;  TITLE OF INVENTION:  Metastasis
```

```
;  FILE REFERENCE: 5013.1007
;  CURRENT APPLICATION NUMBER: US/10/240,485
;  CURRENT FILING DATE:  2002-10-02
;  PRIOR APPLICATION NUMBER: PCT/EP01/03970
;  PRIOR FILING DATE: 2001-04-06
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 202
; SEQ ID NO 164
;   LENGTH: 20486
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-485-164

  Query Match             4.6%;  Score 49.6;  DB 15;  Length 20486;
  Best Local Similarity   47.7%;  Pred. No. 1.8;
  Matches  145;  Conservative    0;  Mismatches  159;  Indels    0;  Gaps    0;

Qy        231 TTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAA 290
              ||||| |||   | | | |   ||| | | ||| | |   ||       |  || | |
Db      12378 TTTTTATAATTTAACAAATATTTATAAAACAATTTACATTAATTAACCCCTTAAAACACA 12319

Qy        291 GGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAA 350
              | |||       |   | || |  ||     |  || |  || |  | | | |
Db      12318 AAAAAAAACCTTCAAAACATCATCAAAAATATTCTAAATACAATAAAATAACATATATAT 12259

Qy        351 AATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGT 410
              ||| |       ||  || |  | ||  |  | |  ||   ||  |  ||||   ||
Db      12258 AATAAAATATATTTTAATAAAATCATAATAAAATATAAAACAATTAACCAACAATAATCA 12199

Qy        411 ATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAA 470
              | |  ||||   ||||  ||     |  || |  ||||||  | || || ||||||  ||
Db      12198 AAATAACCTAACTTCTCTTCCACTTTTAATTTCAAATTAAAATAATTTATATTTTTTAAC 12139

Qy        471 TAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTG 530
              |    | |||  | |||||| || |  | |||| || | |       ||||  | |
Db      12138 ATAAAAAACAAATTAAAATTTTTAATCTTATTTTAATACTTATTACCGATAAATCATTTT 12079

Qy        531 TCAT 534
              ||
Db      12078 ATAT 12075


RESULT 28
US-10-067-514-1/c
; Sequence 1, Application US/10067514
; Publication No. US20030054531A1
; GENERAL INFORMATION:
;  APPLICANT: Gretarsdottir, Solveig
;  APPLICANT:  Jonsdottir, Sif
;  APPLICANT:  Reynisdottir, Sigridur Th.
;  TITLE OF INVENTION: HUMAN STROKE GENE
;  FILE REFERENCE: 2345.2010-003
;  CURRENT APPLICATION NUMBER: US/10/067,514
;  CURRENT FILING DATE:  2002-02-04
;  PRIOR APPLICATION NUMBER: US 09/811/352
;  PRIOR FILING DATE: 2001-03-19
;  NUMBER OF SEQ ID NOS: 84
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 1691139
;   TYPE: DNA
;   ORGANISM: Human
US-10-067-514-1

  Query Match             4.6%;  Score 49.6;  DB 14;  Length 1691139;
  Best Local Similarity   46.8%;  Pred. No. 15;
  Matches  189;  Conservative    0;  Mismatches  214;  Indels    1;  Gaps    1;

Qy        130 CATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGA 189
              |||||   |||   ||||||||||   |  |    ||| ||  |||  |  | |  | |
Db    1458464 CATATATATTTATATGAGAAATATATTTATTATATAATAAGCATATATATGAGAAATACA 1458405

Qy        190 TCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTAC 249
              |   ||  |||| |  |   ||| || ||||       |  |||| | ||  | |  ||
Db    1458404 TTATATAATAAATAGTTATATTTATTTATATGAGAATTATATTTATATATTATATATTTA 1458345

Qy        250 AGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGG-TATAAAGAAATGCATT 308
                 ||| |  |||    || | | |  ||   | | | | |  |||||   |    |
Db    1458344 TATAATAAATATATAATATTATATATATTTATATAATATATGTATATAATATATAATACA 1458285

Qy        309 CAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACAT 368
              | | ||  |  || |||      |  | | || || | |  ||||  ||  ||| || ||
Db    1458284 TATATAATAAATATGTATATAATATATAATAAATATATAATAAATATGTATATAATATAT 1458225

Qy        369 TACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTA 428
              |  |     |       | | | | |||  |      |                | | | |
Db    1458224 AATAAATATATATTTATACAATTTATAAGAAATATATATTTATATAAGAAATATATATAA 1458165

Qy        429 TCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAA 488
              |  |   | ||| ||  ||  | | ||| ||| ||  | ||  | |    | | |||
Db    1458164 TTTTATAAAATATATAATATATAATATATTATATATTATATATTTTGTATATATATATAT 1458105

Qy        489 TTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGTC 532
              | |   |  ||| || | || || ||    |   ||| |||
Db    1458104 AATATATATTTATATAAATATATATATTTAAAGAGGTTTCAGTC 1458061


RESULT 29
US-10-419-723-1/c
; Sequence 1, Application US/10419723
; Publication No. US20040014099A1
; GENERAL INFORMATION:
;  APPLICANT: Gretarsdottir, Solveig
;  APPLICANT:  Jonsdottir, Sif
;  APPLICANT:  Reynisdottir, Sigridur Th.
;  APPLICANT:  Thorleifsson, Gudmar
;  APPLICANT:  Gulcher, Jeffrey
;  TITLE OF INVENTION: SUSCEPTIBILITY GENE FOR HUMAN STROKE;
;  TITLE OF INVENTION:  METHODS OF TREATMENT
;  FILE REFERENCE: 2345.2010-005
;  CURRENT APPLICATION NUMBER: US/10/419,723
;  CURRENT FILING DATE:  2003-04-18
;  PRIOR APPLICATION NUMBER: 10/255,120
;  PRIOR FILING DATE: 2002-09-25
;  PRIOR APPLICATION NUMBER: 10/067,514
;  PRIOR FILING DATE: 2002-02-04
;  PRIOR APPLICATION NUMBER: 09/811,352
;  PRIOR FILING DATE: 2001-03-19
;  NUMBER OF SEQ ID NOS: 87
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 1691139
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-419-723-1

  Query Match             4.6%;  Score 49.6;  DB 17;  Length 1691139;
  Best Local Similarity   46.8%;  Pred. No. 15;
  Matches  189;  Conservative    0;  Mismatches  214;  Indels    1;  Gaps    1;

Qy        130 CATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGA 189
              |||||   |||   ||||||||||   |  |    ||| ||  |||  |  | |  | |
Db    1458464 CATATATATTTATATGAGAAATATATTTATTATATAATAAGCATATATATGAGAAATACA 1458405

Qy        190 TCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTAC 249
              |   ||  |||| |  |   ||| || ||||       |  |||| | ||  | |  ||
Db    1458404 TTATATAATAAATAGTTATATTTATTTATATGAGAATTATATTTATATATTATATATTTA 1458345
```

```
Qy        250 AGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGG-TATAAAGAAATGCATT 308
              ||| |  |||    || | | |  ||   | | | | |  |||||   |    |
Db    1458344 TATAATAAATATATAATATTATATATATTTATATAATATATGTATATAATATATAATACA 1458285

Qy        309 CAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACAT 368
              | | ||  |  || |||    |  | | || || | |  ||||  ||  ||| || ||
Db    1458284 TATATAATAAATATGTATATAATATATAATAAATATATAATAAATATGTATATAATATAT 1458225

Qy        369 TACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTA 428
              |   |    |      | | | | |||  |     |                  | | | |
Db    1458224 AATAAATATATATTTATACAATTTATAAGAAATATATATTTATATAAGAAATATATATAA 1458165

Qy        429 TCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAA 488
              |   |    | ||| ||  ||  | | ||| ||| ||  | ||  | |    | | |||
Db    1458164 TTTTATAAAATATATAATATATAATATATTATATATTATATATTTTGTATATATATATAT 1458105

Qy        489 TTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGTC 532
              | |    |  ||| || | || || ||    |   ||| |||
Db    1458104 AATATATATTTATATAAATATATATATTTAAAGAGGTTTCAGTC 1458061


RESULT 30
US-09-938-842A-4474/c
; Sequence 4474, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
;  APPLICANT: Harper, Jeff
;  APPLICANT:  Kreps, Joel
;  APPLICANT:  Wang, Xun
;  APPLICANT:  Zhu, Tong
;  TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
;  TITLE OF INVENTION:  SAME, AND METHODS OF USE
;  FILE REFERENCE: SCRIP1300-3
;  CURRENT APPLICATION NUMBER: US/09/938,842A
;  CURRENT FILING DATE:  2001-08-24
;  PRIOR APPLICATION NUMBER: US 60/227,866
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: US 60/264,647
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/300,111
;  PRIOR FILING DATE: 2001-06-22
;  NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 4474
;   LENGTH: 2000
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
US-09-938-842A-4474

  Query Match             4.5%;  Score 49.4;  DB 9;  Length 2000;
  Best Local Similarity   45.4%;  Pred. No. 0.64;
  Matches  214;  Conservative    0;  Mismatches  256;  Indels    1;  Gaps    1;

Qy        120 TGAAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTT 179
              | || |   | |||  |   | || | || |   |  ||||   |     |||||||||
Db       1375 TAAAATTAGTAATACAAAAATCGTGGTGATAATATTTTCATACTATTAATTTTAATAGTT 1316

Qy        180 TTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAA 239
              ||  ||         | |||||  | |   | | | |||||  |    |  || | | ||
Db       1315 TTTGTGATCGGTAACTATTAAAACCGTATATATATATATATA-TATATATATATATATAT 1257

Qy        240 AAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAG 299
              |       || | | ||   |  |  | | | |||||     ||  |  ||||   ||
Db       1256 ATATATATATATATATTTTGTAAACCGTATATTTTTTACCAGTAAATTAAAGAATTATCA 1197

Qy        300 AAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGC 359
              |  ||    || || |||| | |   ||      |||| | || |  |     | | |
Db       1196 ATGTGTTAACATAGGAAGAACGTTCCTTTTATTATTGAGATTTGATAATCCTTTTACCCC 1137

Qy        360 ATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCT 419
              ||  | |    ||  ||           || |   |  |  ||||  ||
Db       1136 TGGTTCTAGTTTAGAAGTTTATCAATGTTCTTATATCCTTCAGTTAAACATAGGAACAAC 1077

Qy        420 GCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAG 479
              | |   ||     |      | |   || |      |||  |  | |  |||| |||| |
Db       1076 GTTCCTTTTATTCTTGAGATTTGATAAATCAATTAAACATATTATCAGAAAAAATCTAAT 1017

Qy        480 AAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGTCATTCAGG 539
              ||      | | ||||| | ||  | |||| ||   ||    ||||||  | | ||| |
Db       1016 AATCATACAATATTTATTTTCTTATTAAACGTTATTTTCAGTAATATTTTCTTATTAAAC 957

Qy        540 AGTATGGAAGTATCGATGGCTTCTTTAAAATGAAGCAGGAGGGTCTGGCAG 590
              || |  ||||    |   || |  |||   | |||  | |  |  |||
Db        956 TGTTTTTTAGTAATATTTTTTTTTGATAAAATTAACAGATGTGATTTTCAG 906


RESULT 31
US-09-938-842A-4474/c
; Sequence 4474, Application US/09938842A
; Publication No. US20040009476A9
; GENERAL INFORMATION:
;  APPLICANT: Harper, Jeff
;  APPLICANT:  Kreps, Joel
;  APPLICANT:  Wang, Xun
;  APPLICANT:  Zhu, Tong
;  TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
;  TITLE OF INVENTION:  SAME, AND METHODS OF USE
;  FILE REFERENCE: SCRIP1300-3
;  CURRENT APPLICATION NUMBER: US/09/938,842A
;  CURRENT FILING DATE:  2001-08-24
;  PRIOR APPLICATION NUMBER: US 60/227,866
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: US 60/264,647
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/300,111
;  PRIOR FILING DATE: 2001-06-22
;  NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 4474
;   LENGTH: 2000
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
US-09-938-842A-4474

  Query Match             4.5%;  Score 49.4;  DB 11;  Length 2000;
  Best Local Similarity   45.4%;  Pred. No. 0.64;
  Matches  214;  Conservative    0;  Mismatches  256;  Indels    1;  Gaps    1;

Qy        120 TGAAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTT 179
              | || |   | |||  |   | || | || |   |  ||||   |     |||||||||
Db       1375 TAAAATTAGTAATACAAAAATCGTGGTGATAATATTTTCATACTATTAATTTTAATAGTT 1316

Qy        180 TTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAA 239
              ||  ||         | |||||  | |   | | | |||||  |    |  || | | ||
Db       1315 TTTGTGATCGGTAACTATTAAAACCGTATATATATATATATA-TATATATATATATATAT 1257

Qy        240 AAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAG 299
              |       || | | ||   |  |  | | | |||||     ||  |  ||||   ||
Db       1256 ATATATATATATATATTTTGTAAACCGTATATTTTTTACCAGTAAATTAAAGAATTATCA 1197

Qy        300 AAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGC 359
              |  ||    || || |||| | |   ||      |||| | || |  |     | | |
Db       1196 ATGTGTTAACATAGGAAGAACGTTCCTTTTATTATTGAGATTTGATAATCCTTTTACCCC 1137

Qy        360 ATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCT 419
              ||  | |    ||  ||           || |   |  |  ||||  ||
Db       1136 TGGTTCTAGTTTAGAAGTTTATCAATGTTCTTATATCCTTCAGTTAAACATAGGAACAAC 1077

Qy        420 GCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAG 479
              | |   ||     |      | |   || |      |||  |  | |  |||| |||| |
Db       1076 GTTCCTTTTATTCTTGAGATTTGATAAATCAATTAAACATATTATCAGAAAAAATCTAAT 1017

Qy        480 AAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGTCATTCAGG 539
              ||      | | ||||| | ||  | |||| ||   ||    ||||||  | | ||| |
Db       1016 AATCATACAATATTTATTTTCTTATTAAACGTTATTTTCAGTAATATTTTCTTATTAAAC 957

Qy        540 AGTATGGAAGTATCGATGGCTTCTTTAAAATGAAGCAGGAGGGTCTGGCAG 590
```

```
                 || |   ||||    |   || |   |||   | |||  | |  |  |||
Db        956 TGTTTTTTAGTAATATTTTTTTTTGATAAAATTAACAGATGTGATTTTCAG 906


RESULT 32
US-10-311-455-1472
; Sequence 1472, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1472
;   LENGTH: 6102
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1472

  Query Match             4.5%;  Score 49.2;  DB 15;  Length 6102;
  Best Local Similarity   47.4%;  Pred. No. 1.2;
  Matches  147;  Conservative    0;  Mismatches  163;  Indels    0;  Gaps    0;

Qy        261 TATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATC 320
              | ||  |  ||| |   | | |  ||||||  |||  ||||  |   ||   | ||| |
Db       5758 TGTTTGTAAGTTGTATTTATATATAAAAAATTTATGGAGAATTTTTTTTTTAATAAAAAGA 5817

Qy        321 ATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTG 380
              |  |||||    ||| ||  |     |    || || | |||||   ||  |   |  ||
Db       5818 AAGTATTAAGTTGTTTAATTTATTTAAATTTATTAGAGAATATTTAGTTGGTGTTTAATG 5877

Qy        381 CAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAA 440
              |   |  |||||      | | |   |||       |  | ||| |  | | | |
Db       5878 GGTTTTAGTTTAAAAGAGGGAATTGAGTTTATATATTGAGTAATTTTTTTTTTAATAGATA 5937

Qy        441 TAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTAC 500
              || |||||   |   |  | ||||  |||   ||   ||      |||  | |||||  |
Db       5938 TATAAATTTATATGGAAATGATTTAAGAATATATAAAAGTGTGAATATATATTTATTTGT 5997

Qy        501 TATTAAAACATTTACTTTCCAAATATTCTGTCATTCAGGAGTATGGAAGTATCGATGGCT 560
              |   ||| | |  |||  | ||||  | |  |   |  ||  | || |    | |
Db       5998 TTAAAAGAAAGTATTTTTATATATATATTTTTTTGTTGTGTTAGTGGAGAAGGTTTAGAA 6057

Qy        561 TCTTTAAAAT 570
              |  ||| ||
Db       6058 GCGATAATAT 6067


RESULT 33
US-10-311-455-457/c
; Sequence 457, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 457
;   LENGTH: 6244
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-457

  Query Match             4.5%;  Score 49.2;  DB 15;  Length 6244;
  Best Local Similarity   47.4%;  Pred. No. 1.2;
  Matches  181;  Conservative    0;  Mismatches  198;  Indels    3;  Gaps    1;

Qy        146 AGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACA 205
              | | ||||  ||   |||| ||   |  |||  | | || |    |  |  |  |
Db       2605 ATATATATATATACATACATATATATATATATATATATATATACATACATATATATAT 2546

Qy        206 TTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTC 265
              | | |   ||       |  | || | | || |    | |||| | ||| |  |||
Db       2545 ATATATATATATATACATATATATATATATATATATATACATACATATATATATATAT 2486

Qy        266 ATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATA 325
              ||   |  |        |  | | |  |||| | | ||   ||   |   | ||| |  |
Db       2485 ATACATACATATACATATATATATATATATATATATATTTTATAATCCATATCTTTAA 2426

Qy        326 TTAGCCAGTTGAAAATTAAACA---CAAAATGAGTGCATATTACATTACTTAATCTTGCA 382
              | | | |  | ||||  ||| |    |||||       | |||  ||||  || |||
Db       2425 TCATCAAAATAAAAAAAAAAAAAACAAAAATTTTAAAAAATTTAATTAAATATTCTAAAT 2366

Qy        383 GTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATA 442
              |||  |||||| || | | |     || | ||         |||   ||    | | |
Db       2365 TAAAAACTTAAAAAATCCATCAATTATAATTCCACAATAAAAATTAAAACAACCAATAAA 2306

Qy        443 GAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTA 502
              |||  ||  |   | |||  |    ||||| ||  |  ||||  || | | ||    |
Db       2305 TAAAATAATAACTAATTTCATACTTAAATATTCAAACTAAACTAAAATCTATAACCAAAC 2246

Qy        503 TTAAAACATTTACTTTCCAAAT 524
              | |||| |  ||  || || ||
Db       2245 TAAAAAAAAATATATTACACAT 2224


RESULT 34
US-10-433-793-89/c
; Sequence 89, Application US/10433793
; Publication No. US20040142334A1
; GENERAL INFORMATION:
;  APPLICANT: Epigenomics AG
;  TITLE OF INVENTION: Diagnose von mit Angiogenese assoziierten Krankheiten
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/10/433,793
;  CURRENT FILING DATE:  2003-06-06
;  NUMBER OF SEQ ID NOS: 212
; SEQ ID NO 89
;   LENGTH: 34688
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-433-793-89

  Query Match             4.5%;  Score 49.2;  DB 18;  Length 34688;
  Best Local Similarity   43.6%;  Pred. No. 2.9;
  Matches  277;  Conservative    0;  Mismatches  353;  Indels    6;  Gaps    1;

Qy        131 ATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGAT 190
```

```
Db      16782 ATAAAATATACTTTCAATATTAAATACAAATAAAATTACTATATATAATTCGAAAAAAAA 16723

Qy        191 CTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACA 250

Db      16722 ATCTTCAAAACTCTAACACTTCAATTAAATTTATATAACATTATAATTTCATAAATCAAA 16663

Qy        251 GACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCA 310

Db      16662 ACCATAAAAATCCTCTCTAATATTTAAATTATCTTAAATCCACTTAAAAAA------TCA 16609

Qy        311 GAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTA 370

Db      16608 CAAAATAAAAAAATAAAAACAAACTTAATAAATATATAAATATAAAAAAATAAATATTTC 16549

Qy        371 CTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATC 430

Db      16548 TAATAATTTCACCATTTAATTAAAACAACGACAAAATACAATAAAAATTATTTTTTAAAA 16489

Qy        431 GCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATT 490

Db      16488 AAAACACCCTCTTAAATTCACCATAAACTAAATTTTAAAATATATACATATAAACAAAAC 16429

Qy        491 TTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGTCATTCAGGAGTATGGAAGT 550

Db      16428 TAAAAACTAATATAAAAACATTAAATTTTAAAAAAATAAAACAATAAATAATTTTTAAAA 16369

Qy        551 ATCGATGGCTTCTTTAAAATGAAGCAGGAGGGTCTGGCAGAGAGTATCTATGAAATAAGT 610

Db      16368 TTCTCTTTAATCATATTTTTAATAAAAAAAATTTTCTTAAATACAATAAACACATAATTT 16309

Qy        611 TCCTCTGACCTTCACGCTTAATTTTCTGAATGGAGTGGAGCAAATTACTTCAAGCTTCAC 670

Db      16308 TCATTCATATACTAAAATAAACATCATAAACCTTATTCAACTAAATTATAACAATAAATC 16249

Qy        671 TTAACTTGCATATGAAATGAACCGTACAAAAATACAAGAGTGTCAGGAGAAAGTTATGCT 730

Db      16248 CTATTTATCTTAAATAATTAATATCTCAAAATATCCCATATTTCAAAAAAAATTAACAAT 16189

Qy        731 CTGGTAAATATTTTGCAAAACAGATAAAAGATAATA 766

Db      16188 CAAATCTTAAATTTCTTAAAAATTTAAAAAAAAATA 16153


RESULT 35
US-10-311-455-1389/c
; Sequence 1389, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1389
;   LENGTH: 6265
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1389

  Query Match             4.5%;  Score 48.8;  DB 15;  Length 6265;
  Best Local Similarity   50.9%;  Pred. No. 1.5;
  Matches  116;  Conservative    0;  Mismatches  112;  Indels    0;  Gaps    0;

Qy        262 ATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCA 321

Db       1700 AATCAATTCCTTATTTTTTAAACGAATACTCTAAAAATAAATAAAATTACCCAAAAAAAA 1641

Qy        322 TATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGC 381

Db       1640 ATCAAAAACACTTTTATAAACACTAACTTCATACAAATAAATAACACTAACTCCTATAAA 1581

Qy        382 AGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAAT 441

Db       1580 CCTTAAACTCACAAACTCATATTTAAAAAAATTTAAAATCTATATAATATAATTATACAC 1521

Qy        442 AGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAAT 489

Db       1520 TAAATTTATATCAAATTTTATTTCTATAATAATCCTTCAAAAAACAAT 1473


RESULT 36
US-10-311-455-1731/c
; Sequence 1731, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determi
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1731
;   LENGTH: 7455
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1731

  Query Match             4.5%;  Score 48.8;  DB 15;  Length 7455;
  Best Local Similarity   45.0%;  Pred. No. 1.7;
  Matches  182;  Conservative    0;  Mismatches  222;  Indels    0;  Gaps    0;

Qy        149 AATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTG 208

Db       5837 ATTTTATACCTATAAATTTTATTAAAAATATAAAATCAAAAAATACTCATAAAAAAATAA 5778

Qy        209 TTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATT 268

Db       5777 TTTTATTTTACTAACCTTAATAAACTAAAAAATAAATATATAAACATAAATATATAAATA 5718

Qy        269 TGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTA 328

Db       5717 TATATATCATAATTTTAAAAAAATATTTCCCCAATAAAAAATAATAAAAATATAAAATAA 5658

Qy        329 GCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAA 388

Db       5657 TAAAATCTTTATTTTCTACCATATCACAACTAAATAAAAACAACTTAAATAATATAAAAC 5598

Qy        389 GGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATT 448

Db       5597 AATAAAACAATAAAACATAACTATTATACAACATTTTTTATTATACAAAATAACATAACT 5538

Qy        449 ACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAA 508
```

```
Db       5537 ATTTAAAATAATAAAAAAATCAATAAAAAAATAATAAAAAATATTTACTTATAAAAAAAA 5478

Qy        509 CATTTACTTTCCAAATATTCTGTCATTCAGGAGTATGGAAGTAT 552
Db       5477 AATAAATTTTTAAAAATTAAACTTATAAAACAATATATACATTT 5434


RESULT 37
US-10-311-455-739
; Sequence 739, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 739
;   LENGTH: 19659
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-739

  Query Match             4.5%;  Score 48.8;  DB 15;  Length 19659;
  Best Local Similarity   48.5%;  Pred. No. 2.7;
  Matches  193;  Conservative    0;  Mismatches  202;  Indels    3;  Gaps    2;

Qy        111 TCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAG 170
Db      11664 TTTATTTGTTTAAAAGGTTTTTTTTAGTTTTATTAAGATAAATGAAGTGTTATAAAAGTA 11723

Qy        171 TTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGA 230
Db      11724 TTTAAAGTAATGTATATTATAATGGTTAATATAAAATAGTATTGAGTATTTGATAGTGAA 11783

Qy        231 TTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAA 290
Db      11784 ATTTTTTATTTAAATTATAGTTTTTATAATTGTTTAATAATAAGTTTTTAGTTTTTTATT 11843

Qy        291 GGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAA 350
Db      11844 GGGATTTAATAAGAAAATTAGAAAATTATATTTTATTTATGAAAAAAAAATTTTTGTTAA 11903

Qy        351 AATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGT 410
Db      11904 TATATTATTATAGTAGATAAG-TATTATAAAATTTTTTTGGTTATATTAAAAATTAATAT 11962

Qy        411 ATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAA 470
Db      11963 ATAAATATTTATTTTTTATAATTTTAAAAAGATAAATTATT--AATTTTTGAAATAGAAA 12020

Qy        471 TAATCTCAGAAAACATAATTTTTTATGTACTATTAAAA 508
Db      12021 TTTTTTTATATAAGATAATGTAATTTGTTGGATTTTAA 12058


RESULT 38
US-10-322-281-676/c
; Sequence 676, Application US/10322281
; Publication No. US20040126762A1
; GENERAL INFORMATION:
;  APPLICANT: David W. Morris
;  APPLICANT:  Marc S. Malandro
;  TITLE OF INVENTION: Novel Compositions and Methods in Cancer
;  FILE REFERENCE: 529452001000
;  CURRENT APPLICATION NUMBER: US/10/322,281
;  CURRENT FILING DATE:  2002-12-17
;  NUMBER OF SEQ ID NOS: 866
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 676
;   LENGTH: 93544
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-322-281-676

  Query Match             4.5%;  Score 48.6;  DB 18;  Length 93544;
  Best Local Similarity   47.4%;  Pred. No. 6.4;
  Matches  176;  Conservative    0;  Mismatches  194;  Indels    1;  Gaps    1;

Qy         33 TCTCTATGTTCTTATAAGCAAATGTTACAATGAGATATGAGTTTTAAAGCCAGATCTTCC 92
Db      89520 TCAGTTTCTACATACTGCCTTGTATTTGATTTTTTTTTTTTTGGTAATGTACCATTTTTA 89461

Qy         93 TTATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATA 152
Db      89460 TTTTCTTCTTTCTTTGCCTGTATATTTTTAGTTATTTTCAGTTCCTATCATTTTAAAATA 89401

Qy        153 TTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTT 212
Db      89400 TTTATCCCCAAATGTAATCCCTTCGAGTT-TATTATTTATAATGTCAGTGGTTTTATTTT 89342

Qy        213 GTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTT 272
Db      89341 TTAAATATTCTAGATTCTTACATTGATTCTTTCTGATTGACAAATTCATCCATATCAGAA 89282

Qy        273 TTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCA 332
Db      89281 TTTTAAGTGTTAACACAAGCCCTCAAGCAAGGCTGCTGAAGAATTTGCATACTTTTCTAT 89222

Qy        333 GTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTA 392
Db      89221 GTTCCAAACTTGAAATTTCCAAAGTTTTTAATGCATTACTTTAGACCCAAGAGAAAGGTA 89162

Qy        393 AAAAGTCAACC 403
Db      89161 CAACATTTTCC 89151


RESULT 39
US-10-087-192-1504
; Sequence 1504, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
;  APPLICANT: Morris, David W.
;  APPLICANT:  Engelhard, Eric K.
;  TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
;  TITLE OF INVENTION:  CANCER
;  FILE REFERENCE: 529452000122
;  CURRENT APPLICATION NUMBER: US/10/087,192
;  CURRENT FILING DATE:  2002-03-01
;  PRIOR APPLICATION NUMBER: US 09/747,377
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: US 09/798,586
;  PRIOR FILING DATE: 2001-03-02
;  NUMBER OF SEQ ID NOS: 2059
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1504
;   LENGTH: 142318
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(142318)
```

```
;   OTHER INFORMATION: n = A,T,C or G
US-10-087-192-1504

  Query Match             4.5%;  Score 48.6;  DB 13;  Length 142318;
  Best Local Similarity   47.4%;  Pred. No. 7.9;
  Matches  176;  Conservative    0;  Mismatches  194;  Indels    1;  Gaps    1;

Qy         33 TCTCTATGTTCTTATAAGCAAATGTTACAATGAGATATGAGTTTTAAAGCCAGATCTTCC 92
              ||  | | | | ||    |  | ||  | |    | |   |  ||| |    || ||
Db     137461 TCAGTTTCTACATACTGCCTTGTATTTGATTTTTTTTTTTTTGGTAATGTACCATTTTTA 137520

Qy         93 TTATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATA 152
              || |||   | |     || ||  | || |  | | |    |  | |  ||   |||||
Db     137521 TTTTCTTCTTTCTTTGCCTGTATATTTTTAGTTATTTTCAGTTCCTATCATTTTAAAATA 137580

Qy        153 TTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTT 212
              || |||   | |  | |     | |  || ||  |  | | || | |  |   || ||||
Db     137581 TTTATCCCCAAATGTAATCCCTTCGAGTT-TATTATTTATAATGTCAGTGGTTTTATTTT 137639

Qy        213 GTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTT 272
              | ||||| ||   ||  |   || |      || |  |||| |    |   ||  |
Db     137640 TTAAATATTCTAGATTCTTACATTGATTCTTTCTGATTGACAAATTCATCCATATCAGAA 137699

Qy        273 TTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCA 332
              |||    |||| | | |||   | ||| || ||      ||||    ||||  ||
Db     137700 TTTTAAGTGTTAACACAAGCCCTCAAGCAAGGCTGCTGAAGAATTTGCATACTTTTCTAT 137759

Qy        333 GTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTA 392
              |||  ||| |  | |        |||   || | |||||||| |       ||  |||||||
Db     137760 GTTCCAAACTTGAAATTTCCAAAGTTTTTAATGCATTACTTTAGACCCAAGAGAAAGGTA 137819

Qy        393 AAAAGTCAACC 403
              ||  |   ||
Db     137820 CAACATTTTCC 137830


RESULT 40
US-10-311-455-1369
; Sequence 1369, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1369
;   LENGTH: 5641
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1369

  Query Match             4.4%;  Score 48.4;  DB 15;  Length 5641;
  Best Local Similarity   47.4%;  Pred. No. 1.8;
  Matches  145;  Conservative    0;  Mismatches  161;  Indels    0;  Gaps    0;

Qy        171 TTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGA 230
              ||| ||  ||| | | || | |  | |||     ||| | || ||| |    |   ||
Db       2111 TTATTATATTTTTTGTAGTTATTTTTTTAGTTTAATTTTGTTTTTATTTAGTTTTTTTAT 2170

Qy        231 TTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAA 290
              || |     | |  |     | ||| |  | || || |||| | | |||    | | |
Db       2171 GTTGTAATCGATAAATATGTTATATATATATTTTTATATGTTATGGTTTTAACGATATAT 2230

Qy        291 GGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAA 350
              ||||  | |       |  | | |  |    || || ||  |   |||||   || | ||
Db       2231 TATATATATATTATTTTATATAAATGGTGTTTAAATAAGTTATAAGAAAAAATAAGAAAA 2290

Qy        351 AATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGT 410
              ||||  |  |||||     || |||||        |   ||  ||  |  |  |  || || |
Db       2291 TATGAATTTATATTGTTATATTTAATAGATATATGTAAAATATTATATAGATATATAGAT 2350

Qy        411 ATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAA 470
                 |   |   |  ||||     |  | |||   || |    | |||||  | |   |
Db       2351 TATATTAATATATATTTATATATTTATATATATTTATAATATTATATTTTTATATTTTAT 2410

Qy        471 TAATCT 476
              || | |
Db       2411 TATTGT 2416


RESULT 41
US-10-239-676-90/c
; Sequence 90, Application US/10239676
; Publication No. US20030082609A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
;  FILE REFERENCE: 5013.1003
;  CURRENT APPLICATION NUMBER: US/10/239,676
;  CURRENT FILING DATE:  2002-09-24
;  PRIOR APPLICATION NUMBER: PCT/EP01/03968
;    DE 10019058.8
;    DE 10019173.8
;    DE 10032529.7
;    DE 10043826.1
;  PRIOR FILING DATE: 2001-04-06
;    2000-04-06
;    2000-04-07
;    2000-06-30
;    2000-09-01
;  NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 90
;   LENGTH: 5689
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1818, 1930..1931, 1934..1935, 1947..1948, 1966)
US-10-239-676-90

  Query Match             4.4%;  Score 48.4;  DB 14;  Length 5689;
  Best Local Similarity   48.4%;  Pred. No. 1.8;
  Matches  121;  Conservative    0;  Mismatches  129;  Indels    0;  Gaps    0;

Qy        238 AAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAA 297
              |||| |   || |   ||      |||   ||| |||||   || || |||||   |  ||
Db       2030 AAAAAAAAATAAATTTATTTTTATATATTTTTTTTTTTTTATTATTTAAAAATAATTAAA 1971

Qy        298 AGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGT 357
              | ||    |   |   || |  | |  |  |  |  |   ||| | | ||| |
Db       1970 AAAANATAAAAAAATTTAATAANNTTAAAAAAAAANNTANNATAAAAAAAAATAATAATA 1911

Qy        358 GCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTAC 417
              |||  | |||  |||||      | |  ||   |||||     |  ||    |  | ||
Db       1910 ATATAAAAAATTTATTAATAAAAAAATATAAATAAAAAAAAATAAATATTAATTAAATAA 1851

Qy        418 CTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTC 477
                |  |||  |||         | | |  |||||  | ||||| ||    |  ||||||
Db       1850 TTTATTTTATATATATTAAAATAAAATAATTANTAAAAATTATAAAATAAAAATAAATAA 1791
```

```
Qy        478 AGAAAACATA 487
              | |||  |||
Db       1790 AAAAATAATA 1781


RESULT 42
US-10-240-453-100/c
; Sequence 100, Application US/10240453
; Publication No. US20030148326A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA
;  TITLE OF INVENTION:  Transcription
;  TITLE OF INVENTION:  by Means of Assessing the Methylation Status of Genes Associated
;  TITLE OF INVENTION:  With DNA TRanscription
;  FILE REFERENCE: 5013.1009
;  CURRENT APPLICATION NUMBER: US/10/240,453
;  CURRENT FILING DATE:  2002-10-02
;  PRIOR APPLICATION NUMBER: PCT/EP01/03973
;  PRIOR FILING DATE: 2001-04-06
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 350
; SEQ ID NO 100
;   LENGTH: 5689
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1818, 1930..1931, 1934..1935, 1947..1948, 1966)
US-10-240-453-100

  Query Match             4.4%;  Score 48.4;  DB 15;  Length 5689;
  Best Local Similarity   48.4%;  Pred. No. 1.8;
  Matches  121;  Conservative    0;  Mismatches  129;  Indels    0;  Gaps    0;

Qy        238 AAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAA 297
              |||| |   || |   ||     |||   ||| ||||| ||  || |||||   |  ||
Db       2030 AAAAAAAAATAAATTTATTTTTATATATTTTTTTTTTTTTTATTATTTAAAAATAATTAAA 1971

Qy        298 AGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGT 357
              | ||    |   |   || |  |  |  |  |  |   ||| | | ||| |
Db       1970 AAAANATAAAAAAATTTAATAANNTTAAAAAAAAANNTANNATAAAAAAAAATAATAATA 1911

Qy        358 GCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTAC 417
               |||  | |||  |||||     | |  ||   |||||    |  ||    |  | ||
Db       1910 ATATAAAAAATTTATTAATAAAAAAATATAAATAAAAAAAAAATAAATATTAATTAAATAA 1851

Qy        418 CTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTC 477
              |  |||  |||        | | |  ||||  | ||||| ||   |  ||||||
Db       1850 TTTATTTTATATATATTAAAATAAAATAATTANTAAAAATTATAAAATAAAAATAAATAA 1791

Qy        478 AGAAAACATA 487
              | |||  |||
Db       1790 AAAAATAATA 1781


RESULT 43
US-10-221-714A-148/c
; Sequence 148, Application US/10221714A
; Publication No. US20040048254A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with
;  TITLE OF INVENTION:  tumor suppressor genes and oncogenes
;  FILE REFERENCE: 5013.1005
;  CURRENT APPLICATION NUMBER: US/10/221,714A
;  CURRENT FILING DATE:  2003-01-21
;  PRIOR APPLICATION NUMBER: PCT/EP01/02955
;  PRIOR FILING DATE: 2001-03-15
;  PRIOR APPLICATION NUMBER: DE 10013847.0
;  PRIOR FILING DATE: 2000-03-15
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 540
; SEQ ID NO 148
;   LENGTH: 5689
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: 1818, 1930..1931, 1934..1935, 1947..1948, 1966
;   OTHER INFORMATION: n is a or g or c or t
US-10-221-714A-148

  Query Match             4.4%;  Score 48.4;  DB 17;  Length 5689;
  Best Local Similarity   48.4%;  Pred. No. 1.8;
  Matches  121;  Conservative    0;  Mismatches  129;  Indels    0;  Gaps    0;

Qy        238 AAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAA 297
              |||| |   || |   ||     |||   ||| ||||| ||  || |||||   |  ||
Db       2030 AAAAAAAAATAAATTTATTTTTATATATTTTTTTTTTTTTTATTATTTAAAAATAATTAAA 1971

Qy        298 AGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGT 357
              | ||    |   |   || |  |  |  |  |  |   ||| | | ||| |
Db       1970 AAAANATAAAAAAATTTAATAANNTTAAAAAAAAANNTANNATAAAAAAAAATAATAATA 1911

Qy        358 GCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTAC 417
               |||  | |||  |||||     | |  ||   |||||    |  ||    |  | ||
Db       1910 ATATAAAAAATTTATTAATAAAAAAATATAAATAAAAAAAAAATAAATATTAATTAAATAA 1851

Qy        418 CTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTC 477
              |  |||  |||        | | |  ||||  | ||||| ||   |  ||||||
Db       1850 TTTATTTTATATATATTAAAATAAAATAATTANTAAAAATTATAAAATAAAAATAAATAA 1791

Qy        478 AGAAAACATA 487
              | |||  |||
Db       1790 AAAAATAATA 1781


RESULT 44
US-10-311-455-1445/c
; Sequence 1445, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
```

```
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1445
;   LENGTH: 6106
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1445

  Query Match             4.4%;  Score 48.4;  DB 15;  Length 6106;
  Best Local Similarity   45.5%;  Pred. No. 1.9;
  Matches  172;  Conservative    0;  Mismatches  206;  Indels    0;  Gaps    0;

Qy        160 TTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATA 219
              | | ||| ||   || |   | | | ||   |  | ||||| |      | |  | | |
Db       4729 TAAAAAAACAAACAAAAACCTCGAAAAATTACATAACTTAAAAATTAAAATATAATTAAA 4670

Qy        220 TACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTT 279
              | |  |   | ||    ||   | | | | | | | |   |    |   | |    |
Db       4669 AAATAACCAAAATTAAAAAATTTAACCAAAATAAACAAAATAAAAATTAAAATCTAAACT 4610

Qy        280 TGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAA 339
              ||||||   | ||| |||    |||   ||  |      ||  |      || |||
Db       4609 AAAAAAAAAAAAAAAAAAAAAAAAAACTTCTTTTAATAAATTACTAAAAAAACCTTCAAA 4550

Qy        340 ATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTC 399
              |||   |||   |    | ||||  ||| | ||  ||  | |      | |
Db       4549 TAAAAAATAAAATAAAAATTAAATTAATTTATTCAAATTTTAATTTCTCTTTATCCACT 4490

Qy        400 AACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTT 459
              |||  |||  ||    |      ||   | |   |   || |||  | ||   ||||| ||
Db       4489 AACTAAAAAATAATTAATAAATTTCAATTTAAAATACCATATAATATTTCAAACAAAATT 4430

Qy        460 TATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTC 519
              |   |  ||  ||  |   |||  |||| | |   |      ||  || | ||| |||||
Db       4429 AAAAATATAACAAAAATTTAAAATAATAAATATAAAAAATATAAAAATAAATTAACTTTA 4370

Qy        520 CAAATATTCTGTCATTCA 537
              |||  ||||  | |||  |
Db       4369 TAAAAATTCATTTATTTA 4352


RESULT 45
US-10-257-166-113/c
; Sequence 113, Application US/10257166
; Publication No. US20040023230A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Method and Nucleic Acids for Analysing the Methylation of
;  TITLE OF INVENTION:  Genes Implicated in Pharmacogenomics
;  FILE REFERENCE: 5013.1011
;  CURRENT APPLICATION NUMBER: US/10/257,166
;  CURRENT FILING DATE:  2002-10-07
;  PRIOR APPLICATION NUMBER: PCT/EP01/07470
;    DE 10032529.7
;    DE 10043826.1
;  PRIOR FILING DATE: 2001-06-29
;    2000-06-30
;    2000-09-01
;  NUMBER OF SEQ ID NOS: 178
; SEQ ID NO 113
;   LENGTH: 6106
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-257-166-113

  Query Match             4.4%;  Score 48.4;  DB 17;  Length 6106;
  Best Local Similarity   45.5%;  Pred. No. 1.9;
  Matches  172;  Conservative    0;  Mismatches  206;  Indels    0;  Gaps    0;

Qy        160 TTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATA 219
              | | ||| ||   || |   | | | ||   |  | ||||| |      | |  | | |
Db       4729 TAAAAAAACAAACAAAAACCTCGAAAAATTACATAACTTAAAAATTAAAATATAATTAAA 4670

Qy        220 TACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTT 279
              | |  |   | ||    ||   | | | | | | | |   |    |   | |    |
Db       4669 AAATAACCAAAATTAAAAAATTTAACCAAAATAAACAAAATAAAAATTAAAATCTAAACT 4610

Qy        280 TGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAA 339
              ||||||   | ||| |||    |||   ||  |      ||  |      || |||
Db       4609 AAAAAAAAAAAAAAAAAAAAAAAAAACTTCTTTTAATAAATTACTAAAAAAACCTTCAAA 4550

Qy        340 ATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTC 399
              |||   |||   |    | ||||  ||| | ||  ||  | |      | |
Db       4549 TAAAAAATAAAATAAAAATTAAATTAATTTATTCAAATTTTAATTTCTCTTTATCCACT 4490

Qy        400 AACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTT 459
              |||  |||  ||    |      ||   | |   |   || |||  | ||   ||||| ||
Db       4489 AACTAAAAAATAATTAATAAATTTCAATTTAAAATACCATATAATATTTCAAACAAAATT 4430

Qy        460 TATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTC 519
              |   |  ||  ||  |   |||  |||| | |   |      ||  || | ||| |||||
Db       4429 AAAAATATAACAAAAATTTAAAATAATAAATATAAAAAATATAAAAATAAATTAACTTTA 4370

Qy        520 CAAATATTCTGTCATTCA 537
              |||  ||||  | |||  |
Db       4369 TAAAAATTCATTTATTTA 4352


RESULT 46
US-10-221-714A-151/c
; Sequence 151, Application US/10221714A
; Publication No. US20040048254A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with
;  TITLE OF INVENTION:  tumor suppressor genes and oncogenes
;  FILE REFERENCE: 5013.1005
;  CURRENT APPLICATION NUMBER: US/10/221,714A
;  CURRENT FILING DATE:  2003-01-21
;  PRIOR APPLICATION NUMBER: PCT/EP01/02955
;  PRIOR FILING DATE: 2001-03-15
;  PRIOR APPLICATION NUMBER: DE 10013847.0
;  PRIOR FILING DATE: 2000-03-15
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 540
; SEQ ID NO 151
;   LENGTH: 6106
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-714A-151

  Query Match             4.4%;  Score 48.4;  DB 17;  Length 6106;
  Best Local Similarity   45.5%;  Pred. No. 1.9;
  Matches  172;  Conservative    0;  Mismatches  206;  Indels    0;  Gaps    0;

Qy        160 TTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATA 219
```

```
Db       4729 TAAAAAAACAAACAAAAACCTCGAAAAATTACATAACTTAAAAATTAAAATATAATTAAA 4670

Qy        220 TACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTT 279

Db       4669 AAATAACCAAAATTAAAAAATTTAACCAAAATAAACAAAATAAAAATTAAAATCTAAACT 4610

Qy        280 TGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAA 339

Db       4609 AAAAAAAAAAAAAAAAAAAAAAAAAACTTCTTTTAATAAATTACTAAAAAAACCTTCAAA 4550

Qy        340 ATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTC 399

Db       4549 TAAAAAATAAAATAAAAATTAAATTAATTTATTCAAATTTTTAATTTCTCTTTATCCACT 4490

Qy        400 AACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTT 459

Db       4489 AACTAAAAAATAATTAATAAATTTCAATTTAAAATACCATATAATATTTCAAACAAAATT 4430

Qy        460 TATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTC 519

Db       4429 AAAAATATAACAAAAATTTAAAATAATAAATATAAAAAATATAAAAATAAATTAACTTTA 4370

Qy        520 CAAATATTCTGTCATTCA 537

Db       4369 TAAAAATTCATTTATTTA 4352


RESULT 47
US-10-311-455-175/c
; Sequence 175, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 175
;   LENGTH: 8246
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-175

  Query Match             4.4%;  Score 48.4;  DB 15;  Length 8246;
  Best Local Similarity   47.6%;  Pred. No. 2.2;
  Matches  175;  Conservative    0;  Mismatches  191;  Indels    2;  Gaps    1;

Qy        150 ATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGT 209

Db       7840 ATATTTATCCTAATAACTCAAAAAAAAATATTACATTAAAAATCAAATTTAATAATTTTA 7781

Qy        210 TTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTT 269

Db       7780 AATATTCTTATAATATAAAAACCATACAAAAAAAATCCCTAAATAATATAAACTTTAATT 7721

Qy        270 GTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAG 329

Db       7720 ATTTTTCTTTAATTAAAAATTTAAAAAACCATCTATAATTATAATTCAAACAAATTATAT 7661

Qy        330 CCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAG 389

Db       7660 TAAAAAAAACAATAAAATTAAAA--AATTCCAACAAAATTCTAAAATATTTTATAATAAT 7603

Qy        390 GTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTA 449

Db       7602 TAACACTAATAAACTACATAAATATTATAACCTTTCACAATAAAACTACAATACTAATAA 7543

Qy        450 CCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAAC 509

Db       7542 TATACCTTTTCTAAAAATAAAAATAATCAATAAACACTAACTTTTCCTTAATTATTAAAT 7483

Qy        510 ATTTACTT 517

Db       7482 ATAAACTT 7475


RESULT 48
US-10-278-698-255
; Sequence 255, Application US/10278698
; Publication No. US20050037344A1
; GENERAL INFORMATION:
;  APPLICANT: PathoArray GmbH
;  APPLICANT:  Stuhlmuller, Bruno
;  APPLICANT:  Haupl, Thomas
;  TITLE OF INVENTION: Nucleic Acid Array
;  FILE REFERENCE: O30027US
;  CURRENT APPLICATION NUMBER: US/10/278,698
;  CURRENT FILING DATE:  2002-10-23
;  NUMBER OF SEQ ID NOS: 1050
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 255
;   LENGTH: 115218
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-278-698-255

  Query Match             4.4%;  Score 48.4;  DB 19;  Length 115218;
  Best Local Similarity   45.5%;  Pred. No. 7.9;
  Matches  212;  Conservative    0;  Mismatches  251;  Indels    3;  Gaps    1;

Qy        138 TTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCT 197

Db      17014 TTATATTATTATATATTATATATTATATATTATATTATATTATTATATAATAT---ATAT 17070

Qy        198 TAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATAC 257

Db      17071 TATATATTATATTATATTATTATATAATATTTTATATTATATTATATTATTATATAATAT 17130

Qy        258 AGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAG 317

Db      17131 ATATTATATGATATATTATATTATTATATAATATATTATTATATGATATATTATATTACAT 17190

Qy        318 ATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATC 377

Db      17191 TATATATTATATTATATGATATATTATATGATATATTATATTATATGATATATTATATTA 17250

Qy        378 TTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGC 437

Db      17251 TATATTATATAATATAATATAATATATCATATAATATATCATATAATATAATATATCATA 17310

Qy        438 AAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATG 497

Db      17311 TAATATATTATATAATAAAATATATAATATAATATATAATATAATATATTATTATATATA 17370

Qy        498 TACTATTAAAACATTTACTTTCCAAATATTCTGTCATTCAGGAGTATGGAAGTATCGATG 557

Db      17371 TAATATATTATATAACATATTATATATATAATAATATATAATATTATATAATATATTATA 17430

Qy        558 GCTTCTTTAAAATGAAGCAGGAGGGTCTGGCAGAGAGTATCTATGA 603

Db      17431 TTATATATAATATAATATAATATATTATATTATATTATATATTA 17476
```

```
RESULT 49
US-10-278-698-769
; Sequence 769, Application US/10278698
; Publication No. US20050037344A1
; GENERAL INFORMATION:
;  APPLICANT: PathoArray GmbH
;  APPLICANT:  Stuhlmuller, Bruno
;  APPLICANT:  Haupl, Thomas
;  TITLE OF INVENTION: Nucleic Acid Array
;  FILE REFERENCE: O30027US
;  CURRENT APPLICATION NUMBER: US/10/278,698
;  CURRENT FILING DATE:  2002-10-23
;  NUMBER OF SEQ ID NOS: 1050
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 769
;   LENGTH: 115218
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-278-698-769

  Query Match             4.4%;  Score 48.4;  DB 19;  Length 115218;
  Best Local Similarity   45.5%;  Pred. No. 7.9;
  Matches  212;  Conservative    0;  Mismatches  251;  Indels    3;  Gaps    1;

Qy        138 TTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCT 197
              ||   ||   | |||||    |||| | || |  | ||| | || || || ||   || |
Db      17014 TTATATTATTATATATTATATATTATATTATATTATTATTATATAATAT---ATAT 17070

Qy        198 TAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATAC 257
              || | |   | || | ||| ||||      ||   | || ||  | |   |      |||
Db      17071 TATATATTATATTATATTATTATATAATATTTTATATTATATTATATTATTATATAATAT 17130

Qy        258 AGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAG 317
              |  |  |   | | ||   ||  |  | || |  ||| |    ||  |||      | |
Db      17131 ATATTATATGATATATTATATTATTATATAATATATATTATATGATATATTATATTACAT 17190

Qy        318 ATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATC 377
              ||||  ||      |  | |||| |    | || |    |||| | ||    || ||
Db      17191 TATATATTATATTATATGATATATTATATGATATATTATATTATATGATATATTATATTA 17250

Qy        378 TTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGC 437
              |      | |   || |    |  | |    |||   | |     | | ||    |
Db      17251 TATATTATATAATATAATATAATATATCATATAATATATCATATAATATAATATATCATA 17310

Qy        438 AAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATG 497
               |||| |   ||  | ||| | |||  |  ||   ||   | || | || ||| | |||
Db      17311 TAATATATTATATAATAAAATATATAATATAATATATAATATAATATATTATTATATATA 17370

Qy        498 TACTATTAAAACATTTACTTTCCAAATATTCTGTCATTCAGGAGTATGGAAGTATCGATG 557
              || |||   |      |  ||  | ||||  |   ||  |  | |||  ||      ||
Db      17371 TAATATATTATATAACATATTATATATATAATAATATATAATATTATATAATATATTATA 17430

Qy        558 GCTTCTTTAAAATGAAGCAGGAGGGTCTGGCAGAGAGTATCTATGA 603
                | | ||| || |   |  |      |   | | |||| ||| |
Db      17431 TTATATATAATATAATATAATATATATTATATTATATTATATATTA 17476


RESULT 50
US-10-311-455-988/c
; Sequence 988, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 988
;   LENGTH: 17967
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-988

  Query Match             4.4%;  Score 48.2;  DB 15;  Length 17967;
  Best Local Similarity   46.5%;  Pred. No. 3.5;
  Matches  192;  Conservative    0;  Mismatches  218;  Indels    3;  Gaps    1;

Qy        128 CTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAA 187
              || |||| |   |  || |    ||||  |  || | ||  |  || ||    | || |
Db      10059 CTAATATAACAATAATTTATTTTATTTTATTTTCCAAATAAATATTAAACATATACAT 10000

Qy        188 GATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTT 247
              ||      |  |     |    ||      ||       |  |    | |||||  |  |
Db       9999 TATAATCAATACATTTACTAAACTTAAACTTAATAAATATAAAAAAATATAAAACCCAAT 9940

Qy        248 ACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCAT 307
              || |       |  ||  |  | |||         ||  |||||    |||   |    |
Db       9939 TCATAACCTACGAAATCAAAATTATTTAAAAAAATTTTAAAAATACTTAATATATCTTAA 9880

Qy        308 TCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACA 367
              |   ||  |    | ||||  | |  |||||  ||    ||   | |  || || ||
Db       9879 ATAATAAATAACATAAAATTAAAAAATATAAAATCTAATTTTAATCAAATCAATCTTTCA 9820

Qy        368 TTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTT 427
              || | ||||    |   |||   |||||   | |  |||   |  |  |   |||  ||
Db       9819 ATATTAAATCAAATAAATAAACAAAAAAAAAAAATCAAAAACAAAACCATCTAACTTAATT 9760

Qy        428 ATCGCACT---GCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAAC 484
              ||| | |    | || | ||| ||  ||  |  |||  || |   |||||  || | |||
Db       9759 CTCGTAATATCACTAAAAAAAAAATAATACTACCTTTTTTCTTATAATAATAACAAATAAC 9700

Qy        485 ATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGTCATTCA 537
              || | ||    | |     || | ||  |||| | | || |   | |||||  |
Db       9699 ATTAATTAAACTCTTACTATATACCAAATACTCTACTAACAACTTATCATTTA 9647


RESULT 51
US-10-433-793-44/c
; Sequence 44, Application US/10433793
; Publication No. US20040142334A1
; GENERAL INFORMATION:
;  APPLICANT: Epigenomics AG
;  TITLE OF INVENTION: Diagnose von mit Angiogenese assoziierten Krankheiten
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/10/433,793
;  CURRENT FILING DATE:  2003-06-06
;  NUMBER OF SEQ ID NOS: 212
; SEQ ID NO 44
;   LENGTH: 6676
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-433-793-44

  Query Match             4.4%;  Score 48;  DB 18;  Length 6676;
  Best Local Similarity   46.0%;  Pred. No. 2.4;
  Matches  162;  Conservative    0;  Mismatches  190;  Indels    0;  Gaps    0;

Qy        171 TTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGA 230
              | || | || | ||    || | |  | ||| |||     |    |  ||  |  | | |
Db       1313 TAAAAAATTCTATATTTTATTTTAAATAAAACACAAATAATACAAATCATCTTATCATTA 1254

Qy        231 TTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAA 290
```

```
                | | ||  | | | || |   |     |   ||   | | ||| |  ||  || | || |
Db       1253 TATATTATATAAAACTAAATTATCATTATTCATAATTATTTTTCTTATTCTTTAACAATA 1194

Qy        291 GGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAA 350
              | |||| ||    ||  | | ||| | || | | ||  ||    ||||| | | | ||
Db       1193 TTTTTAAAAAACAAAATAAATATAAAAAACAAAAAATAATCAACCAAAAATAACAAAAAA 1134

Qy        351 AATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGT 410
              ||  |    | | | | |||     |  ||      |     |    | |   ||   |
Db       1133 AAATAAAAAAATTAATCTAACTACCCATACCAACTTTAAACTTACCCACTATAAAATCCT 1074

Qy        411 ATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAA 470
              | || |  | ||| | |||| |     || |   |  || |  | |  | ||     |
Db       1073 AAACAATTTACTTACCTATCTAAACCTTAAAAATCAAAACTATCACCTAAAATAACATAT 1014

Qy        471 TAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAA 522
              ||||  |     | | ||  | |||| ||   || | | | | |   || ||
Db       1013 TAATAACTACCTAAAAAAACTCTTATATAAATTTCACAAAATAAAAATCAAA 962


RESULT 52
US-10-311-455-199
; Sequence 199, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 199
;   LENGTH: 6509
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-199

  Query Match             4.4%;  Score 47.8;  DB 15;  Length 6509;
  Best Local Similarity   51.1%;  Pred. No. 2.7;
  Matches  112;  Conservative    0;  Mismatches  107;  Indels    0;  Gaps    0;

Qy        285 AAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAA 344
              ||||||   || | ||||    || | |||||||| || |   |    | | || | ||
Db       2513 AAAAAAAAAATGAGGAAAGAGTTTAATAGAAAGATTATTTGAAAAAGGGGTAAATAATAT 2572

Qy        345 ACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCT 404
              | |||||   |  |||| | |||  ||| | ||  |   ||   ||||||
Db       2573 GAAGAAAATTTATTAATATAAGATTTATTATTTTTTAAAGAAAGATTAAAAAATATGATA 2632

Qy        405 AAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTT 464
              ||  || || |      | |  ||    |  | ||  || |     |  ||||||||
Db       2633 AATTGTTTAATTTTATTTATAATAAAAAAAGAAATATGAAAGTGTTTTCGGTTTTTATTT 2692

Qy        465 TGGAAATAATCTCAGAAAACATAATTTTTTATGTACTAT 503
                |||       || || ||| ||  | || || | |
Db       2693 ATTTAATTGGTAAGGATAAAATATTTGATAATATATTGT 2731


RESULT 53
US-10-311-455-1916/c
; Sequence 1916, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determi
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1916
;   LENGTH: 12138
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1916

  Query Match             4.4%;  Score 47.8;  DB 15;  Length 12138;
  Best Local Similarity   46.5%;  Pred. No. 3.6;
  Matches  154;  Conservative    0;  Mismatches  177;  Indels    0;  Gaps    0;

Qy        197 TTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATA 256
              | ||| |||||   ||  | ||| ||| | || || || ||| | | | | | | | |
Db       4766 TAAAATACATTACATTTATTATACACTTCATTTATATTATTATTATATATAATATATAAT 4707

Qy        257 CAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAA 316
              |   | | ||    ||      |  || ||||    || |||         ||
Db       4706 AAAACAATTATACAATTCACCATAATTTAAAATCAATAAAAACCCTAAACTTATTTTCCT 4647

Qy        317 GATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAAT 376
              |   |||| | ||  |  |||| |||| |  |||| |    | |||| |||  |  |
Db       4646 AAAACTATACCATCCCATCTAAAAATAAAAAACAAATCATCAAACATTAAATTCTTATAA 4587

Qy        377 CTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTG 436
                ||  |||    ||    |||      |  | ||| |   | |||    ||  |
Db       4586 AAAACATACAACCTAAATCCCTCACGTATACAATTTACAATACGATTTACATTCCTATAA 4527

Qy        437 CAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTAT 496
              ||    | |||||  | ||  | |       |   || |||| | || |   |    | ||
Db       4526 AAATATAATATTACTTCTAACCTAACAAAAAACGAAAACTCAAACAATAATACCGATAAT 4467

Qy        497 GTACTATTAAAACATTTACTTTCCAAATATT 527
                |  || |  | |  |||     |||  ||
Db       4466 AAAAAATAACTATAAATACAAATAAAACTTT 4436


RESULT 54
US-10-240-453-210/c
; Sequence 210, Application US/10240453
; Publication No. US20030148326A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA
;  TITLE OF INVENTION:  Transcription
;  TITLE OF INVENTION:  by Means of Assessing the Methylation Status of Genes Associated
;  TITLE OF INVENTION:  With DNA TRanscription
;  FILE REFERENCE: 5013.1009
;  CURRENT APPLICATION NUMBER: US/10/240,453
;  CURRENT FILING DATE:  2002-10-02
;  PRIOR APPLICATION NUMBER: PCT/EP01/03973
;  PRIOR FILING DATE: 2001-04-06
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
```

```
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 350
; SEQ ID NO 210
;   LENGTH: 12138
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-453-210

  Query Match             4.4%;  Score 47.8;  DB 15;  Length 12138;
  Best Local Similarity   46.5%;  Pred. No. 3.6;
  Matches  154;  Conservative    0;  Mismatches  177;  Indels    0;  Gaps    0;

Qy        197 TTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATA 256
              | ||| ||||||   ||  | ||| ||| | || || || ||| | |  | | | | |
Db       4766 TAAAATACATTACATTTATTATACACTTCATTTATATTATTATTATATATAATATATAAT 4707

Qy        257 CAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAA 316
              |   | | ||    ||       |  || ||||    || |||         ||
Db       4706 AAAACAATTATACAATTCACCATAATTTAAAATCAATAAAAACCCTAAACTTATTTTCCT 4647

Qy        317 GATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAAT 376
              |   ||||  | ||  |  |||| |||| |  ||||  |    | |||| |||  |  |
Db       4646 AAAACTATACCATCCCATCTAAAAATAAAAAACAAATCATCAAACATTAAATTCTTATAA 4587

Qy        377 CTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTG 436
              ||  |||    ||    |||      |   |.||| |   | |||    ||  |
Db       4586 AAAACATACAACCTAAATCCCTCACGTATACAATTTACAATACGATTTACATTCCTATAA 4527

Qy        437 CAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTTAT 496
              ||    | |||||  | ||  | |       |   || |||| | || |  |    | ||
Db       4526 AAATATAATATTACTTCTAACCTAACAAAAAACGAAAACTCAAACAATAATACCGATAAT 4467

Qy        497 GTACTATTAAAACATTTACTTTCCAAATATT 527
                |  || |  | |  |||     |||  ||
Db       4466 AAAAAATAACTATAAATACAAATAAAACTTT 4436


RESULT 55
US-10-433-793-179/c
; Sequence 179, Application US/10433793
; Publication No. US20040142334A1
; GENERAL INFORMATION:
;  APPLICANT: Epigenomics AG
;  TITLE OF INVENTION: Diagnose von mit Angiogenese assoziierten Krankheiten
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/10/433,793
;  CURRENT FILING DATE:  2003-06-06
;  NUMBER OF SEQ ID NOS: 212
; SEQ ID NO 179
;   LENGTH: 40324
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-433-793-179

  Query Match             4.4%;  Score 47.6;  DB 18;  Length 40324;
  Best Local Similarity   47.5%;  Pred. No. 7.2;
  Matches  172;  Conservative    0;  Mismatches  189;  Indels    1;  Gaps    1;

Qy        149 AATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTG 208
              ||  ||  ||   | | || |   || | |  | || || |  | | ||  ||| |
Db      14565 AAATTTACTCTCAAAATATAACAAAAAATTAATAAAAAATATATATTTTAAATTACACTT 14506

Qy        209 TTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATT 268
                |  |||| | |      | | ||     ||| |   | | |      |  ||  | ||
Db      14505 AATATTTAAAAACCAATTATTACTTAAAATAAACAAACTTCCAATTCCAACATAACCTTT 14446

Qy        269 TGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTA 328
              | |||||   ||    ||||    || | |  |||    ||  | ||| |   ||| |||
Db      14445 TATTTTTATATTACCAAAAATTTTTAAACAATAATAATATCTAATAAAAAATTTATCTTA 14386

Qy        329 GCCAGTTGAAAATTAAACACAAAATGAGTGCAT-ATTACATTACTTAATCTTGCAGTCAA 387
                  |   |   ||     |  | | |  || | |  ||  ||  |||    | | ||
Db      14385 TATCTATATTACACAACTTTTACCTAAATTAATAAACATTTTCTTTTTTCTAAAAATAAA 14326

Qy        388 AGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAAT 447
              |  |||| | |||| |       |    |  || ||  | | |  || |  |||| || |
Db      14325 ATATAAATACTCAAACCTCCTCTTCCATTTCTACTACCATTTTTCATTTTAAATCGATAA 14266

Qy        448 TACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAA 507
              | |  ||||| |   |    ||||  |  |||  |||    |  |||  || |   || ||
Db      14265 TCCTTCAAATATCTATCATAAAATTAATTCACTAAAATAATATTTAAATTTTAAATAAAC 14206

Qy        508 AC 509
              ||
Db      14205 AC 14204


RESULT 56
US-10-312-841-2/c
; Sequence 2, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
;  APPLICANT: Epigenomics AG
;  TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
;  FILE REFERENCE: E01/1208/WO
;  CURRENT APPLICATION NUMBER: US/10/312,841
;  CURRENT FILING DATE:  2002-12-30
;  NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 2
;   LENGTH: 3673778
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (379615)
US-10-312-841-2

  Query Match             4.4%;  Score 47.6;  DB 16;  Length 3673778;
  Best Local Similarity   46.0%;  Pred. No. 64;
  Matches  161;  Conservative    0;  Mismatches  189;  Indels    0;  Gaps    0;

Qy        178 TTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTT 237
              || |  | || |  | |     ||| | |    |    | | |  || |    |||| | |
Db    1902334 TTATTAAAAAAAAATAAAACAAAATATAAAATAACATATAAAATCTACTAAAATTTATAT 1902275

Qy        238 AAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAA 297
              |||| |    | | | ||| |  ||  ||||| | |||  ||    |  | | |      |
Db    1902274 AAAAAAAAACAAAAAAATATATATACACATTTATATTTACTTACATATACATAAAATATA 1902215

Qy        298 AGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGT 357
              ||   | | |   ||| |  ||   | |      |   |||   || |||||| |
Db    1902214 TTTAAAAAAATAAACAAAAAAATATCCCTAATTACCTATTAAAAATAAAACAAAATAAAA 1902155

Qy        358 GCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTAC 417
              | |  |  |||| |  |         | |||  | |||  | || ||| |   | |
Db    1902154 AAAAAACATTTTACCTTTTAAACATTTTAAATTTTAAATTTTAAACTATATCAAAAAATA 1902095

Qy        418 CTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTC 477
                | ||      |      | | |||   | | ||| || |   |  |||   | |
Db    1902094 AAAAATAATTCCTAAAACAACCAAATAAACCCCAAAAAAATTCAAAATAAAAAACCTTTT 1902035

Qy        478 AGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATT 527
              | |||  |  |   ||| | |||  |||  | | | | |||  |||  ||
Db    1902034 AAAAACTAATAAAATTTTTTTACCTTTATTAAAATAAATTTTAAAAATTT 1901985
```

```
RESULT 57
US-09-991-936-1307/c
; Sequence 1307, Application US/09991936
; Publication No. US20030073827A1
; GENERAL INFORMATION:
;  APPLICANT: Brandt, Kevin S.
;  APPLICANT:  Gaines, Patrick J.
;  APPLICANT:  Stinchcomb, Dan T.
;  APPLICANT:  Wisnewski, Nancy
;  TITLE OF INVENTION: FLEA HEAD, NERVE CORD, HINDGUT AND MALPIGHIAN TUBULE
;  TITLE OF INVENTION:  NUCLEIC ACID MOLECULES, PROTEINS AND USES THEREOF
;  FILE REFERENCE: FC-6-C1
;  CURRENT APPLICATION NUMBER: US/09/991,936
;  CURRENT FILING DATE:  2001-11-21
;  PRIOR APPLICATION NUMBER: US/09/543,668
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: 60/128,704
;  PRIOR FILING DATE: 1999-04-09
;  NUMBER OF SEQ ID NOS: 1959
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1307
;   LENGTH: 549
;   TYPE: DNA
;   ORGANISM: Ctenocephalides felis
US-09-991-936-1307

  Query Match             4.3%;  Score 47.4;  DB 10;  Length 549;
  Best Local Similarity   47.3%;  Pred. No. 0.99;
  Matches  175;  Conservative    0;  Mismatches  192;  Indels    3;  Gaps    1;

Qy        136 AGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCAT 195
              | ||| ||||| |   |   || ||    ||| | ||||||   ||    || | |
Db        451 ATTTTAGTTGAAATCAAAGTATTATACATAATTAATTAATAACATTTACTAATAAATGTA 392

Qy        196 CTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACAT 255
                ||     || | |   ||  ||      |||  | |||   | |     |||  ||
Db        391 GCAAAGTGTNGTGCTGTTAGAACCTAAAAATGTGAAATATTTTTTAAATAAACCAGTGAT 332

Qy        256 ACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAA 315
               ||   |||  | |   |  | |   | ||||| | ||| | |||  ||   || | |
Db        331 TCAAAAATTAGTATTAATCAGATACCTGCAAAATATGTAAATAGATATTTATATACACCT 272

Qy        316 AGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAA 375
              | | ||    |    ||| ||   |   |  | | | |  |  | |  ||  |||
Db        271 AAAGCAATAGTATTTCAGATGGCGACAGATATAACAGTAAACG---AAAAAGTTTTTTAT 215

Qy        376 TCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACT 435
              || | || | |     ||||| | ||  ||     |||      |  ||  |||    |
Db        214 TCATCCAATTATTTAAAAAAAATGAAATTATTTTTATTAAGTTTAATTGTTTAAAATAAA 155

Qy        436 GCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTA 495
               ||  |   | |  ||||||  ||  | |||    |  |    | || | | ||||||
Db        154 TAAATAATTCAATTTCACAAAAATTTGTCTGGTTTAATGCCGGCATAAAAAACTTTTTTG 95

Qy        496 TGTACTATTA 505
              | ||||||||
Db         94 TTTACTATTA 85


RESULT 58
US-10-221-714A-236
; Sequence 236, Application US/10221714A
; Publication No. US20040048254A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with
;  TITLE OF INVENTION:  tumor suppressor genes and oncogenes
;  FILE REFERENCE: 5013.1005
;  CURRENT APPLICATION NUMBER: US/10/221,714A
;  CURRENT FILING DATE:  2003-01-21
;  PRIOR APPLICATION NUMBER: PCT/EP01/02955
;  PRIOR FILING DATE: 2001-03-15
;  PRIOR APPLICATION NUMBER: DE 10013847.0
;  PRIOR FILING DATE: 2000-03-15
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 540
; SEQ ID NO 236
;   LENGTH: 5297
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-714A-236

  Query Match             4.3%;  Score 47.4;  DB 17;  Length 5297;
  Best Local Similarity   44.7%;  Pred. No. 3;
  Matches  183;  Conservative    0;  Mismatches  226;  Indels    0;  Gaps    0;

Qy         87 TCTTCCTTATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGA 146
              | ||  ||   | | ||    |  | ||||| || |    | |  | ||  |||  |
Db       1136 TGTTTTTTTATTTTTGATTTAGTTTTAGTTATTTAGCGATATTTTGTGTTTTTTTTATT 1195

Qy        147 GAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACAT 206
               |||  ||  |  | | | || | || ||| | | | | || |   | | |||    | |
Db       1196 AAAAATTTTGTTTTGAGAGATTATTTTATATTATGGAATAAAAATTGGTATTAGGTATTT 1255

Qy        207 TGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCA 266
              | ||||  |    || |    |   | |   |  ||  ||| | | ||  |  |||  |
Db       1256 TTTTTTTATTGGTTATTTAAATTGATGATAAGACGGATGTTATATATATGTATTTATATA 1315

Qy        267 TTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATAT 326
              ||     |||   |    |  |||| | ||| || |||   | | | || |    |||||
Db       1316 TTGTAGATTGTAGTAGGGATTAAAGTTTTAAGGAGATGTGGTAAAATAATGTAATTATAT 1375

Qy        327 TAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCA 386
              |  | | || |      || | ||          | ||| |||    | |       |
Db       1376 TTTCGAATTAATTGGATAAAAAAAGGATTTGTATTTTTAAATTCGAAAGTAGGTATTTTT 1435

Qy        387 AAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAA 446
                  |||||| |       | | | | |   | ||| |||       |   | | | ||
Db       1436 TGTTTAAAAAATATTTTAGATGAAAGAGTTATTTTTTTATTAGTTTTTTTTGATTTGGAA 1495

Qy        447 TTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTA 495
              |       ||||  | |  |   |||   | |  |  |||   | ||
Db       1496 TCGTTTTGGTTTTTATGTTTATTGAATTAGATATTATTTAAGAATATA 1544


RESULT 59
US-10-311-455-993/c
; Sequence 993, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
```

```
; SEQ ID NO 993
;   LENGTH: 6092
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-993

  Query Match             4.3%;  Score 47.4;  DB 15;  Length 6092;
  Best Local Similarity   49.2%;  Pred. No. 3.2;
  Matches  153;  Conservative    0;  Mismatches  156;  Indels    2;  Gaps    1;

Qy        215 TAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTT 274
              |||||  || |||  |||    |||   ||  || |   |||     || |||| |  |
Db       2524 TAATAATCTTCCTAAATTCAAATAATTAACAATAAAATTATAATTTAATAAATTTTTAAT 2465

Qy        275 TGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGT 334
                ||| ||||||||   ||| ||  ||   ||       || || |  || ||   | |
Db       2464 ATCTTTTTTCAAAAACAATATCAATTAAACAATCATCTTTAAAATAAATTACTAATAATT 2405

Qy        335 TGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAA 394
              |     |||    | ||      | |  | |||||  | | || ||   ||   | ||
Db       2404 TATTTTCTAACTTTTATAATTACATAAAAAAAATTACAAATTATTTCATCTAA--TTTAA 2347

Qy        395 AAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACA 454
              | ||||  |  |    ||  ||||  ||| | ||  ||| || | |  ||  | |||
Db       2346 TATTCAAAATCCATAACTATAACCTTATTTATAATTTTACTAAAATTTAAACCTCCCATC 2287

Qy        455 AATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTA 514
              ||    |  ||          |||   ||| |  ||    ||  ||    |    ||||||
Db       2286 AAAAACAAATTAATCCCCTACTCTTTAAAAAACATAACCTACCTAAATTCTCTACATTTC 2227

Qy        515 CTTTCCAAATA 525
              | |||.|||||
Db       2226 CATTCGAAATA 2216


RESULT 60
US-09-814-353-18587/c
; Sequence 18587, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
;  APPLICANT: Lee, John
;  APPLICANT:  Thompson, Pamela
;  APPLICANT:  Lillie, James
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
;  TITLE OF INVENTION:  IDENTIFICATION, ASSESSMENT, PREVENTION, AND
;  TITLE OF INVENTION:  THERAPY OF OVARIAN CANCER
;  FILE REFERENCE: MRI-006B
;  CURRENT APPLICATION NUMBER: US/09/814,353
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: US 60/191,031
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: US 60/207,124
;  PRIOR FILING DATE: 2000-05-25
;  PRIOR APPLICATION NUMBER: US 60/211,940
;  PRIOR FILING DATE: 2000-06-15
;  PRIOR APPLICATION NUMBER: US 60/216,820
;  PRIOR FILING DATE: 2000-07-07
;  PRIOR APPLICATION NUMBER: US 60/220,661
;  PRIOR FILING DATE: 2000-07-25
;  PRIOR APPLICATION NUMBER: US 60/257,672
;  PRIOR FILING DATE: 2000-12-21
;  NUMBER OF SEQ ID NOS: 22037
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 18587
;   LENGTH: 376
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 209
;   OTHER INFORMATION: n = A,T,C or G
US-09-814-353-18587

  Query Match             4.3%;  Score 47.2;  DB 10;  Length 376;
  Best Local Similarity   50.9%;  Pred. No. 0.91;
  Matches  137;  Conservative    0;  Mismatches  129;  Indels    3;  Gaps    1;

Qy         85 GATCTTCCTTATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTT 144
              | | ||  |||    | |      || | |  |  ||   | | |  | |||   |  ||
Db        328 GGTTTTTTTTAAGGTTTTTAAAAATTTTTTATCTTTTTTTTTTTTTTTTTTGAAAATTTTT 269

Qy        145 GAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGAC 204
              | |||||||  |  ||     ||  ||  |  |||  || || || |  | ||||
Db        268 AAAAAATATTTTTTTTTTTTTTCTTTTTTTTTTTTAATAAAATATTTTTTTTTAATTTN 209

Qy        205 ATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATT 264
              || | || || | |||   | || | |  ||   || |  ||    | ||  |  ||||
Db        208 TTTTTCTTTTTTAAATAAATCTTTAAATAATTATTAAAAGTTTTTGTATATTAAAGTATT 149

Qy        265 CATTTGTTTT---TGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCA 321
              |  | ||||    |  ||  || | || || |||| | ||||  | | | | ||| |  |
Db        148 TAAATATTTTATATATTTAATTAAGAATAGATATATATAAATAAAATAAAAAAAAAAAAA 89

Qy        322 TATATTAGCCAGTTGAAAATTAAACACAA 350
              | ||    ||   ||||  ||| | ||
Db         88 AAAATATATCAAAAAAAAAAAAAAAAAA 60


RESULT 61
US-09-814-353-18778/c
; Sequence 18778, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
;  APPLICANT: Lee, John
;  APPLICANT:  Thompson, Pamela
;  APPLICANT:  Lillie, James
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
;  TITLE OF INVENTION:  IDENTIFICATION, ASSESSMENT, PREVENTION, AND
;  TITLE OF INVENTION:  THERAPY OF OVARIAN CANCER
;  FILE REFERENCE: MRI-006B
;  CURRENT APPLICATION NUMBER: US/09/814,353
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: US 60/191,031
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: US 60/207,124
;  PRIOR FILING DATE: 2000-05-25
;  PRIOR APPLICATION NUMBER: US 60/211,940
;  PRIOR FILING DATE: 2000-06-15
;  PRIOR APPLICATION NUMBER: US 60/216,820
;  PRIOR FILING DATE: 2000-07-07
;  PRIOR APPLICATION NUMBER: US 60/220,661
;  PRIOR FILING DATE: 2000-07-25
;  PRIOR APPLICATION NUMBER: US 60/257,672
;  PRIOR FILING DATE: 2000-12-21
;  NUMBER OF SEQ ID NOS: 22037
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 18778
;   LENGTH: 457
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-814-353-18778

  Query Match             4.3%;  Score 47.2;  DB 10;  Length 457;
  Best Local Similarity   49.2%;  Pred. No. 1;
  Matches  124;  Conservative    0;  Mismatches  128;  Indels    0;  Gaps    0;

Qy        170 GTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTG 229
              |||  |  |||| | ||| |     | ||     |  || ||||  ||  |    |||
Db        328 GTTTTTTTTTTTTTTGAAAAAAAATTTTTTTTCAATTTTTTTTTATATTTATATCCAAAA 269

Qy        230 ATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAA 289
              ||||||||        ||| | |   | | |  ||  ||| |||   |||| |||  |||
Db        268 ATTTTTTTTTCCTCTTTTATATAATAAAACCCTTTTTTTTTTTTGGTGTTTTTTCCCAAA 209
```

```
Qy        290 AGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACA 349
              |   | |||  ||   |   | | ||| |  | | |      |    ||||  ||| | |
Db        208 AAAAAAAAATTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAAAAAAAAAAAAA 149

Qy        350 AAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGG 409
              |||  |    | | || |  |   ||       |   |||    ||||    |   |||
Db        148 AAAAAAAAAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAATCCGAAAAAAATAAAAAAAAA 89

Qy        410 TATACTACCTGC 421
              || | |||||||
Db         88 TAAAGTACCTGC 77


RESULT 62
US-10-311-455-666/c
; Sequence 666, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 666
;   LENGTH: 6222
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-666

  Query Match             4.3%;  Score 47.2;  DB 15;  Length 6222;
  Best Local Similarity   48.2%;  Pred. No. 3.6;
  Matches  133;  Conservative    0;  Mismatches  143;  Indels    0;  Gaps    0;

Qy        252 ACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAG 311
              | | |||  ||  |      ||   | |  | ||||||   | |||  |   |
Db       3469 ATACACATATACCCTAAAACTTAAAATATAATAAAAAAAAAAAAAAAATATATTAAATTA 3410

Qy        312 AGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTAC 371
              | ||    || ||| ||     || |   || || | ||||| |    ||   |  |||
Db       3409 ATAACTCCCAAATAATACAATTTTAATTCTTTAAAAAAAAATTAAAAAATCACAATATAC 3350

Qy        372 TTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCG 431
              |||| ||   |   |||  ||||   | ||  |||    | |     |     |||
Db       3349 CTAATTTTAAAAATAAAACTAAATTATAAATATAAAAAACTTCCCATTATAACCTTTATC 3290

Qy        432 CACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTT 491
              | |  |  ||| | ||    | | |||||  |||      | | |    |||| || ||||
Db       3289 CCCAACTCATAAATATATATATATATTTTCTTTCCTTTTTTAACATTTAAAAAATTATTT 3230

Qy        492 TTTATGTACTATTAAAACATTTACTTTCCAAATATT 527
              || | ||  | |||||  | |   || |  | | |
Db       3229 ATTTTTTAAAAATAAAATCTCTCTATTACCCAAACT 3194


RESULT 63
US-10-240-589C-21/c
; Sequence 21, Application US/10240589C
; Publication No. US20040076956A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with
;  TITLE OF INVENTION:  DNA repair
;  FILE REFERENCE: 5013.1008
;  CURRENT APPLICATION NUMBER: US/10/240,589C
;  CURRENT FILING DATE:  2003-09-02
;  PRIOR APPLICATION NUMBER: PCT/EP01/03972
;  PRIOR FILING DATE: 2001-04-06
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 148
; SEQ ID NO 21
;   LENGTH: 18060
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-589C-21

  Query Match             4.3%;  Score 47.2;  DB 17;  Length 18060;
  Best Local Similarity   50.4%;  Pred. No. 6.1;
  Matches  115;  Conservative    0;  Mismatches  113;  Indels    0;  Gaps    0;

Qy        281 GTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAA 340
              ||   |||||   | ||| |||   | | |   ||| |     | |      |    | |
Db       2503 GTCTCAAAAACAAACAAAAAAAACAAATAATTTAAATACATACTTTCCCAAAAAAAATAT 2444

Qy        341 TTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCA 400
              ||| | |||| |   |||  | | |||| ||  |  |     |    |||||  ||
Db       2443 ACAAATAAAAAACAAACATATAAAAAAATACTCAACATCACTAATCATCAAAAAAATACA 2384

Qy        401 ACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTT 460
              |   |||   | | ||     |  || |   | ||  |||||   | || || | | |
Db       2383 AATCAAAACCACAATAAAATATCCCTAAAACCCCTTAAAATAACCACTATCAAATAATAA 2324

Qy        461 ATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAA 508
              || ||   || ||| | | |||||  ||    ||||  || |||||||
Db       2323 ATATTAATAAAAATATAAAAAAACTAAAACCCTTATACACGATTAAAA 2276


RESULT 64
US-10-357-930-8821/c
; Sequence 8821, Application US/10357930
; Publication No. US20040259086A1
; GENERAL INFORMATION:
;  APPLICANT: Schlegel. Robert
;  APPLICANT:  Endege, Wilson
;  APPLICANT:  Monahan, John
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
;  TITLE OF INVENTION:  IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF
;  TITLE OF INVENTION:  HUMAN PROSTATE CANCER
;  FILE REFERENCE: MRI-007BCN
;  CURRENT APPLICATION NUMBER: US/10/357,930
;  CURRENT FILING DATE:  2003-02-04
;  PRIOR APPLICATION NUMBER: 09/785,276
;  PRIOR FILING DATE: 2003-02-16
;  PRIOR APPLICATION NUMBER: 60/183,319
;  PRIOR FILING DATE: 2000-02-17
;  PRIOR APPLICATION NUMBER: 60/189,862
;  PRIOR FILING DATE: 2000-03-16
;  PRIOR APPLICATION NUMBER: 60/207,454
;  PRIOR FILING DATE: 2000-05-25
;  PRIOR APPLICATION NUMBER: 60/211,314
;  PRIOR FILING DATE: 2000-06-09
;  PRIOR APPLICATION NUMBER: 60/219,007
;  PRIOR FILING DATE: 2000-07-18
;  PRIOR APPLICATION NUMBER: 60/255,281
```

```
;  PRIOR FILING DATE: 2000-12-13
;  NUMBER OF SEQ ID NOS: 62232
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8821
;   LENGTH: 420
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 31, 102, 161, 167, 172, 185, 186, 187, 188, 189, 190, 196,
;   LOCATION: 199, 204, 207, 212, 217, 218, 220, 227, 229, 235, 239, 240,
;   LOCATION: 241, 243, 265, 270, 275, 276, 278, 280, 282, 283, 285, 286,
;   LOCATION: 294, 296, 297, 301, 308, 317, 326, 327, 332, 335, 339
;   OTHER INFORMATION: n = A,T,C or G
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 342, 350, 355, 357, 362, 364, 365, 369, 380, 390, 398, 400,
;   LOCATION: 402, 409
;   OTHER INFORMATION: n = A,T,C or G
US-10-357-930-8821

  Query Match             4.3%;  Score 47;  DB 18;  Length 420;
  Best Local Similarity   39.3%;  Pred. No. 1.1;
  Matches  149;  Conservative    0;  Mismatches  230;  Indels    0;  Gaps    0;

Qy         35 TCTATGTTCTTATAAGCAAATGTTACAATGAGATATGAGTTTTAAAGCCAGATCTTCCTT 94
              | | | | | | |   |  | | | || | ||      |||  |   | |  ||  |
Db        408 TATCTTNTNTNTTTTTCTNTTATAAAAANAAAATCATCTNTTTNNANGAAANTNATCANT 349

Qy         95 ATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATATT 154
              | | | |     |   |   ||   | | |  |||     | | |  | |||||
Db        348 TTTTTTNTTNTTTNTGNAGGGNNTTTTTTTTNTTTTTTATNGAAAGGNTGGNNANATATT 289

Qy        155 GATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGT 214
                | |   |   |||  |   ||| |  || |     |    |          ||||
Db        288 TTNNANNANTNANNGGTTNTTTTNTTTTTTTAAAAAAAAAATAAANTNNNGAGNTTTTTN 229

Qy        215 TAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTT 274
               | | |    ||   |||| |  || |  | |           ||     |       |
Db        228 ANAAAAAANGNNTTTTNTTTTNTTNAATANATNATTTTNNNNNNGCGTAAAAAAAANAAT 169

Qy        275 TGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGT 334
                  | ||||||   | ||| |||   |   | | ||| |  | | |  |   |
Db        168 ANAAAAANTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 109

Qy        335 TGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAA 394
               ||||  ||| | ||||  |    | |  | |  |   ||       |   ||| | |||
Db        108 AAAAAANAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAAAA 49

Qy        395 AAGTCAACCTAAAGGTATA 413
              ||   ||   |||   | |
Db         48 AAAAAAAAAAAAAAAAANA 30


RESULT 65
US-10-198-846-735
; Sequence 735, Application US/10198846
; Publication No. US20030099974A1
; GENERAL INFORMATION:
;  APPLICANT: Lillie, James
;  APPLICANT:  Xu, Yongyao
;  APPLICANT:  Wang, Youzhen
;  APPLICANT:  Steinmann, Kathleen
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
;  TITLE OF INVENTION:  FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
;  TITLE OF INVENTION:  THERAPY OF BREAST CANCER
;  FILE REFERENCE: MRI-049
;  CURRENT APPLICATION NUMBER: US/10/198,846
;  CURRENT FILING DATE:  2002-07-18
;  PRIOR APPLICATION NUMBER: 60/306,220
;  PRIOR FILING DATE: 2001-07-18
;  NUMBER OF SEQ ID NOS: 14084
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 735
;   LENGTH: 804
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 4, 5, 6, 10, 15, 680, 745, 756, 769, 773, 801
;   OTHER INFORMATION: n = A,T,C or G
US-10-198-846-735

  Query Match             4.3%;  Score 47;  DB 14;  Length 804;
  Best Local Similarity   49.8%;  Pred. No. 1.5;
  Matches  119;  Conservative    0;  Mismatches  120;  Indels    0;  Gaps    0;

Qy        250 AGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTC 309
              | |  ||   | |  | | || || |   | ||   | |   | |||   |||  ||
Db        424 ACAACTATCTCAAGCCCTGTGATTATAAATGGTGTTATAGGCTTGTAATCCTATGGGTTT 483

Qy        310 AGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATT 369
              |||  ||  |  | | ||    | ||||||  |||||||  | ||| | ||| |  ||
Db        484 AGACTAAAGTATTTTCTTCATAAATTGAAAGAGAAACACACTAGGAATTCAGAAGATTTT 543

Qy        370 ACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTAT 429
              | ||||  |||     |||  | ||  |  |   ||||||    ||  ||   |  || |
Db        544 AATTAAAGTTGTTCAAAAAATTTAATGGGAATTTTAAAGGCTGGCTCACTAAGTAGTTTT 603

Qy        430 CGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAA 488
                | || |  ||     ||  |  ||  || ||| || || | |   ||  |  ||
Db        604 GATAATGAATTTAATTTATAGAAAGAAAGTTCTTTGGGCAAAATTTTGGTAATGCCAAA 662


RESULT 66
US-10-311-455-106/c
; Sequence 106, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determi
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 106
;   LENGTH: 5421
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: 3355, 3371, 3386, 3455, 3591, 3612, 3620, 3644, 3752, 3813
;   OTHER INFORMATION: n is a or g or c or t
US-10-311-455-106

  Query Match             4.3%;  Score 47;  DB 15;  Length 5421;
  Best Local Similarity   45.9%;  Pred. No. 3.7;
  Matches  158;  Conservative    0;  Mismatches  186;  Indels    0;  Gaps    0;

Qy        209 TTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATT 268
              ||   |||| |   |  ||     ||||| |   |      |  | | | |  ||| | ||
Db       4113 TTCCTTTAAAACTTTTACTACTACTTTTTCATTAAAAAATCCAATAAAAACTTATCCTTT 4054

Qy        269 TGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTA 328
                |      ||  ||||  ||| |||||   ||    |   | | |||  |   |  |||
```

```
Db       4053 CATACGACCTTCTTTCATCAAACGTATATCAAACCTAAAAAAAAAAAAACAATTTTAATTC 3994

Qy        329 GCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAA 388

Db       3993 AAAAAATATCAATTTCAATATAAATAAAAATATATTAAATAATCTAACATTTTACTATTT 3934

Qy        389 GGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATT 448

Db       3933 CAAATAAAAATAACTACACTATTTAATAAACACGTCAATTATTCAATTTTCTAAAAATAT 3874

Qy        449 ACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAA 508

Db       3873 ATTTCATATTCTACACTAAAATTAATTCTACAAATAAATAATATTTATTTAAAATATACT 3814

Qy        509 CATTTACTTTCCAAATATTCTGTCATTCAGGAGTATGGAAGTAT 552

Db       3813 NTTAACTATTCTCACTATCATAAAAATATATAATTTAAAAATAT 3770


RESULT 67
US-10-311-455-460
; Sequence 460, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 460
;   LENGTH: 17183
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-460

  Query Match             4.3%;  Score 47;  DB 15;  Length 17183;
  Best Local Similarity   45.2%;  Pred. No. 6.6;
  Matches  213;  Conservative    0;  Mismatches  255;  Indels    3;  Gaps    1;

Qy        123 AGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTG 182

Db       5486 AGTATAAGAAATTAAAGTATATAAGAGATTTTGTAAAAGTTTAGTATATATATGATTTTG 5545

Qy        183 TAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAG 242

Db       5546 AAAATTATTTTAATGAATTTATATAATTATTTTAGTTATTAAATCGTTAATGTTTAGAGA 5605

Qy        243 ACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAA 302

Db       5606 GGAGAAAATATGTTGATTATTTTTTATTATTTTATTTTGTTAATTTAGTAGAGTAATATG 5665

Qy        303 TGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATA 362

Db       5666 TTATTAAAGAGAATTTAGTAAATATTTTTTGATGGGAATTAGA---TAAAGGGATAGTTA 5722

Qy        363 TTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCT 422

Db       5723 TTTTAATATAGATTTTTGTGATTTAAAGTAATAATATTATTATTAATAATGGAATAGAAG 5782

Qy        423 TTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAA 482

Db       5783 TTATTTTAGGAGATAGGTAAGAAATTATAGTAATGATATATTTAGGGATATAGTTGGAAA 5842

Qy        483 ACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGTCATTCAGGAGT 542

Db       5843 AGTTATTTTTTTATGTATGAGAACGAGTTTAAGTTTTAACGTTTTAAATTGATAAGTAGA 5902

Qy        543 ATGGAAGTATCGATGGCTTCTTTAAAATGAAGCAGGAGGGTCTGGCAGAGA 593

Db       5903 AAATAAGAAGTTAATTTTTTTTTTAGATTAGATGGTAGTTTAGGTAAAAAA 5953


RESULT 68
US-10-221-714A-514/c
; Sequence 514, Application US/10221714A
; Publication No. US20040048254A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with
;  TITLE OF INVENTION:  tumor suppressor genes and oncogenes
;  FILE REFERENCE: 5013.1005
;  CURRENT APPLICATION NUMBER: US/10/221,714A
;  CURRENT FILING DATE:  2003-01-21
;  PRIOR APPLICATION NUMBER: PCT/EP01/02955
;  PRIOR FILING DATE: 2001-03-15
;  PRIOR APPLICATION NUMBER: DE 10013847.0
;  PRIOR FILING DATE: 2000-03-15
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 540
; SEQ ID NO 514
;   LENGTH: 61020
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-714A-514

  Query Match             4.3%;  Score 47;  DB 17;  Length 61020;
  Best Local Similarity   43.2%;  Pred. No. 12;
  Matches  271;  Conservative    0;  Mismatches  355;  Indels    1;  Gaps    1;

Qy        136 AGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCAT 195

Db      34581 ACTTTAAATAATAATAATACATCGATATAAATTCATAAATTATAATAAATATACATTTCT 34522

Qy        196 CTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACAT 255

Db      34521 AATAAA-ACATATTAATAATAAAAAATACTATACATATAACTAAATAAATATATATAAAA 34463

Qy        256 ACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAA 315

Db      34462 TAACTCTATATTTTATTTTACTATAAACCAAAAAATATCCTTAAAAATATCTTAAAAAAA 34403

Qy        316 AGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAA 375

Db      34402 ATAAAAAATTTATTAAATTTAAAAATAATAATAAAATCGAAAAAACATTATAAACAAAAA 34343

Qy        376 TCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACT 435

Db      34342 AAACAACTTAAAAAAAAATAACCATAAAATAAAACAAATTAAAACTTACAAAAAATATTT 34283

Qy        436 GCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTA 495

Db      34282 ACAAAAAAACTAAAAAACAAAAACAAAAATATTAAAAAACGAATAAACAATAAAATATAA 34223
```

```
Qy        496 TGTACTATTAAAACATTTACTTTCCAAATATTCTGTCATTCAGGAGTATGGAAGTATCGA 555
              |  |  |||  |  ||    | ||  ||| | |   |      |       ||  | | |
Db      34222 AAAATAACAAAATAAAATAACAACAAATAATTATTTACATTTAAAAATAAAAAAAAACCA 34163

Qy        556 TGGCTTCTTTAAAATGAAGCAGGAGGGTCTGGCAGAGAGTATCTATGAAATAAGTTCCTC 615
              |   |      |||   |  |      |       | | ||   || | ||    |  |
Db      34162 TTAATCAAAATAAAATCAAAATTTAATTAATTTCAATAATAATAATAATATCCAATAATA 34103

Qy        616 TGACCTTCACGCTTAATTTTCTGAATGGAGTGGAGCAAATTACTTCAAGCTTCACTTAAC 675
              |   |   |||    ||  |    ||||  ||   |  | | | |||
Db      34102 ATAATAATAACTACCTTTTAACCAAACAATATATACAAAACACGATATTCATAATTTATA 34043

Qy        676 TTGCATATGAAATGAACCGTACAAAAATACAAGAGTGTCAGGAGAAAGTTATGCTCTGGT 735
              ||    ||   || ||    ||||||| | | |   |  | |||  |||  | |
Db      34042 TTCATCATTTCATTAAAAACACAAAAAAAATAAAATCCTATAAAAAAAATATAATTTAAC 33983

Qy        736 AAATATTTTGCAAAACAGATAAAAGAT 762
              |||||       || |   |||| ||
Db      33982 TAATATAAAAACTAAAATTAAAAAAT 33956


RESULT 69
US-09-790-988-1/c
; Sequence 1, Application US/09790988
; Patent No. US20020127687A1
; GENERAL INFORMATION:
;  APPLICANT: SHIGENOBU, SHUJI
;  APPLICANT:  WATANABE, HIDEMI
;  APPLICANT:  HATTORI, MASAHIRA
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
;  FILE REFERENCE: 081356/0159
;  CURRENT APPLICATION NUMBER: US/09/790,988
;  CURRENT FILING DATE:  2001-02-23
;  PRIOR APPLICATION NUMBER: JP2000-107160
;  PRIOR FILING DATE: 2000-04-07
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 640681
;   TYPE: DNA
;   ORGANISM: Buchnera sp.
US-09-790-988-1

  Query Match             4.3%;  Score 47;  DB 9;  Length 640681;
  Best Local Similarity   46.7%;  Pred. No. 39;
  Matches  149;  Conservative    0;  Mismatches  170;  Indels    0;  Gaps    0;

Qy        190 TCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTAC 249
              |||  | |||      |   | |  |  ||| |      | ||||||||||||     ||
Db     581884 TCTTTTTTTATTATTAGACTATAATGTATACAAATAATAGATTTTTTTAAATATTATTTT 581825

Qy        250 AGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTC 309
              |  || | |     |  ||| | | |  | || | |       | | ||    ||
Db     581824 AATCAGAAATAGCTATTTTTTTATATAATATTTTAATATTCATCTTTTAAAATATTATAT 581765

Qy        310 AGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATT 369
              | |    | || ||||    |   ||| || |     |||| | |  |  |  |  ||
Db     581764 TTATATTTTTGATTTATTTTAGAAAGGAATATAATTTTAAAAATTATTAAACTTATAATA 581705

Qy        370 ACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTAT 429
              |  | ||  |  |   ||   ||||    |  | | |||  ||||      |  |||     |
Db     581704 ATGTTATTATTTAAAAAATTATAAATGTTTTAACAAAATTTATAGGTAATTTTTTACATT 581645

Qy        430 CGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAAT 489
              ||  | |||      | |  || ||| |||||||  ||||    |   | ||| || ||
Db     581644 GAAACATCGAATTATTTTCAATACGAATATTATTTTTCAAATGTACGATGTAAAAATTAT 581585

Qy        490 TTTTTATGTACTATTAAAA 508
              | | |    || |||
Db     581584 AAAATTTTTTTATTTTAAA 581566


RESULT 70
US-10-311-455-1734/c
; Sequence 1734, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1734
;   LENGTH: 7851
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1734

  Query Match             4.3%;  Score 46.8;  DB 15;  Length 7851;
  Best Local Similarity   44.8%;  Pred. No. 5;
  Matches  180;  Conservative    0;  Mismatches  222;  Indels    0;  Gaps    0;

Qy        131 ATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGAT 190
              ||||  | |  || | | ||||  ||   |  |  | || |  | |     |
Db       7161 ATATATATATTTTTTAAATATATAAATACATCATATATATTTTAATATATAAATATATTA 7102

Qy        191 CTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACA 250
              || ||| | |   ||  | | | ||||||  |    |  | | | ||  || |  | |
Db       7101 TATATATTATATATTTTAATATATAAATATATTATATATAATATATTTTAATATATAAAT 7042

Qy        251 GACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCA 310
              |||  |  |  |  |||  | |      | ||  | | |   || |   | |  | |
Db       7041 ATTATATATATTATATTTTAATATATAAATATTATATATATTAATTATATATTTTAATAT 6982

Qy        311 GAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTA 370
              | | ||  |||||||   | |  ||||||  | |  |  || |    |||  || ||||
Db       6981 ATAATATATATTATATTATATAATAATAAATTATATATTATATTATATAATAATAAATTA 6922

Qy        371 CTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATC 430
              || | |   |   ||   || ||  | |      ||  |||| ||    | | ||||
Db       6921 TATATTATACTATATAATAATATAATATAATATATAATATATATTATAATATATATTATA 6862

Qy        431 GCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATT 490
              |    | ||   | |    | |   |      |   |   || | | | |      ||
Db       6861 TTATATTATATTATATTATATATTATAATATAAATTATATATATTTTATATATATATTTT 6802

Qy        491 TTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGTC 532
              | |||    ||||  || || ||      ||| | |   ||
Db       6801 ATATATTATATATTTTAATATATAAAACAAAAAAACTTCTTC 6760


RESULT 71
US-09-960-352-11234
; Sequence 11234, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
;  APPLICANT: Warren, Wesley C.
;  APPLICANT:  Tao, Nengbing
;  APPLICANT:  Byatt, John C.
;  APPLICANT:  Mathialagan, Nagappan
;  TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
;  TITLE OF INVENTION:  MUSCLE AND FAT DEPOSITION
;  FILE REFERENCE: 16511.006/37-21(10298)C
```

```
;  CURRENT APPLICATION NUMBER: US/09/960,352
;  CURRENT FILING DATE:  2001-09-24
;  NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 11234
;   LENGTH: 419
;   TYPE: DNA
;   ORGANISM: Bos taurus
;   OTHER INFORMATION: Clone ID: 48-LIB3058-052-Q1-K1-D8
US-09-960-352-11234

  Query Match             4.3%;  Score 46.6;  DB 9;  Length 419;
  Best Local Similarity   49.2%;  Pred. No. 1.3;
  Matches  155;  Conservative    0;  Mismatches  154;  Indels    6;  Gaps    1;

Qy        206 TTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTC 265
              || |||| ||  | | |   ||  |||||||         ||  |     |      | ||
Db         96 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTATTTTTTTTTTTTTT 155

Qy        266 ATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATA 325
              ||| |||||  ||| ||   ||||   ||||  |||    || | | |   | || ||
Db        156 TTTTTTTTTTTTTTTTTTTTAAAAATAATAATAAAAAAATTTTAAAAAGTTAAAATTTA 215

Qy        326 TTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTC 385
              | |     || || || ||| | ||  |   |  ||  ||  ||  | |||  |
Db        216 TAATTTTATTAAATATAAAAAAAAATTTTTTTATATTATATTTTTTTAAATTTAAATTAA 275

Qy        386 AAAGGTAAAAAGTC------AACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAA 439
              |    |||||  |       ||    ||  || |  |  |  |  | |       |  ||
Db        276 ATTTTTAAAATTTTTTAAAAAATTATAATTTAAATAAAATAATAAATAAAAAATTTTAAA 335

Qy        440 ATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTA 499
              | | |  ||   | |||| || || |  |||||| |    | | |||||||| | | |
Db        336 AAAAATTTTTTAAAAAATATTTTTATAAAAATAATTTTTATATAAATAATTTTATTTTTT 395

Qy        500 CTATTAAAACATTTA 514
               ||   | | |||||
Db        396 TTAAATATATATTTA 410


RESULT 72
US-10-311-455-909/c
; Sequence 909, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 909
;   LENGTH: 6535
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-909

  Query Match             4.3%;  Score 46.6;  DB 15;  Length 6535;
  Best Local Similarity   47.2%;  Pred. No. 5.1;
  Matches  175;  Conservative    0;  Mismatches  194;  Indels    2;  Gaps    1;

Qy        144 TGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGA 203
              | | ||  |||   ||  |  | | |  |||||  |       |  | |   ||| || |

Db       2362 TAATAATAATTACACAACATTACTAATATAATAAATACTACCGAATTATACACTTCAAAA 2303

Qy        204 CATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTAT 263
              | |         || ||  |   ||      | ||  || ||   | |  || | |   |
Db       2302 TAATTAAAATCGAAAATTTTATATTACAAATATTTTAACACAAAAAATTCAAAAAATAAA 2243

Qy        264 TCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATA 323
              | | |    ||  ||| || ||||||   | ||| |||    |   | | ||| |  | |
Db       2242 TAAATAAACTTATTTTTATTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2183

Qy        324 TATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAG 383
              |  |   |    |||| ||| | ||||  |    | |  | | |    ||| |   |
Db       2182 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACTCAACAATATATTAT 2123

Qy        384 TCAAAGGTAAAAAGTCAACCT--AAAGGTATACTACCTGCTTTCTTATCGCACTGCAAAT 441
              |||||  |||| |   ||      |||  |||| |       ||||||||   | |  |||
Db       2122 TCAAACATAAATATCAAAAAATCAAAACTATATTTTTAAATTTCTTATTAAAATAAAAAA 2063

Qy        442 AGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACT 501
              ||     | ||  ||   | |  ||| ||  | | |||| |  ||     |   ||
Db       2062 CCCAAAATATAAAATATTACATCTCAAAAAAAAATAAAAAAAAAACATAAAAAAAAAACA 2003

Qy        502 ATTAAAACATT 512
              |||||  | |
Db       2002 CTTAAATAAAT 1992


RESULT 73
US-10-311-455-1956/c
; Sequence 1956, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determi
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1956
;   LENGTH: 8592
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1956

  Query Match             4.3%;  Score 46.6;  DB 15;  Length 8592;
  Best Local Similarity   46.9%;  Pred. No. 5.8;
  Matches  177;  Conservative    0;  Mismatches  199;  Indels    1;  Gaps    1;

Qy        279 TTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAA 338
              || | || | ||  | ||||  | |       |   | || | ||  |    || || |
Db        542 TTTTACATATAACTTTTAAAATATTTATAAAAATAATCGAACCTAACTCCTTCATTTTAC 483

Qy        339 AATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGT 398
              || |||||| | ||| | | || |  | || |   |||  |  | |||||   | | |
Db        482 AAATAAACATATAATAAATCCA-AAAAAATAAAAAAATTATTTACTCAAAATCACACAAC 424

Qy        399 CAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATT 458
              || |||||  || |       || |    ||  | |   |       ||  ||| | ||
Db        423 TAAACTAAAACTAAAAACAAATCTCTTAATTCCTAATAATATATTCTTTTCACACTATTT 364

Qy        459 TTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTT 518
                | |||  ||    |  | ||| | |||   |||     | |  ||| ||  |  ||
```

```
Db        363 AAACTTTCTAACCTTTACAAACAAATAAAATACGTTACAACTTTTCTAAAAATACAAATT 304

Qy        519 CCAAATATTCTGTCATTCAGGAGTATGGAAGTATCGATGGCTTCTTTAAAATGAAGCAGG 578

Db        303 TTTAAAATTTTTAAAAAAAATATTATAAAACTCAAATAATTTACCTTAAAATCATAATTC 244

Qy        579 AGGGTCTGGCAGAGAGTATCTATGAAATAAGTTCCTCTGACCTTCACGCTTAATTTTCTG 638

Db        243 CCTTTCTCTCACCCACTATCCCTATAATTTTTAATATATAAATATAACTTTAATTTTTTT 184

Qy        639 AATGGAGTGGAGCAAAT 655

Db        183 TATTTAATAAAAAAAAT 167


RESULT 74
US-10-482-823-5/c
; Sequence 5, Application US/10482823
; Publication No. US20050003463A1
; GENERAL INFORMATION:
;  APPLICANT: Adorjan, Peter
;  APPLICANT:  Burger, Matthias
;  APPLICANT:  Maier, Sabine
;  APPLICANT:  Lesche, Ralph
;  APPLICANT:  Cottrell, Susan
;  APPLICANT:  De Vos, Theo
;  TITLE OF INVENTION: Method and nucleic acids for the analysis of colorectal cell
;  TITLE OF INVENTION:  proliferative disorders
;  FILE REFERENCE: 47675-55/P119US
;  CURRENT APPLICATION NUMBER: US/10/482,823
;  CURRENT FILING DATE:  2003-12-30
;  PRIOR APPLICATION NUMBER: PCT/EP03/01457
;  PRIOR FILING DATE: 2003-02-13
;  PRIOR APPLICATION NUMBER: EP 02004551.4
;  PRIOR FILING DATE: 2002-02-27
;  NUMBER OF SEQ ID NOS: 15
; SEQ ID NO 5
;   LENGTH: 29993
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-482-823-5

  Query Match             4.3%;  Score 46.6;  DB 18;  Length 29993;
  Best Local Similarity   46.6%;  Pred. No. 11;
  Matches  185;  Conservative    0;  Mismatches  209;  Indels    3;  Gaps    1;

Qy        134 TGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTC 193

Db      13597 TAATTTTCTAATAAAAAAACAAATATATCCAACATACATTAATATTTTCAATAAACTAAC 13538

Qy        194 ATCTTAAAGACA---TTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACA 250

Db      13537 ATTTAAAAAACAAATTTCTCATTAAAATTTATAATCTTCATAATTCTAATTTTCTCCAAA 13478

Qy        251 GACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCA 310

Db      13477 AACAAACCTTAATCTCTTTAATTACCTTCTACTAAAATCAATTTCATATTATTTTAAAAA 13418

Qy        311 GAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTA 370

Db      13417 ATTTTTAACCCTCAAAATTATAAATTTCATTTAAAAAAAAACATATTCTATATTTATTTA 13358

Qy        371 CTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATC 430

Db      13357 TATCAATATTTCAAAATATAAAAACCAAAAATCCCAAATATATCTATTAAAAATATACAT 13298

Qy        431 GCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATT 490

Db      13297 AACCTTTTACTAAAAAAAATTACATCTTTTTTTTCAAACACAATTAAAAAAATAAAAAAA 13238

Qy        491 TTTTATGTACTATTAAAACATTTACTTTCCAAATATT 527

              |||||| || | || ||  ||||| |||  ||   ||
Db      13237 TTTTATTTAATTTTTAATAATTTATTTTTAAACATTT 13201


RESULT 75
US-10-312-841-1/c
; Sequence 1, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
;  APPLICANT: Epigenomics AG
;  TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
;  FILE REFERENCE: E01/1208/WO
;  CURRENT APPLICATION NUMBER: US/10/312,841
;  CURRENT FILING DATE:  2002-12-30
;  NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 1
;   LENGTH: 3673778
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (3294164)
US-10-312-841-1

  Query Match             4.3%;  Score 46.6;  DB 16;  Length 3673778;
  Best Local Similarity   45.3%;  Pred. No. 1.1e+02;
  Matches  169;  Conservative    0;  Mismatches  204;  Indels    0;  Gaps    0;

Qy        198 TAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATAC 257

Db    1713797 TATATAAAATATATATAAAATATATAAAATATATATAAAATATATATAATATATATATAA 1713738

Qy        258 AGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAG 317

Db    1713737 TATATATATAATATATATATTATATATATATAATATATAAAATATATATAAAATATATAT 1713678

Qy        318 ATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATC 377

Db    1713677 AAAATATATAAAATATATATAAAATATATATAAAATATATATAAAATATATAAAATATAT 1713618

Qy        378 TTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGC 437

Db    1713617 ATAAAATATATATAAAATATATAAAATATATATAAAATATATAAAATATATATAAAATAT 1713558

Qy        438 AAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATG 497

Db    1713557 ATATATAAAATATATATAAAATATATATATAAATTATATATATAAAATATATATATAAA 1713498

Qy        498 TACTATTAAAACATTTACTTTCCAAATATTCTGTCATTCAGGAGTATGGAAGTATCGATG 557

Db    1713497 TATATATAAAATATATATATAAAATATATATATAAAATATATATAAAATATATATATAAA 1713438

Qy        558 GCTTCTTTAAAAT 570

Db    1713437 ATATATATAAAAT 1713425


RESULT 76
US-10-425-114-21407/c
; Sequence 21407, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
;  APPLICANT: Liu, Jingdong
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Screen, Steven E
;  APPLICANT:  Tabaska, Jack E
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53313)B
```

```
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 21407
;   LENGTH: 2635
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: LIB3332-032-E10_FLI
US-10-425-114-21407

  Query Match             4.2%;  Score 46.2;  DB 17;  Length 2635;
  Best Local Similarity   48.9%;  Pred. No. 4;
  Matches  154;  Conservative    0;  Mismatches  158;  Indels    3;  Gaps    1;

Qy        146 AGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACA 205
              ||| | ||     || |  ||||    || || || || ||  | || ||| || | | |
Db        815 AGAGAGATATTCTATAAAGAATCTTACAAAAGATTGGTTGAGTAGCTTATCATACAAATA 756

Qy        206 TTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTC 265
              | ||||| |   | |  || ||   |||||     | | | | | |   | |  ||||
Db        755 TGGTTTTTTCTGTTTTTTCTCTACCTTTTTGGCCTACAGCCTCCTGTATTGTAAATATTT 696

Qy        266 ATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATA 325
              ||||  || || |       ||| |  | ||  |   | |       |  | || | | |
Db        695 CTTTGCTTCTGTTAGTGAAGAAATACATGTAGGGGCCTCCCCCACCATTCACATAAAAAA 636

Qy        326 TTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTC 385
              ||  | | |     || ||||| | || ||      |||    | || |  ||    ||
Db        635 TTGACAAATATGCTATAAAACAAAGAAGGAAGCTGGATTTATATCTTTCAGTTTT---TC 579

Qy        386 AAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAA 445
              ||||  |  |  |  | |   | | ||| ||   |||   |    | | ||  |||   |
Db        578 AAAGCAATGATCTGCAACAGCATGAATAATAGGAGCTACATGTGGGAATTGTCAATTCCA 519

Qy        446 ATTACCACAAATTTT 460
              | ||    ||  |||
Db        518 AGTAAGTAAATATTT 504


RESULT 77
US-10-425-115-158426
; Sequence 158426, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(53222)B
;  CURRENT APPLICATION NUMBER: US/10/425,115
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 158426
;   LENGTH: 4119
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: MRT4577_76064C.1
US-10-425-115-158426

  Query Match             4.2%;  Score 46.2;  DB 18;  Length 4119;
  Best Local Similarity   48.9%;  Pred. No. 5;
  Matches  154;  Conservative    0;  Mismatches  158;  Indels    3;  Gaps    1;

Qy        146 AGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACA 205
              ||| | ||     || |  ||||    || || || || ||  | || ||| || | | |
Db       3297 AGAGAGATATTCTATAAAGAATCTTACAAAAGATTGGTTGAGTAGCTTATCATACAAATA 3356

Qy        206 TTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTC 265
              | ||||| |   | |  || ||   |||||     | | | | | |   | |  ||||
Db       3357 TGGTTTTTTCTGTTTTTTCTCTACCTTTTTGGCCTACAGCCTCCTGTATTGTAAATATTT 3416

Qy        266 ATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATA 325
              ||||  || || |       ||| |  | ||  |   | |       |  | || | | |
Db       3417 CTTTGCTTCTGTTAGTGAAGAAATACATGTAGGGGCCTCCCCCACCATTCACATAAAAAA 3476

Qy        326 TTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTC 385
              ||  | | |     || ||||| | || ||      |||    | || |  ||    ||
Db       3477 TTGACAAATATGCTATAAAACAAAGAAGGAAGCTGGATTTATATCTTTCAGTTTT---TC 3533

Qy        386 AAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAA 445
              ||||  |  |  |  | |   | | ||| ||   |||   |    | | ||  |||   |
Db       3534 AAAGCAATGATCTGCAACAGCATGAATAATAGGAGCTACATGTGGGAATTGTCAATTCCA 3593

Qy        446 ATTACCACAAATTTT 460
              | ||    ||  |||
Db       3594 AGTAAGTAAATATTT 3608


RESULT 78
US-10-311-455-951/c
; Sequence 951, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 951
;   LENGTH: 6963
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-951

  Query Match             4.2%;  Score 46.2;  DB 15;  Length 6963;
  Best Local Similarity   48.9%;  Pred. No. 6.5;
  Matches  161;  Conservative    0;  Mismatches  158;  Indels   10;  Gaps    1;

Qy        206 TTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTC 265
              ||  ||| || |||||||| ||  || | ||||||| |  || ||    |    | | |
Db       4889 TTACTTTCTTTATATACTCTCTCCATATATTTAAAAAATTTTTCATTTTTTTTCATCATT 4830

Qy        266 ATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATA 325
              |  | |||||     |    |  ||||  | | |  |  |  || |  |||      |||
Db       4829 ACCTATTTTTCACCTACACTCAAAACCTTTTATAATTTTTTTTTAATTAAAATCTTCATA 4770

Qy        326 TTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTC 385
              |     | | |   |   | | ||   | ||||    |   | | ||     ||
Db       4769 AAATTTTAATAACACAAATTTAACATATCTATTTATATATATTACGTATACCTCTACTTC 4710

Qy        386 AAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAA 445
              |  ||||||  |          | | |     |||| |  ||  | |  || || ||
Db       4709 ACAAATAAAAACT----------ACAAAATTTACTCTTTTTATTCAAAATAAAATTAAAA 4660

Qy        446 ATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTA 505
              |||||  | ||| || |  |   |  |||  | | |||  || |   |||||  |    |||
Db       4659 ATTACAAAAAAATTAAACTAAAATCTAAATTTAAAAACTATTAACATTTATTTTAATTTA 4600

Qy        506 AAACATTTACTTTCCAAATATTCTGTCAT 534
              | | || || ||   |||||   | | ||
```

```
Db       4599 ATAAATATATTTATAAAATAACTTTTAAT 4571


RESULT 79
US-10-221-714A-9/c
; Sequence 9, Application US/10221714A
; Publication No. US20040048254A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with
;  TITLE OF INVENTION:  tumor suppressor genes and oncogenes
;  FILE REFERENCE: 5013.1005
;  CURRENT APPLICATION NUMBER: US/10/221,714A
;  CURRENT FILING DATE:  2003-01-21
;  PRIOR APPLICATION NUMBER: PCT/EP01/02955
;  PRIOR FILING DATE: 2001-03-15
;  PRIOR APPLICATION NUMBER: DE 10013847.0
;  PRIOR FILING DATE: 2000-03-15
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 540
; SEQ ID NO 9
;   LENGTH: 8169
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: 2047, 2136, 7481, 7492, 7502..7503, 7505, 7511
;   OTHER INFORMATION: n is a or g or c or t
US-10-221-714A-9

  Query Match             4.2%;  Score 46.2;  DB 17;  Length 8169;
  Best Local Similarity   49.3%;  Pred. No. 7;
  Matches  175;  Conservative    0;  Mismatches  178;  Indels    2;  Gaps    2;

Qy        175 TAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTT 234
              ||  ||| ||     | |     |||| || |  | ||   ||  ||     ||  ||
Db       1043 TAACTTTATAACTCTTATATAAATAAACACCTCATATTAACAAACTATATAATTATTTAA 984

Qy        235 TTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTA 294
              ||||||| | |||| |   |||       | ||| |  | |     | || | ||||  ||
Db        983 TTTAAAAAAACTTAAAATAATAACCAACTCCATCTAATATCTCAATTTTTATAAAAAATA 924

Qy        295 TAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATG 354
              |||| | |   ||| | | | | || |  ||| |   |  | || ||| |  |  |||
Db        923 TAAATATAATTATTAAAAAATAAATTAACTATAAATAAAATAAATATTTATAAATAAACT 864

Qy        355 AGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATAC 414
                |  | ||| | | |||||  ||  |    ||   | | |  |||  |||| | ||
Db        863 CATATAATTTAAAAT-CTTAAAATTTTATATTAAATCATATATACAAAATAAATATCTAT 805

Qy        415 TACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAAT 474
                | |  |    ||    |    || ||   ||  |||||  ||| |       |||.| |
Db        804 ATCATAATCAAATAAAATAAAATAATTAATTATC-CCACATTTTTAAAAAAAAAAAATT 746

Qy        475 CTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCT 529
               | | |  | | |||      |  |  | |||||  || ||   |  ||| |  ||
Db        745 ATAAAACTAAAAAATAAAAATTAAAAAAATAAAAAATCTAAAATTAAAAAAAACT 691


RESULT 80
US-10-312-841-2
; Sequence 2, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
;  APPLICANT: Epigenomics AG
;  TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
;  FILE REFERENCE: E01/1208/WO
;  CURRENT APPLICATION NUMBER: US/10/312,841
;  CURRENT FILING DATE:  2002-12-30
;  NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 2
;   LENGTH: 3673778
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (379615)
US-10-312-841-2

  Query Match             4.2%;  Score 46.2;  DB 16;  Length 3673778;
  Best Local Similarity   58.3%;  Pred. No. 1.3e+02;
  Matches   81;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy        151 TATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTT 210
              |||||| | |   | | | |  ||| ||| | |       | || ||| | | ||| ||
Db    2977346 TATTGATGAATTGTATTAAATATATAATTTAGAAATTATATTAATTTTATATATATTTTT 2977405

Qy        211 TTGTTAATATACTCCCTTGATTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTG 270
              | |  | || | |   | | ||||||||   |  ||||  |  ||   ||||  || |
Db    2977406 TAGAAATTAAAGTAGGTAGTTTTTTTAATTTAGTTTACGAAGTTAGTTATATTAGTTCG 2977465

Qy        271 TTTTTGGTTTGTTCAAAAA 289
              |||||   ||||| | |||
Db    2977466 TTTTTATATTGTTGATAAA 2977484


RESULT 81
US-10-311-455-1037/c
; Sequence 1037, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determi
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1037
;   LENGTH: 6074
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1037

  Query Match             4.2%;  Score 46;  DB 15;  Length 6074;
  Best Local Similarity   48.8%;  Pred. No. 6.8;
  Matches  124;  Conservative    0;  Mismatches  130;  Indels    0;  Gaps    0;

Qy        239 AAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAA 298
              | | |||     | |  || |  ||  |||| | | |   || ||       |    | |
Db       4228 ATAAACCACCACGCCTAACCGAAATATATTTTTATATACATTATTTTCCTTAAAATTTTA 4169

Qy        299 GAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTG 358
```

```
                ||| || | | ||| || ||  | |          |||   ||| || |||
Db       4168 ACAATTTCTTAAAAAAAATATAATTAACTTTATCAAACAAATAAAAAAACTAAAACTCAA 4109

Qy        359 CATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACC 418
              ||| |  | || || ||| |  | ||| |    ||||   ||    ||    | ||||
Db       4108 AATAATTAAATAATTTATCCTCAAATCACACAACAAAAACAAAAAATAAAACACACTAAA 4049

Qy        419 TGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCA 478
              | |    || |   | |  ||| | ||    ||||||     |  |   ||| || || |
Db       4048 TTCAAAATTTTAAAATTCAAAAAATAATACCCCACAATAAAAACCTCAAAAACAACCTTA 3989

Qy        479 GAAAACATAATTTT 492
              |||| | ||||||
Db       3988 AAAAAAAAAATTTT 3975


RESULT 82
US-10-311-455-1224/c
; Sequence 1224, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1224
;   LENGTH: 6591
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1224

  Query Match             4.2%;  Score 46;  DB 15;  Length 6591;
  Best Local Similarity   47.0%;  Pred. No. 7;
  Matches  142;  Conservative    0;  Mismatches  160;  Indels    0;  Gaps    0;

Qy        225 CCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTC 284
              | |  | | |  |||| | ||| |  || |  | ||||||||  |   |    | |||
Db       3329 CATACAATATAATAAACTAACTTTCGCACTTTAAATTATTCATTAATCATATAATAATTC 3270

Qy        285 AAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAA 344
              ||| ||   ||     ||   | ||   |||       |  ||   | ||  || ||||
Db       3269 AAATAATTCATCTTCCAAAATAAACACTAAAACTAACACTCATAATAAATTATAATTTAA 3210

Qy        345 ACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCT 404
              | |  |||  |      | |    |  |||   |      ||    |   ||  |
Db       3209 AAAATAAAACAACCGCCAATCTCATTAATAAAAATTATAATAATTAAATTTATCAAAAAT 3150

Qy        405 AAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTT 464
              |||  || || | || ||||||  |        || |        || | | | |||||
Db       3149 AAACATACACGATCTTCTTTCTAAAATTTTCTTTAAAATTTTCGCACAAATAATCTATTT 3090

Qy        465 TGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAAT 524
                ||||  |   | ||||  | |  ||||||  |    | |||  |  |  || |||||
Db       3089 CTTAAATTTTAAAAAAAAAACTTAAATTTTATTAATCCCTCAAAACTCAATCTTTCAAAT 3030

Qy        525 AT 526
              ||
Db       3029 TT 3028


RESULT 83
US-10-240-453-322/c
; Sequence 322, Application US/10240453
; Publication No. US20030148326A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA
;  TITLE OF INVENTION:  Transcription
;  TITLE OF INVENTION:  by Means of Assessing the Methylation Status of Genes Associated
;  TITLE OF INVENTION:  With DNA TRanscription
;  FILE REFERENCE: 5013.1009
;  CURRENT APPLICATION NUMBER: US/10/240,453
;  CURRENT FILING DATE:  2002-10-02
;  PRIOR APPLICATION NUMBER: PCT/EP01/03973
;  PRIOR FILING DATE: 2001-04-06
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 350
; SEQ ID NO 322
;   LENGTH: 9095
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-453-322

  Query Match             4.2%;  Score 46;  DB 15;  Length 9095;
  Best Local Similarity   44.6%;  Pred. No. 8.2;
  Matches  232;  Conservative    0;  Mismatches  280;  Indels    8;  Gaps    1;

Qy         33 TCTCTATGTTCTTATAAGCAAATGTTACAATGAGATATGAGTTTTAAAGCCAGATCTTCC 92
              ||| | | | |||  |   || | |   | | |||  | |  ||||   | ||  |
Db       7037 TCTATTTATACTTTAATAAAATTTATCTTCTAATATAAAAATCATAAATAAATATAATTA 6978

Qy         93 TTATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATA 152
              | |     |     | | |  |    | |  || | |  |   | | |||  |
Db       6977 TAACTATAATATAATTTTTTAAACCAATATTTTCTAAAATTTAAAAATAACTCAAAAAAA 6918

Qy        153 TTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTT 212
              |  |||| || |||| | |  ||| |||  || |  ||  |  ||  | | |||||    |
Db       6917 TCTATCAATAAAAATAAATACATACTTTATTACATCATTTTTTCACATAAACATTTCAAT 6858

Qy        213 GTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTT 272
              | ||||     ||   ||   |  | | |  ||  |   | | | | ||||||  |
Db       6857 AATTATATTAAAACTATCTTAAATCTATAAATTTTTTAATTAAAAATCAATTCATCTCAC 6798

Qy        273 TTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCA 332
              | |  |   |  || | || |  |||| | | | | |       |  |  |
Db       6797 TATATTAAACTATCAATATTAATTATCAAATTTAATTATATATTACATTACTTTACCTAA 6738

Qy        333 GTTGAAAATTAAACACAAAATGAGTGCATATTACATTACT--------TAATCTTGCAGT 384
              ||  |||| |||   ||  |     | |  |  ||| ||         |||  || || |
Db       6737 TTTTCAAATAAAAATAAACTTTTTATCTTTATTTATTTCTATAATAAATAAATTTACAAT 6678

Qy        385 CAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGA 444
              ||     ||  |   |    | | |  ||     |||  ||      |    || | |
Db       6677 AAATTACAATTAAATATTAAATAAATTTTATAATCAATTTAATACAAAAAATAAATTTTA 6618

Qy        445 AATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATT 504
              || || || ||   |  | |  | |  |  | |||| ||| ||  | || ||     |
Db       6617 AAATATCAAAATACATTATATATTACTTTTTCTAAAAAAAATATTTCCTCATTTATACAT 6558

Qy        505 AAAACATTTACTTTCCAAATATTCTGTCATTCAGGAGTAT 544
              ||| | | |        | || || | | |||    |||
Db       6557 TAAAAAATAAAAAAAACTAAATACTATAACTCATACCTAT 6518
```

RESULT 84
US-10-311-455-163/c
; Sequence 163, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 163
; LENGTH: 17137
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-163

Query Match 4.2%; Score 46; DB 15; Length 17137;
Best Local Similarity 44.4%; Pred. No. 11;
Matches 184; Conservative 0; Mismatches 230; Indels 0; Gaps 0;

```
Qy        108 ATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAAT 167
              ||   ||  | |  |  | | |   |  |  ||  | | | |  || | || |  |
Db      11499 ATACATATATATATATATATATATAAATATATTTCTACACATAAAAATATAAATATACTT 11440

Qy        168 CAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCT 227
              |    |  | |||  | |   |    |  |||  |      |||    |||||    |
Db      11439 CCTAAACAATTTTCAAATACATTAATTTAATAATAAATACAATTTACATATATAAATACA 11380

Qy        228 TGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAA 287
              | || | | || |  || |   | | ||     ||| ||| | |   |   ||  | | |
Db      11379 TAATATATATATACTACATATTATATATTATTTTATACATATATACCTATATTTATAATA 11320

Qy        288 AAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACA 347
              | |  ||  |  ||||   |    | ||  |  | | | || |      ||||  || |
Db      11319 ATAACTACCATAAAATATTTAACAATAATTAAAAAAAAATACCACAAAAAAAAAAAATCT 11260

Qy        348 CAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAA 407
               |||||      |||||  |||| |||   ||        ||   | |||      ||| |
Db      11259 AAAAATCTAATAATATTTCATTTCTTCGCCTATATAATCAATACATAAACATCCACTACA 11200

Qy        408 GGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGG 467
                  |    |  | | |     |      |      ||| | |  |  |  | | |||
Db      11199 TTACTCTCTACAATTATTTATAAACTTAAATATATACTATTTCAAATACATACAATTTTA 11140

Qy        468 AAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCA 521
              |||  |   |   ||| ||| |||| |    |  ||||||| | | ||| || |
Db      11139 AAAAACAATACACAAAAATATTTTTATTATAAAAATTAAAAAACTAACTCTCTA 11086
```

RESULT 85
US-10-425-115-115706/c
; Sequence 115706, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 115706
; LENGTH: 1121
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1121)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_37006C.1
US-10-425-115-115706

Query Match 4.2%; Score 45.8; DB 18; Length 1121;
Best Local Similarity 49.4%; Pred. No. 3.3;
Matches 119; Conservative 0; Mismatches 122; Indels 0; Gaps 0;

```
Qy        285 AAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAA 344
              ||||||   | ||| |||   |   | | |||||| | |||  |   | |  |||| |||
Db        742 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGATAAAATAAAAAAAAATAAAAAAATAA 683

Qy        345 ACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCT 404
              | | |||| |    | |  |    |   ||   |  |   ||    |||||   ||
Db        682 AAAAAAAAAAAAAAAAAAAAATTAAAAAAAAAAAATAAAAGTAATAAAAAAAAACAAAAAA 623

Qy        405 AAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTT 464
              |||  |||| ||    |  | ||||   |    |||   | | ||  | ||     |
Db        622 AAATATATAATAAAAACAATATTATAAAAAAAAAAAATAACAATAAAAAAACAAAAACAA 563

Qy        465 TGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAAT 524
              |||||||  | | ||| |||||       |     || ||| | |   ||    ||||
Db        562 ATAAAATAATATAAAAAACCATAAAAAAAAAAAAATATAATAACAAAAACACAAAAAAAT 503

Qy        525 A 525
              |
Db        502 A 502
```

RESULT 86
US-10-311-455-164
; Sequence 164, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 164
; LENGTH: 17137
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-164

Query Match 4.2%; Score 45.8; DB 15; Length 17137;
Best Local Similarity 49.4%; Pred. No. 12;
Matches 119; Conservative 0; Mismatches 122; Indels 0; Gaps 0;

```
Qy        207 TGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCA 266
              |||||   |||  || |   ||  || ||||  | ||   ||   |  |    ||||
Db      14656 TGTTTAAGTAAAGTAATTATTTTTTTGTTTTTTATGAAAATAATTAGTTTTTTTTATTGG 14715

Qy        267 TTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATAT 326
              |     ||  ||  || || ||| ||||   |||||   ||   ||||    |   || |
Db      14716 ATATAAGTTTTTTATTTAAAGAAAAGTATTTTGAAATTAGTTTTTAGAATTTTTTAATGT 14775

Qy        327 TAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCA 386
              ||   | || |||  || | |  | | |  |   | ||| | ||  || | ||    | |
Db      14776 TATTAAATTTAAATATATAAATTAGAAGGATTTGTTTTAAAATAAATAGTGTTATTATAA 14835

Qy        387 AAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAA 446
              | |||||||    |       || ||   ||  |  ||  |  | |    | || || | |
Db      14836 TATGTAAAATTGGATTAATTAGTTAATTTAAATTTTTGATAGTTGACGGGAAATTATATA 14895

Qy        447 T 447
              |
Db      14896 T 14896
```

RESULT 87
US-09-754-853A-2
; Sequence 2, Application US/09754853A
; Publication No. US20030005491A1
; GENERAL INFORMATION:
; APPLICANT: Hauge, Brian M.
; APPLICANT: Parnell, Laurence D.
; APPLICANT: Parsons, Jeremy D.
; APPLICANT: Wang, Ming Li
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; TITLE OF INVENTION: Soybean Cyst Nematode Resistance
; FILE REFERENCE: 38-10(15810)B
; CURRENT APPLICATION NUMBER: US/09/754,853A
; CURRENT FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: US 60/174,880
; PRIOR FILING DATE: 2000-01-07
; NUMBER OF SEQ ID NOS: 1119
; SEQ ID NO 2
; LENGTH: 335913
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (45163)..(45314),(45450)..(45509),(46941)..(48763),(48975)..(49573)
; OTHER INFORMATION: Clone ID: 240O17_region_G3
US-09-754-853A-2

Query Match 4.2%; Score 45.8; DB 10; Length 335913;
Best Local Similarity 49.4%; Pred. No. 53;
Matches 206; Conservative 0; Mismatches 202; Indels 9; Gaps 3;

```
Qy        119 TTGAAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGT 178
              || || |     || |    ||  ||   || |  |||| |||  ||| | || |    |
Db      98750 TTAAAATTATATATGTTTTATTTTTTTTTAACTTATGATAATTTTAAAGCTGTAATATTT 98809

Qy        179 TTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTA 238
              ||| | |     | | |  || ||  |||  ||| || |       ||  || |||||
Db      98810 TTTATTGTTTTTATGACTTTGAAATCATATTTTATTTTATGTTTAAAATTTTTTGTTTTA 98869

Qy        239 AAAGACCTTACAGACATACAGCTAT--TCATTTGTTTTTGGTTTGTTCAAAAAAGGTATA 296
              |    |||   || |  |  | |  |  |||  ||||   || ||  |||||  ||||
Db      98870 TGATTTTTTATTTACTTGTATTTTTTATACTTTTATTTTCAATTCTTTTAAAAAATTATA 98929

Qy        297 AAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAG 356
              || || |  |||  |    |  |||       |  |   ||||  ||      || ||
Db      98930 AATAATTTTATTTATTTAATTATTAAATAA-----ATATTTTAATTTTACTAGTTATTAG 98984

Qy        357 TGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTA 416
              |    |||   || ||  | |       |    || ||  |     ||| | |||   |
Db      98985 TTTCATTTAATATATTTTGTATATTTAAAATTTATATAATTTTCTAATAATGTTATTGAA 99044

Qy        417 CCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAA--ATTTTATTTTGGAAATAAT 474
              || |||  || |   | |  ||  | ||||||  || ||  || ||   |     |
Db      99045 GTTGATTTGTTTTGTAAATAAAATAAAAAATTAACATAATGATATTTAATATTGTTTTTC 99104

Qy        475 CTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGT 531
              |   || || ||| |||| |  ||||  |    |   |   |  |||||||  |||
Db      99105 TTTTAAATACTTAAATTTTAAAAAACTAAAATTTTAAAAAAAGTTAAAATATTTTGT 99161
```

RESULT 88
US-09-754-853A-3
; Sequence 3, Application US/09754853A
; Publication No. US20030005491A1
; GENERAL INFORMATION:
; APPLICANT: Hauge, Brian M.
; APPLICANT: Parnell, Laurence D.
; APPLICANT: Parsons, Jeremy D.
; APPLICANT: Wang, Ming Li
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; TITLE OF INVENTION: Soybean Cyst Nematode Resistance
; FILE REFERENCE: 38-10(15810)B
; CURRENT APPLICATION NUMBER: US/09/754,853A
; CURRENT FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: US 60/174,880
; PRIOR FILING DATE: 2000-01-07
; NUMBER OF SEQ ID NOS: 1119
; SEQ ID NO 3
; LENGTH: 335913
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (46798)..(48763),(48975)..(49573)
; OTHER INFORMATION: Clone ID: 240O17_region_G3
US-09-754-853A-3

Query Match 4.2%; Score 45.8; DB 10; Length 335913;
Best Local Similarity 49.4%; Pred. No. 53;
Matches 206; Conservative 0; Mismatches 202; Indels 9; Gaps 3;

```
Qy        119 TTGAAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGT 178
              || || |     || |    ||  ||   || |  |||| |||  ||| | || |    |
Db      98750 TTAAAATTATATATGTTTTATTTTTTTTTAACTTATGATAATTTTAAAGCTGTAATATTT 98809

Qy        179 TTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTA 238
              ||| | |     | | |  || ||  |||  ||| || |       ||  || |||||
Db      98810 TTTATTGTTTTTATGACTTTGAAATCATATTTTATTTTATGTTTAAAATTTTTTGTTTTA 98869

Qy        239 AAAGACCTTACAGACATACAGCTAT--TCATTTGTTTTTGGTTTGTTCAAAAAAGGTATA 296
              |    |||   || |  |  | |  |  |||  ||||   || ||  |||||  ||||
Db      98870 TGATTTTTTATTTACTTGTATTTTTTATACTTTTATTTTCAATTCTTTTAAAAAATTATA 98929

Qy        297 AAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAG 356
              || || |  |||  |    |  |||       |  |   ||||  ||      || ||
Db      98930 AATAATTTTATTTATTTAATTATTAAATAA-----ATATTTTAATTTTACTAGTTATTAG 98984

Qy        357 TGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTA 416
              |    |||   || ||  | |       |    || ||  |     ||| | |||   |
Db      98985 TTTCATTTAATATATTTTGTATATTTAAAATTTATATAATTTTCTAATAATGTTATTGAA 99044

Qy        417 CCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAA--ATTTTATTTTGGAAATAAT 474
              || |||  || |   | |  ||  | ||||||  || ||  || ||   |     |
Db      99045 GTTGATTTGTTTTGTAAATAAAATAAAAAATTAACATAATGATATTTAATATTGTTTTTC 99104

Qy        475 CTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGT 531
              |   || || ||| |||| |  ||||  |    |   |   |  |||||||  |||
Db      99105 TTTTAAATACTTAAATTTTAAAAAACTAAAATTTTAAAAAAAGTTAAAATATTTTGT 99161
```

RESULT 89
US-10-275-323A-13/c
; Sequence 13, Application US/10275323A
; Publication No. US20050079491A1

```
; GENERAL INFORMATION:
;  APPLICANT: DONNE-GOUSSE, CAROLE
;  APPLICANT:  LAUDET, VINCENT
;  APPLICANT:  HANNI, CATHERINE
;  TITLE OF INVENTION: METHOD OF DETECTING AND IDENTIFYING THE PRESENCE OF BIOLOGICAL
;  TITLE OF INVENTION:  MATERIALS DERIVED FROM BIRDS, AND OLIGONUCLEOTIDES FOR CARRYING
;  TITLE OF INVENTION:  IT OUT
;  FILE REFERENCE: 0508-1052
;  CURRENT APPLICATION NUMBER: US/10/275,323A
;  CURRENT FILING DATE:  2002-11-08
;  PRIOR APPLICATION NUMBER: PCT/FR01/01279
;  PRIOR FILING DATE: 2001-04-26
;  PRIOR APPLICATION NUMBER: FR 00/05850
;  PRIOR FILING DATE: 2000-05-09
;  NUMBER OF SEQ ID NOS: 96
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 13
;   LENGTH: 408
;   TYPE: DNA
;   ORGANISM: Artificial
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: DNA fragment amplified
;   OTHER INFORMATION: from the genome of birds
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (43)..(43)
;   OTHER INFORMATION: a, c, g, or t
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (57)..(57)
;   OTHER INFORMATION: a, c, g, or t
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (71)..(71)
;   OTHER INFORMATION: a, c, g, or t
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (134)..(134)
;   OTHER INFORMATION: a, c, g, or t
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (212)..(212)
;   OTHER INFORMATION: a, c, g, or t
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (214)..(214)
;   OTHER INFORMATION: a, c, g, or t
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (217)..(217)
;   OTHER INFORMATION: a, c, g, or t
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (226)..(227)
;   OTHER INFORMATION: a, c, g, or t
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (236)..(236)
;   OTHER INFORMATION: a, c, g, or t
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (245)..(245)
;   OTHER INFORMATION: a, c, g, or t
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (253)..(253)
;   OTHER INFORMATION: a, c, g, or t
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (269)..(269)
;   OTHER INFORMATION: a, c, g, or t
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (324)..(324)
;   OTHER INFORMATION: a, c, g, or t
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (330)..(330)
;   OTHER INFORMATION: a, c, g, or t
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (352)..(352)
;   OTHER INFORMATION: a, c, g, or t
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (355)..(355)
;   OTHER INFORMATION: a, c, g, or t
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (357)..(357)
;   OTHER INFORMATION: a, c, g, or t
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (362)..(362)
;   OTHER INFORMATION: a, c, g, or t
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (364)..(364)
;   OTHER INFORMATION: a, c, g, or t
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (406)..(406)
;   OTHER INFORMATION: a, c, g, or t
US-10-275-323A-13

  Query Match             4.2%;  Score 45.6;  DB 19;  Length 408;
  Best Local Similarity   9.0%;  Pred. No. 2.2;
  Matches   31;  Conservative  191;  Mismatches  124;  Indels    0;  Gaps    0;

Qy        133 ATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCT 192
              :: :::: : : ::: : :::|     |  :::::: ::| : :: ::::  :: : ::::
Db        379 RYTRRYDRDHYRDRBNVNDDKTNKNRANRDBDRRDTWRTRRKDDWYDYYNRDKTDNDWBH 320

Qy        193 CATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGA 252
                :::|:  :    :|||:   ||||| : :   :::::::: |:::::    ::: :
Db        319 DBBBYTRGTRCTGGWGTTYCTMTAATAADKWTRDDHBRHDHHRTKHDWWYNKKWHDGWHK 260

Qy        253 CATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGA 312
              : :: :  :::::  :::     ::|::::   : ::: :: :: : :::: :|::::
Db        259 HKHVRVNDDWDBVWNRDDWRRSRNRWTWDDDRNNDGVRRTWDNRDNDNDDDDWWTDVDDK 200

Qy        313 GAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACT 372
              : :::: ::::::::: :: : ::::::: :: ::::|:::  :: | :::::     :
Db        199 WDKDDDDKDBDDRDWWKAYRHWABRWRDWKKDWWDHWDAWYDGHKRAARWHYDYTGGTTK 140

Qy        373 TAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGC 432
              :: |: :: :::: :::: ::: ::: ::  : ::: | :: ::  :  | || ::
Db        139 HRTTYNKDDDKYSYRDRRRWWRYRSWDRRWRDKWWDWTKKWTWRWRHTATGTCCDDCRAG 80

Qy        433 ACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCA 478
              | ||   :: |:   : : : :|   ::: :||| :::: :::|
Db         79 CATTCACYNRRYARCAYBWDWVNVTRGRDBDGYGGANRKRRCMWYA 34


RESULT 90
US-10-311-455-1643/c
; Sequence 1643, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
```

```
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1643
;   LENGTH: 8446
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1643

  Query Match             4.2%;  Score 45.6;  DB 15;  Length 8446;
  Best Local Similarity   48.2%;  Pred. No. 9.8;
  Matches  160;  Conservative    0;  Mismatches  169;  Indels    3;  Gaps    1;

Qy        294 ATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAAT 353
              ||||| ||  ||   | | | | ||  || || | ||   | | |  |||| | |||
Db       1896 ATAAAATAAACCAAATACAAACAAATACTACATAATCCCACTCACATATAAAATCTAAAC 1837

Qy        354 GAGTGCAT---ATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGT 410
              | |  ||   || | | |||  |||     | | |    ||||| | || | |
Db       1836 AAATTAATTCCATAAAAATACAAAATAAAATAATAATTACCAAAAATTAAAACGATTAAA 1777

Qy        411 ATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAA 470
              | ||          |||       |   |||  |||    || || ||| |||    |||
Db       1776 AAACGTAAAATAAAATTAAAAATATATTAATCAAAAAATACAAAATTTTAATTAAATAAA 1717

Qy        471 TAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTG 530
              ||  | |    | | ||| | ||||  |   ||| |||  | | |    | |||||
Db       1716 AAAAATAATTTCAAAAAATCTATTATACAACGTTATAACTATAATTAAAAATATATTATA 1657

Qy        531 TCATTCAGGAGTATGGAAGTATCGATGGCTTCTTTAAAATGAAGCAGGAGGGTCTGGCAG 590
              |  || |  | ||   || ||   ||    || | | ||  ||  |  |  | |   |
Db       1656 TTCTTAAAAAATACTAAAATAAATATATTCTCCTCACAAAAAATAATAACTATATAAAAT 1597

Qy        591 AGAGTATCTATGAAATAAGTTCCTCTGACCTT 622
              |    || ||| || ||| |   | | | | |
Db       1596 AATACATATATTAATTAATTAAATTTAATCAT 1565


RESULT 91
US-10-311-455-1653/c
; Sequence 1653, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1653
;   LENGTH: 15951
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1653

  Query Match             4.2%;  Score 45.6;  DB 15;  Length 15951;
  Best Local Similarity   45.0%;  Pred. No. 13;
  Matches  171;  Conservative    0;  Mismatches  209;  Indels    0;  Gaps    0;

Qy        143 TTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAG 202
              || | ||| ||  ||  | || |         | | |    || ||   || | || |||
Db       8305 TTTATAAAAATAAATTCTCACTATATTACCCAAACTAACCTAAAACCCCTAAACTCAAAC 8246

Qy        203 ACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTA 262
              |      |||| || |      |         | |  | | | | | | | |||    ||||
Db       8245 AATACTTTTTCTTTAACCTCCAAAAATACTAATATTATAAAAATAAAAATAATACGATAC 8186

Qy        263 TTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCAT 322
              ||  ||| ||||   |   ||  ||  |  | |   | | ||   | | |||
Db       8185 TTTTTTTTTTTTAAATAAATTACAATCAAAAAAATTACACTACAAAAAAAAAAACTATTA 8126

Qy        323 ATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCA 382
              || | ||   | |||||||| | |||   | |  || | |||| | | ||  |   |
Db       8125 CAATAATCCCAATAAAAATTAAAAAAAAACAAATTTAATCTAACATCAATAAAAATACAA 8066

Qy        383 GTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATA 442
              | |||   | |||   ||       |   | | ||       |   | |  | |   || |
Db       8065 ATAAAAAAAATAAATATAAAAATTATTAAAAATACAAAACTATACAACTTATTAATAACA 8006

Qy        443 GAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTA 502
              ||  |     |   | ||  || ||||     ||  |    | |||  |  || |
Db       8005 AAACCCAAATATTACAAACTATTACAACTAATACATTAATTCCCTTATATTTCTAAACGA 7946

Qy        503 TTAAAACATTTACTTTCCAA 522
              | | || ||| |   |  ||
Db       7945 TAATAATATTAATAATATAA 7926


RESULT 92
US-10-240-485-133/c
; Sequence 133, Application US/10240485
; Publication No. US20030148327A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with
;  TITLE OF INVENTION:  Metastasis
;  FILE REFERENCE: 5013.1007
;  CURRENT APPLICATION NUMBER: US/10/240,485
;  CURRENT FILING DATE:  2002-10-02
;  PRIOR APPLICATION NUMBER: PCT/EP01/03970
;  PRIOR FILING DATE: 2001-04-06
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 202
; SEQ ID NO 133
;   LENGTH: 15951
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-485-133

  Query Match             4.2%;  Score 45.6;  DB 15;  Length 15951;
  Best Local Similarity   45.0%;  Pred. No. 13;
  Matches  171;  Conservative    0;  Mismatches  209;  Indels    0;  Gaps    0;

Qy        143 TTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAG 202
              || | ||| ||  ||  | || |         | | |    || ||   || | || |||
Db       8305 TTTATAAAAATAAATTCTCACTATATTACCCAAACTAACCTAAAACCCCTAAACTCAAAC 8246
```

```
Qy        203 ACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTA 262
              |     ||||  ||  |     |       | |  | | | |  | | |   ||||
Db       8245 AATACTTTTTCTTTAACCTCCAAAAATACTAATATTATAAAAATAAAAATAATACGATAC 8186

Qy        263 TTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCAT 322
              ||  |||  ||||    |   ||  ||  |   | |     | | ||   | | |||
Db       8185 TTTTTTTTTTTTTAAATAAATTACAATCAAAAAAATTACACTACAAAAAAAAAAACTATTA 8126

Qy        323 ATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCA 382
              || | ||   | |||||||||  | |||   | |  || | |||| | | ||  |    |
Db       8125 CAATAATCCCAATAAAAATTAAAAAAAAACAAATTTAATCTAACATCAATAAAAATACAA 8066

Qy        383 GTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATA 442
              | |||   | |||   ||      |   | | |||     |   | |  | |   || |
Db       8065 ATAAAAAAAATAAATATAAAAATTATTAAAAATACAAAACTATACAACTTATTAATAACA 8006

Qy        443 GAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTA 502
              ||   |       |     | ||  || ||||     ||  |     | ||| |  || |
Db       8005 AAACCCAAATATTACAAACTATTACAACTAATACATTAATTCCCTTATATTTCTAAACGA 7946

Qy        503 TTAAAACATTTACTTTCCAA 522
              | | || ||| |   |  ||
Db       7945 TAATAATATTAATAATATAA 7926
```

```
RESULT 93
US-10-221-613-219/c
; Sequence 219, Application US/10221613
; Publication No. US20040029123A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, ALexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with Cell Cycle
;  FILE REFERENCE: 5013.1004
;  CURRENT APPLICATION NUMBER: US/10/221,613
;  CURRENT FILING DATE:  2002-09-13
;  PRIOR APPLICATION NUMBER: PCT/EP01/02945
;    DE 10013847.00
;    DE 10019058.8
;    DE 10019173.8
;    DE 10032529.7
;    DE 10043826.1
;  PRIOR FILING DATE: 2001-03-15
;    2000-03-15
;    2000-04-06
;    2000-04-07
;    2000-06-30
;    2000-09-01
;  NUMBER OF SEQ ID NOS: 428
; SEQ ID NO 219
;   LENGTH: 19576
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (5370)
US-10-221-613-219

  Query Match             4.2%;  Score 45.6;  DB 17;  Length 19576;
  Best Local Similarity   45.7%;  Pred. No. 15;
  Matches  233;  Conservative    0;  Mismatches  274;  Indels    3;  Gaps    2;

Qy        125 TGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTA 184
              | ||  |||| | |       | | | | |  | |    |||    | | || ||   |
Db      13678 TATCAAATATAAATCAAAAACATATACAATTTTAAAATAAAAAACTTAATTAATTAAATT 13619

Qy        185 GAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGAC 244
              || ||  |  |||||| |       |       || |    ||  |   | | |||| |
Db      13618 AAAAATAACCCCTTAAAAATCACCAATCCAATTTAAAACAACTAAACAATCTAAAAAAAT 13559

Qy        245 CTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATG 304
              |||  | | | |  |||  | |  || ||       || |||  ||  || ||| || |
Db      13558 TTTAAAAAAAAATAAACATTTAATAATTATTTAAAAATTTAAATAATTTAAAAAAAACTA 13499

Qy        305 CATTCAGAGAAAGAT--CATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATA 362
              |   | | ||| ||  |||||||  |  || |   |||  |  |  |   | |    | |
Db      13498 AAAAAAAAAAAACATTACATATATAAAACATTCCTAAAAAACTCCTACTCTCAAAATAAA 13439

Qy        363 TTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCT-AAAGGTATACTACCTGC 421
              | || ||  |||     | | |||     | | | ||  | |||   ||   | |
Db      13438 AAAAATCACCAAATAAAATAATAAAAAAACTATACTAAAAATCAAAAAAATTTAATTTTA 13379

Qy        422 TTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAA 481
              | || ||   ||| ||||         |||  |||  | |||||   | |||     | |
Db      13378 ATCCTAATTAAACTACAAAACTCTCAAACCCTAAAACTCATTTTCCTATTATATAAAATA 13319

Qy        482 AACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGTCATTCAGGAG 541
              || | | ||  | |   | ||||| ||   || || | | ||  |    | |
Db      13318 AAAAAATTTAATAAATAATTATTATAATCATTTCTATTCTAAACTAAACTAAAAATAATT 13259

Qy        542 TATGGAAGTATCGATGGCTTCTTTAAAATGAAGCAGGAGGGTCTGGCAGAGAGTATCTAT 601
              |     |  | |        || ||||   |  |              | |     | |||
Db      13258 TCCTTTATCACCACAAATACTTTAAAAAATCACTATATTAAAAAACAAAACTCACTATAT 13199

Qy        602 GAAATAAGTTCCTCTGACCTTCACGCTTAA 631
              |||||  |||||||  |  | |  |||||
Db      13198 TAAATACTTTCCTCTATCTATAAATCTTAA 13169
```

```
RESULT 94
US-10-311-455-1576/c
; Sequence 1576, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determi
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1576
;   LENGTH: 6821
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1576

  Query Match             4.2%;  Score 45.4;  DB 15;  Length 6821;
  Best Local Similarity   48.3%;  Pred. No. 9.9;
  Matches  127;  Conservative    0;  Mismatches  136;  Indels    0;  Gaps    0;

Qy        260 CTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGAT 319
              | | | |||| |||||  ||| ||  | | |  |||| | | ||  |   | | ||  |
Db       2344 CAACTAATTTTTTTTTTTTTTTTTTATATATATATATATATATATAAAAATAAAATCTT 2285

Qy        320 CATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTT 379
              | ||||| | |||    ||  ||||||  |  ||          ||     | | |  |
Db       2284 CCTATATAACCCAAACTAATCTTAAACTCCTAAACTCAAATAATCCACCCAACTTAATCT 2225

Qy        380 GCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAA 439
              |    | |  |||   | |||  |||     ||   |   |  | | ||  |  | |
Db       2224 CCCAAAATACTAAAATTATAAACGTAATCCACTATACCTAACCCTATCATTTCTTTTCTC 2165
```

```
Qy        440 ATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTA 499
              | | ||| || ||     |||||  |  ||| ||| | | | |  | ||     |||
Db       2164 AAAAAAAATAACAATTTATTTATCCTACAAAAAATTTTAAATATTAAAAAAAATTAACCT 2105

Qy        500 CTATTAAAACATTTACTTTCCAA 522
                |   | || ||||    .|||
Db       2104 AAAAATATACTTTTAAAAAACAA 2082


RESULT 95
US-10-221-714A-241/c
; Sequence 241, Application US/10221714A
; Publication No. US20040048254A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with
;  TITLE OF INVENTION:  tumor suppressor genes and oncogenes
;  FILE REFERENCE: 5013.1005
;  CURRENT APPLICATION NUMBER: US/10/221,714A
;  CURRENT FILING DATE:  2003-01-21
;  PRIOR APPLICATION NUMBER: PCT/EP01/02955
;  PRIOR FILING DATE: 2001-03-15
;  PRIOR APPLICATION NUMBER: DE 10013847.0
;  PRIOR FILING DATE: 2000-03-15
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 540
; SEQ ID NO 241
;   LENGTH: 15872
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-714A-241

  Query Match             4.2%;  Score 45.4;  DB 17;  Length 15872;
  Best Local Similarity   46.9%;  Pred. No. 15;
  Matches  142;  Conservative    0;  Mismatches  161;  Indels    0;  Gaps    0;

Qy        189 ATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTA 248
              ||||    | ||| | ||  |||| |  | ||||    |     | || ||||| |  |
Db      15081 ATCTTTAATAAAACATATATTTTTCATTTTCTACTTTTTCACATCTTCTAAAAAATATCC 15022

Qy        249 CAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATT 308
              | | |||  | |  |||  |    ||| || |||||    | ||| |||| |||
Db      15021 ATAAAAAACATAAACTACTTTTATAACATTTTTTTAAAAAATCCCAAAAATAAATCCATA 14962

Qy        309 CAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACAT 368
              |  | |||    | | |||    |    || |  ||| || ||   |   ||    |
Db      14961 CTAATAAAACCAAAAAATTTTTAAAAATAATACAAAAAACCAATCAATAACACCGAATCA 14902

Qy        369 TACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTA 428
              ||||  |    |    |    |||||  | ||| |  |   | |    |
Db      14901 AACTTTTTAAAAAATAATACTAAAAAAACTAAACATCCATAAACATACACCTAAAAAAAT 14842

Qy        429 TCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAA 488
              ||  |||  || |  | | | ||   ||   |    |  |||  |  ||| ||||   ||
Db      14841 TCTAACTTAAACTTAATACTTCCTACAAAAATTAACTCAAAAATAAATCATAAAATTAAA 14782

Qy        489 TTT 491
              | |
Db      14781 TAT 14779


RESULT 96
US-10-240-589C-22
; Sequence 22, Application US/10240589C
; Publication No. US20040076956A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with
;  TITLE OF INVENTION:  DNA repair
;  FILE REFERENCE: 5013.1008
;  CURRENT APPLICATION NUMBER: US/10/240,589C
;  CURRENT FILING DATE:  2003-09-02
;  PRIOR APPLICATION NUMBER: PCT/EP01/03972
;  PRIOR FILING DATE: 2001-04-06
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 148
; SEQ ID NO 22
;   LENGTH: 18060
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-589C-22

  Query Match             4.2%;  Score 45.4;  DB 17;  Length 18060;
  Best Local Similarity   49.8%;  Pred. No. 16;
  Matches  115;  Conservative    0;  Mismatches  116;  Indels    0;  Gaps    0;

Qy        278 TTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGA 337
              || |||  |||||   |||||  | |   | | |    || ||    | ||    ||  ||
Db      15555 TTCGTTTTAAAAATAAATAAAAAGAGTAAATGATTTGAATATATATTTTTTTAAAGGAGA 15614

Qy        338 AAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAG 397
              |  |||| | |||||  ||  ||   |   | |||||  ||   |   |     |||
Db      15615 TATGTAAATAGAAAATAGGTATATGAAAAGATGTTTAATATTATTGGTTATTAGGGAAAT 15674

Qy        398 TCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAAT 457
              ||  ||||  |||| |       ||| | |       |   ||| |  |||   | | |
Db      15675 GTAAATTAAAATTATAATGAAATATTTTTAAGATTTTTTAGAATGGTTATTGTTAAATAA 15734

Qy        458 TTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAA 508
              |   | ||||  |  || |     |||     | ||||| |||||  || | |
Db      15735 TAAGTGTTGGTGAAGATGTGGAGAAATTGGAATTTTTGTGTACGGTTGAGA 15785


RESULT 97
US-10-450-826-33
; Sequence 33, Application US/10450826
; Publication No. US20040101818A1
; GENERAL INFORMATION:
;  APPLICANT: Ji, Darren
;  APPLICANT:  Axelrod, Douglas W.
;  APPLICANT:  Cook, Jonathon S.
;  APPLICANT:  Jaiswal, Neelam
;  APPLICANT:  Eistein, Richard
;  APPLICANT:  Houghton, Adam
;  APPLICANT:  Mertz, Lawrence
;  TITLE OF INVENTION: Gene Expression Profiles Associated with Osteoblast Differentiatic
;  FILE REFERENCE: 044921-5039-WO
;  CURRENT APPLICATION NUMBER: US/10/450,826
;  CURRENT FILING DATE:  2003-06-18
;  PRIOR APPLICATION NUMBER: US 60/255,882
;  PRIOR FILING DATE: 2000-12-18
;  PRIOR APPLICATION NUMBER: US 60/285,691
;  PRIOR FILING DATE: 2001-04-24
;  NUMBER OF SEQ ID NOS: 149
```

```
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 33
;   LENGTH: 159400
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: Genbank Accession No. AL121766
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(159400)
;   OTHER INFORMATION: n = a or c or g or t
US-10-450-826-33

  Query Match             4.2%;  Score 45.4;  DB 18;  Length 159400;
  Best Local Similarity   49.8%;  Pred. No. 46;
  Matches  115;  Conservative    0;  Mismatches  116;  Indels    0;  Gaps    0;

Qy        314 AAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTT 373
              | |  | ||| | |     |   |||  | ||   |  |  |   || ||   ||  | |
Db     149168 ATATTTTATACACTTTTGTGGAAAAATCTTAAATTACTAGCATCTCAAATCTGATGGCAT 149227

Qy        374 AATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCA 433
              |   || | | ||| |  | || |    | ||||  |||| ||| |  | |    |    |
Db     149228 TAAACTGTATTAAAATGGTATAAATTGTCTTAAAATTATATTACATTTTCTAAAAAGTTA 149287

Qy        434 CTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTT 493
              ||   |   |||  |  |  ||      ||||    | |||||| |||| |  | | | |
Db     149288 CTATGATATGAATATTACCAAATACAAATTTCTCCATTAATCTAAGAATATTTCAATATA 149347

Qy        494 TATGTACTATTAAAACATTTACTTTCCAAATATTCTGTCATTCAGGAGTAT 544
              ||  | |   |||  ||||   || |  || ||| | ||||  | ||  ||
Db     149348 TAAATTCATATAATTCATTGTATTACTGAAGATTATATCATAAAAGATGAT 149398


RESULT 98
US-10-021-323-11554/c
; Sequence 11554, Application US/10021323
; Publication No. US20040123340A1
; GENERAL INFORMATION:
;  APPLICANT: Deikman, Jill
;  APPLICANT:  Feng, Paul C.C.
;  APPLICANT:  Fincher, Karen L.
;  APPLICANT:  Ziegler, Todd E.
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(52274)B
;  CURRENT APPLICATION NUMBER: US/10/021,323
;  CURRENT FILING DATE:  2001-12-12
;  PRIOR APPLICATION NUMBER: US 60/255, 619
;  PRIOR FILING DATE: 2000-12-14
;  NUMBER OF SEQ ID NOS: 17880
; SEQ ID NO 11554
;   LENGTH: 337
;   TYPE: DNA
;   ORGANISM: Gossypium hirsutum
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(337)
;   OTHER INFORMATION: unsure at all n locations
;   OTHER INFORMATION: Clone ID: LIB3829-002-Q1-N6-D2
US-10-021-323-11554

  Query Match             4.1%;  Score 45.2;  DB 18;  Length 337;
  Best Local Similarity   54.1%;  Pred. No. 2.5;
  Matches   92;  Conservative    0;  Mismatches   78;  Indels    0;  Gaps    0;

Qy        206 TTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTC 265
              ||| ||| | | | |  || |||  |||||| |||||     |   | |   ||  |||
Db        181 TTGGTTTTTAAGTTTTTCTCTTTTTTTTTTAAAAAGGGACCATTTAAAACTAGGAATTT 122

Qy        266 ATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATA 325
              |||  ||||  ||| | | |||||   | ||| |||    |   | | ||| |  | | |
Db        121 TTTTAGTTTTTTTTTTTCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 62

Qy        326 TTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAA 375
              |   |    |||| ||| | |||| |    ||  || |  ||  ||
Db         61 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATTATAAAAAACCCAA 12


RESULT 99
US-10-357-930-48520
; Sequence 48520, Application US/10357930
; Publication No. US20040259086A1
; GENERAL INFORMATION:
;  APPLICANT: Schlegel. Robert
;  APPLICANT:  Endege, Wilson
;  APPLICANT:  Monahan, John
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
;  TITLE OF INVENTION:  IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF
;  TITLE OF INVENTION:  HUMAN PROSTATE CANCER
;  FILE REFERENCE: MRI-007BCN
;  CURRENT APPLICATION NUMBER: US/10/357,930
;  CURRENT FILING DATE:  2003-02-04
;  PRIOR APPLICATION NUMBER: 09/785,276
;  PRIOR FILING DATE: 2003-02-16
;  PRIOR APPLICATION NUMBER: 60/183,319
;  PRIOR FILING DATE: 2000-02-17
;  PRIOR APPLICATION NUMBER: 60/189,862
;  PRIOR FILING DATE: 2000-03-16
;  PRIOR APPLICATION NUMBER: 60/207,454
;  PRIOR FILING DATE: 2000-05-25
;  PRIOR APPLICATION NUMBER: 60/211,314
;  PRIOR FILING DATE: 2000-06-09
;  PRIOR APPLICATION NUMBER: 60/219,007
;  PRIOR FILING DATE: 2000-07-18
;  PRIOR APPLICATION NUMBER: 60/255,281
;  PRIOR FILING DATE: 2000-12-13
;  NUMBER OF SEQ ID NOS: 62232
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 48520
;   LENGTH: 422
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-357-930-48520

  Query Match             4.1%;  Score 45.2;  DB 18;  Length 422;
  Best Local Similarity   51.0%;  Pred. No. 2.8;
  Matches  107;  Conservative    0;  Mismatches  103;  Indels    0;  Gaps    0;

Qy        285 AAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAA 344
              ||||||  || ||| |||   |   | | ||| |  | ||||  |||| |  | ||   |
Db        159 AAAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATATAGGCCACTAAATAAACCA 218

Qy        345 ACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCT 404
              | ||||  | ||   | || || | | |   ||| ||   ||||  |    |  ||| ||
Db        219 AAACAACTTAAGAAGAAATCACTTAAATACCTCTGGCTTCCAAATTTCTTCATGCAAACT 278

Qy        405 AAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTT 464
              |  |  ||     ||  ||     ||   | ||   | | |||     | ||   || ||
Db        279 ATTGAAATGTATGCTAATTGAACATGCAAATGGGGAAAAAAAAAATGTAGAATGATTCTTG 338

Qy        465 TGGAAATAATCTCAGAAAACATAATTTTTT 494
                  | | | || |||| ||   | |||
Db        339 CTATTTTTAGCACAAAAAAAATGTCTGTTT 368


RESULT 100
US-10-433-793-119/c
; Sequence 119, Application US/10433793
; Publication No. US20040142334A1
; GENERAL INFORMATION:
;  APPLICANT: Epigenomics AG
;  TITLE OF INVENTION: Diagnose von mit Angiogenese assoziierten Krankheiten
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/10/433,793
;  CURRENT FILING DATE:  2003-06-06
```

; NUMBER OF SEQ ID NOS: 212
; SEQ ID NO 119
; LENGTH: 3873
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (34..35, 37, 47..48, 50..51, 65..66, 68, 73..74, 76, 81..82)
; OTHER INFORMATION: unknown base
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (86, 88, 117..118, 127..128, 130..131, 133..134, 3515..)
; OTHER INFORMATION: unknown base
US-10-433-793-119

Query Match 4.1%; Score 45.2; DB 18; Length 3873;
Best Local Similarity 47.4%; Pred. No. 8.3;
Matches 166; Conservative 0; Mismatches 183; Indels 1; Gaps 1;

```
Qy        146 AGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACA 205
              | | || || |  || | || ||| | ||    |   || ||     ||| |   | ||
Db       2067 AAACATCTTAAAAATAAAAACTCAATAAAATAATAAATACAAATATACATTTATTAAACC 2008

Qy        206 TTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTC 265
              ||   |   ||       |  |||  | |  || | |   |||  | |  | |  || | |
Db       2007 TTTAATAAATACCCATTTTACTTCCTCTCATTTATAACTCTTTTACATTTTCTTCTTTAC 1948

Qy        266 ATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATA 325
              ||. |||||   |  |||  || |   || || |||    | |  |  ||  | | | |
Db       1947 TTTATTTTTTATATATTTCTTAATAAACATTAAAAAAATTTTACCCAATAAATTTAAAAA 1888

Qy        326 TTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTC 385
              |  |  || | ||||  |   || | |  | ||  |   ||  |||  |      |
Db       1887 AATTTCTCTATAACAATAAAATCCCTATAACTAAAAATACCCACACA-AATAATAACCAC 1829

Qy        386 AAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAA 445
                |  ||||| |  || | |   |  |        |||  |    | |  |  || ||
Db       1828 TTAACAAAAAACTATACATCACATTTAAACCTTCATTTTTAAACTAAATTATATCTAAAA 1769

Qy        446 ATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTA 495
              | |    ||||||||||||  ||||||   || |||    | | | |||
Db       1768 ACTTTTTCAAATTTTATTTAAAAAATAAAAACAAAAATTTCACTATATTA 1719
```

Search completed: April 22, 2005, 07:12:52
Job time : 497.843 secs

This Page Blank (uspto)

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 20, 2005, 22:49:00 ; Search time 1982.96 Seconds
(without alignments)
18171.562 Million cell updates/sec

Title: US-09-939-209A-3_COPY_4970_11056
Perfect score: 6087
Sequence: 1 ccagtaggaaacagttcttc.....catcatcagaaaggcactat 6087

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_16Dec04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	6087	100.0	20300	6	ABK47337	Abk47337 Genomic n
	2	277.6	4.6	1238	4	AAL26766	Aal26766 Human bre
c	3	277.6	4.6	2752	11	ACN92693	Acn92693 Breast ca
	4	276	4.5	1531	11	ACN88713	Acn88713 Breast ca
	5	274.4	4.5	3165	8	ABX63113	Abx63113 Human cDN
	6	274.4	4.5	3167	10	ADD12660	Add12660 Human cDN
	7	226.4	3.7	3733	13	ADR07336	Adr07336 Full leng
c	8	181	3.0	160198	10	ADL13962	Adl13962 Osteoarth
	9	180.2	3.0	1902	4	AAK94664	Aak94664 Human ful
	10	180.2	3.0	1902	12	ADL31624	Adl31624 Full leng
c	11	173.8	2.9	693	7	ADS31123	Ads31123 Human gen
	12	171	2.8	4366	12	ADQ64492	Adq64492 Novel hum
	13	170.2	2.8	110000	6	ABT00010_10	Continuation (11 o
	14	170.2	2.8	110000	6	ABT01503_10	Continuation (11 o
	15	170.2	2.8	110000	12	ADH77486_10	Continuation (11 o
	16	169.6	2.8	110000	4	AAK95240_10	Continuation (11 o
	17	169.6	2.8	110000	4	AAK96733_10	Continuation (11 o
	18	168.4	2.8	7457	4	AAK86211	Aak86211 Human imm
	19	168	2.8	2769	12	ADP21408	Adp21408 Gene RGS4
	20	167.8	2.8	401	4	AAK96142	Aak96142 Human neu
	21	167.8	2.8	401	4	AAK97635	Aak97635 Human neu
	22	167.8	2.8	401	6	ABT00912	Abt00912 Human neu
	23	167.8	2.8	401	6	ABT02405	Abt02405 Human neu
c	24	167.8	2.8	182328	12	ADL08128	Adl08128 Human gen
	25	166.2	2.7	847	4	AAK91687	Aak91687 Human cDN
	26	166.2	2.7	847	4	AAK93880	Aak93880 Human cDN
	27	166.2	2.7	847	12	ADL30307	Adl30307 3' end of
	28	166.2	2.7	847	12	ADL28114	Adl28114 5' end of
	29	166	2.7	3769	11	ADM02449	Adm02449 Human cDN
	30	165.4	2.7	73742	12	ADQ97824	Adq97824 Human can
	31	163.8	2.7	2238	13	ACN39085	Acn39085 Tumour-as
	32	163.2	2.7	34337	6	ABL55857	Abl55857 Human GAB
	33	163.2	2.7	34337	10	ADG88376	Adg88376 Human tra
	34	162.2	2.7	414	9	ACH30357	Ach30357 Human tes
	35	161.4	2.7	228835	12	ADQ97421	Adq97421 Human can
	36	158.2	2.6	110000	6	ABS55320_2	Continuation (3 of
	37	158.2	2.6	110000	12	ADQ59398_2	Continuation (3 of
	38	158.2	2.6	341511	6	ABS55200	Abs55200 Genomic D
	39	155.2	2.5	84607	2	AAX90847	Aax90847 Human PAC
	40	155	2.5	43445	12	ADH18345	Adh18345 Human apo
	41	155	2.5	43445	12	ADO32886	Ado32886 Human apo
	42	154.8	2.5	16310	3	AAA34964	Aaa34964 Human ade
	43	154.8	2.5	16310	3	AAF21086	Aaf21086 Human low
	44	154.8	2.5	16310	10	ABZ96780	Abz96780 Human nuc
	45	154.8	2.5	16310	11	ABD20629	Abd20629 Human pul

ALIGNMENTS

RESULT 1
ABK47337
ID ABK47337 standard; DNA; 20300 BP.
XX
AC ABK47337;
XX
DT 18-JUN-2002 (first entry)
XX
DE Genomic nucleotide sequence encoding human RGS-4 protein.
XX
KW RGS-4; schizophrenia; human; regulator of G protein signalling 4;
KW neuroleptic; polymorphism; schizotypy; schizoid; chromosome 1q21-22;
KW gene; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT variation replace(4121,T)
FT /*tag= a
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace(4123,A)
FT /*tag= b
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace(4368,C)
FT /*tag= c
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace(4621,C)
FT /*tag= d
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace(4790,T)
FT /*tag= e
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace(4816,T)
FT /*tag= f
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace(4970,T)
FT /*tag= g
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace(5055,G)
FT /*tag= h
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace(5295,A)
FT /*tag= i

```
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(5695,A)
FT                   /*tag= j
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(7375,T)
FT                   /*tag= k
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(7759,A)
FT                   /*tag= l
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(8596,A)
FT                   /*tag= m
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(9602. .9610,CA)
FT                   /*tag= n
FT                   /note= "Deletion polymorphism"
FT   variation       replace(9892,A)
FT                   /*tag= o
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(9963,A)
FT                   /*tag= p
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(10132,A)
FT                   /*tag= q
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(11056,C)
FT                   /*tag= r
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(11091,T)
FT                   /*tag= s
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(11106,A)
FT                   /*tag= t
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(11774,T)
FT                   /*tag= u
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(12143,A)
FT                   /*tag= v
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(12145,T)
FT                   /*tag= w
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(14367,G)
FT                   /*tag= x
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(17027. .17029,GA)
FT                   /*tag= y
FT                   /note= "Deletion polymorphism"
FT   variation       replace(17630,T)
FT                   /*tag= Z
FT                   /standard_name= "Single-nucleotide polymorphism"
XX
PN   WO200216653-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US026622.
XX
PR   24-AUG-2000; 2000US-0228021P.
XX
PA   (UYPI-) UNIV PITTSBURGH.
XX
PI   Levitt PR,  Mirnics K,  Kodavali VC,  Nimgaonkar VL;
XX
DR   WPI; 2002-292070/33.
XX
PT   Diagnosing, assessing susceptibility and treating schizophrenia, involves
PT   observing regulator of G-protein signaling 4, RGS4 levels in a subject.
XX
PS   Claim 1; Page 20-33; 112pp; English.
XX
CC   This invention relates to a novel method for diagnosing schizophrenia or
CC   determining susceptibility to schizophrenia in a human. The method
CC   comprises obtaining from a patient a DNA sample and detecting variations
CC   in the regulator of G-protein signalling 4 (RGS4) gene. Alternatively,
CC   the method involves measuring RGS4 mRNA or protein levels in a tissue
CC   sample from the patient and determining if there is a reduced level. The
CC   method of the invention is useful for diagnosing and determining
CC   susceptibility to schizophrenia. The invention also comprises a method
CC   that is useful for treating schizophrenia which includes a prophylactic
CC   treatment. The method of genotyping polymorphic variants in the RGS-4
CC   gene is applied to diagnosing pathologies of the schizophrenic spectrum,
CC   such as in particular schizotypy, schizoid individuals, etc. This method
CC   offers the possibility of diagnosing schizophrenia by a biological test
CC   and no longer exclusively by clinical evaluations. The present sequence
CC   represents the genomic DNA encoding the human regulator of G-protein
CC   signalling 4 (RGS4) protein used in the method of the invention. The gene
CC   for the RGS4 protein is located on human chromosome 1q21-22
XX
SQ   Sequence 20300 BP; 6157 A; 4102 C; 3775 G; 6266 T; 0 U; 0 Other;

  Query Match            100.0%;  Score 6087;  DB 6;  Length 20300;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 6087;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCAGTAGGAAACAGTTCTTCTAAAGCCTATTGTCAACAATACCTTCCAGATGTTAGCATT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4970 CCAGTAGGAAACAGTTCTTCTAAAGCCTATTGTCAACAATACCTTCCAGATGTTAGCATT 5029

Qy         61 TTACAATTTAAGGAACTTAAAATAGCCTTCAAACTTTTTGCCAGTTTCTCTGATATCCAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5030 TTACAATTTAAGGAACTTAAAATAGCCTTCAAACTTTTTGCCAGTTTCTCTGATATCCAA 5089

Qy        121 TCTATTCTTTTACTCTGCCTCCCAAGCTTTCTTTCTAGAATGCTAACCTGATCGGCTTAA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5090 TCTATTCTTTTACTCTGCCTCCCAAGCTTTCTTTCTAGAATGCTAACCTGATCGGCTTAA 5149

Qy        181 GTACTTGAACTACCTCTTCTCCTCCATTAACTACAGAGTAAATTCTGGTCTTCAGAGTAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5150 GTACTTGAACTACCTCTTCTCCTCCATTAACTACAGAGTAAATTCTGGTCTTCAGAGTAA 5209

Qy        241 CAAGAAACACCCTTTAGTTCTCAGCATATTCGTGCACCTTCATTTATCTCTCCTTCTCTC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5210 CAAGAAACACCCTTTAGTTCTCAGCATATTCGTGCACCTTCATTTATCTCTCCTTCTCTC 5269

Qy        301 TCAAAGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5270 TCAAAGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGC 5329

Qy        361 AACCAAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5330 AACCAAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTA 5389

Qy        421 ATCAAAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5390 ATCAAAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAA 5449

Qy        481 CTGTCTGTTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5450 CTGTCTGTTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTA 5509

Qy        541 GACAAAAGGATATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5510 GACAAAAGGATATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTA 5569

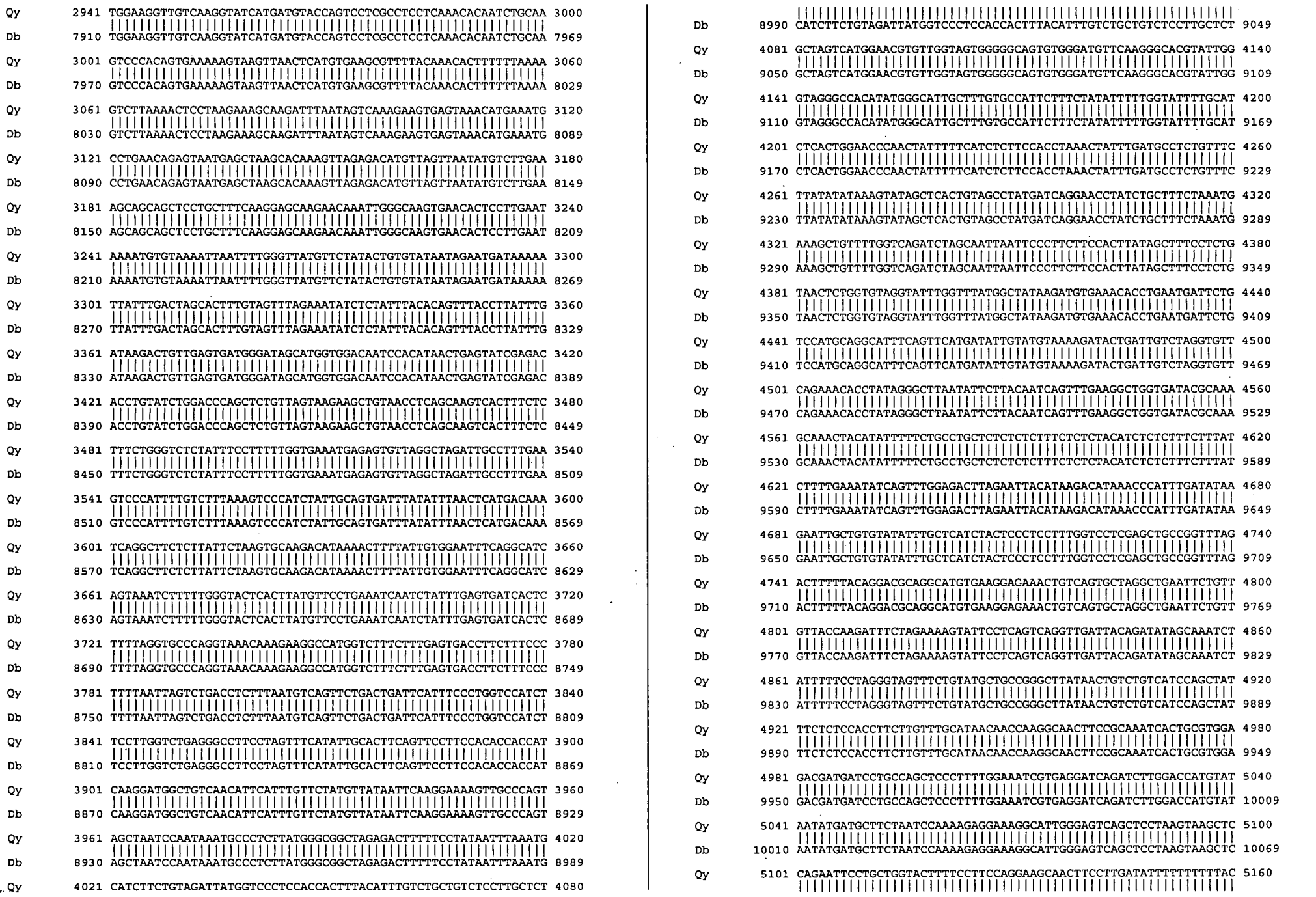
Qy        601 TTCAGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5570 TTCAGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGG 5629

Qy        661 GCAGGACTCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5630 GCAGGACTCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGAT 5689

Qy        721 TTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      5690 TTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGG 5749
Qy       781 AATTCTGGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGG 840
Db      5750 AATTCTGGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGG 5809
Qy       841 AGATGGTCAGAAAGAGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAGT 900
Db      5810 AGATGGTCAGAAAGAGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAGT 5869
Qy       901 ACTTCTTAGCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAACACTGC 960
Db      5870 ACTTCTTAGCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAACACTGC 5929
Qy       961 AATCATGCTAAACTCCTCTATGACCTTCAAACATTCCACTTGCTTTTATTCTTTATGGTT 1020
Db      5930 AATCATGCTAAACTCCTCTATGACCTTCAAACATTCCACTTGCTTTTATTCTTTATGGTT 5989
Qy      1021 GTGATGGCATAGAGGTCAATAGCAAAGACCCTGGAGTCCCACTGTCTGAGCTGGCATAAC 1080
Db      5990 GTGATGGCATAGAGGTCAATAGCAAAGACCCTGGAGTCCCACTGTCTGAGCTGGCATAAC 6049
Qy      1081 ATTACTACCACTTAATCAATGTGTAAGCTCAGGTAAGTACTTAAGTCCTCTATGCTTCAT 1140
Db      6050 ATTACTACCACTTAATCAATGTGTAAGCTCAGGTAAGTACTTAAGTCCTCTATGCTTCAT 6109
Qy      1141 CTGTAAAATGAGAATCATTGAAGAACATTCTCTCAGGATGGATCATGAGGAATAAGTGAA 1200
Db      6110 CTGTAAAATGAGAATCATTGAAGAACATTCTCTCAGGATGGATCATGAGGAATAAGTGAA 6169
Qy      1201 TTAACTGGCATATAGTGCTTAAACCAGTGCCTTGCTCAGTTAGTGACAGATAAAATCATC 1260
Db      6170 TTAACTGGCATATAGTGCTTAAACCAGTGCCTTGCTCAGTTAGTGACAGATAAAATCATC 6229
Qy      1261 TGTTATTACTGTGCCCACTATTGTGATGCTCTTCTCTTCTTTGTACAACGACTACATCTC 1320
Db      6230 TGTTATTACTGTGCCCACTATTGTGATGCTCTTCTCTTCTTTGTACAACGACTACATCTC 6289
Qy      1321 TATTTATCATTTTAGGGTCTCCTTGTGAAAAACCACTCCAGATTCAAAAGATTGAGTTTA 1380
Db      6290 TATTTATCATTTTAGGGTCTCCTTGTGAAAAACCACTCCAGATTCAAAAGATTGAGTTTA 6349
Qy      1381 ATCTCTATCCTCTGTGCTTTCCTGGAGTTTTGTAAAGTAAATCTTCACTTGACATCATGG 1440
Db      6350 ATCTCTATCCTCTGTGCTTTCCTGGAGTTTTGTAAAGTAAATCTTCACTTGACATCATGG 6409
Qy      1441 ATAGGTTCTTGGAAACTACAACTTCAAGTGAAAGGACATAACTAAACCAATTTTTTTCTC 1500
Db      6410 ATAGGTTCTTGGAAACTACAACTTCAAGTGAAAGGACATAACTAAACCAATTTTTTTCTC 6469
Qy      1501 ATCAACGTTATAATGAAATGGCATTGATGAAATGATGGCATTCAAGGACCTGCTGTACCT 1560
Db      6470 ATCAACGTTATAATGAAATGGCATTGATGAAATGATGGCATTCAAGGACCTGCTGTACCT 6529
Qy      1561 TGTTTCACTTAAAGTCACTGTTTCCAATAATCTATTGATGACATTGAGGACTTACTATAT 1620
Db      6530 TGTTTCACTTAAAGTCACTGTTTCCAATAATCTATTGATGACATTGAGGACTTACTATAT 6589
Qy      1621 AATAATAAATATATATATAATCGACGAAACAGGAATCAAACTGCTAACTCTGCTAACTGG 1680
Db      6590 AATAATAAATATATATATAATCGACGAAACAGGAATCAAACTGCTAACTCTGCTAACTGG 6649
Qy      1681 TCTCCCTGCTTCCACACTCTGCCCACTCATCTCAGTCTTTCTTTCACAAGAGTCAGAATG 1740
Db      6650 TCTCCCTGCTTCCACACTCTGCCCACTCATCTCAGTCTTTCTTTCACAAGAGTCAGAATG 6709
Qy      1741 ATCAGATGAGACCCCTCCTCTGCTTCTGTTTCTTCCATGGATTTCCACTGCACTCTGATA 1800
Db      6710 ATCAGATGAGACCCCTCCTCTGCTTCTGTTTCTTCCATGGATTTCCACTGCACTCTGATA 6769
Qy      1801 AAGTCCAGCCTCTTGACCACAGCCTACAAATCCTTGCACGATCTATCGTTTACTTTTCCA 1860
Db      6770 AAGTCCAGCCTCTTGACCACAGCCTACAAATCCTTGCACGATCTATCGTTTACTTTTCCA 6829

Qy      1861 TCTCCTTTTATGCTACTTTCATCTTGTTCTCAATTCTCTAGCTATGCTGGCCCCTTCTTG 1920
Db      6830 TCTCCTTTTATGCTACTTTCATCTTGTTCTCAATTCTCTAGCTATGCTGGCCCCTTCTTG 6889
Qy      1921 TTCTTTCCCATTTTTTTTTAATTTTTAAAATTTGTATATATTTATGGGTTATAAGTGAAA 1980
Db      6890 TTCTTTCCCATTTTTTTTTAATTTTTAAAATTTGTATATATTTATGGGTTATAAGTGAAA 6949
Qy      1981 TCTTTTTAGATGCATAGGTTGTATAGTGATAAAATCAGGGCTTTTAGGGTATTCATCACC 2040
Db      6950 TCTTTTTAGATGCATAGGTTGTATAGTGATAAAATCAGGGCTTTTAGGGTATTCATCACC 7009
Qy      2041 TGAATGATGTACATTGTACCCCTTAAGTAATTTCTCACCATCCGCTGACTTCTTGCCCCC 2100
Db      7010 TGAATGATGTACATTGTACCCCTTAAGTAATTTCTCACCATCCGCTGACTTCTTGCCCCC 7069
Qy      2101 TGGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGC 2160
Db      7070 TGGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGC 7129
Qy      2161 TGACTTCTTGCCCCCTGGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAA 2220
Db      7130 TGACTTCTTGCCCCCTGGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAA 7189
Qy      2221 TTTCTCACCATCCGCTGACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCAT 2280
Db      7190 TTTCTCACCATCCGCTGACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCAT 7249
Qy      2281 TCCACACTCTACATCTATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGC 2340
Db      7250 TCCACACTCTACATCTATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGC 7309
Qy      2341 AATATTTGTCTTTCTGTGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCT 2400
Db      7310 AATATTTGTCTTTCTGTGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCT 7369
Qy      2401 AGGCTGCTGCAAAAGGCATGATTTCATTCTTTTTTATGGCTATGTTCTTTCCCAATTTAG 2460
Db      7370 AGGCTGCTGCAAAAGGCATGATTTCATTCTTTTTTATGGCTATGTTCTTTCCCAATTTAG 7429
Qy      2461 ATAAAGAACACTCGCACTTGCTCTTACTTCTATTTGGAATACTAATTCCTAGGCTTCTTG 2520
Db      7430 ATAAAGAACACTCGCACTTGCTCTTACTTCTATTTGGAATACTAATTCCTAGGCTTCTTG 7489
Qy      2521 CATTGCTTTCTCCTTCTCACCCATCAAATCTCATTTTAGATACCACCTCTTCAAAGAGGG 2580
Db      7490 CATTGCTTTCTCCTTCTCACCCATCAAATCTCATTTTAGATACCACCTCTTCAAAGAGGG 7549
Qy      2581 CTTTCCTGACCACCTTGGCTGAATTAGCCCTTCACCATCTGATTACTCTCTAGCACATCA 2640
Db      7550 CTTTCCTGACCACCTTGGCTGAATTAGCCCTTCACCATCTGATTACTCTCTAGCACATCA 7609
Qy      2641 CCTGCCCATTTTATTCATGGTACAGGTCAAAATCTGGAATCACCTGATTTGTTTATTTTC 2700
Db      7610 CCTGCCCATTTTATTCATGGTACAGGTCAAAATCTGGAATCACCTGATTTGTTTATTTTC 7669
Qy      2701 TGACTCCTTCTACTGAGATGAAAACTCTACTAGAGCGGAGATTTTATCTGCTTGTATCAG 2760
Db      7670 TGACTCCTTCTACTGAGATGAAAACTCTACTAGAGCGGAGATTTTATCTGCTTGTATCAG 7729
Qy      2761 GTACTGCTTCAAACAGCACCTGATACAGAGTAGGTGGTCAAAAGATATTTCTTAAACAAA 2820
Db      7730 GTACTGCTTCAAACAGCACCTGATACAGAGTAGGTGGTCAAAAGATATTTCTTAAACAAA 7789
Qy      2821 TGAACAAATAAAAAGTAGATCTTTTGAGAGTAAAGCTCTTCCACACTACCAGAGTCATTC 2880
Db      7790 TGAACAAATAAAAAGTAGATCTTTTGAGAGTAAAGCTCTTCCACACTACCAGAGTCATTC 7849
Qy      2881 AGGAATGACAAATCATAGAATAACAGAATTTGATGCTTTGTGCATATCAGAGAAAGAAGG 2940
Db      7850 AGGAATGACAAATCATAGAATAACAGAATTTGATGCTTTGTGCATATCAGAGAAAGAAGG 7909
```


Qy 2941 TGGAAGGTTGTCAAGGTATCATGATGTACCAGTCCTCGCCTCCTCAAACACAATCTGCAA 3000
Db 7910 TGGAAGGTTGTCAAGGTATCATGATGTACCAGTCCTCGCCTCCTCAAACACAATCTGCAA 7969
Qy 3001 GTCCCACAGTGAAAAAGTAAGTTAACTCATGTGAAGCGTTTTACAAACACTTTTTTAAAA 3060
Db 7970 GTCCCACAGTGAAAAAGTAAGTTAACTCATGTGAAGCGTTTTACAAACACTTTTTTAAAA 8029
Qy 3061 GTCTTAAAACTCCTAAGAAAGCAAGATTTAATAGTCAAAGAAGTGAGTAAACATGAAATG 3120
Db 8030 GTCTTAAAACTCCTAAGAAAGCAAGATTTAATAGTCAAAGAAGTGAGTAAACATGAAATG 8089
Qy 3121 CCTGAACAGAGTAATGAGCTAAGCACAAAGTTAGAGACATGTTAGTTAATATGTCTTGAA 3180
Db 8090 CCTGAACAGAGTAATGAGCTAAGCACAAAGTTAGAGACATGTTAGTTAATATGTCTTGAA 8149
Qy 3181 AGCAGCAGCTCCTGCTTTCAAGGAGCAAGAACAAATTGGGCAAGTGAACACTCCTTGAAT 3240
Db 8150 AGCAGCAGCTCCTGCTTTCAAGGAGCAAGAACAAATTGGGCAAGTGAACACTCCTTGAAT 8209
Qy 3241 AAAATGTGTAAAATTAATTTTGGGTTATGTTCTATACTGTGTATAATAGAATGATAAAAA 3300
Db 8210 AAAATGTGTAAAATTAATTTTGGGTTATGTTCTATACTGTGTATAATAGAATGATAAAAA 8269
Qy 3301 TTATTTGACTAGCACTTTGTAGTTTAGAAATATCTCTATTTACACAGTTTACCTTATTTG 3360
Db 8270 TTATTTGACTAGCACTTTGTAGTTTAGAAATATCTCTATTTACACAGTTTACCTTATTTG 8329
Qy 3361 ATAAGACTGTTGAGTGATGGGATAGCATGGTGGACAATCCACATAACTGAGTATCGAGAC 3420
Db 8330 ATAAGACTGTTGAGTGATGGGATAGCATGGTGGACAATCCACATAACTGAGTATCGAGAC 8389
Qy 3421 ACCTGTATCTGGACCCAGCTCTGTTAGTAAGAAGCTGTAACCTCAGCAAGTCACTTTCTC 3480
Db 8390 ACCTGTATCTGGACCCAGCTCTGTTAGTAAGAAGCTGTAACCTCAGCAAGTCACTTTCTC 8449
Qy 3481 TTTCTGGGTCTCTATTTCCTTTTTGGTGAAATGAGAGTGTTAGGCTAGATTGCCTTTGAA 3540
Db 8450 TTTCTGGGTCTCTATTTCCTTTTTGGTGAAATGAGAGTGTTAGGCTAGATTGCCTTTGAA 8509
Qy 3541 GTCCCATTTTGTCTTTAAAGTCCCATCTATTGCAGTGATTTATATTTAACTCATGACAAA 3600
Db 8510 GTCCCATTTTGTCTTTAAAGTCCCATCTATTGCAGTGATTTATATTTAACTCATGACAAA 8569
Qy 3601 TCAGGCTTCTCTTATTCTAAGTGCAAGACATAAAACTTTTATTGTGGAATTTCAGGCATC 3660
Db 8570 TCAGGCTTCTCTTATTCTAAGTGCAAGACATAAAACTTTTATTGTGGAATTTCAGGCATC 8629
Qy 3661 AGTAAATCTTTTTGGGTACTCACTTATGTTCCTGAAATCAATCTATTTGAGTGATCACTC 3720
Db 8630 AGTAAATCTTTTTGGGTACTCACTTATGTTCCTGAAATCAATCTATTTGAGTGATCACTC 8689
Qy 3721 TTTTAGGTGCCCAGGTAAACAAAGAAGGCCATGGTCTTTCTTTGAGTGACCTTCTTTCCC 3780
Db 8690 TTTTAGGTGCCCAGGTAAACAAAGAAGGCCATGGTCTTTCTTTGAGTGACCTTCTTTCCC 8749
Qy 3781 TTTTAATTAGTCTGACCTCTTTAATGTCAGTTCTGACTGATTCATTTCCCTGGTCCATCT 3840
Db 8750 TTTTAATTAGTCTGACCTCTTTAATGTCAGTTCTGACTGATTCATTTCCCTGGTCCATCT 8809
Qy 3841 TCCTTGGTCTGAGGGCCTTCCTAGTTTCATATTGCACTTCAGTTCCTTCCACACCACCAT 3900
Db 8810 TCCTTGGTCTGAGGGCCTTCCTAGTTTCATATTGCACTTCAGTTCCTTCCACACCACCAT 8869
Qy 3901 CAAGGATGGCTGTCAACATTCATTTGTTCTATGTTATAATTCAAGGAAAAGTTGCCCAGT 3960
Db 8870 CAAGGATGGCTGTCAACATTCATTTGTTCTATGTTATAATTCAAGGAAAAGTTGCCCAGT 8929
Qy 3961 AGCTAATCCAATAAATGCCCTCTTATGGGCGGCTAGAGACTTTTTCCTATAATTTAAATG 4020
Db 8930 AGCTAATCCAATAAATGCCCTCTTATGGGCGGCTAGAGACTTTTTCCTATAATTTAAATG 8989
Qy 4021 CATCTTCTGTAGATTATGGTCCCTCCACCACTTTACATTTGTCTGCTGTCTCCTTGCTCT 4080
Db 8990 CATCTTCTGTAGATTATGGTCCCTCCACCACTTTACATTTGTCTGCTGTCTCCTTGCTCT 9049
Qy 4081 GCTAGTCATGGAACGTGTTGGTAGTGGGGGCAGTGTGGGATGTTCAAGGGCACGTATTGG 4140
Db 9050 GCTAGTCATGGAACGTGTTGGTAGTGGGGGCAGTGTGGGATGTTCAAGGGCACGTATTGG 9109
Qy 4141 GTAGGGCCACATATGGGCATTGCTTTGTGCCATTCTTTCTATATTTTTGGTATTTTGCAT 4200
Db 9110 GTAGGGCCACATATGGGCATTGCTTTGTGCCATTCTTTCTATATTTTTGGTATTTTGCAT 9169
Qy 4201 CTCACTGGAACCCAACTATTTTTCATCTCTTCCACCTAAACTATTTGATGCCTCTGTTTC 4260
Db 9170 CTCACTGGAACCCAACTATTTTTCATCTCTTCCACCTAAACTATTTGATGCCTCTGTTTC 9229
Qy 4261 TTATATATAAAGTATAGCTCACTGTAGCCTATGATCAGGAACCTATCTGCTTTCTAAATG 4320
Db 9230 TTATATATAAAGTATAGCTCACTGTAGCCTATGATCAGGAACCTATCTGCTTTCTAAATG 9289
Qy 4321 AAAGCTGTTTTGGTCAGATCTAGCAATTAATTCCCTTCTTCCACTTATAGCTTTCCTCTG 4380
Db 9290 AAAGCTGTTTTGGTCAGATCTAGCAATTAATTCCCTTCTTCCACTTATAGCTTTCCTCTG 9349
Qy 4381 TAACTCTGGTGTAGGTATTTGGTTTATGGCTATAAGATGTGAAACACCTGAATGATTCTG 4440
Db 9350 TAACTCTGGTGTAGGTATTTGGTTTATGGCTATAAGATGTGAAACACCTGAATGATTCTG 9409
Qy 4441 TCCATGCAGGCATTTCAGTTCATGATATTGTATGTAAAAGATACTGATTGTCTAGGTGTT 4500
Db 9410 TCCATGCAGGCATTTCAGTTCATGATATTGTATGTAAAAGATACTGATTGTCTAGGTGTT 9469
Qy 4501 CAGAAACACCTATAGGGCTTAATATTCTTACAATCAGTTTGAAGGCTGGTGATACGCAAA 4560
Db 9470 CAGAAACACCTATAGGGCTTAATATTCTTACAATCAGTTTGAAGGCTGGTGATACGCAAA 9529
Qy 4561 GCAAACTACATATTTTTCTGCCTGCTCTCTCTCTTTCTCTCTACATCTCTCTTTCTTTAT 4620
Db 9530 GCAAACTACATATTTTTCTGCCTGCTCTCTCTCTTTCTCTCTACATCTCTCTTTCTTTAT 9589
Qy 4621 CTTTTGAAATATCAGTTTGGAGACTTAGAATTACATAAGACATAAACCCATTTGATATAA 4680
Db 9590 CTTTTGAAATATCAGTTTGGAGACTTAGAATTACATAAGACATAAACCCATTTGATATAA 9649
Qy 4681 GAATTGCTGTGTATATTTGCTCATCTACTCCCTCCTTTGGTCCTCGAGCTGCCGGTTTAG 4740
Db 9650 GAATTGCTGTGTATATTTGCTCATCTACTCCCTCCTTTGGTCCTCGAGCTGCCGGTTTAG 9709
Qy 4741 ACTTTTTACAGGACGCAGGCATGTGAAGGAGAAACTGTCAGTGCTAGGCTGAATTCTGTT 4800
Db 9710 ACTTTTTACAGGACGCAGGCATGTGAAGGAGAAACTGTCAGTGCTAGGCTGAATTCTGTT 9769
Qy 4801 GTTACCAAGATTTCTAGAAAAGTATTCCTCAGTCAGGTTGATTACAGATATAGCAAATCT 4860
Db 9770 GTTACCAAGATTTCTAGAAAAGTATTCCTCAGTCAGGTTGATTACAGATATAGCAAATCT 9829
Qy 4861 ATTTTTCCTAGGGTAGTTTCTGTATGCTGCCGGGCTTATAACTGTCTGTCATCCAGCTAT 4920
Db 9830 ATTTTTCCTAGGGTAGTTTCTGTATGCTGCCGGGCTTATAACTGTCTGTCATCCAGCTAT 9889
Qy 4921 TTCTCTCCACCTTCTTGTTTGCATAACAACCAAGGCAACTTCCGCAAATCACTGCGTGGA 4980
Db 9890 TTCTCTCCACCTTCTTGTTTGCATAACAACCAAGGCAACTTCCGCAAATCACTGCGTGGA 9949
Qy 4981 GACGATGATCCTGCCAGCTCCCTTTTGGAAATCGTGAGGATCAGATCTTGGACCATGTAT 5040
Db 9950 GACGATGATCCTGCCAGCTCCCTTTTGGAAATCGTGAGGATCAGATCTTGGACCATGTAT 10009
Qy 5041 AATATGATGCTTCTAATCCAAAAGAGGAAAGGCATTGGGAGTCAGCTCCTAAGTAAGCTC 5100
Db 10010 AATATGATGCTTCTAATCCAAAAGAGGAAAGGCATTGGGAGTCAGCTCCTAAGTAAGCTC 10069
Qy 5101 CAGAATTCCTGCTGGTACTTTTCCTTCCAGGAAGCAACTTCCTTGATATTTTTTTTTTAC 5160

```
Db      10070 CAGAATTCCTGCTGGTACTTTTCCTTCCAGGAAGCAACTTCCTTGATATTTTTTTTTTAC 10129

Qy       5161 AGGCATATGAATAAAAACTATATTTTGCAGCATTGTACACTTTTTTTCCTTTTCTAGAAA 5220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10130 AGGCATATGAATAAAAACTATATTTTGCAGCATTGTACACTTTTTTTCCTTTTCTAGAAA 10189

Qy       5221 TTCTAAACCTCTGACATTGGTGGAGACATTGAGTACATTTTTTCCCATATCCCTACTTTT 5280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10190 TTCTAAACCTCTGACATTGGTGGAGACATTGAGTACATTTTTTCCCATATCCCTACTTTT 10249

Qy       5281 CAGAAGGATTTTCTCTGCTCGTTCACTTAACATTGCTGATGCGTCAGTCTTTTCTTCCTC 5340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10250 CAGAAGGATTTTCTCTGCTCGTTCACTTAACATTGCTGATGCGTCAGTCTTTTCTTCCTC 10309

Qy       5341 ATCTCTTTCAGGGGCTGGAGAGGCAGAGGGAGACAGAGGAGCTGGTACTGCAGAGCGGTC 5400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10310 ATCTCTTTCAGGGGCTGGAGAGGCAGAGGGAGACAGAGGAGCTGGTACTGCAGAGCGGTC 10369

Qy       5401 GTCTGATTGGCTGGACGGTCGTAGCTGGGCTATAAAAGAGACCCCTACAGGCTTAGCAGG 5460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10370 GTCTGATTGGCTGGACGGTCGTAGCTGGGCTATAAAAGAGACCCCTACAGGCTTAGCAGG 10429

Qy       5461 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 5520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10430 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 10489

Qy       5521 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 5580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10490 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 10549

Qy       5581 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAGGTAAGATTGCTT 5640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10550 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAGGTAAGATTGCTT 10609

Qy       5641 TCAGCCATTAACCATATTAAACTTTTGGCTAGACTTTCTCAGTTATTTACATGTTGTACT 5700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10610 TCAGCCATTAACCATATTAAACTTTTGGCTAGACTTTCTCAGTTATTTACATGTTGTACT 10669

Qy       5701 TACTAACCTAGTTCTGTGCAATTAGAAACAGTGTGGTCAGGAGAGCACGACTTTCTAACT 5760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10670 TACTAACCTAGTTCTGTGCAATTAGAAACAGTGTGGTCAGGAGAGCACGACTTTCTAACT 10729

Qy       5761 TTCCTCCAAGACTAGCTAGATATTGTGACTTAAGACATGTGCTCCCCAAATTTCAGCCCT 5820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10730 TTCCTCCAAGACTAGCTAGATATTGTGACTTAAGACATGTGCTCCCCAAATTTCAGCCCT 10789

Qy       5821 TATGTGTTGTTTTGTGTGACCTCAGTTTTGAGAACTGTTCTATTCTTTAAGCCAGGTCTA 5880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10790 TATGTGTTGTTTTGTGTGACCTCAGTTTTGAGAACTGTTCTATTCTTTAAGCCAGGTCTA 10849

Qy       5881 AGAAAGCTAGTTTTAATTAAGAAGCGAGATGAGGTTTGAGGCTATGTACAGTGATCTGTA 5940
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10850 AGAAAGCTAGTTTTAATTAAGAAGCGAGATGAGGTTTGAGGCTATGTACAGTGATCTGTA 10909

Qy       5941 ATATCTCCATCTGTGATTACTACTGCTATTTGAGCATCCCTGGAGTACATAGAAGCCTGG 6000
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10910 ATATCTCCATCTGTGATTACTACTGCTATTTGAGCATCCCTGGAGTACATAGAAGCCTGG 10969

Qy       6001 CTCTGGGCTTTCTGATTGTATGCTACAACTTGTTTCAGGAAAGGTACCCCAGAATGAGGT 6060
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10970 CTCTGGGCTTTCTGATTGTATGCTACAACTTGTTTCAGGAAAGGTACCCCAGAATGAGGT 11029

Qy       6061 TTGGCTCCATCATCAGAAAGGCACTAT 6087
              |||||||||||||||||||||||||||
Db      11030 TTGGCTCCATCATCAGAAAGGCACTAT 11056
```

RESULT 2
AAL26766
ID AAL26766 standard; cDNA; 1238 BP.
XX
AC AAL26766;
XX
DT 07-DEC-2001 (first entry)
XX
DE Human breast cancer expressed polynucleotide 19223.
XX
KW Human; breast cancer; cell marker; cytostatic; ss.
XX
OS Homo sapiens.
XX
PN WO200151628-A2.
XX
PD 19-JUL-2001.
XX
PF 10-JAN-2001; 2001WO-US000798.
XX
PR 14-JAN-2000; 2000US-0176077P.
PR 14-MAR-2000; 2000US-0189167P.
PR 24-MAR-2000; 2000US-0192099P.
PR 29-MAR-2000; 2000US-0193480P.
PR 15-MAY-2000; 2000US-0205230P.
PR 09-JUN-2000; 2000US-0211315P.
PR 25-JUL-2000; 2000US-0220534P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX
DR WPI; 2001-451856/48.
XX
PT New peptide useful as a marker for the diagnosis of breast cancer.
XX
PS Claim 1; Page 3683-3684; 3695pp; English.
XX
CC The invention relates to human breast cancer expressed polynucleotides
CC (AAL07544-AAL26789) and methods of assessing whether a patient is
CC afflicted with breast cancer by examining the correlation between the
CC expression of certain markers and the cancerous state of breast cells.
CC The polynucleotides and encoded polypeptides are potential markers for
CC detecting, diagnosing, monitoring, characterising treating and
CC potentially preventing breast cancer. The polynucleotides and encoded
CC polypeptides are also useful for isolating compounds with cytostatic
CC activity
XX
SQ Sequence 1238 BP; 386 A; 265 C; 307 G; 274 T; 0 U; 6 Other;

Query Match 4.6%; Score 277.6; DB 4; Length 1238;
Best Local Similarity 98.6%; Pred. No. 3.6e-57;
Matches 280; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
Qy       5345 CTTTCAGGGGCTGGAGAGGCAGAGGGAGACAGAGGAGCTGGTACTGCAGAGCGGTCGTCT 5404
              ||    | ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         58 CTCCTAAGGGCTGGAGAGGCAGAGGGAGACAGAGGAGCTGGTACTGCAGAGCGGTCGTCT 117

Qy       5405 GATTGGCTGGACGGTCGTAGCTGGGCTATAAAAGAGACCCCTACAGGCTTAGCAGGAAGA 5464
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        118 GATTGGCTGGACGGTCGTAGCTGGGCTATAAAAGAGACCCCTACAGGCTTAGCAGGAAGA 177

Qy       5465 CGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGCCGA 5524
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        178 CGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGCCGA 237

Qy       5525 AGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAATAAG 5584
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        238 AGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAATAAG 297

Qy       5585 ATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
              |||||||||||||||||||||||||||||||||||||||||||||
Db        298 ATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 341
```

RESULT 3
ACN92693/c
ID ACN92693 standard; DNA; 2752 BP.
XX

AC ACN92693;
XX
DT 02-DEC-2004 (first entry)
XX
DE Breast cancer related marker, seq id 13843.
XX
KW Cancer; breast; tumour; cytostatic; marker; detection; therapy; ds.
XX
OS Homo sapiens.
XX
PN US2003099974-A1.
XX
PD 29-MAY-2003.
XX
PF 18-JUL-2002; 2002US-00198846.
XX
PR 18-JUL-2001; 2001US-0306220P.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX
DR WPI; 2003-787014/74.
XX
PT Novel isolated polypeptide associated with breast cancer, useful for
PT detecting presence of polypeptide in sample, as a marker for breast
PT cancer.
XX
PS Disclosure; SEQ ID NO 13843; 36pp; English.
XX
CC The invention relates to an isolated polypeptide (I) associated with
CC breast cancer which is encoded by a nucleic acid molecule comprising a
CC nucleotide sequence (S1). Further disclosed is an antibody that binds to
CC the polypeptide of the invention. The activity of the polypeptide of the
CC invention may be described as cytostatic. The antibody is useful for
CC detecting the presence of (I) in a sample. Nucleic acid molecules of the
CC invention are useful in the detection of breast tumours. (I) is useful as
CC a marker for breast cancer and in breast cancer therapy. Sequences given
CC in records ACN78851-ACN92934 represent nucleic acid markers associated
CC with breast cancer. Note: The sequence listing does not form part of the
CC specification but may be obtained in electronic format from the USPTO web
CC site at seqdata.uspto.gov/sequence.html?DocID=20030099974
XX
SQ Sequence 2752 BP; 816 A; 602 C; 553 G; 781 T; 0 U; 0 Other;

Query Match 4.6%; Score 277.6; DB 11; Length 2752;
Best Local Similarity 98.6%; Pred. No. 5.2e-57;
Matches 280; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
Qy       5345 CTTTCAGGGGCTGGAGAGGCAGAGGGAGACAGAGGAGCTGGTACTGCAGAGCGGTCGTCT 5404
             ||   | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2696 CTCCTAAGGGCTGGAGAGGCAGAGGGAGACAGAGGAGCTGGTACTGCAGAGCGGTCGTCT 2637

Qy       5405 GATTGGCTGGACGGTCGTAGCTGGGCTATAAAAGAGACCCCTACAGGCTTAGCAGGAAGA 5464
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2636 GATTGGCTGGACGGTCGTAGCTGGGCTATAAAAGAGACCCCTACAGGCTTAGCAGGAAGA 2577

Qy       5465 CGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGCCGA 5524
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2576 CGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGCCGA 2517

Qy       5525 AGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAATAAG 5584
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2516 AGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAATAAG 2457

Qy       5585 ATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             |||||||||||||||||||||||||||||||||||||||||||||
Db       2456 ATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 2413
```

RESULT 4
ACN88713
ID ACN88713 standard; DNA; 1531 BP.

XX
AC ACN88713;
XX
DT 02-DEC-2004 (first entry)
XX
DE Breast cancer related marker, seq id 9863.
XX
KW Cancer; breast; tumour; cytostatic; marker; detection; therapy; ds.
XX
OS Homo sapiens.
XX
PN US2003099974-A1.
XX
PD 29-MAY-2003.
XX
PF 18-JUL-2002; 2002US-00198846.
XX
PR 18-JUL-2001; 2001US-0306220P.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX
DR WPI; 2003-787014/74.
XX
PT Novel isolated polypeptide associated with breast cancer, useful for
PT detecting presence of polypeptide in sample, as a marker for breast
PT cancer.
XX
PS Disclosure; SEQ ID NO 9863; 36pp; English.
XX
CC The invention relates to an isolated polypeptide (I) associated with
CC breast cancer which is encoded by a nucleic acid molecule comprising a
CC nucleotide sequence (S1). Further disclosed is an antibody that binds to
CC the polypeptide of the invention. The activity of the polypeptide of the
CC invention may be described as cytostatic. The antibody is useful for
CC detecting the presence of (I) in a sample. Nucleic acid molecules of the
CC invention are useful in the detection of breast tumours. (I) is useful as
CC a marker for breast cancer and in breast cancer therapy. Sequences given
CC in records ACN78851-ACN92934 represent nucleic acid markers associated
CC with breast cancer. Note: The sequence listing does not form part of the
CC specification but may be obtained in electronic format from the USPTO web
CC site at seqdata.uspto.gov/sequence.html?DocID=20030099974
XX
SQ Sequence 1531 BP; 439 A; 319 C; 363 G; 405 T; 0 U; 5 Other;

Query Match 4.5%; Score 276; DB 11; Length 1531;
Best Local Similarity 98.2%; Pred. No. 9.7e-57;
Matches 279; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```
Qy       5345 CTTTCAGGGGCTGGAGAGGCAGAGGGAGACAGAGGAGCTGGTACTGCAGAGCGGTCGTCT 5404
             ||   | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         58 CTCCTAAGGGCTGGAGAGGCAGAGGGAGACAGAGGAGCTGGTACTGCAGAGCGGTCGTCT 117

Qy       5405 GATTGGCTGGACGGTCGTAGCTGGGCTATAAAAGAGACCCCTACAGGCTTAGCAGGAAGA 5464
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        118 GATTGGCTGGACGGTCGTAGCTGGGCTATAAAAGAGACCCCTACAGGCTTAGCAGGAACA 177

Qy       5465 CGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGCCGA 5524
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        178 CGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGCCGA 237

Qy       5525 AGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAATAAG 5584
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        238 AGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAATAAG 297

Qy       5585 ATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             |||||||||||||||||||||||||||||||||||||||||||||
Db        298 ATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 341
```

RESULT 5
ABX63113

ID ABX63113 standard; cDNA; 3165 BP.
XX
AC ABX63113;
XX
DT 25-FEB-2003 (first entry)
XX
DE Human cDNA #113 differentially expressed in activated vascular tissue.
XX
KW Human; gene; ss; vascular tissue; cytostatic; atherosclerosis; cardiant;
KW hypotensive; antidiabetic; gynaecological; vasotropic; cerebroprotective;
KW gene therapy; vascular disease; cancer; coronary; artery disease;
KW hypertension; diabetes; pre-eclampsia; restenosis;
KW ischaemia-reperfusion injury; stroke.
XX
OS Homo sapiens.
XX
PN US2002137081-A1.
XX
PD 26-SEP-2002.
XX
PF 08-JAN-2002; 2002US-00044090.
XX
PR 28-JUL-2000; 2000US-0222469P.
PR 08-JAN-2001; 2001US-0260483P.
XX
PA (BAND/) BANDMAN O.
XX
PI Bandman O;
XX
DR WPI; 2003-110597/10.
XX
PT Combination for diagnosing, staging, treating, or monitoring the
PT progression of treatment of a vascular disease, e.g. atherosclerosis,
PT comprises several cDNAs that are differentially expressed in activated
PT vascular tissue.
XX
PS Claim 1; Page; 18pp; English.
XX
CC This invention relates to a combination comprising several cDNAs that are
CC differentially expressed in activated vascular tissue. The invention also
CC discloses a high throughput method for detecting differentially expressed
CC cDNAs in a sample. The cDNAs of the invention may have
CC antiarteriosclerotic; cytostatic; cardiant; hypotensive; antidiabetic;
CC gynaecological; vasotropic and cerebroprotective activities and may be
CC used in gene therapy. The cDNAs of the invention may be used in a high-
CC throughput methods for detecting differential expression of one or more
CC cDNAs in a sample, or screening several molecules or compounds to
CC identify a molecule or compound that specifically binds a cDNA of the
CC invention. A protein encoded by the cDNA may be used to screen several
CC molecules or compounds to identify a ligand that specifically binds to
CC the protein, or to produce or purify an antibody to the protein that can
CC be used to detect a protein in a sample or purify a natural or
CC recombinant protein from a sample. The nucleotides may be useful for
CC diagnosing, staging, treating, or monitoring the progression of treatment
CC of a vascular disease, e.g. atherosclerosis, cancer, coronary artery
CC disease, hypertension, diabetes, pre-eclampsia, ischaemia- reperfusion
CC injury, restenosis, or stroke. The cDNAs can also be used for large-scale
CC genetic or gene expression analysis of several new nucleic acid
CC molecules. Antibodies to the proteins encoded by the cDNAs are useful for
CC diagnosing pre-pathologic disorders, and chronic or acute diseases
CC associated with abnormalities in the expression, amount or distribution
CC of the protein. The present sequence represents a cDNA of the invention
CC that is differentially expressed in activated vascular tissue. Note: The
CC sequence data for this patent did not form part of the specification, but
CC was obtained in electronic format directly from USPTO at
CC http.seqdata.uspto.gov/sequence.html?DocID=20020137081
XX
SQ Sequence 3165 BP; 889 A; 636 C; 672 G; 967 T; 0 U; 1 Other;

Query Match 4.5%; Score 274.4; DB 8; Length 3165;
Best Local Similarity 97.9%; Pred. No. 3.4e-56;
Matches 278; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 5345 CTTTCAGGGGCTGGAGAGGCAGAGGGAGACAGAGGAGCTGGTACTGCAGAGCGGTCGTCT 5404
Db 63 CGTAGAGGGCTCGGAGAGGCAGAGGGAGACAGAGGAGCTGGTACTGCAGAGCGGTCGTCT 122

Qy 5405 GATTGGCTGGACGGTCGTAGCTGGGCTATAAAAGAGACCCCTACAGGCTTAGCAGGAAGA 5464
Db 123 GATTGGCTGGACGGTCGTAGCTGGGCTATAAAAGAGACCCCTACAGGCTTAGCAGGAAGA 182

Qy 5465 CGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGCCGA 5524
Db 183 CGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGCCGA 242

Qy 5525 AGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAATAAG 5584
Db 243 AGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAATAAG 302

Qy 5585 ATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
Db 303 ATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 346

RESULT 6
ADD12660
ID ADD12660 standard; cDNA; 3167 BP.
XX
AC ADD12660;
XX
DT 01-JAN-2004 (first entry)
XX
DE Human cDNA expressed during adipocyte differentiation #6.
XX
KW ss; gene; human; adipocyte differentiation; obesity; type II diabetes;
KW lipodystrophy; hyperinsulinaemia.
XX
OS Homo sapiens.
XX
PN US2003113720-A1.
XX
PD 19-JUN-2003.
XX
PF 30-JUL-2001; 2001US-00918624.
XX
PR 28-JUL-2000; 2000US-0222470P.
XX
PA (SCHE/) SCHEBYE X M.
PA (SORN/) SORNASSE T.
XX
PI Schebye XM, Sornasse T;
XX
DR WPI; 2003-810888/76.
XX
PT Novel isolated cDNAs expressed in adipocyte differentiation useful for
PT treating subject with disorder such as obesity, type II diabetes,
PT lipodystrophy or hyperinsulinemia.
XX
PS Claim 1; SEQ ID NO 6; 105pp; English.
XX
CC The invention relates to an isolated cDNA expressed during adipocyte
CC differentiation. The cDNA is useful for treating a subject with a
CC disorder such as obesity, type II diabetes, lipodystrophy or
CC hyperinsulinaemia. The nucleic acid is useful for a high throughput
CC method of using a cDNA to screen several molecules or compounds to
CC identify a ligand which specifically binds the cDNA which involves
CC combining the nucleic acid with several molecules or compounds under
CC conditions to allow specific binding, and detecting specific binding
CC between each cDNA and at least one molecule or compound, thus identifying
CC a ligand that specifically binds to each cDNA. The several molecules or
CC compounds are chosen from DNA molecules, RNA molecules, peptide nucleic
CC acid molecules, mimetics, peptides, transcription factors, repressors and
CC regulatory proteins. The present sequence represents a human cDNA
CC expressed during adipocyte differentiation.
XX
SQ Sequence 3167 BP; 877 A; 625 C; 671 G; 970 T; 0 U; 24 Other;

```
  Query Match             4.5%;  Score 274.4;  DB 10;  Length 3167;
  Best Local Similarity   97.9%;  Pred. No. 3.4e-56;
  Matches  278;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       5345 CTTTCAGGGGCTGGAGAGGCAGAGGGAGACAGAGGAGCTGGTACTGCAGAGCGGTCGTCT 5404
              | |  ||||   ||||||||||||||||||||||||||||||||||||||||||||||||
Db         63 CGTAGAGGGCTCGGAGAGGCAGAGGGAGACAGAGGAGCTGGTACTGCAGAGCGGTCGTCT 122

Qy       5405 GATTGGCTGGACGGTCGTAGCTGGGCTATAAAAGAGACCCCTACAGGCTTAGCAGGAAGA 5464
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        123 GATTGGCTGGACGGTCGTAGCTGGGCTATAAAAGAGACCCCTACAGGCTTAGCAGGAAGA 182

Qy       5465 CGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGCCGA 5524
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        183 CGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGCCGA 242

Qy       5525 AGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAATAAG 5584
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        243 AGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAATAAG 302

Qy       5585 ATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
              |||||||||||||||||||||||||||||||||||||||||||||
Db        303 ATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 346


RESULT 7
ADR07336
ID   ADR07336 standard; cDNA; 3733 BP.
XX
AC   ADR07336;
XX
DT   04-NOV-2004  (first entry)
XX
DE   Full length human cDNA useful for treating neurological disease Seq 842.
XX
KW   gene; ss; human; oligo-capping method; diagnostic marker; gene therapy;
KW   osteoporosis; neurological disease; Alzheimer's disease;
KW   Parkinson's disease; dementia; short memory; cancer;
KW   sense or motor function; emotional reaction; fear response; panic;
KW   osteopathic; neuroprotective; nootropic; antiparkinsonian; cytostatic;
KW   tranquiliser.
XX
OS   Homo sapiens.
XX
PN   EP1447413-A2.
XX
PD   18-AUG-2004.
XX
PF   12-FEB-2004; 2004EP-00003145.
XX
PR   14-FEB-2003; 2003JP-00102207.
PR   09-MAY-2003; 2003JP-00131452.
XX
PA   (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI   Isogai T,  Yamamoto J,  Nishikawa T,  Isono Y,  Sugiyama T,  Otsuki T;
PI   Wakamatsu A,  Ishii S,  Nagai K,  Irie R;
XX
DR   WPI; 2004-583265/57.
DR   P-PSDB; ADR09292.
XX
PT   New 1995 cDNA, useful for treating osteoporosis, neurological diseases,
PT   Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.
XX
PS   Claim 1; SEQ ID NO 842; 2686pp; English.
XX
CC   This invention relates to novel, isolated full length human cDNA
CC   molecules and the encoded proteins thereof. Specifically, it refers to
CC   cDNA clones obtained by an oligo-capping method, where none of these
CC   clones are identical to any known human mRNAs. The present invention
CC   describes an immunoassay to identify agonists and antagonists, as well as
CC   antibodies, antisense molecules and siRNAs that can all be used to bind
CC   to and modulate expression of the cDNA molecules. As such, these
CC   molecules are useful for diagnostic markers or therapeutic targets for
CC   the various diseases or morbid states. In particular, they are useful in
CC   gene therapy for treating osteoporosis, neurological disease, Alzheimer's
CC   disease, Parkinson's disease, dementia, short memory and various cancers,
CC   as well as for maintaining equilibrium of sense or motor function, and
CC   for treating emotional reaction, fear response and panic. Accordingly,
CC   they exhibit osteopathic, neuroprotective, nootropic, antiparkinsonian,
CC   cytostatic and tranquiliser activities. This polynucleotide is a full
CC   length human cDNA sequence of the invention. NOTE: This sequence is not
CC   given in the sequence listing of the specification but can be obtained on
CC   CD-ROM from the European Patent Office, Vienna Sub-office.
XX
SQ   Sequence 3733 BP; 981 A; 845 C; 1063 G; 844 T; 0 U; 0 Other;

  Query Match             3.7%;  Score 226.4;  DB 13;  Length 3733;
  Best Local Similarity   68.4%;  Pred. No. 2.1e-44;
  Matches  449;  Conservative    0;  Mismatches  161;  Indels   46;  Gaps    8;

Qy        307 CTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAA 366
              | || ||||||| |  || ||||||| | ||| ||   || || |||| |||    ||||
Db        821 CAGCTGTAGGGGCGGGAAGGTGTGATCCCTTTCCTTACCCGTCGTAAGAGTCATGGCCAA 880

Qy        367 AACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAA 426
              |||  ||||  |||||| ||| ||| || || | | | ||||||||||||||||||||
Db        881 CACT--TATAACAAAAGAAAGGCTAACAAAAGAAGAGCGTAACAAATTTATTTAATCAAA 938

Qy        427 GTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCT 486
              ||| ||  |||||||  ||  ||||||| | ||||| ||||||||||||||  ||    |
Db        939 GTTGTAAGTGACATGAGAACCTTCAGAGAAGAAGAACCAAAGACCCAGGG--AAACTGCC 996

Qy        487 GTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAA 546
              |||||  | || ||||| |||||||||| || |||    |   |    || | ||||||
Db        997 TATTTTTATACTTAGGTTTGATGAAGAATAGACAGCCATGAAGAAATGGGACTGGACAAA 1056

Qy        547 AGGATATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGA 606
              || | |   ||                | || ||| ||||||||||||||| ||||||
Db       1057 GGGGTGTAGGCT---------------GAGAGGGAACCCAGCAAGGCCTGTCTGTTCAGA 1101

Qy        607 TTCTTCTTGATCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGA 666
              |||||  ||  |||||        ||| ||||||| ||||||||   ||||||||||
Db       1102 TTCTTTGTGGCCTCTC--------TGTGTAGCATTCCTTCCTCCTAGATATGGGGCAGGA 1153

Qy        667 CTCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATG 726
              ||   ||   |||||||||||||||||||||  || ||  |    |     || | |
Db       1154 CTTCCCTGGCATGAGGGTCTTCAAGGGAGAAGG--GACAGGAGAGAGTAACCTTTCTCGG 1211

Qy        727 GCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCT 786
              |||||| |||  |||||| |||||||||| ||||||  |||||||||||| | ||||
Db       1212 CTTTGCTTTGGGAGAGAGGAATTCTAGTTTCTGTGACCCGCCTTGGGGAAGAGAAGTTCT 1271

Qy        787 GGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGG 846
              || ||||||||||||           | |    | || |||| ||||||||| ||||
Db       1272 GGATTCTGTGACTT-------CTGGGGAGAGAGGGTGAGAGACAGGAGGGCAAGAGAAAT 1324

Qy        847 TCAGAAAGAGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAGTACTTCT 906
              ||| | |        |  |||||||||| |||||  |||||||||||||| |||||   |
Db       1325 TCACAGA-------GCTTGCTTCTGAGGCCTTCCAATCTCCTTTAGTTCAAAGTA---CC 1374

Qy        907 TAGCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAACACTGCAA 962
              ||||  ||||||| |  ||| || |  ||| || |||||||| | | ||| || |
Db       1375 CAGCACGCCAAAGCCCCTTACATTTGGGTATTGTGTTCTGAGCCCCAGCACAGCCA 1430


RESULT 8
ADL13962/c
ID   ADL13962 standard; DNA; 160198 BP.
XX
AC   ADL13962;
XX
DT   06-MAY-2004  (first entry)
XX
DE   Osteoarthritis-associated polymorphic nucleotide #494.
```

```
XX
KW   ds; gene; osteopathic; antiinflammatory; antiarthritic; gene therapy;
KW   joint space narrowing; osteophyte development; joint pain;
KW   osteoarthritis; SNP; single nucleotide polymorphism.
XX
OS   Homo sapiens.
XX
PN   WO2003054166-A2.
XX
PD   03-JUL-2003.
XX
PF   19-DEC-2002; 2002WO-US041225.
XX
PR   20-DEC-2001; 2001US-0342603P.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Jones KA,  Schafer A;
XX
DR   WPI; 2003-559141/52.
XX
PT   Determining susceptibility of an individual to joint space narrowing,
PT   osteophyte development and/or joint pain comprises identifying whether
PT   the individual has at least one polymorphism in a polynucleotide encoding
PT   a protein.
XX
PS   Disclosure; SEQ ID NO 494; 297pp; English.
XX
CC   The invention relates to a method of determining susceptibility of an
CC   individual to joint space narrowing and/or osteophyte development and/or
CC   joint pain comprising identifying whether the individual has at least one
CC   polymorphism in a polynucleotide encoding at least one of the protein
CC   listed in the specification. The methods, composition and agent are
CC   useful for modulating the susceptibility of an individual to joint space
CC   narrowing and/or osteophyte development and/or joint pain that is
CC   associated with a disease, preferably osteoarthritis. The cell line and
CC   the non-human animal are useful for screening for an agent for diagnosing
CC   an individual having susceptibility to joint space narrowing and/or
CC   osteophyte development and/or joint pain. This sequence corresponds to
CC   the polynucleotide encoding a protein listed in the specification. (Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification but was obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences).
XX
SQ   Sequence 160198 BP; 48087 A; 32398 C; 32827 G; 46852 T; 0 U; 34 Other;

  Query Match             3.0%;  Score 181;  DB 10;  Length 160198;
  Best Local Similarity   69.2%;  Pred. No. 1.6e-32;
  Matches  359;  Conservative    0;  Mismatches  130;  Indels   30;  Gaps    7;

Qy        301 TCAAAGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGC 360
              |||  |||    ||||| || ||| | ||||||| ||| || | ||||||||||
Db     156691 TCATTGCTTGTTTAGGGATGGAAAGGCGTGATACCTTTCCT-TACCATCATAAGG----- 156638

Qy        361 AACCAAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTA 420
                |||| |||||||||   |||| |||| ||| || || ||| | | | ||||||||||
Db     156637 --CCAACACTCCTATA-ACAAAGGCAGGCTAACAAAAGAAAAGCATTATAAATTTATTTA 156581

Qy        421 ATCAAAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAA 480
              |||||| |||||  |||||||||||| || |||||||||||||| |||||||||| || |||
Db     156580 ATCAAAATTTTATGTGACATGGGAGCCTACAGAAATGAAGACTCAAAGACCCAAGGAAAA 156521

Qy        481 CTGTCTGTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTA 540
               |||    ||||| || | ||||| |||||||||| ||  ||         |    ||||
Db     156520 TTGT---ATTTTTATGTTTAGGTTTGATGAAGAATAGACGGCCATGTAGAAATGTGATTG 156464

Qy        541 GACAAAAGGATATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTA 600
              ||||||          |  |  | ||  |    |||||  | ||| ||||||||||
Db     156463 GACAAATCATGTGATCTCATGGCAATAGACCAAGTGGGGAGGCCCAGTAAGGCCTGTCTG 156404

Qy        601 TTCAGATTCTTCTTGATCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGG 660
              |  ||||||||||||  ||||||         || || || || ||||| ||| || ||||
Db     156403 TCTAGATTCTTCTTGGCCTCTCT--------GTGTAACACTCCTTCCTTCTGGGTGTGGG 156352

Qy        661 GCAGGACTCTTCTTCAATGAGGGTCTTCAAGGGAGAAG--GGAGAAAGTGGCCTTTTTAG 718
              ||||||||||||  |||||||||| |||||||||||  ||| | || | ||||| |||
Db     156351 GCAGGACTCTTCTGGAATGAGGGTCTGCAAGGGAGAAGGAGGACAGAGAGACCTTTCTAG 156292

Qy        719 ATTTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGA 778
              ||||||  || |||||    || |||||||         ||| | ||    |||| ||||||
Db     156291 GTTTTATAGCGTGCTTTTTGGGAGAGGAG--------TCTGTAACTAGACTTGAGGAAGA 156240

Qy        779 GGAATTCTGGTTTCTGTGACTTGCTTTCATGAAGAAAGA 817
              |||||| |||||   ||  |||  |||  |    |||||
Db     156239 GGAATTTTGGTTCTGGTTGCTTTTTTTTTTTTTTAAAGA 156201


RESULT 9
AAK94664
ID   AAK94664 standard; cDNA; 1902 BP.
XX
AC   AAK94664;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human full-length cDNA, SEQ ID NO: 3657.
XX
KW   Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1130094-A2.
XX
PD   05-SEP-2001.
XX
PF   07-JUL-2000; 2000EP-00114089.
XX
PR   08-JUL-1999;   99JP-00194486.
PR   11-JAN-2000; 2000JP-00118774.
PR   02-MAY-2000; 2000JP-00183765.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Nishikawa T,  Isogai T,  Hayashi K,  Ishii S,  Kawai Y;
PI   Wakamatsu A,  Sugiyama T,  Nagai K,  Kojima S,  Otsuki T,  Koga H;
XX
DR   WPI; 2001-524255/58.
XX
PT   830 Primers useful for synthesizing full length cDNA clones and their use
PT   in genetic manipulation.
XX
PS   Claim 8; SEQ ID NO 3657; 1380pp + Sequence Listing; English.
XX
CC   The invention relates to primers for synthesising full length cDNA
CC   clones. 830 cDNA molecules encoding a human protein have been isolated
CC   and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have
CC   been determined. Primers for synthesising the full length cDNA are useful
CC   for clarifying the function of the protein encoded by the cDNA. The full
CC   length clones were obtained by construction of full length enriched cDNA
CC   libraries that were synthesised by the oligo-capping method. The primers
CC   enable the production of the full length cDNA easily without any special
CC   methods. The present sequence is a full length human cDNA of the
CC   invention. Note: The sequence data for this patent did not form part of
CC   the printed specification, but was obtained in CD-ROM format directly
CC   from EPO
XX
SQ   Sequence 1902 BP; 576 A; 333 C; 440 G; 553 T; 0 U; 0 Other;

  Query Match             3.0%;  Score 180.2;  DB 4;  Length 1902;
  Best Local Similarity   66.8%;  Pred. No. 3.1e-33;
  Matches  404;  Conservative    0;  Mismatches  168;  Indels   33;  Gaps    9;

Qy        385 CAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTACATGACATGGGA 444
              || |||||||| || ||    |||||||     |||||||||| ||||  ||||    ||
Db        213 CAAGTTAATAACAGAAAGGTATAACAAA-----TTAATCAAAGATTTATGTGACCCAAGA 267
```

```
Qy        445 GTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTTTTTTGCTGAGGTT 504
              | ||||| ||||||||||||||||||    |||| || || |||||||||  |     | |||
Db        268 GCCTTCAAAAATGAAGACCCAAAGATGTAGGGAAAGCTATCTGTTTTTCT--GCTAAGTT 325

Qy        505 CGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGATATGATCTAGTGGT 564
                | | ||||||||  |||         |    || | ||||||||||||||||  |  ||
Db        326 TAACGTAGAATGGACAGCCGTCTAGAAATGCGACTGGACAAAAGGATATGATTTGCTGAC 385

Qy        565 AAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTTCTTGATCTCTCTC 624
              ||  |||| || | |   || | | | ||| | || | ||||  ||||||| ||||||
Db        386 AATAGACTGAGAGTGTGGCAAACCGA-GCCAGGCTGTCCAGACACTTCTTGGTCTCTC-- 442

Qy        625 TCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTTCTTCAATGAGGGT 684
                  ||| || ||||  |||||||   |||| |||||   | |  ||  |||||||||
Db        443 ------TGTGTAACATTTCTTCCTCCCTGGTATAGGGCACAGCCCCACTGGAATGAGGGT 496

Qy        685 CTTCAAGGGAGAAGGGAGAAAGTGG--------CCTTTTTAGATTTTATGGCTTGCTTCG 736
              |||||||| |||||||||| | |         | ||   |  ||||||| ||||||  |
Db        497 CTTCAAGGAAGAAGGGAGAAGGGAGAGAGTAACCTTTCAGAAGTTTTATGACTTGCTGTG 556

Qy        737 GGGA-AGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCTGGTTTCTGT 795
              || | |||||  ||| ||| |||||||||   ||||| ||||| ||||| |  || || |
Db        557 GGAAGAGAGGGATTCAAGTGTCTATGACCTGCCTTGGAGAAGACGAATTTTATTTCCTAT 616

Qy        796 GA-CTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGGTCAGAAAG 854
              || ||  || |  | ||||| |||| |  || |||| ||   ||||| || |||| ||
Db        617 GAACTCACTATGGGGGAGAAAGGGGAGCAGAAGACAGGGGGCTGGGAGAAGGGCAGACAG 676

Qy        855 AGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAGTACTTCTTAGCATAC 914
              || ||| ||    | ||||||||||   | ||||||||||| |||   | ||| | || |
Db        677 AGGCTTTGC----TTTGAGGGCTTCTAATTTCCTTTAGTTCAAAG---TCCTTGGGATGC 729

Qy        915 CAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAACACTGCAATCATGCTAAACT 974
              ||||||| |||| |||||  |||  | |||| |||   ||||    ||  |    ||| |
Db        730 CAAAGCATTATATTTTGGAGTATTCTCTTCTAAGCCACAACAAACTAAAAAAATGAAAAT 789

Qy        975 CCTCT 979
              | | |
Db        790 CATTT 794


RESULT 10
ADL31624
ID   ADL31624 standard; cDNA; 1902 BP.
XX
AC   ADL31624;
XX
DT   20-MAY-2004  (first entry)
XX
DE   Full length human cDNA clone SeqID 3657.
XX
KW   human; medicine; signal transduction; glycoprotein; transcription;
KW   oligo-capping method; ss.
XX
OS   Homo sapiens.
XX
PN   EP1396543-A2.
XX
PD   10-MAR-2004.
XX
PF   07-JUL-2000; 2003EP-00025638.
XX
PR   08-JUL-1999;   99JP-00194486.
PR   11-JAN-2000; 2000JP-00118774.
PR   02-MAY-2000; 2000JP-00183865.
PR   07-JUL-2000; 2000EP-00114089.
XX
PA   (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI   Ota T,  Nishikawa T,  Isogai T,  Hayashi K,  Ishii S,  Kawai Y;
PI   Wakamatsu A,  Sugiyama T,  Nagai K,  Kojima S,  Otsuki T,  Koga H;
XX
DR   WPI; 2004-204755/20.
XX
PT   New oligonucleotide primers (830 cDNAs) useful for synthesizing full
PT   length human cDNAs.
XX
PS   Example 1; SEQ ID NO 3657; 1340pp; English.
XX
CC   This invention relates to a novel primers useful for synthesising full
CC   length cDNA molecules that encode human proteins. Specifically, it refers
CC   to secretory or membrane proteins that are potential therapeutic agents/
CC   target molecules in the field of medicine, and in particular genes
CC   encoding proteins that are associated with signal transduction,
CC   glycoproteins and transcription. The present invention describes a method
CC   for efficiently cloning a full length human cDNA from both the 5' and 3'
CC   ends using the oligo-capping method. This polynucleotide sequence is a
CC   full length human cDNA clone of the invention.
XX
SQ   Sequence 1902 BP; 576 A; 333 C; 440 G; 553 T; 0 U; 0 Other;

  Query Match             3.0%;  Score 180.2;  DB 12;  Length 1902;
  Best Local Similarity   66.8%;  Pred. No. 3.1e-33;
  Matches  404;  Conservative    0;  Mismatches  168;  Indels   33;  Gaps    9;

Qy        385 CAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTACATGACATGGGA 444
              || |||||||| || ||    |||||||     |||||||||| ||||  ||||    ||
Db        213 CAAGTTAATAACAGAAAGGTATAACAAA-----TTAATCAAAGATTTATGTGACCCAAGA 267

Qy        445 GTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTTTTTTGCTGAGGTT 504
              | ||||| ||||||||||||||||||    |||| || || |||||||||  |     | |||
Db        268 GCCTTCAAAAATGAAGACCCAAAGATGTAGGGAAAGCTATCTGTTTTTCT--GCTAAGTT 325

Qy        505 CGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGATATGATCTAGTGGT 564
                | | ||||||||  |||         |    || | ||||||||||||||||  |  ||
Db        326 TAACGTAGAATGGACAGCCGTCTAGAAATGCGACTGGACAAAAGGATATGATTTGCTGAC 385

Qy        565 AAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTTCTTGATCTCTCTC 624
              ||  |||| || | |   || | | | ||| | || | ||||  ||||||| ||||||
Db        386 AATAGACTGAGAGTGTGGCAAACCGA-GCCAGGCTGTCCAGACACTTCTTGGTCTCTC-- 442

Qy        625 TCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTTCTTCAATGAGGGT 684
                  ||| || ||||  |||||||   |||| |||||   | |  ||  |||||||||
Db        443 ------TGTGTAACATTTCTTCCTCCCTGGTATAGGGCACAGCCCCACTGGAATGAGGGT 496

Qy        685 CTTCAAGGGAGAAGGGAGAAAGTGG--------CCTTTTTAGATTTTATGGCTTGCTTCG 736
              |||||||| |||||||||| | |         | ||   |  ||||||| ||||||  |
Db        497 CTTCAAGGAAGAAGGGAGAAGGGAGAGAGTAACCTTTCAGAAGTTTTATGACTTGCTGTG 556

Qy        737 GGGA-AGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCTGGTTTCTGT 795
              || | |||||  ||| ||| |||||||||   ||||| ||||| ||||| |  || || |
Db        557 GGAAGAGAGGGATTCAAGTGTCTATGACCTGCCTTGGAGAAGACGAATTTTATTTCCTAT 616

Qy        796 GA-CTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGGTCAGAAAG 854
              || ||  || |  | ||||| |||| |  || |||| ||   ||||| || |||| ||
Db        617 GAACTCACTATGGGGGAGAAAGGGGAGCAGAAGACAGGGGGCTGGGAGAAGGGCAGACAG 676

Qy        855 AGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAGTACTTCTTAGCATAC 914
              || ||| ||    | ||||||||||   | ||||||||||| |||   | ||| | || |
Db        677 AGGCTTTGC----TTTGAGGGCTTCTAATTTCCTTTAGTTCAAAG---TCCTTGGGATGC 729

Qy        915 CAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAACACTGCAATCATGCTAAACT 974
              ||||||| |||| |||||  |||  | |||| |||   ||||    ||  |    ||| |
Db        730 CAAAGCATTATATTTTGGAGTATTCTCTTCTAAGCCACAACAAACTAAAAAAATGAAAAT 789

Qy        975 CCTCT 979
              | | |
Db        790 CATTT 794


RESULT 11
ADS31123/c
ID   ADS31123 standard; DNA; 693 BP.
XX
AC   ADS31123;
XX
```

DT 18-NOV-2004 (first entry)
XX
DE Human genome high complexity repeat found in the HIRA gene #156.
XX
KW Human; ds;
KW histone cell cycle regulation defective, S. cerevisiae homologue A; HIRA;
KW high complexity repeat; in situ hybridisation; Southern blot;
KW chromosome breakpoint; inherited genetic disease; neoplastic disorder;
KW chromosome 22; DiGeorge syndrome; Velo-Cardio-facial syndrome.
XX
OS Homo sapiens.
XX
PN US2003224356-A1.
XX
PD 04-DEC-2003.
XX
PF 14-MAY-2001; 2001US-00854867.
XX
PR 16-MAY-2000; 2000US-00573080.
XX
PA (KNOL/) KNOLL J H M.
PA (ROGA/) ROGAN P K.
XX
PI Knoll JHM, Rogan PK;
XX
DR WPI; 2002-062378/08.
XX
PT Single copy genomic hybridization probes for detecting specific nucleic
PT acid sequences in sample by in situ hybridization useful for detection of
PT acquired or inherited genetic diseases.
XX
PS Example 1; SEQ ID NO 156; 30pp; English.
XX
CC The invention relates to a nucleic acid hybridisation probe comprising a
CC labelled, single copy nucleic acids of at least 50 nucleotides, which
CC will hybridise to a deduced single copy sequence interval in target
CC nucleic acid (TNA) of known sequence. The single copy sequence is deduced
CC by comparing the target nucleic acid (e.g. a disease causing gene) with a
CC collection of high and low complexity repeat sequences as found in the
CC genome of the organism from containing the target nucleic acid. The probe
CC is generated by PCR on the target sequence. The probe is essentially free
CC of blocking nucleic acid sequences which will hybridise to repeat
CC sequences within the genome of which the TNA is a part, and is labelled
CC with a label selected from fluorochrome-responsive labels, fluorochromes,
CC calorimetric chemical, conjugated proteins, antibodies, antigens and
CC their mixtures. The probe is useful in a hybridisation method, where the
CC hybridisation method is from in situ hybridisation, Southern blot, and
CC other methods in which nucleic acid is immobilised, where the method
CC further comprises selecting a single copy nucleic acid which will
CC hybridise to a duplicon or triplicon sequence domain. The probe is useful
CC for determining the existence of previously unknown repeat sequence
CC families in a genome. The method comprises reacting a labelled probe with
CC the genome, causing the probe to hybridise and ascertaining if the probe
CC hybridises to the genome at more than three preferably ten different
CC locations as a determination of new repeat sequence family, where the
CC determining step comprises selecting the single copy sequence from a
CC duplicon or triplicon sequence domain. The probe is useful for
CC determining a chromosome breakpoint and is useful in the fields for
CC cytogenetics and molecular genetics for determining the presence of
CC specific nucleic acid sequences in a sample of eukaryotic origin, e.g.
CC the probes may be used to analyse specific chromosomal locations by in
CC situ hybridisation as a detection of acquired or inherited genetic
CC diseases especially for detection of genetic or neoplastic disorders.
CC Unlike prior art techniques, the probe permits more precise chromosomal
CC breakpoint determinations by in situ hybridisation. The genomic sequence
CC comprising the human HIRA gene (histone cell cycle regulation defective,
CC S. cerevisiae, homologue A) was analysed for single copy sequence
CC intervals for use as probes of the invention. HIRA is located on
CC chromosome 22 as a duplicate, deletions of 1 copy lead to DiGeorge and
CC Velo-Cardio-facial syndromes. The present sequence is a high complexity
CC repeat found within the human genome used to analyse the HIRA gene for
CC repeat regions. Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from USPTO at seqdata.uspto.gov/sequence.html?DocID=20030224356.
XX
SQ Sequence 693 BP; 165 A; 157 C; 127 G; 203 T; 0 U; 41 Other;

Query Match 2.9%; Score 173.8; DB 7; Length 693;
Best Local Similarity 72.8%; Pred. No. 7.1e-32;
Matches 283; Conservative 13; Mismatches 78; Indels 15; Gaps 6;

```
Qy      302 CAAAGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCA 361
            ||      | |||||||   ||| |||||||| | ||  ||   ||||||||||||| |
Db      690 CATTANNNCTGTAGGGGCAGAAAGGTGTGATNCCTTCCCT-ACCCATCATAAGGGTCACG 632

Qy      362 ACCAAAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAA 421
            |  | ||||||||| ||||| |:|||||  ||||| ||| | |||:| ||||||||||
Db      631 GCNNACACTCCTATAATAAAANAYAGGTTAGCAAGAGAAAAGCATAAYAGATTTATTTAA 572

Qy      422 TCAAAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAAC 481
            |||||||||||  |||||||:||| ||||||||||||||:||||:|| ||||| ||||
Db      571 TCAAAGTTTTATGTGACATGRGAGCCTTCAGAAATGAAGACMCAAAKACTCAGGGAAAAC 512

Qy      482 TGTCTGTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAG 541
            | |    ||||| ||:| || || |||||||||||||| |||         |    ||||:|
Db      511 TCT---ATTTTTATGYTTAGATTTGATGAAGAATGGANAGCTGCGCAGAAATGTGATTRG 455

Qy      542 ACAAAA-GGATATGATCTAGTGGTAAAGGACTCAG-GGGGAAACACAGCAAGGCCTGTCT 599
            |||||| || :|||||||| ||||:|  :||| :| | ||:||| |||||||| |||||
Db      454 ACAAAAGGGGWATGATCTAATGGTMATARACTGRGTGTGGRAACCCAGCAAGGCTTGTCT 395

Qy      600 ATTCAGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGG 659
            |||  |||||||||| |||:||        |||   |||||| ||||||| | |||| |
Db      394 GTTCGTATTCTTCTTGGTCTMTC--------TGTGCGGCATTCCTTCCTCCCGGGTATAG 343

Qy      660 GGCAGGA-CTCTTCTTCAATGAGGGTCTT 687
            ||| ||| | |||||  ||||||||||||
Db      342 GGCGGGACCCCTTCTGGAATGAGGGTCTT 314
```

RESULT 12
ADQ64492
ID ADQ64492 standard; cDNA; 4366 BP.
XX
AC ADQ64492;
XX
DT 07-OCT-2004 (first entry)
XX
DE Novel human cDNA sequence #1653.
XX
KW ss; gene; osteopathic; neuroprotective; nootropic; antiparkinsonian;
KW cytostatic; gene therapy; diagnostic marker; morbid state; osteoporosis;
KW neurological disease; Alzheimer's disease; Parkinson's disease; dementia;
KW cancer.
XX
OS Homo sapiens.
XX
PN EP1440981-A2.
XX
PD 28-JUL-2004.
XX
PF 21-JAN-2004; 2004EP-00001196.
XX
PR 21-JAN-2003; 2003JP-00102206.
PR 09-MAY-2003; 2003JP-00131392.
XX
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Nagai K, Irie R;
XX
DR WPI; 2004-535376/52.
DR P-PSDB; ADQ66680.
XX
PT Novel 2495 cDNA, useful for treating osteoporosis, neurological diseases,
PT Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.

```
XX
PS   Claim 1; SEQ ID NO 1653; 2449pp; English.
XX
CC   The invention relates to 2495 novel polynucleotides (I) and their encoded
CC   polypeptides, sequences hybridizing to these nucleotides, sequences
CC   encoding partial polypeptides and sequences having 70% or 90% identity to
CC   the nucleotide and protein sequences. The nucleotides and polypeptides
CC   are useful as diagnostic markers or therapeutic target for the diseases
CC   or morbid states. They are also useful for treating osteoporosis,
CC   neurological diseases, Alzheimer's diseases, Parkinson's diseases,
CC   dementia and various cancers. This sequence corresponds to a nucleotide
CC   sequence of the invention.
XX
SQ   Sequence 4366 BP; 1287 A; 779 C; 946 G; 1354 T; 0 U; 0 Other;

  Query Match             2.8%;  Score 171;  DB 12;  Length 4366;
  Best Local Similarity   64.3%;  Pred. No. 8.1e-31;
  Matches  403;  Conservative    0;  Mismatches  180;  Indels   44;  Gaps    8;

Qy        366 AAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAA 425
              | |||||||||| | ||||||||||||| ||||  ||| |    |||||||||| | |||
Db        544 ACACTCCTATAATGAAAGACAGGTTAACAAGACAAAAGCACGGCAAATTTATTCAGTCAG 603

Qy        426 AGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTC 485
              | |||| | ||||| || ||  ||||  ||||||||| |||||||||| ||||||
Db        604 AATTTTGCGTGACACGGAAGCATTCA-CAATGAAGACTCAAAGACCCAAGGGAAAACTGT 662

Qy        486 TGTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAA 545
                  ||||| |||| || ||| |||| || |||| | |          |    |||| |||||
Db        663 CCATTTTTATGCTTAGATTCAATGAGGAGTGGACAACCATGTAAAAATGGGATTGGACAA 722

Qy        546 AAGGATATGATCTAGTGGTAAAGGACTCAGGGGGA--AACACAGCAAGGCCTGTCTATTC 603
              || |  |  |       ||||  || | | |||||  ||| ||||||||||| |||||
Db        723 AAAGGAAATAATCTAATGTAATAGATTGAAGGGGAACAACCCAGCAAGGCCTGACTATTT 782

Qy        604 AGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCA 663
              |||||||||||  |||||        ||| | |||||| |||||||   |||| | |||
Db        783 GGATTCTTCTTGGCCTCTC--------TGTGTGGCATTCCTTCCTCCCTGGTATAGAGCA 834

Qy        664 GGACTCTTCTTCAATGAGGGTCTTCAAG----------GGAGAAGGGAGAAAGTGGCCTT 713
              |||| | |  |  | | || || | |           || |||| |||     |  ||
Db        835 GGACCCCTTCTGGAATAAGGGTTTTATGATCTACTATCAGACAAGGTAGACCAGAGAATT 894

Qy        714 TTTAGATTTTATGGCTTGCTTC----------------GGGGAAGAGGAGTTCTAGTTT 756
              | |  ||      |      | |               | ||  |  |  | ||||||
Db        895 TCTTTATGGCCAGCTCATATGCAGAAAGGCAATGGAAGTTGAGAGAAATATGTTTAGTTT 954

Qy        757 CTATGACCCATCTTGGGGAAGAGGAATTCTGGTTTCTGTGACTTGCTTTCATGAAGAAAG 816
              ||||||||||   ||||| |||||||||||| ||||| || | ||| ||   | |||| |
Db        955 CTATGACCCACTGTGGGGCAGAGGAATTCTGATTTCTATGGCCTGCCTTAGGGTAGAAGG 1014

Qy        817 AGGAGTAAGAGGCAGGAGGGCAGGAGATGGTCAGAAAGAGACTTGGCTGCTTCTGAGGGC 876
              |||  | ||| |||||||||||||| |  |||| |||||||| | | ||   | |  |
Db       1015 GGGACCAGGAGACAGGAGGGCAGGAGAAGCCCAGAGAGAGACTTTGTTTCTA--GGGTCC 1072

Qy        877 TTCCGCTCTCCTTTAGTTCCAAGTACTTCTTAGCATACCAAAGCACTATACTTTGGCATA 936
              ||||  | |||||  | || |||||||    ||||||  ||||||||  ||||||||   |
Db       1073 TTCCAGTATCCTTCGGCTCAAAGTACTC---AGCATATCAAAGCACCTTACTTTGGGGCA 1129

Qy        937 -TGGTTTTCTGAGCTCTAACACTGCAA 962
               || ||||||||||||  ||  ||| ||
Db       1130 TTGTTTTTCTGAGCTCCAATGCTGTAA 1156


RESULT 13
ABT00010_10
Continuation (11 of 15) of ABT00010 from base 1000001 (Human neuregulin 1 gene. )
WP Sequence split into 15 fragments  LOCUS ABT00010 Accession Abt00010
WP     Fragment Name      Begin     End
WP     ABT00010_00            1    110000
WP     ABT00010_01       100001    210000
WP     ABT00010_02       200001    310000
WP     ABT00010_03       300001    410000
WP     ABT00010_04       400001    510000
WP     ABT00010_05       500001    610000
WP     ABT00010_06       600001    710000
WP     ABT00010_07       700001    810000
WP     ABT00010_08       800001    910000
WP     ABT00010_09       900001   1010000
WP     ABT00010_10      1000001   1110000
WP     ABT00010_11      1100001   1210000
WP     ABT00010_12      1200001   1310000
WP     ABT00010_13      1300001   1410000
WP     ABT00010_14      1400001   1503841

  Query Match             2.8%;  Score 170.2;  DB 6;  Length 110000;
  Best Local Similarity   68.8%;  Pred. No. 5.7e-30;
  Matches  232;  Conservative    1;  Mismatches  104;  Indels    0;  Gaps    0;

Qy       2146 TTTCTCACCATCCGCTGACTTCTTGCCCCCTGGGTATTCATCACCTGAATGATGTGCATT 2205
              ||| | ||        | |||      |    | | |||  ||||||   |||  ||| | ||
Db      13790 TTTGTTACATGGATATAACTCTGGTCTTTCAGTGTAACCATCACTCAAATAGTGTACCTT 13849

Qy       2206 GTACCCCTTAAGTAATTTCTCACCATCCGCTGACTTCTTGCCCCCTCATCCTTCTGAGGC 2265
              |||||  ||| |||| || |||     | |   | ||   ||| | |||| ||:|||| |
Db      13850 GTACCTATTACGTAACTTGTCATTCCTCACCCTCCTCCCACCCTCCCATCTTTYTGAGTC 13909

Qy       2266 TCCATTGTCCATCATTCCACACTCTACATCTATGTGTACACATTATTTAGCTCCTACTTA 2325
              |||| |||| || |||||||| |||||   || |||||||| |||||||||| || ||||||
Db      13910 TCCAATGTCTATTATTCCACCCTCTATGTCCATGTGTACTCATTATTTAGTTCTTACTTA 13969

Qy       2326 TAAGTGATAACATGCAATATTTGTCTTTCTGTGTCTGTCTTGTTTTACTTATGATAATGG 2385
              ||||| | |||||||||||||||| |||||||| ||||  || ||| ||||| |||||| |
Db      13970 TAAGTAAGAACATGCAATATTTGACTTTCTGTTTCTGAGTTATTTCACTTAAGATAATAG 14029

Qy       2386 CCCCCAGTTCTATCTAGGCTGCTGCAAAAGGCATGATTTCATTCTTTTTTATGGCTATGT 2445
              ||  ||| || ||| | | |||||||||| ||||| || |||||||| |||||||   ||
Db      14030 CCTTCAGGTCCATCCATGTTGCTGCAAAAGACATGAATTGATTCTTTTGTATGGCTGAGT 14089

Qy       2446 TCTTTCCCAATTTAGATAAAGAACACTCGCACTTGCT 2482
                | | |||   |  ||| |    ||     |||| |
Db      14090 AGTATTCCATGGTGTATATACGTTACATTTTCTTGAT 14126


RESULT 14
ABT01503_10
Continuation (11 of 15) of ABT01503 from base 1000001 (Human neuregulin 1 gene. )
WP Sequence split into 15 fragments  LOCUS ABT01503 Accession Abt01503
WP     Fragment Name      Begin     End
WP     ABT01503_00            1    110000
WP     ABT01503_01       100001    210000
WP     ABT01503_02       200001    310000
WP     ABT01503_03       300001    410000
WP     ABT01503_04       400001    510000
WP     ABT01503_05       500001    610000
WP     ABT01503_06       600001    710000
WP     ABT01503_07       700001    810000
WP     ABT01503_08       800001    910000
WP     ABT01503_09       900001   1010000
WP     ABT01503_10      1000001   1110000
WP     ABT01503_11      1100001   1210000
WP     ABT01503_12      1200001   1310000
WP     ABT01503_13      1300001   1410000
WP     ABT01503_14      1400001   1503841

  Query Match             2.8%;  Score 170.2;  DB 6;  Length 110000;
  Best Local Similarity   68.8%;  Pred. No. 5.7e-30;
  Matches  232;  Conservative    1;  Mismatches  104;  Indels    0;  Gaps    0;

Qy       2146 TTTCTCACCATCCGCTGACTTCTTGCCCCCTGGGTATTCATCACCTGAATGATGTGCATT 2205
              ||| | ||        | |||      |    | | |||  ||||||   |||  ||| | ||
Db      13790 TTTGTTACATGGATATAACTCTGGTCTTTCAGTGTAACCATCACTCAAATAGTGTACCTT 13849

Qy       2206 GTACCCCTTAAGTAATTTCTCACCATCCGCTGACTTCTTGCCCCCTCATCCTTCTGAGGC 2265
              |||||  ||| |||| || |||     | |   | ||   ||| | |||| ||:|||| |
Db      13850 GTACCTATTACGTAACTTGTCATTCCTCACCCTCCTCCCACCCTCCCATCTTTYTGAGTC 13909
```

```
Qy       2266 TCCATTGTCCATCATTCCACACTCTACATCTATGTGTACACATTATTTAGCTCCTACTTA 2325
             |||| |||| || ||||||| |||||  || |||||||| |||||||||| || ||||||
Db      13910 TCCAATGTCTATTATTCCCACCCTCTATGTCCATGTGTACTCATTATTTAGTTCTTACTTA 13969

Qy       2326 TAAGTGATAACATGCAATATTTGTCTTTCTGTGTCTGTCTTGTTTTACTTATGATAATGG 2385
             ||||| | |||||||||||||||| |||||||| ||||  || ||| ||||| |||||| |
Db      13970 TAAGTAAGAACATGCAATATTTGACTTTCTGTTTCTGAGTTATTTCACTTAAGATAATAG 14029

Qy       2386 CCCCCAGTTCTATCTAGGCTGCTGCAAAAGGCATGATTTCATTCTTTTTTATGGCTATGT 2445
             ||  ||| || ||| | | |||||||||||| ||||| || |||||||| |||||||  ||
Db      14030 CCTTCAGGTCCATCCATGTTGCTGCAAAAGACATGAATTGATTCTTTTGTATGGCTGAGT 14089

Qy       2446 TCTTTCCCAATTTAGATAAAGAACACTCGCACTTGCT 2482
                | | |||   |  ||| |    ||     |||| |
Db      14090 AGTATTCCATGGTGTATATACGTTACATTTTCTTGAT 14126


RESULT 15
ADH77486_10
Continuation (11 of 15) of ADH77486 from base 1000001 (Human neuregulin gene. )
WP Sequence split into 15 fragments  LOCUS ADH77486 Accession Adh77486
WP     Fragment Name      Begin     End
WP     ADH77486_00            1    110000
WP     ADH77486_01       100001    210000
WP     ADH77486_02       200001    310000
WP     ADH77486_03       300001    410000
WP     ADH77486_04       400001    510000
WP     ADH77486_05       500001    610000
WP     ADH77486_06       600001    710000
WP     ADH77486_07       700001    810000
WP     ADH77486_08       800001    910000
WP     ADH77486_09       900001   1010000
WP     ADH77486_10      1000001   1110000
WP     ADH77486_11      1100001   1210000
WP     ADH77486_12      1200001   1310000
WP     ADH77486_13      1300001   1410000
WP     ADH77486_14      1400001   1503841

  Query Match             2.8%;  Score 170.2;  DB 12;  Length 110000;
  Best Local Similarity   68.8%;  Pred. No. 5.7e-30;
  Matches  232;  Conservative    1;  Mismatches  104;  Indels    0;  Gaps    0;

Qy       2146 TTTCTCACCATCCGCTGACTTCTTGCCCCCTGGGTATTCATCACCTGAATGATGTGCATT 2205
             ||| | ||       | |||     |   | | |||  ||||||   |||  ||| | ||
Db      13790 TTTGTTACATGGATATAACTCTGGTCTTTCAGTGTAACCATCACTCAAATAGTGTACCTT 13849

Qy       2206 GTACCCCTTAAGTAATTTCTCACCATCCGCTGACTTCTTGCCCCCTCATCCTTCTGAGGC 2265
             |||||  ||| |||| || |||     | |   | ||   ||| | |||| ||:|||| |
Db      13850 GTACCTATTACGTAACTTGTCATTCCTCACCCTCCTCCCACCCTCCCATCTTTYTGAGTC 13909

Qy       2266 TCCATTGTCCATCATTCCACACTCTACATCTATGTGTACACATTATTTAGCTCCTACTTA 2325
             |||| |||| || ||||||| |||||  || |||||||| |||||||||| || ||||||
Db      13910 TCCAATGTCTATTATTCCACCCTCTATGTCCATGTGTACTCATTATTTAGTTCTTACTTA 13969

Qy       2326 TAAGTGATAACATGCAATATTTGTCTTTCTGTGTCTGTCTTGTTTTACTTATGATAATGG 2385
             ||||| | |||||||||||||||| |||||||| ||||  || ||| ||||| |||||| |
Db      13970 TAAGTAAGAACATGCAATATTTGACTTTCTGTTTCTGAGTTATTTCACTTAAGATAATAG 14029

Qy       2386 CCCCCAGTTCTATCTAGGCTGCTGCAAAAGGCATGATTTCATTCTTTTTTATGGCTATGT 2445
             ||  ||| || ||| | | |||||||||||| ||||| || |||||||| |||||||  ||
Db      14030 CCTTCAGGTCCATCCATGTTGCTGCAAAAGACATGAATTGATTCTTTTGTATGGCTGAGT 14089

Qy       2446 TCTTTCCCAATTTAGATAAAGAACACTCGCACTTGCT 2482
                | | |||   |  ||| |    ||     |||| |
Db      14090 AGTATTCCATGGTGTATATACGTTACATTTTCTTGAT 14126


RESULT 16
AAK95240_10
Continuation (11 of 15) of AAK95240 from base 1000001 (Human neuregulin-1 gene. )
WP Sequence split into 15 fragments  LOCUS AAK95240 Accession Aak95240
WP     Fragment Name      Begin     End
WP     AAK95240_00            1    110000
WP     AAK95240_01       100001    210000
WP     AAK95240_02       200001    310000
WP     AAK95240_03       300001    410000
WP     AAK95240_04       400001    510000
WP     AAK95240_05       500001    610000
WP     AAK95240_06       600001    710000
WP     AAK95240_07       700001    810000
WP     AAK95240_08       800001    910000
WP     AAK95240_09       900001   1010000
WP     AAK95240_10      1000001   1110000
WP     AAK95240_11      1100001   1210000
WP     AAK95240_12      1200001   1310000
WP     AAK95240_13      1300001   1410000
WP     AAK95240_14      1400001   1503900

  Query Match             2.8%;  Score 169.6;  DB 4;  Length 110000;
  Best Local Similarity   68.8%;  Pred. No. 8.1e-30;
  Matches  232;  Conservative    0;  Mismatches  105;  Indels    0;  Gaps    0;

Qy       2146 TTTCTCACCATCCGCTGACTTCTTGCCCCCTGGGTATTCATCACCTGAATGATGTGCATT 2205
             ||| | ||       | |||     |   | | |||  ||||||   |||  ||| | ||
Db      13849 TTTGTTACATGGATATAACTCTGGTCTTTCAGTGTAACCATCACTCAAATAGTGTACCTT 13908

Qy       2206 GTACCCCTTAAGTAATTTCTCACCATCCGCTGACTTCTTGCCCCCTCATCCTTCTGAGGC 2265
             |||||  ||| |||| || |||     | |   | ||   ||| | |||| ||:|||| |
Db      13909 GTACCTATTACGTAACTTGTCATTCCTCACCCTCCTCCCACCCTCCCATCTTTNTGAGTC 13968

Qy       2266 TCCATTGTCCATCATTCCACACTCTACATCTATGTGTACACATTATTTAGCTCCTACTTA 2325
             |||| |||| || ||||||| |||||  || |||||||| |||||||||| || ||||||
Db      13969 TCCAATGTCTATTATTCCACCCTCTATGTCCATGTGTACTCATTATTTAGTTCTTACTTA 14028

Qy       2326 TAAGTGATAACATGCAATATTTGTCTTTCTGTGTCTGTCTTGTTTTACTTATGATAATGG 2385
             ||||| | |||||||||||||||| |||||||| ||||  || ||| ||||| |||||| |
Db      14029 TAAGTAAGAACATGCAATATTTGACTTTCTGTTTCTGAGTTATTTCACTTAAGATAATAG 14088

Qy       2386 CCCCCAGTTCTATCTAGGCTGCTGCAAAAGGCATGATTTCATTCTTTTTTATGGCTATGT 2445
             ||  ||| || ||| | | |||||||||||| ||||| || |||||||| |||||||  ||
Db      14089 CCTTCAGGTCCATCCATGTTGCTGCAAAAGACATGAATTGATTCTTTTGTATGGCTGAGT 14148

Qy       2446 TCTTTCCCAATTTAGATAAAGAACACTCGCACTTGCT 2482
                | | |||   |  ||| |    ||     |||| |
Db      14149 AGTATTCCATGGTGTATATACGTTACATTTTCTTGAT 14185


RESULT 17
AAK96733_10
Continuation (11 of 15) of AAK96733 from base 1000001 (Human neuregulin-1 gene. )
WP Sequence split into 15 fragments  LOCUS AAK96733 Accession Aak96733
WP     Fragment Name      Begin     End
WP     AAK96733_00            1    110000
WP     AAK96733_01       100001    210000
WP     AAK96733_02       200001    310000
WP     AAK96733_03       300001    410000
WP     AAK96733_04       400001    510000
WP     AAK96733_05       500001    610000
WP     AAK96733_06       600001    710000
WP     AAK96733_07       700001    810000
WP     AAK96733_08       800001    910000
WP     AAK96733_09       900001   1010000
WP     AAK96733_10      1000001   1110000
WP     AAK96733_11      1100001   1210000
WP     AAK96733_12      1200001   1310000
WP     AAK96733_13      1300001   1410000
WP     AAK96733_14      1400001   1503900

  Query Match             2.8%;  Score 169.6;  DB 4;  Length 110000;
  Best Local Similarity   68.8%;  Pred. No. 8.1e-30;
  Matches  232;  Conservative    0;  Mismatches  105;  Indels    0;  Gaps    0;

Qy       2146 TTTCTCACCATCCGCTGACTTCTTGCCCCCTGGGTATTCATCACCTGAATGATGTGCATT 2205
             ||| | ||       | |||     |   | | |||  ||||||   |||  ||| | ||
Db      13849 TTTGTTACATGGATATAACTCTGGTCTTTCAGTGTAACCATCACTCAAATAGTGTACCTT 13908
```

```
Qy       2206 GTACCCCTTAAGTAATTTCTCACCATCCGCTGACTTCTTGCCCCCTCATCCTTCTGAGGC 2265
             |||||  ||| |||| || |||     | |   | ||   ||| | |||| || |||| |
Db      13909 GTACCTATTACGTAACTTGTCATTCCTCACCCTCCTCCCACCCTCCCATCTTTNTGAGTC 13968

Qy       2266 TCCATTGTCCATCATTCCACACTCTACATCTATGTGTACACATTATTTAGCTCCTACTTA 2325
             |||| |||| || ||||||| |||||  || |||||||| |||||||||| || ||||||
Db      13969 TCCAATGTCTATTATTCCACCCTCTATGTCCATGTGTACTCATTATTTAGTTCTTACTTA 14028

Qy       2326 TAAGTGATAACATGCAATATTTGTCTTTCTGTGTCTGTCTTGTTTTACTTATGATAATGG 2385
             ||||| | |||||||||||||| |||||||| ||||  || ||| ||||| |||||| |
Db      14029 TAAGTAAGAACATGCAATATTTGACTTTCTGTTTCTGAGTTATTTCACTTAAGATAATAG 14088

Qy       2386 CCCCCAGTTCTATCTAGGCTGCTGCAAAAGGCATGATTTCATTCTTTTTTATGGCTATGT 2445
             ||  ||| || ||| | | |||||||||||| ||||| || |||||||| |||||||  ||
Db      14089 CCTTCAGGTCCATCCATGTTGCTGCAAAAGACATGAATTGATTCTTTTGTATGGCTGAGT 14148

Qy       2446 TCTTTCCCAATTTAGATAAAGAACACTCGCACTTGCT 2482
                | | |||   |  ||| |    ||     |||| |
Db      14149 AGTATTCCATGGTGTATATACGTTACATTTTCTTGAT 14185
```

RESULT 18
AAK86211
ID AAK86211 standard; DNA; 7457 BP.
XX
AC AAK86211;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:41023.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001354.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.

```
PR   08-NOV-2000; 2000US-0246613P.
PR   17-NOV-2000; 2000US-0249207P.
PR   17-NOV-2000; 2000US-0249208P.
PR   17-NOV-2000; 2000US-0249209P.
PR   17-NOV-2000; 2000US-0249210P.
PR   17-NOV-2000; 2000US-0249211P.
PR   17-NOV-2000; 2000US-0249212P.
PR   17-NOV-2000; 2000US-0249213P.
PR   17-NOV-2000; 2000US-0249214P.
PR   17-NOV-2000; 2000US-0249215P.
PR   17-NOV-2000; 2000US-0249216P.
PR   17-NOV-2000; 2000US-0249217P.
PR   17-NOV-2000; 2000US-0249218P.
PR   17-NOV-2000; 2000US-0249244P.
PR   17-NOV-2000; 2000US-0249245P.
PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250160P.
PR   01-DEC-2000; 2000US-0250391P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS   Disclosure; SEQ ID NO 41023; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting the
CC   nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention
XX
SQ   Sequence 7457 BP; 2326 A; 1283 C; 1300 G; 2548 T; 0 U; 0 Other;

  Query Match             2.8%;  Score 168.4;  DB 4;  Length 7457;
  Best Local Similarity   67.2%;  Pred. No. 4.5e-30;
  Matches  238;  Conservative    0;  Mismatches  116;  Indels    0;  Gaps    0;

Qy       2159 GCTGACTTCTTGCCCCCTGGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGT 2218
             | |||  ||  | |  ||||    |      |  ||| || | |||||||||| |||| |
Db       5567 GGTGAAGTCCAGGCTTCTGGTGTATTCATCACCAAATAATATACATTGTACCCATTAAAT 5626

Qy       2219 AATTTCTCACCATCCGCTGACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATC 2278
              ||||||||     ||      | |||  | | |||| ||||| ||| || || |||| ||
Db       5627 AATTTCTCATTCCCCATCCCCCTCTGACTCTCTCACCCTTCCGAGTCTTCAGTGTCTATT 5686

Qy       2279 ATTCCACACTCTACATCTATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACAT 2338
              | |||||||| ||  || |||||||| |||| ||||||||  ||| |||| ||| |||||
Db       5687 ACTCCACACTGTATGTCCATGTGTACCCATTTTTTAGCTCTCACTGATAAATGAGAACAT 5746

Qy       2339 GCAATATTTGTCTTTCTGTGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTAT 2398
              ||| |||||| |||||||| ||||| |||||| |||||||||||| | | ||||||| ||
Db       5747 GCAGTATTTGACTTTCTGTTTCTGTGTTGTTTCACTTATGATAATAGACTCCAGTTCCAT 5806

Qy       2399 CTAGGCTGCTGCAAAAGGCATGATTTCATTCTTTTTTATGGCTATGTTCTTTCCCAATTT 2458
              | | | |||||||||| |||| |||||||| |||||||||   |  | ||   | ||
Db       5807 CCATGTTGCTGCAAAAGACATGGTTTCATTCTCTTTTATGGCTGAATAATATCATGAGTT 5866

Qy       2459 AGATAAAGAACACTCGCACTTGCTCTTACTTCTATTTGGAATACTAATTCCTAG 2512
                |  | |    ||    ||  | |   ||  |  ||  | | |     ||||
Db       5867 CATTTTAAATGCATCAATTTTAATATGTATTTCAAATGTTAAAATGCACACTAG 5920


RESULT 19
ADP21408
ID   ADP21408 standard; DNA; 2769 BP.
XX
AC   ADP21408;
XX
DT   09-SEP-2004  (first entry)
XX
DE   Gene RGS4 for screening for cardiac therapeutic preparation.
XX
KW   ds; cardiant; gene therapy; cardiac therapeutic preparation;
KW   beta-adrenergic receptor antagonist; endothelial receptor antagonist;
KW   calcium channel antagonist; phosphodiesterase inhibitor;
KW   angiotensin converting enzyme inhibitor; heart failure.
XX
OS   Homo sapiens.
XX
PN   WO2004050894-A2.
XX
PD   17-JUN-2004.
XX
PF   26-NOV-2003; 2003WO-US037927.
XX
PR   27-NOV-2002; 2002US-0429379P.
PR   31-DEC-2002; 2002US-0437051P.
PR   31-DEC-2002; 2002US-0437102P.
XX
PA   (ARTE-) ARTESIAN THERAPEUTICS INC.
XX
PI   Bednarik DP;
XX
DR   WPI; 2004-450738/42.
XX
PT   Screening potential therapeutic compounds for cardiac therapeutic
PT   preparations, useful for treating heart failure in a subject, comprises
PT   contacting a sample of cell or tissue with a compound and detecting gene
PT   expression level.
XX
PS   Claim 10; SEQ ID NO 82; 152pp; English.
XX
CC   The invention relates to a method of screening potential therapeutic
CC   compounds for cardiac therapeutic preparations by contacting a sample
CC   comprising a cell or tissue with a potential therapeutic compound and
CC   detecting a level of expression of a gene that codes for a product
CC   encoded by a nucleic acid selected from 98 sequences given in the
CC   specification and its conservative variants, comparing the level of the
CC   expression of the gene to the level of expression of the gene in the
CC   absence of the compound, and identifying a potential therapeutic compound
CC   for use as a cardiac therapeutic preparation if the potential therapeutic
CC   compound affects the level of expression of the gene. The pharmaceutical
CC   composition or active agents (i.e., beta-adrenergic receptor antagonist,
CC   endothelial receptor antagonist, calcium channel antagonist,
```

```
CC   phosphodiesterase inhibitor, or angiotensin converting enzyme inhibitor)
CC   is useful for treating heart failure in a subject. This sequence
CC   represents one of the genes whose expression may be altered by the
CC   compounds detected by the screen of the invention.
XX
SQ   Sequence 2769 BP; 783 A; 555 C; 576 G; 855 T; 0 U; 0 Other;

  Query Match             2.8%;  Score 168;  DB 12;  Length 2769;
  Best Local Similarity   100.0%;  Pred. No. 3.6e-30;
  Matches  168;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5461 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 5520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAG 60

Qy       5521 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 5580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCGAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAA 120

Qy       5581 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             ||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TAAGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 168


RESULT 20
AAK96142
ID   AAK96142 standard; DNA; 401 BP.
XX
AC   AAK96142;
XX
DT   17-DEC-2001  (first entry)
XX
DE   Human neuregulin gene single nucleotide polymorphism SNP8NRG1013903.
XX
KW   Human; neuregulin-1 associated gene 1; NRG1AG1; Schizophrenia gene;
KW   gene therapy; single nucleotide polymorphism; SNP; ds.
XX
OS   Homo sapiens.
XX
PN   WO200164876-A2.
XX
PD   07-SEP-2001.
XX
PF   28-FEB-2001; 2001WO-US006376.
XX
PR   28-FEB-2000; 2000US-00515715.
XX
PA   (DECO-) DECODE GENETICS EHF.
XX
PI   Stefansson H,  Steinthorsdottir V,  Gulcher JR;
XX
DR   WPI; 2001-550179/61.
XX
PT   Neuregulin-1 associated gene 1 nucleic acids and fragments, useful for
PT   preventing diagnosing and treating schizophrenia.
XX
PS   Disclosure; Page 659; 750pp; English.
XX
CC   This sequence represents a single nucleotide polymorphism (SNP) of the
CC   human neuregulin-1 associated gene 1 (NRG1AG1) of the invention. The
CC   NRG1AG1 gene is also referred to as the human Schizophrenia gene. The
CC   invention also relates to fragments or variants of the gene and the
CC   NRG1AG1 polypeptides they encode. The NRG1AG1 nucleic acids and
CC   polypeptides may be used in the prevention, diagnosis and treatment of
CC   diseases associated with inappropriate NRG1AG1 expression. For example,
CC   they may be used to treat disorders associated with decreased expression
CC   by rectifying mutations or deletions in a patient's genome that affect
CC   the activity of NRG1AG1 by expressing inactive proteins or to supplement
CC   the patients own production of NRG1AG1. Additionally, the gene may be
CC   used to produce NRG1AG1 polypeptides, by inserting the nucleic acids into
CC   a host cell and culturing the cell to express the protein. The gene may
CC   also be used as DNA probes and primers in diagnostic assays to detect and
CC   quantitate the presence of similar nucleic acids in samples, and
CC   therefore which patients may be in need of restorative therapy. The
CC   NRG1AG1 polypeptides may also be used as antigens in the production of
CC   antibodies against NRG1AG1 and in assays to identify modulators of
CC   NRG1AG1 expression and activity. Anti-NRG1AG1 antibodies and antagonists
CC   may also be used to down regulate expression and activity. Anti-NRG1AG1
CC   antibodies may also be used as diagnostic agents for detecting the
CC   presence of NRG1AG1 polypeptides in samples. NRG1AG1 Is associated with
CC   schizophrenia which may be prevented, diagnosed and/or treated by the
CC   above methods
XX
SQ   Sequence 401 BP; 98 A; 79 C; 62 G; 161 T; 0 U; 1 Other;

  Query Match             2.8%;  Score 167.8;  DB 4;  Length 401;
  Best Local Similarity   71.2%;  Pred. No. 1.6e-30;
  Matches  220;  Conservative    1;  Mismatches   88;  Indels    0;  Gaps    0;

Qy       2146 TTTCTCACCATCCGCTGACTTCTTGCCCCCTGGGTATTCATCACCTGAATGATGTGCATT 2205
             ||| | ||       | |||     |   | | |||  ||||||   |||  ||| | ||
Db         88 TTTGTTACATGGATATAACTCTGGTCTTTCAGTGTAACCATCACTCAAATAGTGTACCTT 147

Qy       2206 GTACCCCTTAAGTAATTTCTCACCATCCGCTGACTTCTTGCCCCCTCATCCTTCTGAGGC 2265
             |||||  ||| |||| || |||     | |   | ||   ||| | |||| ||:|||| |
Db        148 GTACCTATTACGTAACTTGTCATTCCTCACCCTCCTCCCACCCTCCCATCTTTYTGAGTC 207

Qy       2266 TCCATTGTCCATCATTCCACACTCTACATCTATGTGTACACATTATTTAGCTCCTACTTA 2325
             |||| |||| || |||||| |||||  || |||||||| |||||||||| || ||||||
Db        208 TCCAATGTCTATTATTCCACCCTCTATGTCCATGTGTACTCATTATTTAGTTCTTACTTA 267

Qy       2326 TAAGTGATAACATGCAATATTTGTCTTTCTGTGTCTGTCTTGTTTTACTTATGATAATGG 2385
             ||||| | ||||||||||||||| |||||||| ||||  || ||| ||||| |||||| |
Db        268 TAAGTAAGAACATGCAATATTTGACTTTCTGTTTCTGAGTTATTTCACTTAAGATAATAG 327

Qy       2386 CCCCCAGTTCTATCTAGGCTGCTGCAAAAGGCATGATTTCATTCTTTTTTATGGCTATGT 2445
             ||  ||| || ||| | | |||||||||||| ||||| || |||||||| |||||||  ||
Db        328 CCTTCAGGTCCATCCATGTTGCTGCAAAAGACATGAATTGATTCTTTTGTATGGCTGAGT 387

Qy       2446 TCTTTCCCA 2454
                | | |||
Db        388 AGTATTCCA 396


RESULT 21
AAK97635
ID   AAK97635 standard; DNA; 401 BP.
XX
AC   AAK97635;
XX
DT   17-DEC-2001  (first entry)
XX
DE   Human neuregulin gene single nucleotide polymorphism SNP8NRG1013903.
XX
KW   Human; neuregulin 1 gene; schizophrenia; gene therapy; SNP;
KW   single nucleotide polymorphism; ds.
XX
OS   Homo sapiens.
XX
PN   WO200164877-A2.
XX
PD   07-SEP-2001.
XX
PF   28-FEB-2001; 2001WO-US006377.
XX
PR   28-FEB-2000; 2000US-00515716.
XX
PA   (DECO-) DECODE GENETICS EHF.
XX
PI   Stefansson H,  Steinthorsdottir V,  Gulcher JR;
XX
DR   WPI; 2001-514841/56.
XX
PT   Neuregulin 1 nucleic acids and proteins useful for diagnosing preventing
PT   and treating schizophrenia.
XX
```

PS Disclosure; Page 244; 756pp; English.
XX
CC This sequence represents a single nucleotide polymorphism (SNP) from the
CC human neuregulin 1 gene of the invention. The invention also relates to
CC fragments or variants of the neuregulin 1 gene. The gene and its proteins
CC may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate neuregulin 1 expression, such as
CC schizophrenia. For example they may be used to treat disorders associated
CC with decreased neuregulin 1 expression by rectifying mutations or
CC deletions in a patient's genome that affect the activity of neuregulin 1
CC by expressing inactive proteins or to supplement the patients own
CC production of polypeptides. Additionally, the gene may be used to produce
CC the neuregulin 1 protein, by inserting the nucleic acids into a host cell
CC and culturing the cell to express the protein. The gene and its
CC complementary sequences may also be used as DNA probes in diagnostic
CC assays to detect and quantitate the presence of similar nucleic acids in
CC samples, and therefore which patients may be in need of restorative
CC therapy. The protein may also be used as antigens in the production of
CC antibodies against neuregulin 1 and in assays to identify modulators of
CC neuregulin 1 expression and activity. The antibodies and antagonists may
CC also be used to down regulate expression and activity. The antibodies may
CC also be used as diagnostic agents for detecting the presence of
CC neuregulin 1 in samples
XX
SQ Sequence 401 BP; 98 A; 79 C; 62 G; 161 T; 0 U; 1 Other;

Query Match 2.8%; Score 167.8; DB 4; Length 401;
Best Local Similarity 71.2%; Pred. No. 1.6e-30;
Matches 220; Conservative 1; Mismatches 88; Indels 0; Gaps 0;

```
Qy       2146 TTTCTCACCATCCGCTGACTTCTTGCCCCCTGGGTATTCATCACCTGAATGATGTGCATT 2205
             ||| | ||       | |||     |   | | |||  ||||||  |||  ||| | ||
Db         88 TTTGTTACATGGATATAACTCTGGTCTTTCAGTGTAACCATCACTCAAATAGTGTACCTT 147

Qy       2206 GTACCCCTTAAGTAATTTCTCACCATCCGCTGACTTCTTGCCCCCTCATCCTTCTGAGGC 2265
             |||||  ||| |||| || |||     | |   | ||   ||| | |||| ||:|||| |
Db        148 GTACCTATTACGTAACTTGTCATTCCTCACCCTCCTCCCACCCTCCCATCTTTYTGAGTC 207

Qy       2266 TCCATTGTCCATCATTCCACACTCTACATCTATGTGTACACATTATTTAGCTCCTACTTA 2325
             |||| |||| || |||||| |||||  || ||||||| |||||||||| || ||||||
Db        208 TCCAATGTCTATTATTCCACCCTCTATGTCCATGTGTACTCATTATTTAGTTCTTACTTA 267

Qy       2326 TAAGTGATAACATGCAATATTTGTCTTTCTGTGTCTGTCTTGTTTTACTTATGATAATGG 2385
             ||||| | |||||||||||||| |||||||| ||||  || ||| |||||  |||||| |
Db        268 TAAGTAAGAACATGCAATATTTGACTTTCTGTTTCTGAGTTATTTCACTTAAGATAATAG 327

Qy       2386 CCCCCAGTTCTATCTAGGCTGCTGCAAAAGGCATGATTTCATTCTTTTTTATGGCTATGT 2445
             ||  ||| || ||| | | |||||||||||| ||||| || |||||||| |||||||  ||
Db        328 CCTTCAGGTCCATCCATGTTGCTGCAAAAGACATGAATTGATTCTTTTGTATGGCTGAGT 387

Qy       2446 TCTTTCCCA 2454
                | | |||
Db        388 AGTATTCCA 396
```

RESULT 22
ABT00912
ID ABT00912 standard; DNA; 401 BP.
XX
AC ABT00912;
XX
DT 07-NOV-2002 (first entry)
XX
DE Human neuregulin-1-associated gene 1 fragment SEQ ID NO: 941.
XX
KW Human; neuregulin 1; neuregulin-1-associated gene 1; NRG1; NRG1AG1;
KW schizophrenia; chromosome 8p12; single nucleotide polymorphism; SNP;
KW neuroleptic; gene therapy; splice variant; gene; ds.
XX
OS Homo sapiens.
XX
PN US2002045577-A1.
XX
PD 18-APR-2002.
XX
PF 28-FEB-2001; 2001US-00795668.
XX
PR 28-FEB-2000; 2000US-00515716.
XX
PA (DECO-) DECODE GENETICS EHF.
XX
PI Stefansson H, Steinthorsdottir V, Gulcher JR;
XX
DR WPI; 2002-425447/45.
XX
PT New neuregulin 1 gene, schizophrenia gene residing in 1.5 Mb segment on
PT human chromosome 8p12, useful for diagnosing susceptibility to or
PT treating schizophrenia and for screening schizophrenia treating agents.
XX
PS Disclosure; Page 600; 700pp; English.
XX
CC The present invention relates to the human neuregulin 1 gene (NRG1),
CC single nucleotide polymorphisms within which were identified as being
CC associated with an increased susceptibility to schizophrenia, which is
CC located on chromosome 8p12. Also found within the same sequence is the
CC neuregulin-1-associated gene 1 (NRG1AG1). The gene is useful for treating
CC schizophrenia in an individual, for diagnosing susceptibility to
CC schizophrenia, and for screening for agents useful in the treatment of
CC the disease. The present sequence is a fragment of the NRG1AG1 gene of
CC the invention containing a polymorphic site
XX
SQ Sequence 401 BP; 98 A; 79 C; 62 G; 161 T; 0 U; 1 Other;

Query Match 2.8%; Score 167.8; DB 6; Length 401;
Best Local Similarity 71.2%; Pred. No. 1.6e-30;
Matches 220; Conservative 1; Mismatches 88; Indels 0; Gaps 0;

```
Qy       2146 TTTCTCACCATCCGCTGACTTCTTGCCCCCTGGGTATTCATCACCTGAATGATGTGCATT 2205
             ||| | ||       | |||     |   | | |||  ||||||  |||  ||| | ||
Db         88 TTTGTTACATGGATATAACTCTGGTCTTTCAGTGTAACCATCACTCAAATAGTGTACCTT 147

Qy       2206 GTACCCCTTAAGTAATTTCTCACCATCCGCTGACTTCTTGCCCCCTCATCCTTCTGAGGC 2265
             |||||  ||| |||| || |||     | |   | ||   ||| | |||| ||:|||| |
Db        148 GTACCTATTACGTAACTTGTCATTCCTCACCCTCCTCCCACCCTCCCATCTTTYTGAGTC 207

Qy       2266 TCCATTGTCCATCATTCCACACTCTACATCTATGTGTACACATTATTTAGCTCCTACTTA 2325
             |||| |||| || |||||| |||||  || ||||||| |||||||||| || ||||||
Db        208 TCCAATGTCTATTATTCCACCCTCTATGTCCATGTGTACTCATTATTTAGTTCTTACTTA 267

Qy       2326 TAAGTGATAACATGCAATATTTGTCTTTCTGTGTCTGTCTTGTTTTACTTATGATAATGG 2385
             ||||| | |||||||||||||| |||||||| ||||  || ||| |||||  |||||| |
Db        268 TAAGTAAGAACATGCAATATTTGACTTTCTGTTTCTGAGTTATTTCACTTAAGATAATAG 327

Qy       2386 CCCCCAGTTCTATCTAGGCTGCTGCAAAAGGCATGATTTCATTCTTTTTTATGGCTATGT 2445
             ||  ||| || ||| | | |||||||||||| ||||| || |||||||| |||||||  ||
Db        328 CCTTCAGGTCCATCCATGTTGCTGCAAAAGACATGAATTGATTCTTTTGTATGGCTGAGT 387

Qy       2446 TCTTTCCCA 2454
                | | |||
Db        388 AGTATTCCA 396
```

RESULT 23
ABT02405
ID ABT02405 standard; DNA; 401 BP.
XX
AC ABT02405;
XX
DT 07-NOV-2002 (first entry)
XX
DE Human neuregulin-1-associated gene 1 fragment SEQ ID NO: 941.
XX
KW Human; neuregulin 1; neuregulin-1-associated gene 1; NRG1; NRG1AG1;
KW schizophrenia; chromosome 8p12; single nucleotide polymorphism; SNP;
KW neuroleptic; gene therapy; splice variant; gene; ds.
XX

OS Homo sapiens.
XX
PN US2002094954-A1.
XX
PD 18-JUL-2002.
XX
PF 28-FEB-2001; 2001US-00795686.
XX
PR 28-FEB-2000; 2000US-00515715.
XX
PA (DECO-) DECODE GENETICS EHF.
XX
PI Stefansson H, Steinthorsdottir V, Gulcher JR;
XX
DR WPI; 2002-665799/71.
XX
PT NRG1AG1, useful for treating or diagnosing susceptibility to
PT schizophrenia, or for assaying a sample for the presence of NRG1AG1
PT nucleic acid.
XX
PS Disclosure; Page 599; 700pp; English.
XX
CC The present invention relates to the human neuregulin-1-associated gene
CC 1(NRG1AG1), single nucleotide polymorphisms within which were identified
CC as being associated with an increased susceptibility to schizophrenia,
CC which is located on chromosome 8p12. Also found within the same sequence
CC is the neuregulin 1 gene (NRG1). The gene is useful for treating
CC schizophrenia in an individual, for diagnosing susceptibility to
CC schizophrenia, and for screening for agents useful in the treatment of
CC the disease. The present sequence is a fragment of the NRG1AG1 gene of
CC the invention coontaining a polymorphic site'
XX
SQ Sequence 401 BP; 98 A; 79 C; 62 G; 161 T; 0 U; 1 Other;

Query Match 2.8%; Score 167.8; DB 6; Length 401;
Best Local Similarity 71.2%; Pred. No. 1.6e-30;
Matches 220; Conservative 1; Mismatches 88; Indels 0; Gaps 0;

Qy 2146 TTTCTCACCATCCGCTGACTTCTTGCCCCCTGGGTATTCATCACCTGAATGATGTGCATT 2205
Db 88 TTTGTTACATGGATATAACTCTGGTCTTTCAGTGTAACCATCACTCAAATAGTGTACCTT 147

Qy 2206 GTACCCCTTAAGTAATTTCTCACCATCCGCTGACTTCTTGCCCCCTCATCCTTCTGAGGC 2265
Db 148 GTACCTATTACGTAACTTGTCATTCCTCACCCTCCTCCCACCCTCCCATCTTTYTGAGTC 207

Qy 2266 TCCATTGTCCATCATTCCACACTCTACATCTATGTGTACACATTATTTAGCTCCTACTTA 2325
Db 208 TCCAATGTCTATTATTCCACCCTCTATGTCCATGTGTACTCATTATTTAGTTCTTACTTA 267

Qy 2326 TAAGTGATAACATGCAATATTTGTCTTTCTGTGTCTGTCTTGTTTTACTTATGATAATGG 2385
Db 268 TAAGTAAGAACATGCAATATTTGACTTTCTGTTTCTGAGTTATTTCACTTAAGATAATAG 327

Qy 2386 CCCCCAGTTCTATCTAGGCTGCTGCAAAAGGCATGATTTCATTCTTTTTTATGGCTATGT 2445
Db 328 CCTTCAGGTCCATCCATGTTGCTGCAAAAGACATGAATTGATTCTTTTGTATGGCTGAGT 387

Qy 2446 TCTTTCCCA 2454
Db 388 AGTATTCCA 396

RESULT 24
ADL08128/c
ID ADL08128 standard; DNA; 182328 BP.
XX
AC ADL08128;
XX
DT 20-MAY-2004 (first entry)
XX
DE Human gene associated with low HDL-C PRCP.
XX
KW Human; ds; SNP; single nucleotide polymorphism;
KW high density lipoprotein-C; HDL-C; vascular disease; metabolic disease;
KW coronary artery disease; gene.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT variation replace(63956,C)
FT /*tag= a
FT /standard_name= "Single nucleotide polymorphism"
XX
PN US2004043389-A1.
XX
PD 04-MAR-2004.
XX
PF 04-SEP-2002; 2002US-00235192.
XX
PR 04-SEP-2002; 2002US-00235192.
XX
PA (VITI-) VITIVITY INC.
XX
PI Mccarthy J;
XX
DR WPI; 2004-214170/20.
XX
PT Determining whether a subject has, or is at risk of developing, an
PT abnormally low high density lipoprotein-C (HDL-C) level comprises
PT detecting an allelic variant of a polymorphic region from any of a set of
PT 27 genes.
XX
PS Disclosure; SEQ ID NO 47; 37pp; English.
XX
CC The invention relates to determining whether a subject has, or is at risk
CC of developing, an abnormally low high density lipoprotein-C (HDL-C) level
CC comprises determining whether the subject has an allelic variant of a
CC polymorphic region from any of 27 genes (alleles listed in Table 5 of the
CC specification). Also included are determining whether a male subject has,
CC or is at risk of developing, an abnormally low HDL-C level, comprising
CC determining whether the male subject has an allelic variant of a
CC polymorphic region listed in Table 5 which is associated with abnormally
CC low HDL-C levels in males, and determining whether a female subject has,
CC or is at risk of developing, an abnormally low HDL-C level, comprising
CC determining whether the female subject has an allelic variant of a
CC polymorphic region listed in Table 5 which is associated with abnormally
CC low HDL-C levels in females. The allelic variant in determining whether a
CC subject has, or is at risk of developing, an abnormally low HDL-C level
CC is APOA 1 CC, CD14 1 CT, COL5A2 1 GG, EDNRB 1 AG or AA, FABP3 1 CT, GBE1
CC 1 AG or GG, LIPC 5 AA, MTHFR 1 CC, VWF 2 GG, or their complements. The
CC allelic variant in determining whether a male subject has, or is at risk
CC of developing, an abnormally low HDL-C level, LRP1 3 CC or CT, PAI2 4 GG,
CC or PPARG 1 CG, or their complements. The allelic variants are also COL5A2
CC 1 GG, CD14 1 CT or CC, and FABP3 1 CT, in combination, or their
CC complements. The methods are useful for diagnosing (a predisposition to)
CC abnormally low levels of low high density lipoprotein-C (HDL-C) in a
CC subject, The methods are useful in diagnosing (a predisposition to) or
CC prognosticating diseases and disorders associated with abnormal lipid
CC levels such as vascular and metabolic diseases, e.g., coronary artery
CC disease. The present sequence is a human gene containing a SNP (single
CC nucleotide polymorphism associated with low high density lipoprotein-C
CC (HDL-C) levels.
XX
SQ Sequence 182328 BP; 56638 A; 36576 C; 34943 G; 54171 T; 0 U; 0 Other;

Query Match 2.8%; Score 167.8; DB 12; Length 182328;
Best Local Similarity 62.9%; Pred. No. 2.8e-29;
Matches 397; Conservative 0; Mismatches 187; Indels 47; Gaps 7;

Qy 312 GTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTC 371
Db 127069 GTATGGGCAGAAAGGTGCGATACTTTTTCTCACCCATCATAAGGGTCATGGTCAATGCCC 127010

Qy 372 CTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTT 431

```
Db     127009 CTATAATAAAATACAGTTTAACAAGAGAAAAGCATAACAAGTGTATTTAAGCACAGTTTT 126950

Qy        432 ACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTT 491
              |  ||| ||||||||||||||||||||||||||||||| | |    ||| ||      | |||
Db     126949 ATGTGATATGGGAGTCTTCAGAAATGAAGACCCAAAAAACTCAGGGGAAACTCTTCATTT 126890

Qy        492 TTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAA--GG 549
              ||  ||| | ||| |||||||||||||| ||          |    |||   ||||||  ||
Db     126889 TTAGGCTTACGTTTGATGAAGAATGGACAGTCCTGTAGAAAAGTGATGGAACAAAAAGGG 126830

Qy        550 ATATGATCTAGTGGTAAAGGACT-CAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATT 608
              | |    |||||||||  | ||  | ||||||| |||||||| || ||| ||| ||||
Db     126829 AAAGATCCTAGTGGTAATAGTCTGAATGGGGAAACTCAGCAAGGTCTATCTGTTCTGATT 126770

Qy        609 CTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTAT-GGGGCAGGAC 667
              |||||||  ||||          ||| | ||||||||||||||||  | | | | |||  |
Db     126769 CTTCTTGGTCTC----------TGTGTGGCATTCTTTCCTCCTGGATGTAGTGTCAGACC 126720

Qy        668 TCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATGG 727
              |||  |  ||||| ||||||  |    |      | || |  |     |  | | ||| |
Db     126719 CCTTTTGGAATGACGGTCTTAAGATCAACTATGGGACAAGGTAGGTCAGATAATTTCTTT 126660

Qy        728 CTTGC--------------------------TTCGGGGAAGAGGAGTTCTAGTTTCTAT 760
              |  |                          |    || | || |||  |  ||
Db     126659 GTAACCAGCTCCTATACAGAAAGGAAGGGGAAGGTTAGATTTAGTATTTTTAGGCTTCAT 126600

Qy        761 GACCCATCTTGGGGAAGAGGAATTCTGGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGA 820
              | |      ||  |||||  |  | || |||| |||  ||   | ||||||||||
Db     126599 GGCTGGCTTTCAGGAAGAAGGGTCCTAGTTTCTATGACCTGCCTTGGGGGAGAAAGAGGA 126540

Qy        821 GTAAGAGGCAGGAGGGCAGGAGATGGTCAGAAAGAGACTTGGCTGCTTCTGAGGGCTTCC 880
              ||| |||  || | | ||| ||| |    |||||||||| ||| ||      |||||
Db     126539 GTAGGAGAGAGAAAGACAGAAGAGG----GAAAGAGACTTTGCTTCT--GAGACTCTTCC 126486

Qy        881 GCTCTCCTTTAGTTCCAAGTACTTCTTAGCA 911
              | | ||| ||||| ||||||| |  || |
Db     126485 AGTGTTCTTCAGTTCAAAGTACTCCCAAGGA 126455
```

RESULT 25
AAK91687
ID AAK91687 standard; cDNA; 847 BP.
XX
AC AAK91687;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human cDNA 5'-end sequence, SEQ ID NO: 147.
XX
KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
XX
OS Homo sapiens.
XX
PN EP1130094-A2.
XX
PD 05-SEP-2001.
XX
PF 07-JUL-2000; 2000EP-00114089.
XX
PR 08-JUL-1999; 99JP-00194486.
PR 11-JAN-2000; 2000JP-00118774.
PR 02-MAY-2000; 2000JP-00183765.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
DR WPI; 2001-524255/58.
XX
PT 830 Primers useful for synthesizing full length cDNA clones and their use
PT in genetic manipulation.
XX
PS Claim 2; SEQ ID NO 147; 1380pp + Sequence Listing; English.
XX
CC The invention relates to primers for synthesising full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been isolated
CC and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have
CC been determined. Primers for synthesising the full length cDNA are useful
CC for clarifying the function of the protein encoded by the cDNA. The full
CC length clones were obtained by construction of full length enriched cDNA
CC libraries that were synthesised by the oligo-capping method. The primers
CC enable the production of the full length cDNA easily without any special
CC methods. The present sequence is the nucleotide sequence of the 5'-end of
CC a cDNA provided in the invention. Note: The sequence data for this patent
CC did not form part of the printed specification, but was obtained in CD-
CC ROM format directly from EPO
XX
SQ Sequence 847 BP; 272 A; 143 C; 199 G; 229 T; 0 U; 4 Other;

Query Match 2.7%; Score 166.2; DB 4; Length 847;
Best Local Similarity 66.3%; Pred. No. 5.7e-30;
Matches 386; Conservative 0; Mismatches 164; Indels 32; Gaps 9;

```
Qy        385 CAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTACATGACATGGGA 444
              || |||||||| || ||    |||||||     |||||||||| ||||  ||||    ||
Db        213 CAAGTTAATAACAGAAAGGTATAACAAA-----TTAATCAAAGATTTATGTGACCCAAGA 267

Qy        445 GTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTTTTTTGCTGAGGTT 504
              | ||||| ||||||||||||||||||    |||| || || |||||||| |     | |||
Db        268 GCCTTCAAAAATGAAGACCCAAAGATGTAGGGAAAGCTATCTGTTTTTCT--GCTAAGTT 325

Qy        505 CGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGATATGATCTAGTGGT 564
                | | |||||||| |||          |    || | ||||||||||||||||  |  ||
Db        326 TAACGTAGAATGGACAGCCGTCTAGAAATGCGACTGGACAAAAGGATATGATTTGCTGAC 385

Qy        565 AAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTTCTTGATCTCTCTC 624
              ||  |||| || | |   || | | | ||| | || | ||||  |||||| ||||||
Db        386 AATAGACTGAGAGTGTGGCAAACCGA-GCCAGGCTGTCCAGACACTTCTTGGTCTCTC-- 442

Qy        625 TCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTTCTTCAATGAGGGT 684
                  ||| || ||||  |||||||   |||| |||||   | |  ||  |||||||||
Db        443 ------TGTGTAACATTTCTTCCTCCCTGGTATAGGGCACAGCCCCACTGGAATGAGGGT 496

Qy        685 CTTCAAGGGAGAAGGGAGAAAGTGG--------CCTTTTTAGATTTTATGGCTTGCTTCG 736
              |||||||| |||||||||| | |          | ||    |  |||||| ||||| |
Db        497 CTTCAAGGAAGAAGGGAGAAGGGAGAGAGTAACCTTTCAGAAGTTTTATGACTTGCTGTG 556

Qy        737 GGGA-AGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCTGGTTTCTGT 795
              || | |||||  ||| ||| |||||||||   ||||| ||||| ||||| |  || || |
Db        557 GGAAGAGAGGGATTCAAGTGTCTATGACCTGCCTTGGAGAAGACGAATTTTATTTCCTAT 616

Qy        796 G-ACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGGTCAGAAAG 854
              | |||  || |   | |||||| |||| |  || |||| ||   |||||  |  |  ||
Db        617 GAACTCACTATGGGGGAGAAAGGGGAGCAGAAGACAGGGGGCTGGGAGAAAG---GGCAG 673

Qy        855 AGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAGTACTTCTTAGCATAC 914
              | |   | |   | |||||||||   | |||||||||| |||   | ||| | || |
Db        674 ACAGANGCTTTGCTTTGAGGGCTTCTAATTTCCTTTAGTTCAAAG---TCCTTGGGATGC 730

Qy        915 CAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAACA 956
              ||||||| |||| |||||  |||  | |||| |||   | ||
Db        731 CAAAGCATTATATTTTGGAGTATTCTCTTCTAAGCCACACCA 772
```

RESULT 26
AAK93880
ID AAK93880 standard; cDNA; 847 BP.
XX
AC AAK93880;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human cDNA clone representative sequence, SEQ ID NO: 2340.
XX
KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.

```
XX
OS   Homo sapiens.
XX
PN   EP1130094-A2.
XX
PD   05-SEP-2001.
XX
PF   07-JUL-2000; 2000EP-00114089.
XX
PR   08-JUL-1999;   99JP-00194486.
PR   11-JAN-2000; 2000JP-00118774.
PR   02-MAY-2000; 2000JP-00183765.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Nishikawa T,  Isogai T,  Hayashi K,  Ishii S,  Kawai Y;
PI   Wakamatsu A,  Sugiyama T,  Nagai K,  Kojima S,  Otsuki T,  Koga H;
XX
DR   WPI; 2001-524255/58.
XX
PT   830 Primers useful for synthesizing full length cDNA clones and their use
PT   in genetic manipulation.
XX
PS   Example 11; SEQ ID NO 2340; 1380pp + Sequence Listing; English.
XX
CC   The invention relates to primers for synthesising full length cDNA
CC   clones. 830 cDNA molecules encoding a human protein have been isolated
CC   and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have
CC   been determined. Primers for synthesising the full length cDNA are useful
CC   for clarifying the function of the protein encoded by the cDNA. The full
CC   length clones were obtained by construction of full length enriched cDNA
CC   libraries that were synthesised by the oligo-capping method. The primers
CC   enable the production of the full length cDNA easily without any special
CC   methods. The present sequence was used as the representative sequence
CC   from a human clone which was used in homology searches to identify the
CC   clone. Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in CD-ROM format directly from
CC   EPO
XX
SQ   Sequence 847 BP; 272 A; 143 C; 199 G; 229 T; 0 U; 4 Other;

  Query Match             2.7%;  Score 166.2;  DB 4;  Length 847;
  Best Local Similarity   66.3%;  Pred. No. 5.7e-30;
  Matches  386;  Conservative    0;  Mismatches  164;  Indels   32;  Gaps    9;

Qy        385 CAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTACATGACATGGGA 444
              || |||||||| || ||    |||||||     |||||||||| ||||  ||||    ||
Db        213 CAAGTTAATAACAGAAAGGTATAACAAA-----TTAATCAAAGATTTATGTGACCCAAGA 267

Qy        445 GTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTTTTTTTGCTGAGGTT 504
              | ||||| ||||||||||||||||    |||| || || |||||||||  |      | |||
Db        268 GCCTTCAAAAATGAAGACCCAAAGATGTAGGGAAAGCTATCTGTTTTTCT--GCTAAGTT 325

Qy        505 CGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGATATGATCTAGTGGT 564
                | | |||||||| |||         |    || | ||||||||||||||||  |  ||
Db        326 TAACGTAGAATGGACAGCCGTCTAGAAATGCGACTGGACAAAAGGATATGATTTGCTGAC 385

Qy        565 AAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTTCTTGATCTCTCTC 624
              ||  |||| || | |    || | | | ||| | || | ||||  |||||||  ||||||
Db        386 AATAGACTGAGAGTGTGGCAAACCGA-GCCAGGCTGTCCAGACACTTCTTGGTCTCTC-- 442

Qy        625 TCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTTCTTCAATGAGGGT 684
                  ||| || ||||  |||||||    |||| |||||    | |  ||  |||||||||
Db        443 ------TGTGTAACATTTCTTCCTCCCTGGTATAGGGCACAGCCCCACTGGAATGAGGGT 496

Qy        685 CTTCAAGGGAGAAGGGAGAAAGTGG--------CCTTTTTAGATTTTATGGCTTGCTTCG 736
              ||||||||  |||||||||||  |  |        | ||   |  |||||||  ||||||  |
Db        497 CTTCAAGGAAGAAGGGAGAAGGGAGAGAGTAACCTTTCAGAAGTTTTATGACTTGCTGTG 556

Qy        737 GGGA-AGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCTGGTTTCTGT 795
              || | |||||  ||| ||| |||||||||    |||||  |||||  |||||  |  || || |
Db        557 GGAAGAGAGGGATTCAAGTGTCTATGACCTGCCTTGGAGAAGACGAATTTTATTTCCTAT 616

Qy        796 G-ACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGGTCAGAAAG 854
              | |||  || |   | ||||||  ||||  |  || ||||  ||   |||||  |   |  ||
Db        617 GAACTCACTATGGGGGAGAAAGGGGAGCAGAAGACAGGGGGCTGGGAGAAAG---GGCAG 673

Qy        855 AGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAGTACTTCTTAGCATAC 914
              | |   |  |   | ||||||||||    | ||||||||||| |||   | ||| | || |
Db        674 ACAGANGCTTTGCTTTGAGGGCTTCTAATTTCCTTTAGTTCAAAG---TCCTTGGGATGC 730

Qy        915 CAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAACA 956
              ||||||| |||| |||||  |||  | |||| |||   | ||
Db        731 CAAAGCATTATATTTTGGAGTATTCTCTTCTAAGCCACACCA 772


RESULT 27
ADL30307
ID   ADL30307 standard; cDNA; 847 BP.
XX
AC   ADL30307;
XX
DT   20-MAY-2004  (first entry)
XX
DE   3' end of a representative human cDNA cluster SeqID 2340.
XX
KW   human; medicine; signal transduction; glycoprotein; transcription;
KW   oligo-capping method; ss.
XX
OS   Homo sapiens.
XX
PN   EP1396543-A2.
XX
PD   10-MAR-2004.
XX
PF   07-JUL-2000; 2003EP-00025638.
XX
PR   08-JUL-1999;   99JP-00194486.
PR   11-JAN-2000; 2000JP-00118774.
PR   02-MAY-2000; 2000JP-00183865.
PR   07-JUL-2000; 2000EP-00114089.
XX
PA   (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI   Ota T,  Nishikawa T,  Isogai T,  Hayashi K,  Ishii S,  Kawai Y;
PI   Wakamatsu A,  Sugiyama T,  Nagai K,  Kojima S,  Otsuki T,  Koga H;
XX
DR   WPI; 2004-204755/20.
XX
PT   New oligonucleotide primers (830 cDNAs) useful for synthesizing full
PT   length human cDNAs.
XX
PS   Example 18; SEQ ID NO 2340; 1340pp; English.
XX
CC   This invention relates to a novel primers useful for synthesising full
CC   length cDNA molecules that encode human proteins. Specifically, it refers
CC   to secretory or membrane proteins that are potential therapeutic agents/
CC   target molecules in the field of medicine, and in particular genes
CC   encoding proteins that are associated with signal transduction,
CC   glycoproteins and transcription. The present invention describes a method
CC   for efficiently cloning a full length human cDNA from both the 5' and 3'
CC   ends using the oligo-capping method. This polynucleotide sequence is the
CC   3' end of a representative human DNA cluster of the invention.
XX
SQ   Sequence 847 BP; 272 A; 143 C; 199 G; 229 T; 0 U; 4 Other;

  Query Match             2.7%;  Score 166.2;  DB 12;  Length 847;
  Best Local Similarity   66.3%;  Pred. No. 5.7e-30;
  Matches  386;  Conservative    0;  Mismatches  164;  Indels   32;  Gaps    9;

Qy        385 CAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTACATGACATGGGA 444
              || |||||||| || ||    |||||||     |||||||||| ||||  ||||    ||
Db        213 CAAGTTAATAACAGAAAGGTATAACAAA-----TTAATCAAAGATTTATGTGACCCAAGA 267

Qy        445 GTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTTTTTTTGCTGAGGTT 504
              | ||||| ||||||||||||||||    |||| || || |||||||||  |      | |||
```

```
Db        268 GCCTTCAAAAATGAAGACCCAAAGATGTAGGGAAAGCTATCTGTTTTTCT--GCTAAGTT 325

Qy        505 CGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGATATGATCTAGTGGT 564

Db        326 TAACGTAGAATGGACAGCCGTCTAGAAATGCGACTGGACAAAAGGATATGATTTGCTGAC 385

Qy        565 AAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTTCTTGATCTCTCTC 624

Db        386 AATAGACTGAGAGTGTGGCAAACCGA-GCCAGGCTGTCCAGACACTTCTTGGTCTCTC-- 442

Qy        625 TCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTTCTTCAATGAGGGT 684

Db        443 ------TGTGTAACATTTCTTCCTCCCTGGTATAGGGCACAGCCCCACTGGAATGAGGGT 496

Qy        685 CTTCAAGGGAGAAGGGAGAAAGTGG--------CCTTTTTAGATTTTATGGCTTGCTTCG 736

Db        497 CTTCAAGGAAGAAGGGAGAAGGGAGAGAGTAACCTTTCAGAAGTTTTATGACTTGCTGTG 556

Qy        737 GGGA-AGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCTGGTTTCTGT 795

Db        557 GGAAGAGAGGGATTCAAGTGTCTATGACCTGCCTTGGAGAAGACGAATTTTATTTCCTAT 616

Qy        796 G-ACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGGTCAGAAAG 854

Db        617 GAACTCACTATGGGGGAGAAAGGGGAGCAGAAGACAGGGGGCTGGGAGAAAG---GGCAG 673

Qy        855 AGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAGTACTTCTTAGCATAC 914

Db        674 ACAGANGCTTTGCTTTGAGGGCTTCTAATTTCCTTTAGTTCAAAG---TCCTTGGGATGC 730

Qy        915 CAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAACA 956

Db        731 CAAAGCATTATATTTTGGAGTATTCTCTTCTAAGCCACACCA 772
```

RESULT 28
ADL28114
ID ADL28114 standard; cDNA; 847 BP.
XX
AC ADL28114;
XX
DT 20-MAY-2004 (first entry)
XX
DE 5' end of a human cDNA molecule SeqID 147.
XX
KW human; medicine; signal transduction; glycoprotein; transcription;
KW oligo-capping method; ss.
XX
OS Homo sapiens.
XX
PN EP1396543-A2.
XX
PD 10-MAR-2004.
XX
PF 07-JUL-2000; 2003EP-00025638.
XX
PR 08-JUL-1999; 99JP-00194486.
PR 11-JAN-2000; 2000JP-00118774.
PR 02-MAY-2000; 2000JP-00183865.
PR 07-JUL-2000; 2000EP-00114089.
XX
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
DR WPI; 2004-204755/20.
XX
PT New oligonucleotide primers (830 cDNAs) useful for synthesizing full
PT length human cDNAs.
XX
PS Disclosure; SEQ ID NO 147; 1340pp; English.
XX
CC This invention relates to a novel primers useful for synthesising full
CC length cDNA molecules that encode human proteins. Specifically, it refers
CC to secretory or membrane proteins that are potential therapeutic agents/
CC target molecules in the field of medicine, and in particular genes
CC encoding proteins that are associated with signal transduction,
CC glycoproteins and transcription. The present invention describes a method
CC for efficiently cloning a full length human cDNA from both the 5' and 3'
CC ends using the oligo-capping method. This polynucleotide sequence is the
CC 5' end of a full length human cDNA sequence of the invention.
XX
SQ Sequence 847 BP; 272 A; 143 C; 199 G; 229 T; 0 U; 4 Other;

  Query Match   2.7%; Score 166.2; DB 12; Length 847;
  Best Local Similarity   66.3%; Pred. No. 5.7e-30;
  Matches 386; Conservative   0; Mismatches 164; Indels   32; Gaps   9;

```
Qy        385 CAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTTACATGACATGGGA 444

Db        213 CAAGTTAATAACAGAAAGGTATAACAAA-----TTAATCAAAGATTTATGTGACCCAAGA 267

Qy        445 GTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTTTTTTGCTGAGGTT 504

Db        268 GCCTTCAAAAATGAAGACCCAAAGATGTAGGGAAAGCTATCTGTTTTTCT--GCTAAGTT 325

Qy        505 CGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAGGATATGATCTAGTGGT 564

Db        326 TAACGTAGAATGGACAGCCGTCTAGAAATGCGACTGGACAAAAGGATATGATTTGCTGAC 385

Qy        565 AAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCTTCTTGATCTCTCTC 624

Db        386 AATAGACTGAGAGTGTGGCAAACCGA-GCCAGGCTGTCCAGACACTTCTTGGTCTCTC-- 442

Qy        625 TCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGACTCTTCTTCAATGAGGGT 684

Db        443 ------TGTGTAACATTTCTTCCTCCCTGGTATAGGGCACAGCCCCACTGGAATGAGGGT 496

Qy        685 CTTCAAGGGAGAAGGGAGAAAGTGG--------CCTTTTTAGATTTTATGGCTTGCTTCG 736

Db        497 CTTCAAGGAAGAAGGGAGAAGGGAGAGAGTAACCTTTCAGAAGTTTTATGACTTGCTGTG 556

Qy        737 GGGA-AGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCTGGTTTCTGT 795

Db        557 GGAAGAGAGGGATTCAAGTGTCTATGACCTGCCTTGGAGAAGACGAATTTTATTTCCTAT 616

Qy        796 G-ACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGGTCAGAAAG 854

Db        617 GAACTCACTATGGGGGAGAAAGGGGAGCAGAAGACAGGGGGCTGGGAGAAAG---GGCAG 673

Qy        855 AGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTCCAAGTACTTCTTAGCATAC 914

Db        674 ACAGANGCTTTGCTTTGAGGGCTTCTAATTTCCTTTAGTTCAAAG---TCCTTGGGATGC 730

Qy        915 CAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAACA 956

Db        731 CAAAGCATTATATTTTGGAGTATTCTCTTCTAAGCCACACCA 772
```

RESULT 29
ADM02449
ID ADM02449 standard; cDNA; 3769 BP.
XX
AC ADM02449;
XX
DT 20-MAY-2004 (first entry)
XX
DE Human cDNA of the invention SEQ ID NO:1134.
XX
KW ss; gene; human; gene therapy; diagnostic marker; pharmaceutical.
XX
OS Homo sapiens.
XX
PN EP1347046-A1.
XX

```
PD   24-SEP-2003.
XX
PF   12-APR-2002; 2002EP-00008400.
XX
PR   22-MAR-2002; 2002JP-00137785.
XX
PA   (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI   Isogai T,  Sugiyama T,  Otsuki T,  Wakamatsu A,  Sato H,  Ishii S;
PI   Yamamoto J,  Isono Y,  Hio Y,  Otsuka K,  Nagai K,  Irie R,  Tamechika I;
PI   Seki N,  Yoshikawa T,  Otsuka M,  Nagahari K,  Masuho Y;
XX
DR   WPI; 2003-723558/69.
DR   P-PSDB; ADM04892.
XX
PT   New polynucleotides and polypeptides are useful in gene therapy, for
PT   developing a diagnostic marker or medicines for regulating their
PT   expression and activity, or as a target of gene therapy.
XX
PS   Claim 1; SEQ ID NO 1134; 305pp; English.
XX
CC   The invention relates to a novel human polynucleotide and the encoded
CC   polypeptide. A polynucleotide of the invention may have a use in gene
CC   therapy. An oligonucleotide of the invention ADM06202-ADM06773 is useful
CC   as a primer for synthesizing the polynucleotide or as a probe for
CC   detecting the polynucleotide. The polynucleotides ADM01316-ADM03758 are
CC   useful in gene therapy, for developing a diagnostic marker or medicines
CC   for regulating their expression and activity, or as a target of gene
CC   therapy. The proteins ADM03759-ADM06201 encoded by the polynucleotides
CC   are useful as pharmaceutical agents. The present sequence represents a
CC   cDNA sequence of the invention.
XX
SQ   Sequence 3769 BP; 1080 A; 766 C; 731 G; 1192 T; 0 U; 0 Other;

  Query Match             2.7%;  Score 166;  DB 11;  Length 3769;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-29;
  Matches  166;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5463 GACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGCC 5522
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GACGCTCAGAGGATTCTGACAATATCTTTACCGGAGAAGAGGCAAAGTACGCTCAAAGCC 60

Qy       5523 GAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAATA 5582
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GAAGCCACAGCTCCTCCTGCCGCATTTCTTTCCTGCTTGCGAATTCCAAGCTGTTAAATA 120

Qy       5583 AGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 5628
             |||||||||||||||||||||||||||||||||||||||||||||
Db        121 AGATGTGCAAAGGGCTTGCAGGTCTGCCGGCTTCTTGCTTGAGGAG 166


RESULT 30
ADQ97824
ID   ADQ97824 standard; DNA; 73742 BP.
XX
AC   ADQ97824;
XX
DT   07-OCT-2004  (first entry)
XX
DE   Human cancer associated sequence HD11-004, SEQ ID 801.
XX
KW   Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Human; ds.
XX
OS   Homo sapiens.
XX
PN   WO2004060304-A2.
XX
PD   22-JUL-2004.
XX
PF   22-DEC-2003; 2003WO-US041389.
XX
PR   27-DEC-2002; 2002US-00330773.
XX
PA   (SAGR-) SAGRES DISCOVERY INC.
XX
PI   Morris DW,  Malandro MS;
XX
DR   WPI; 2004-543781/52.
XX
PT   New isolated cancer associated nucleic acids comprising at least 10
PT   contiguous nucleotides, useful for diagnosing, preventing and/or treating
PT   cancers such as leukemia and lymphoma.
XX
PS   Claim 1; SEQ ID NO 801; 199pp; English.
XX
CC   The present invention relates to cancer associated sequences (ADQ97025-
CC   ADQ98004). The sequences are useful for the diagnosis, prevention and/or
CC   treatment of cancer, such as leukemia and lymphoma. Note: The sequence
CC   data for this patent did not form part of the printed specification, but
CC   was obtained in electronic formate directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 73742 BP; 22386 A; 13902 C; 13572 G; 23707 T; 0 U; 175 Other;

  Query Match             2.7%;  Score 165.4;  DB 12;  Length 73742;
  Best Local Similarity   64.4%;  Pred. No. 7.1e-29;
  Matches  471;  Conservative    0;  Mismatches  161;  Indels   99;  Gaps   11;

Qy        307 CTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAA 366
              || | |||||||||  ||| ||||||||  | |||||    | ||||||||||||    | |
Db      39099 CTACTGTAGGGGTGGAAAGGTGTGATA-ACTTTCTTCCCTATCATAAGGGTCATGGCTGA 39157

Qy        367 AACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAA 426
              |||||||||  |||||||||||||  | ||||  |||||  ||||||    ||||||||||
Db      39158 CACTCCTATAATAAAAGACAGGTTGACAAGACAAAAACATAACAA----ATTTAATCAAA 39213

Qy        427 GTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCT 486
              ||| | ||||||||||||         |  |||||||||||||||||||||| |||
Db      39214 GTTCTGCATGACATGGGA--------ACCTGAAGACCCAAAGACCCAGGGAAAAAACATC 39265

Qy        487 GTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAA 546
              |||||  |||| || || || ||     ||     |||||||  ||||  | |     ||
Db      39266 TGTTTTTCTGCTTAGATTTGAAGA--TTTGACAGCCATGTAGAAATGT--AATTACACAA 39321

Qy        547 AGGATATGATCTAGTGGTAAAGGACTCAGGGG-GAAACACAGCAAGGCCTGTCTATTCAG 605
              ||| |||||||||  ||||||  |||| |||||  ||||||  |||||  || | || ||| |
Db      39322 AGGGTATGATCTAATGGTAATAGACTGAGGGGAGAAACATAGCAAAACCCGCCTGTTCGG 39381

Qy        606 ATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGG 665
              ||||||||||          |||||  |||||||  |||  |||  || |   |||   |||||
Db      39382 ATTCTTCTTG--------GTCTCTTTGTATAGTATTTCTTCTTCTTTGATATACAGCAGG 39433

Qy        666 ACTCTTCTTCAATGAGGGTCTT----------CAAGGGAGAAGGGAGAA----------- 704
              || | |||  ||||||||||||          |||||  ||    |||||
Db      39434 ACCCCTCTGGAATGAGGGTCTTATGATCTAGACAAGGTAGGTCAGAGAATTTCTTTATAG 39493

Qy        705 ----------------------------------------AGTGGCCTTTTTAGATTTTA 724
                                                     |||  |   | ||| ||| |
Db      39494 CCAGCTCCTACACAGAACGTGGAGGGTTGGGGGTTGTTGGAGTAACTGATCTAGGTTTCA 39553

Qy        725 TGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATT 784
              ||||| | || | |||||||||  |||||||  || || || || |||||| | ||||||||||
Db      39554 TGGCTGGTTTTGAGGAAGAGGGATTCTAGTGTCCATAACTCATCTTTGAGAAGAGGAATT 39613

Qy        785 CTGGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGAT 844
              ||  |||||||| | ||||||    | ||||||  ||||    || ||  | |||| |||
Db      39614 CTAATTTCTGTGGCCTGCTTTGTGGGAGAAAGGTGAGTGGAAGCCATCAAGGCATGAGGA 39673

Qy        845 GGTCAGAAAGAGACTT---------GGCTGCTTCTGAGGGCTTCCGCTCTCCTTTAGTTC 895
              ||||  | ||||||||          ||||||||||||| |||||  |||||| ||||||
Db      39674 AGTCAAAGAGAGACTTTCCTTCTGAGGCTGCTTCTGAGGCCTTCCAATCTCCTATAGTTC 39733

Qy        896 CAAGTACTTCTTAGCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTCTAAC 955
              |||||||    ||||  || |||  |  || |||||  |||  |||||||| | | |
Db      39734 AAAGTACTC---AGCACGCCGAAGTGCCGTAATTTGGGGTATCATTTTCTGACCCCAACA 39790
```

```
Qy        956 ACTGCAATCAT 966
              ||  | ||
Db      39791 TTGGCTTTTAT 39801


RESULT 31
ACN39085
ID   ACN39085 standard; cDNA; 2238 BP.
XX
AC   ACN39085;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Tumour-associated antigenic target (TAT) cDNA DNA325348, SEQ ID NO:3002.
XX
KW   Tumour-associated antigenic target; TAT; human; overexpression; cancer;
KW   tumour; diagnosis; cell proliferative disorder; breast cancer;
KW   colorectal cancer; lung cancer; ovarian cancer; liver cancer;
KW   central nervous system cancer; bladder cancer; pancreatic cancer;
KW   cervical cancer; melanoma; leukaemia; hybridisation probe;
KW   chromosome identification; chromosome mapping; gene mapping;
KW   gene therapy; cytostatic; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO2004030615-A2.
XX
PD   15-APR-2004.
XX
PF   29-SEP-2003; 2003WO-US028547.
XX
PR   02-OCT-2002; 2002US-0414971P.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Wu TD,  Zhang Z,  Zhou Y;
XX
DR   WPI; 2004-347921/32.
DR   P-PSDB; ABM81161.
XX
PT   New tumor-associated antigenic target polypeptides and nucleic acids,
PT   useful in preparing a medicament for treating or detecting a
PT   proliferative disorder, e.g. breast, lung, colorectal, ovarian or
PT   prostate cancer or tumor.
XX
PS   Claim 1; SEQ ID NO 3002; 7273pp; English.
XX
CC   The invention relates to human tumour-associated antigenic target (TAT)
CC   polypeptides, and their related nucleic acids. The TAT polypeptides are
CC   overexpressed in cancer tissues compared to normal tissues, and may thus
CC   serve as effective targets for the diagnosis and treatment of cancer in
CC   mammals. The invention also relates to nucleic acid and polypeptide
CC   sequences at least 80% identical to the TAT nucleic acids and
CC   polypeptides; expression vectors and host cells comprising a TAT nucleic
CC   acid; an antibody specific for a TAT polypeptide; a peptide or organic
CC   molecule which binds to a TAT polypeptide; fusion proteins comprising a
CC   TAT polypeptide; and methods and compositions for the treatment or
CC   diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,
CC   antibodies, antagonists, binding molecules and compositions are useful
CC   for diagnosing or treating a cell proliferative disorder associated with
CC   increased TAT expression, particularly cancers such as breast cancer,
CC   colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder
CC   cancer, pancreatic cancer, cervical cancer, cancers of the central
CC   nervous system, melanoma and leukaemia. TAT nucleic acids may further be
CC   used as hybridisation probes, in chromosome and gene mapping, in
CC   chromosome identification and in gene therapy. The present sequence
CC   represents a TAT nucleic acid of the invention
XX
SQ   Sequence 2238 BP; 512 A; 559 C; 488 G; 679 T; 0 U; 0 Other;

  Query Match             2.7%;  Score 163.8;  DB 13;  Length 2238;
  Best Local Similarity   74.9%;  Pred. No. 3.4e-29;
  Matches  218;  Conservative    0;  Mismatches   72;  Indels    1;  Gaps    1;

Qy       2177 GGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCT 2236
              | |||  |||||||  ||| || | | |||||||| |||||| |||||||| |    | |
Db       1224 GTGTAACCATCACCCTAATAATATACGTTGTACCCATTAAGTTATTTCTCATCCCTCACC 1283

Qy       2237 GACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTCTACATCT 2296
                || |   ||   ||| ||||||||| ||||| |||| || ||||||||||| |  ||
Db       1284 CCCTCCCA-CCTTGTCACCCTTCTGAGTCTCCAATGTCTATTATTCCACACTCCATGTCC 1342

Qy       2297 ATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGTCTTTCTG 2356
              | ||||||||||||||||||||| ||||| |||||| ||||||    |||||| |||||||
Db       1343 ACGTGTACACATTATTTAGCTCCCACTTACAAGTGAGAACATGTGGTATTTGACTTTCTG 1402

Qy       2357 TGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAAGG 2416
              | | ||  || ||| |||||  |||||| || ||||||  ||| | | ||||||||||
Db       1403 TTTTTGAGTTATTTCACTTAAAATAATGACCTCCAGTTTCATCCATGTTGCTGCAAAAGA 1462

Qy       2417 CATGATTTCATTCTTTTTTATGGCTATGTTCTTTCCCAATTTAGATAAAGA 2467
              |||||| || |||||||||||||  |||  | | ||| |  | ||| | |
Db       1463 CATGATCTCGTTCTTTTTTATGGCAGTGTAGTATTCCATTCCATATACATA 1513


RESULT 32
ABL55857
ID   ABL55857 standard; DNA; 34337 BP.
XX
AC   ABL55857;
XX
DT   15-JUL-2002  (first entry)
XX
DE   Human GABA transporter protein gene.
XX
KW   Human; GABA; transporter protein; drug screening; therapeutic target;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       replace(701,C/T)
FT                   /*tag=  a
FT                   /standard_name= "Single nucleotide polymorphism"
FT   CDS             2001. .32337
FT                   /*tag=  b
FT                   /product= "GABA transporter protein"
FT                   /note= "Contains 10 introns"
FT   exon            2001. .2289
FT                   /*tag=  c
FT                   /number= 1
FT   intron          2290. .8039
FT                   /*tag=  d
FT                   /number= 1
FT   variation       replace(4738,C)
FT                   /*tag=  e
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(6156,A/G/C)
FT                   /*tag=  f
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(6425,C)
FT                   /*tag=  g
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(6444,G)
FT                   /*tag=  h
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(7538,A)
FT                   /*tag=  i
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            8040. .8197
FT                   /*tag=  j
FT                   /number= 2
FT   intron          8198. .8564
FT                   /*tag=  k
FT                   /number= 2
FT   exon            8547. .8673
```

```
FT                   /*tag= l
FT                   /number= 3
FT   intron          8674. .10067
FT                   /*tag= m
FT                   /number= 3
FT   variation       replace(9127,G)
FT                   /*tag= n
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(9256,G)
FT                   /*tag= o
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(9421,G)
FT                   /*tag= p
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            10068. .10249
FT                   /*tag= q
FT                   /number= 4
FT   intron          10250. .17500
FT                   /*tag= r
FT                   /number= 4
FT   variation       replace(13248. .13250,TC)
FT                   /*tag= s
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(13377,G)
FT                   /*tag= t
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(13434,C)
FT                   /*tag= u
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(13672,A)
FT                   /*tag= v
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(14502,T)
FT                   /*tag= w
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(14774. .14776,AAA/AA)
FT                   /*tag= x
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(17289,A)
FT                   /*tag= y
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            17501. .17611
FT                   /*tag= z
FT                   /number= 5
FT   intron          17612. .20639
FT                   /*tag= aa
FT                   /number= 5
FT   variation       replace(17660,G)
FT                   /*tag= ab
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19702,G)
FT                   /*tag= ac
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            20640. .20881
FT                   /*tag= ad
FT                   /number= 6
FT   intron          20882. .21173
FT                   /*tag= ae
FT                   /number= 6
FT   exon            21174. .21366
FT                   /*tag= af
FT                   /number= 7
FT   variation       replace(21263,T)
FT                   /*tag= ag
FT                   /standard_name= "Single nucleotide polymorphism"
FT   intron          21367. .23297
FT                   /*tag= ah
FT                   /number= 7
FT   exon            23298. .23490
FT                   /*tag= ai
FT                   /number= 8
FT   intron          23491. .26774
FT                   /*tag= aj
FT                   /number= 8
FT   variation       replace(24502. .24504,TG)
FT                   /*tag= ak
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(24642,C)
FT                   /*tag= al
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            26775. .26934
FT                   /*tag= am
FT                   /number= 9
FT   intron          26935. .30367
FT                   /*tag= an
FT                   /number= 9
FT   variation       replace(28543,T)
FT                   /*tag= ao
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(30303,C)
FT                   /*tag= ap
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            30368. .30530
FT                   /*tag= aq
FT                   /number= 10
FT   intron          30531. .31962
FT                   /*tag= ar
FT                   /number= 10
FT   exon            31963. .32337
FT                   /*tag= as
FT                   /number= 11
XX
PN   US2002031800-A1.
XX
PD   14-MAR-2002.
XX
PF   21-DEC-2000; 2000US-00741149.
XX
PR   05-SEP-2000; 2000US-0229529P.
XX
PA   (LIZZ/) LI Z.
PA   (CHAT/) CHATURVEDI K.
PA   (ZHUS/) ZHU S.
PA   (WOOD/) WOODAGE T.
PA   (GUEG/) GUEGLER K.
PA   (WEBS/) WEBSTER M.
PA   (DFRA/) DI FRANCESCO V.
PA   (BEAS/) BEASLEY E M.
XX
PI   Li Z, Chaturvedi K, Zhu S, Woodage T, Guegler K, Webster M;
PI   Di Francesco V, Beasley EM;
XX
DR   WPI; 2002-361179/39.
DR   P-PSDB; ABB77168.
XX
PT   New peptides and nucleic acid sequences related to the GABA transporter
PT   subfamily, useful in developing diagnostic compositions, as well as in
PT   drug screening, particularly as models for the development of human
PT   therapeutic targets.
XX
PS   Claim 2; Fig 3; 69pp; English.
XX
CC   The sequence encodes a novel human transporter protein of the GABA
CC   transporter subfamily. The invention relates to a novel isolated human
CC   peptide of the GABA transporter subfamily. The peptide and nucleic acids
CC   are useful in developing human therapeutics and diagnostic compositions.
CC   These are also useful in drug screening, particularly as models for the
CC   development of human therapeutic targets
XX
SQ   Sequence 34337 BP; 10987 A; 5399 C; 6061 G; 11588 T; 0 U; 302 Other;

  Query Match             2.7%;  Score 163.2;  DB 6;  Length 34337;
  Best Local Similarity   72.8%;  Pred. No. 1.7e-28;
  Matches  268;  Conservative    0;  Mismatches   88;  Indels   12;  Gaps    4;

Qy        489 TTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAG 548
```

```
                |||||  ||||  ||  ||   |||||  ||  |  |||         |    ||||  ||||||||
Db       4905 TTTTTATGCTTAGATTACCTGAAGTATAGGCAGCCATATAGAAATGTGATTGGACAAAAG 4964

Qy        549 GATATGATCTAGTGGTAAAGGACTCAGGG--GGAAACACAGCAAGGCCTGTCTATTCAGA 606
              | ||  |||||||||   ||   ||||  |  ||   | ||||  |||||||  |||||||  ||||||
Db       4965 GGTA-GATCTAGTGAGAATAGACTAAAGGAAGCAAACCCAGCAAGTCCTGTCTGTTCAGA 5023

Qy        607 TTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGA 666
              ||  ||||||   |||||         ||   |  ||||||  |||||||||   ||||    |||||
Db       5024 TTTTTCTTGGCCTCTC--------TGAGTGGCATTCCTTCCTCCTGGATATGAAACAGGA 5075

Qy        667 CTCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATG 726
              |    ||    |||||||  ||||||||||||||||||||||   ||  |||||  |||  ||||  ||
Db       5076 CGTCTCCGGAATGAGGATCTTCAAGGGAGAAGGGAGAAGATGACCTTTCTAGGTTTTGTG 5135

Qy        727 GCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCT 786
              ||||||||  ||||||||||   ||||||||||||||||||||  ||||  ||   ||  |||  |||
Db       5136 GCTTGCTTTGGGGAAGAGGCATTCTAGTTTCTATGACCCACCTTGAGG-GGAAGAAATCT 5194

Qy        787 GGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGG 846
              ||  ||||||||||||||    ||  |||  |  ||||  |||  ||||||   ||  |  |  ||
Db       5195 TGTGTCTGTGACTTGCTTCAGGGACAAAAAAAGAGTGGGAGACAGGAGTAAAGAAAAAGG 5254

Qy        847 TCAGAAAG 854
              |||  |||
Db       5255 CCAGGAAG 5262
```

```
RESULT 33
ADG88376
ID   ADG88376 standard; DNA; 34337 BP.
XX
AC   ADG88376;
XX
DT   11-MAR-2004  (first entry)
XX
DE   Human transporter gene.
XX
KW   Human; transporter protein; GABA transporter; therapeutic protein;
KW   pharmacogenomic analysis; immune response; biological fluid;
KW   foetal heart tissue; gene therapy; gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1001. .31067
FT                   /*tag=  b
FT                   /product= "Human transporter protein"
FT   exon            1001. .1289
FT                   /*tag=  a
FT   intron          1290. .7053
FT                   /*tag=  c
FT   exon            7054. .7211
FT                   /*tag=  d
FT   intron          7212. .7560
FT                   /*tag=  e
FT   exon            7561. .7687
FT                   /*tag=  f
FT   intron          7688. .9081
FT                   /*tag=  g
FT   exon            9082. .9263
FT                   /*tag=  h
FT   intron          9264. .16514
FT                   /*tag=  i
FT   exon            16515. .16625
FT                   /*tag=  j
FT   intron          16626. .19366
FT                   /*tag=  k
FT   exon            19367. .19608
FT                   /*tag=  l
FT   intron          19609. .19900
FT                   /*tag=  m
FT   exon            19901. .20093
FT                   /*tag=  n
FT   intron          20094. .22026
FT                   /*tag=  o
FT   exon            22027. .22219
FT                   /*tag=  p
FT   intron          22220. .25504
FT                   /*tag=  q
FT   exon            25505. .25664
FT                   /*tag=  r
FT   intron          25665. .29097
FT                   /*tag=  s
FT   exon            29098. .29260
FT                   /*tag=  t
FT   intron          29261. .30692
FT                   /*tag=  u
FT   exon            30693. .31067
FT                   /*tag=  v
XX
PN   US2003157649-A1.
XX
PD   21-AUG-2003.
XX
PF   12-MAR-2003; 2003US-00385614.
XX
PR   05-SEP-2000; 2000US-0229529P.
PR   21-DEC-2000; 2000US-00741149.
PR   05-SEP-2001; 2001WO-US027403.
XX
PA   (APPL-) APPLERA CORP.
XX
PI   Li Z,  Chaturvedi K,  Zhu S,  Woodage T,  Guegler K,  Webster M;
PI   Di Francesco V,  Beasley EM;
XX
DR   WPI; 2003-777995/73.
DR   P-PSDB; ADG88375.
XX
PT   New human transporter proteins and nucleic acid molecules, useful for
PT   diagnosing, preventing or treating disorders associated with aberrant
PT   expression of the transporter proteins, or in pharmacogenomic analysis.
XX
PS   Claim 4; SEQ ID NO 3; 59pp; English.
XX
CC   The invention relates to human transporter protein and its corresponding
CC   nucleic acid sequence related to GABA transporter subfamily. The
CC   invention is useful as models for the development of human therapeutic
CC   targets, aids in the identification of therapeutic proteins, serves as
CC   targets for the development of human therapeutic agents that modulate
CC   transporter activity in cells and tissues that express the transporter,
CC   or in pharmacogenomic analysis. The transporter peptide may be used in
CC   substantial and specific assays, in raising antibodies or in eliciting
CC   another immune response, as a reagent in assays designed to
CC   quantitatively determine the levels of the protein in biological fluids,
CC   and as markers for tissues in which the protein is expressed. The nucleic
CC   acid molecules may be used as probes, primers, chemical intermediates and
CC   in biological assays. These may also be used in diagnosing, preventing or
CC   treating disorders associated with human transporter protein, such as
CC   those expressed in foetal heart tissues. The invention is useful in gene
CC   therapy. The present sequence is human transporter gene.
XX
SQ   Sequence 34337 BP; 10987 A; 5399 C; 6061 G; 11588 T; 0 U; 302 Other;
```

```
  Query Match             2.7%;  Score 163.2;  DB 10;  Length 34337;
  Best Local Similarity   72.8%;  Pred. No. 1.7e-28;
  Matches  268;  Conservative    0;  Mismatches   88;  Indels   12;  Gaps    4;

Qy        489 TTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATTAGACAAAAG 548
              |||||  ||||  ||  ||   |||||  ||  |  |||         |    ||||  ||||||||
Db       4905 TTTTTATGCTTAGATTACCTGAAGTATAGGCAGCCATATAGAAATGTGATTGGACAAAAG 4964

Qy        549 GATATGATCTAGTGGTAAAGGACTCAGGG--GGAAACACAGCAAGGCCTGTCTATTCAGA 606
              | ||  |||||||||   ||   ||||  |  ||   | ||||  |||||||  |||||||  ||||||
Db       4965 GGTA-GATCTAGTGAGAATAGACTAAAGGAAGCAAACCCAGCAAGTCCTGTCTGTTCAGA 5023
```

```
Qy        607 TTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGA 666
              || ||||||  |||||        ||  | ||||||  |||||||||  ||||   |||||
Db       5024 TTTTTCTTGGCCTCTC--------TGAGTGGCATTCCTTCCTCCTGGATATGAAACAGGA 5075

Qy        667 CTCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTATG 726
              |    ||   ||||||| ||||||||||||||||||||||  || ||||| ||| |||| ||
Db       5076 CGTCTCCGGAATGAGGATCTTCAAGGGAGAAGGGAGAAGATGACCTTTCTAGGTTTTGTG 5135

Qy        727 GCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATTCT 786
              ||||||||  |||||||||  ||||||||||||||||||||  |||| ||  || ||| |||
Db       5136 GCTTGCTTTGGGGAAGAGGCATTCTAGTTTCTATGACCCACCTTGAGG-GGAAGAAATCT 5194

Qy        787 GGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGG 846
              || ||||||||||||||     ||  ||| | ||||  ||| ||||||   || | | ||
Db       5195 TGTGTCTGTGACTTGCTTCAGGGACAAAAAAAGAGTGGGAGACAGGAGTAAAGAAAAAGG 5254

Qy        847 TCAGAAAG 854
              ||| |||
Db       5255 CCAGGAAG 5262
```

RESULT 34
ACH30357
ID ACH30357 standard; cDNA; 414 BP.
XX
AC ACH30357;
XX
DT 13-OCT-2003 (first entry)
XX
DE Human testis cDNA #743.
XX
KW Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;
KW genome mapping; biodiversity; genetic disorder.
XX
OS Homo sapiens.
XX
PN US2003073623-A1.
XX
PD 17-APR-2003.
XX
PF 30-JUL-2001; 2001US-00918995.
XX
PR 30-JUL-2001; 2001US-00918995.
XX
PA (DRMA/) DRMANAC R T.
PA (LABA/) LABAT I.
PA (STAC/) STACHE-CRAIN B.
PA (DICK/) DICKSON M C.
PA (JONE/) JONES L W.
XX
PI Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;
XX
DR WPI; 2003-615964/58.
XX
PT New polynucleotide sequences obtained from various cDNA libraries, useful
PT as hybridization probes, as oligomers for PCR, for chromosome and gene
PT mapping, in the recombinant production of protein, or in generating
PT antisense DNA or RNA.
XX
PS Claim 1; SEQ ID NO 17569; 44pp; English.
XX
CC The invention relates to an isolated polynucleotide comprising any one of
CC 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
CC determined by the technique of SBH (sequencing by hybridisation). Also
CC included is a purified polypeptide comprising a sequence corresponding to
CC a reading frame of the novel polynucleotide. The nucleic acid sequences
CC are useful in diagnostics as expressed sequence tags (EST) for
CC identifying expressed genes or for physical mapping of the human genome,
CC in forensics, in assessing biodiversities, or in identifying mutations
CC responsible for genetic disorders and other traits. The nucleotide
CC sequences are also useful as hybridisation probes, as oligomers for PCR,
CC for chromosome and gene mapping, in the recombinant production of
CC protein, or in generating antisense DNA or RNA. The purified polypeptide
CC is useful for generating antibodies specific for it. The present sequence
CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=20030073623
XX
SQ Sequence 414 BP; 119 A; 85 C; 64 G; 145 T; 0 U; 1 Other;

```
  Query Match             2.7%;  Score 162.2;  DB 9;  Length 414;
  Best Local Similarity   77.9%;  Pred. No. 3.9e-29;
  Matches  208;  Conservative    0;  Mismatches   58;  Indels    1;  Gaps    1;

Qy       2180 TATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCTGAC 2239
              ||  ||| |||  ||| |||||||||||||||| ||| ||||||||||| |     || |
Db         28 TAACCATTACCCAAATAATGTGCATTGTACCCTTTAGGTAATTTCTCATCCCTGACTCCC 87

Qy       2240 TTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTCTACATCTATG 2299
              ||||| ||| |||| |||||  ||  ||||| ||| || ||||| ||||||   ||  ||
Db         88 TTCTTCCCCTCTCACCCTTCCAAGTTTCCATAGTCTATTATTCCGCACTCTGTGTCCGTG 147

Qy       2300 TGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGTCTTT-CTGTG 2358
              |||||||||| |||||||||||||||| |||| |||||||| |||||| |||| |||
Db        148 TGTACACATTGTTTAGCTCCTACTTATACGTGAGAACATGCAGTATTTGACTTTGCTGCT 207

Qy       2359 TCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAAGGCA 2418
              ||||  |||||| ||||| |||||| ||  |||| || ||| | | |||| ||||||  |
Db        208 TCTGAGTTGTTTCACTTAAGATAATAGCTACCAGCTCCATCCATGTTGCTTCAAAAGATA 267

Qy       2419 TGATTTCATTCTTTTTTATGGCTATGT 2445
              ||||| ||||||||||||||||  ||
Db        268 AGATTTTATTCTTTTTTATGGCTGCGT 294
```

RESULT 35
ADQ97421
ID ADQ97421 standard; DNA; 228835 BP.
XX
AC ADQ97421;
XX
DT 07-OCT-2004 (first entry)
XX
DE Human cancer associated sequence HD08-046, SEQ ID 398.
XX
KW Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Human; ds.
XX
OS Homo sapiens.
XX
PN WO2004060304-A2.
XX
PD 22-JUL-2004.
XX
PF 22-DEC-2003; 2003WO-US041389.
XX
PR 27-DEC-2002; 2002US-00330773.
XX
PA (SAGR-) SAGRES DISCOVERY INC.
XX
PI Morris DW, Malandro MS;
XX
DR WPI; 2004-543781/52.
XX
PT New isolated cancer associated nucleic acids comprising at least 10
PT contiguous nucleotides, useful for diagnosing, preventing and/or treating
PT cancers such as leukemia and lymphoma.
XX
PS Claim 1; SEQ ID NO 398; 199pp; English.
XX
CC The present invention relates to cancer associated sequences (ADQ97025-
CC ADQ98004). The sequences are useful for the diagnosis, prevention and/or
CC treatment of cancer, such as leukemia and lymphoma. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic formate directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

```
XX
SQ   Sequence 228835 BP; 65103 A; 49297 C; 49161 G; 65274 T; 0 U; 0 Other;

  Query Match             2.7%;  Score 161.4;  DB 12;  Length 228835;
  Best Local Similarity   74.4%;  Pred. No. 1.2e-27;
  Matches  230;  Conservative    0;  Mismatches   76;  Indels    3;  Gaps    2;

Qy       2180 TATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCTGAC 2239
             ||| |||||| ||||| || | |||||||||| ||||||||||||||    ||| ||  |
Db     126155 TATCCATCACTTGAATAATATACATTGTACCCATTAAGTAATTTCTCTTGGTCCCCTCCC 126214

Qy       2240 TTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTCTACATCTATG 2299
             |||   |||||||  |||||  || |||||||||| ||||||| || |||||  || |||
Db     126215 TTCCCACCCCCTCTCCCTTCCAAGTCTCCATTGTCAATCATTCTACTCTCTATGTCCATG 126274

Qy       2300 TGTACA--CATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGTCTTTCTGT 2357
               ||||    || ||||||||| |||||| ||||| ||||||   |||||| || |||||
Db     126275 CATACATGGTTTTTTTAGCTCCCACTTATGAGTGAGAACATGTGGTATTTGACTCTCTGT 126334

Qy       2358 GTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAAGGC 2417
             ||||| ||||||| ||||| |||||||| || |||||| ||| |   ||||||||||| |
Db     126335 GTCTGGCTTGTTTCACTTAAGATAATGG-CCTCAGTTCCATCCATTTGCTGCAAAAGAC 126393

Qy       2418 ATGATTTCATTCTTTTTTATGGCTATGTTCTTTCCCAATTTAGATAAAGAACACTCGCAC 2477
             ||||||| |||||||||||||||||      | | ||| |  |  || | | |||    |
Db     126394 ATGATTTTATTCTTTTTTATGGCTGAAGAGTATTCCATTGCATTTATATACCACATTTTC 126453

Qy       2478 TTGCTCTTA 2486
                ||  || |
Db     126454 TTTACCTAA 126462


RESULT 36
ABS55320_2
Continuation (3 of 5) of ABS55320 from base 200001 (Human transporter protein genomic DN
WP Sequence split into 5 fragments  LOCUS ABS55320 Accession Abs55320
WP     Fragment Name      Begin     End
WP     ABS55320_0             1    110000
WP     ABS55320_1        100001    210000
WP     ABS55320_2        200001    310000
WP     ABS55320_3        300001    410000
WP     ABS55320_4        400001    506286

  Query Match             2.6%;  Score 158.2;  DB 6;  Length 110000;
  Best Local Similarity   71.5%;  Pred. No. 5e-27;
  Matches  208;  Conservative    0;  Mismatches   83;  Indels    0;  Gaps    0;

Qy       2177 GGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCT 2236
             | |||  ||| |||||| | || | ||||||    | ||||||||||||||| |    | |
Db      78893 GTGTAACCATGACCTGAGTAATATACATTGTGATCGTTAAGTAATTTCTCATCCCTCACC 78952

Qy       2237 GACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTCTACATCT 2296
                ||   |||   || ||| |  || ||| | |||| ||||| ||||||||||  |
Db      78953 CCTCTCCCACCCTGGCACCCTCCCAAGTCTCTAATGTCTATCATGCCACACTCTAGGTTC 79012

Qy       2297 ATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGTCTTTCTG 2356
             ||||  |||||||| |||||| | | |||||||||| ||| ||   |||||| |||||||
Db      79013 ATGTACACACATTACTTAGCTACGAGTTATAAGTGAGAACTTGTGGTATTTGACTTTCTG 79072

Qy       2357 TGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAAGG 2416
             | ||||  |||||| |||||||||| ||||||| ||||| | ||| |||||||
Db      79073 TTTCTGCGTTGTTTCACTTAAGATAATGGCCTCCAGTTCCATCTATGTTGCAGCAAAAGA 79132

Qy       2417 CATGATTTCATTCTTTTTTATGGCTATGTTCTTTCCCAATTTAGATAAAGA 2467
             ||| |||| ||| ||| | ||||||   |  | | ||| | |   || | |
Db      79133 CATTATTTTATTGTTTATAATGGCTGAATGGTATTCCATTGTGTGTATACA 79183


RESULT 37
ADQ59398_2
Continuation (3 of 4) of ADQ59398 from base 200001 (Human cancer-associated (CA) gene se
WP Sequence split into 4 fragments  LOCUS ADQ59398 Accession Adq59398
WP     Fragment Name      Begin     End
WP     ADQ59398_0             1    110000
WP     ADQ59398_1        100001    210000
WP     ADQ59398_2        200001    310000
WP     ADQ59398_3        300001    357652

  Query Match             2.6%;  Score 158.2;  DB 12;  Length 110000;
  Best Local Similarity   71.5%;  Pred. No. 5e-27;
  Matches  208;  Conservative    0;  Mismatches   83;  Indels    0;  Gaps    0;

Qy       2177 GGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCT 2236
             | |||  ||| |||||| | || | ||||||    | ||||||||||||||| |    | |
Db      87029 GTGTAACCATGACCTGAGTAATATACATTGTGATCGTTAAGTAATTTCTCATCCCTCACC 87088

Qy       2237 GACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTCTACATCT 2296
                ||   |||   || ||| |  || ||| | |||| ||||| ||||||||||  |
Db      87089 CCTCTCCCACCCTGGCACCCTCCCAAGTCTCTAATGTCTATCATGCCACACTCTAGGTTC 87148

Qy       2297 ATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGTCTTTCTG 2356
             ||||  |||||||| |||||| | | |||||||||| ||| ||   |||||| |||||||
Db      87149 ATGTACACACATTACTTAGCTACGAGTTATAAGTGAGAACTTGTGGTATTTGACTTTCTG 87208

Qy       2357 TGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAAGG 2416
             | ||||  |||||| |||||| |||||||||| ||||||| ||||| | ||| |||||||
Db      87209 TTTCTGCGTTGTTTCACTTAAGATAATGGCCTCCAGTTCCATCTATGTTGCAGCAAAAGA 87268

Qy       2417 CATGATTTCATTCTTTTTTATGGCTATGTTCTTTCCCAATTTAGATAAAGA 2467
             ||| |||| ||| ||| | ||||||   |  | | ||| | |   || | |
Db      87269 CATTATTTTATTGTTTATAATGGCTGAATGGTATTCCATTGTGTGTATACA 87319


RESULT 38
ABS55200
ID   ABS55200 standard; DNA; 341511 BP.
XX
AC   ABS55200;
XX
DT   06-JAN-2003  (first entry)
XX
DE   Genomic DNA encoding human transporter protein.
XX
KW   Human; transporter protein; hypoglycaemia; antidiabetic; gene therapy;
KW   gene; ds; single nucleotide polymorphism; SNP.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             2001. .339512
FT                   /*tag= a
FT                   /product= "Human transporter protein"
FT   exon            2001. .2073
FT                   /*tag= b
FT                   /number= 1
FT   intron          2074. .30237
FT                   /*tag= c
FT                   /number= 1
FT   variation       replace(2557,T)
FT                   /*tag= w
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(5492,C)
FT                   /*tag= x
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(5930,A)
FT                   /*tag= y
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(7269,C)
FT                   /*tag= z
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(9650,G/T)
FT                   /*tag= aa
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(10664,T)
FT                   /*tag= ab
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
```

```
FT   variation       replace(11043,A/G/C)
FT                   /*tag= ac
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(13109,T)
FT                   /*tag= ad
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(13392,A)
FT                   /*tag= ae
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(13395. .13397,GA)
FT                   /*tag= af
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(13563,C)
FT                   /*tag= ag
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(13694,A)
FT                   /*tag= ah
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(14075,A)
FT                   /*tag= ai
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(14573,A)
FT                   /*tag= aj
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(15013,T)
FT                   /*tag= ak
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(15073. .15075,GT)
FT                   /*tag= al
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(15848,T)
FT                   /*tag= am
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(16346,T)
FT                   /*tag= an
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(16367,A)
FT                   /*tag= ao
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(16639,G)
FT                   /*tag= ap
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(18149,A/T)
FT                   /*tag= aq
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(18243,C/A)
FT                   /*tag= ar
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(18293,G/A)
FT                   /*tag= as
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(19185,C)
FT                   /*tag= at
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(24728. .24730,AA)
FT                   /*tag= au
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(24790,T)
FT                   /*tag= av
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(24950,T)
FT                   /*tag= aw
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(28608,G)
FT                   /*tag= ax
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(29244,C)
FT                   /*tag= ay
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(29651,T)
FT                   /*tag= az
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   exon            30238. .30361
FT                   /*tag= d
FT                   /number= 2
FT   intron          30362. .40719
FT                   /*tag= e
FT                   /number= 2
FT   variation       replace(34034,T)
FT                   /*tag= ba
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(36671,G)
FT                   /*tag= bb
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(36836,T)
FT                   /*tag= bc
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(37010,A)
FT                   /*tag= bd
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(38461,A/C)
FT                   /*tag= be
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(38588,A/C)
FT                   /*tag= bf
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(39673,A)
FT                   /*tag= bg
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(39854,C)
FT                   /*tag= bh
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   exon            40720. .40826
FT                   /*tag= f
FT                   /number= 3
FT   intron          40827. .45732
FT                   /*tag= g
FT                   /number= 3
FT   variation       replace(41325,A)
FT                   /*tag= bi
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(41412,A)
FT                   /*tag= bj
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(41475,G)
FT                   /*tag= bk
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(41541,G)
FT                   /*tag= bl
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(41649,G)
FT                   /*tag= bm
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(41887,T)
FT                   /*tag= bn
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(42790,A)
FT                   /*tag= bo
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   exon            45733. .45861
FT                   /*tag= h
FT                   /number= 4
FT   intron          45862. .59818
FT                   /*tag= i
FT                   /number= 4.
FT   variation       replace(46178,T)
FT                   /*tag= bp
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(46282,C)
FT                   /*tag= bq
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(46382,C)
FT                   /*tag= br
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(46550,A)
FT                   /*tag= bs
```

```
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(49271. .49273,AT)
FT                   /*tag= bt
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(49586,A)
FT                   /*tag= bu
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(49835,G)
FT                   /*tag= bv
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(52080,G)
FT                   /*tag= bw
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
```

Query Match 2.6%; Score 158.2; DB 6; Length 341511;
Best Local Similarity 71.5%; Pred. No. 8.4e-27;
Matches 208; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

```
Qy       2177 GGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCT 2236
Db     278893 GTGTAACCATGACCTGAGTAATATACATTGTGATCGTTAAGTAATTTCTCATCCCTCACC 278952

Qy       2237 GACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTCTACATCT 2296
Db     278953 CCTCTCCCACCCTGGCACCCTCCCAAGTCTCTAATGTCTATCATGCCACACTCTAGGTTC 279012

Qy       2297 ATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGTCTTTCTG 2356
Db     279013 ATGTACACACATTACTTAGCTACGAGTTATAAGTGAGAACTTGTGGTATTTGACTTTCTG 279072

Qy       2357 TGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAAGG 2416
Db     279073 TTTCTGCGTTGTTTCACTTAAGATAATGGCCTCCAGTTCCATCTATGTTGCAGCAAAAGA 279132

Qy       2417 CATGATTTCATTCTTTTTTATGGCTATGTTCTTTCCCAATTTAGATAAAGA 2467
Db     279133 CATTATTTTATTGTTTATAATGGCTGAATGGTATTCCATTGTGTGTATACA 279183
```

RESULT 39
AAX90847
ID AAX90847 standard; DNA; 84607 BP.
XX
AC AAX90847;
XX
DT 13-JAN-2000 (first entry)
XX
DE Human PACAP genomic sequence.
XX
KW Pituitary adenylate cyclase activating polypeptide; PACAP gene; probe;
KW neuropsychiatric disorder; bipolar affective disorder; BAD; BP-I; UTR;
KW severe bipolar affective disorder; schizoaffective disorder manic type;
KW SAD-M; chromosome 18p11; retinal ganglion cell; STS marker; ADCYAP1;
KW dinucleotide repeat polymorphism marker; W3440; detection; mutation;
KW clinical observation; subjective report; compound; mapping; diagnosis;
KW treatment; genomic sequence; ds.
XX
OS Homo sapiens.
XX

```
FH   Key             Location/Qualifiers
FT   5'UTR           72668. .72991
FT                   /*tag= a
FT   CDS             72993. .77243
FT                   /*tag= b
FT                   /product= "PACAP protein"
FT                   /note= "Includes exons and introns"
FT   3'UTR           77244. .78651
FT                   /*tag= c
FT                   /note= "Comprises a dinucleotide repeat polymorphism
FT                   marker, W3440 and an STS marker, ADCYAP1"
```

XX
PN WO9951762-A1.
XX
PD 14-OCT-1999.
XX
PF 02-APR-1999; 99WO-US007401.
XX
PR 03-APR-1998; 98US-0080570P.
XX
PA (MILL-) MILLENNIUM PHARM INC.
PA (REGC) UNIV CALIFORNIA.
XX
PI Chen H, Freimer NB;
XX
DR WPI; 1999-611046/52.
XX
PT Use of a gene linked to the region of a chromosome associated with
PT neuropsychiatric disorders, especially for diagnosis of bi-polar
PT affective disorder and schizoaffective disorder manic type.
XX
PS Disclosure; Fig 2A-2U; 99pp; English.
XX
CC The present genomic sequence is that of human pituitary adenylate cyclase
CC activating polypeptide (PACAP). This gene is localised to the short arm
CC of chromosome 18p11 and synthesised in the retinal ganglion cells. The
CC genomic DNA is isolated from lymphocytes of individuals. The coding
CC region of the gene has a dinucleotide repeat polymorphism marker known as
CC W3440 and an STS marker, ADCYAP1 is located in the 3'UTR. These markers
CC are useful to map the gene physically and genetically to the chromosome
CC 18. This sequence can be used to design probes for detecting the presence
CC or absence of a mutation in the wild type PACAP gene. It is also useful
CC for genetic diagnosis of neuropsychiatric disorders like bi- polar
CC affective disorder(BP-I) and schizoaffective disorder manic type (SAD-M),
CC which are more valid than current clinical diagnosis based on clinical
CC observation and subjective reports. It is used for identifying compounds
CC that modulate PACAP gene expression which are useful for treatment of
CC PACAP disorders
XX
SQ Sequence 84607 BP; 23055 A; 17765 C; 18980 G; 24744 T; 0 U; 63 Other;

Query Match 2.5%; Score 155.2; DB 2; Length 84607;
Best Local Similarity 75.7%; Pred. No. 2.4e-26;
Matches 218; Conservative 0; Mismatches 68; Indels 2; Gaps 2;

```
Qy       2175 CTGGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCG 2234
Db      55989 CAGTGTAACCTTCACTTGAATAATGTGTGTTGTACGCATTAAGTAATTTTTCATT-TCTC 56047

Qy       2235 CTGACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTC-TACA 2293
Db      56048 ACCCCATCCCACCATCCCACCACTCTGAGTCTCCACCATCTATAATTCCACACTCTTATG 56107

Qy       2294 TCTATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGTCTTT 2353
Db      56108 TCCATGTGTACACATTGTTTAGCTCCCACTTATGAGTGAGAACATGTAGTATTTGACTTT 56167

Qy       2354 CTGTGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTGCAAA 2413
Db      56168 CTGTTTCTGAGCTATTTCACTTAAGATAATGGCCTCCAGTTCTGTCCATGTTGCTGCAAA 56227

Qy       2414 AGGCATGATTTCATTCTTTTTTATGGCTATGTTCTTTCCCAATTTAGA 2461
Db      56228 AGACATGATTTCATTCTTTTTTATGGCTGAGTAGTATTCCATTGTACA 56275
```

RESULT 40
ADH18345
ID ADH18345 standard; DNA; 43445 BP.
XX
AC ADH18345;
XX
DT 11-MAR-2004 (first entry)
XX
DE Human apolipoprotein B (ApoB) DNA 3 - SEQ ID 334.
XX
KW apolipoprotein B; ApoB; antiarteriosclerotic; cardiant; antidiabetic;

```
KW   anorectic; lipid; cholesterol metabolism; atherosclerosis;
KW   diabetes Type 2; obesity; hyperlipidaemia; cardiovascular; gene therapy;
KW   human; ds.
XX
OS   Homo sapiens.
XX
PN   WO2003097662-A1.
XX
PD   27-NOV-2003.
XX
PF   15-MAY-2003; 2003WO-US015493.
XX
PR   15-MAY-2002; 2002US-00147196.
PR   13-NOV-2002; 2002US-0426324P.
XX
PA   (ISIS-) ISIS PHARM INC.
XX
PI   Crooke RM,  Graham MJ;
XX
DR   WPI; 2004-022840/02.
XX
PT   New antisense compound, useful for preparing a composition for treating
PT   abnormal lipid or cholesterol metabolism, atherosclerosis, diabetes Type
PT   2, obesity, hyperlipidemia or cardiovascular disease.
XX
PS   Example 33; SEQ ID NO 334; 405pp; English.
XX
CC   The invention relates to a novel antisense compound targeted to a nucleic
CC   acid molecule encoding human apolipoprotein B (ApoB) which specifically
CC   hybridises with and inhibits the expression of human apolipoprotein B.
CC   The compound of the invention demonstrates antiarteriosclerotic,
CC   cardiant, antidiabetic and anorectic activities and may be useful for
CC   preparing a composition for treating abnormal lipid or cholesterol
CC   metabolism, atherosclerosis, diabetes Type 2, obesity, hyperlipidaemia or
CC   cardiovascular disease. Furthermore, the compound has gene therapy
CC   applications. The current sequence is that of the human ApoB DNA of the
CC   invention.
XX
SQ   Sequence 43445 BP; 12661 A; 9313 C; 9323 G; 12148 T; 0 U; 0 Other;

  Query Match             2.5%;  Score 155;  DB 12;  Length 43445;
  Best Local Similarity   76.6%;  Pred. No. 2e-26;
  Matches  203;  Conservative    0;  Mismatches   60;  Indels    2;  Gaps    1;

Qy       2170 GCCCCCTGGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACC 2229
             ||   | ||||||  |||||||   |  |  |  | |||||||  ||||||||||||||||
Db      12940 GCTTTCAGGGTATCCATCACCCAGACAACATATAGTGTACCCACTAAGTAATTTCTCACC 12999

Qy       2230 ATCCGCTGACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTC 2289
             ||||     | ||      ||     ||||||||| ||| |||||| |||||||||||| |
Db      13000 ATCCATCTCCCTCCACTTCCACA--CCTTCTGAGTCTCAATTGTCTATCATTCCACACAC 13057

Qy       2290 TACATCTATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGT 2349
             ||  ||  ||||| |||||||||  |||| |||||||| ||| |||| ||||||||||
Db      13058 TATGTCCTTGTGTGCACATTATTTCACTCCCACTTATAAATGACAACACGCAATATTTGT 13117

Qy       2350 CTTTCTGTGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTG 2409
             ||||||||| ||||| ||||| ||||| || |||| || ||||||| ||| | | |||||
Db      13118 CTTTCTGTGACTGTCCTGTTTCACTTAAGACAATGACCTCCAGTTCCATCCATGTTGCTG 13177

Qy       2410 CAAAAGGCATGATTTCATTCTTTTT 2434
             |||| | |||||||| |||||||||
Db      13178 CAAATGACATGATTTTATTCTTTTT 13202


RESULT 41
ADO32886
ID   ADO32886 standard; DNA; 43445 BP.
XX
AC   ADO32886;
XX
DT   12-AUG-2004  (first entry)
XX
DE   Human apolipoprotein B (ApoB) gDNA - SEQ ID 334.
XX
KW   apolipoprotein B; ApoB; cardiovascular; antiarteriosclerotic;
KW   antilipaemic; antidiabetic; anorectic; cardiant; vasotropic; hypotensive;
KW   anabolic; eating disorder; cytostatic; endocrine; vasotropic;
KW   neuroprotective; nootropic; lipid; cholesterol metabolism;
KW   hyperlipoproteinaemia; hyperlipidaemia; hypercholesterolaemia;
KW   Von Gierke's disease; lipodystrophy; Cushing's syndrome;
KW   sexual ateliotic dwarfism; hyperthyroidism; hypertension;
KW   anorexia nervosa; Werner's syndrome; hepatoma; multiple myeloma; uraemia;
KW   impotence; obstructive liver disease; Alzheimer's; dementia; diabetes;
KW   obesity; atherosclerosis; human; ds; chromosome 2p23-2p24.
XX
OS   Homo sapiens.
XX
PN   WO2004044181-A2.
XX
PD   27-MAY-2004.
XX
PF   13-NOV-2003; 2003WO-US036411.
XX
PR   13-NOV-2002; 2002US-0426234P.
PR   15-MAY-2003; 2003WO-US015493.
XX
PA   (ISIS-) ISIS PHARM INC.
XX
PI   Crooke R,  Graham M,  Lemonidis-Tarbet K,  Dobie KW;
XX
DR   WPI; 2004-420321/39.
XX
PT   Antisense oligonucleotide compound that inhibits expression of mRNA
PT   encoding human apolipoprotein B, useful for treating hyperlipidemia,
PT   diabetes, obesity, von Gierke's disease, lipodystrophies, Cushing's
PT   syndrome.
XX
PS   Example 33; SEQ ID NO 334; 483pp; English.
XX
CC   The invention relates to a novel antisense compound where the compound
CC   hybridises to and inhibits expression of mRNA encoding human
CC   apolipoprotein B (ApoB) after 16-24 hours by at least 30% in 80%
CC   confluent HepG2 cells in culture at a concentration of 150 nM. The
CC   compound of the invention demonstrates cardiovascular,
CC   antiarteriosclerotic, antilipaemic, antidiabetic, anorectic, cardiant,
CC   vasotropic, hypotensive, anabolic, eating disorder-related, cytostatic,
CC   endocrine, vasotropic, neuroprotective and nootropic activities and may
CC   be useful for inhibiting the expression of apolipoprotein B in cells or
CC   tissues in vivo in order to address a condition associated with abnormal
CC   lipid or cholesterol metabolism. The compound may be useful for
CC   decreasing circulating lipoprotein levels, triglyceride levels,
CC   cholesterol levels, lipid levels, fatty acid levels, acute phase
CC   reactants and chylomicrons and thus may be utilised during treatment of
CC   hyperlipoproteinaemia, hyperlipidaemia, hypercholesterolaemia,
CC   cardiovascular disorders, Von Gierke's disease, lipodystrophy, Cushing's
CC   syndrome, sexual ateliotic dwarfism, hyperthyroidism, hypertension,
CC   anorexia nervosa, Werner's syndrome, hepatoma, multiple myeloma, uraemia,
CC   impotence, obstructive liver disease, Alzheimer's disease, dementia,
CC   diabetes, obesity and atherosclerosis. The current sequence is that of
CC   the human apolipoprotein B (ApoB) gDNA - SEQ ID 334 of the invention
CC   which is located at chromosome 2p23-2p24.
XX
SQ   Sequence 43445 BP; 12661 A; 9314 C; 9322 G; 12148 T; 0 U; 0 Other;

  Query Match             2.5%;  Score 155;  DB 12;  Length 43445;
  Best Local Similarity   76.6%;  Pred. No. 2e-26;
  Matches  203;  Conservative    0;  Mismatches   60;  Indels    2;  Gaps    1;

Qy       2170 GCCCCCTGGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACC 2229
             ||   | ||||||  |||||||   |  |  |  | |||||||  ||||||||||||||||
Db      12940 GCTTTCAGGGTATCCATCACCCAGACAACATATAGTGTACCCACTAAGTAATTTCTCACC 12999

Qy       2230 ATCCGCTGACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTC 2289
             ||||     | ||      ||     ||||||||| ||| |||||| |||||||||||| |
Db      13000 ATCCATCTCCCTCCACTTCCACA--CCTTCTGAGTCTCAATTGTCTATCATTCCACACAC 13057
```

```
Qy       2290 TACATCTATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGT 2349
              ||  ||  ||||| ||||||||||  |||| ||||||||| ||| |||| ||||||||||||
Db      13058 TATGTCCTTGTGTGCACATTATTTCACTCCCACTTATAAATGACAACACGCAATATTTGT 13117

Qy       2350 CTTTCTGTGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTG 2409
              ||||||||| ||||| ||||| ||||| || |||| || ||||||| ||| | | |||||
Db      13118 CTTTCTGTGACTGTCCTGTTTCACTTAAGACAATGACCTCCAGTTCCATCCATGTTGCTG 13177

Qy       2410 CAAAAGGCATGATTTCATTCTTTTT 2434
              |||| | |||||||| |||||||||
Db      13178 CAAATGACATGATTTTATTCTTTTT 13202
```

RESULT 42
AAA34964
ID AAA34964 standard; DNA; 16310 BP.
XX
AC AAA34964;
XX
DT 28-JUL-2000 (first entry)
XX
DE Human adenosine receptor related polynucleotide SEQ ID NO:2653.
XX
KW Human; adenosine receptor; low adenosine antisense oligonucleotide;
KW phosphorothioate; impaired respiration; inflammation; allergy;
KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX
OS Homo sapiens.
XX
PN WO200009525-A2.
XX
PD 24-FEB-2000.
XX
PF 03-AUG-1999; 99WO-US017712.
XX
PR 03-AUG-1998; 98US-0095212P.
XX
PA (UYEC-) UNIV EAST CAROLINA.
XX
PI Nyce JW;
XX
DR WPI; 2000-205971/18.
XX
PT New antisense oligonucleotides useful for treating e.g. pulmonary
PT vasoconstruction, inflammation, allergies, asthma, hypertension,
PT bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT cancers.
XX
PS Disclosure; Page 815-819; 1343pp; English.
XX
CC The present invention describes a new composition comprising an antisense
CC oligonucleotide (ON) with low adenosine (up to 15%), which targets
CC nucleic acids involved in bronchoconstriction, allergies, and/or
CC inflammation. The ON can have antiinflammatory, antiallergic,
CC antiasthmatic, cytostatic and analgesic activities. The compositions are
CC useful for the treatment of diseases associated with inflammation,
CC impaired airways, including lung disease and diseases whose secondary
CC effects afflict the lungs of a subject. They can be used for treating
CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies, asthma,
CC impeded respiration, respiratory distress syndrome, pain, cystic
CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
CC pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
CC carcinomas, and cancers which may metastasise to the lungs, including
CC breast and prostate cancer. The reduction of the adenosine content of the
CC ONs reduces side effects. The A-containing ONs break down with the
CC release of deoxyadenosine which activates adenosine receptors causing
CC bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
CC nucleotide sequences given in the sequence listing from the present
CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last 185
CC sequences are also called SEQ ID NO:1 to 185, but the sequences differ
CC from the previously named sequences. SEQ ID NO:11 to 1680 (AAA32323 to
CC AAA33992) are specifically claimed ONs from the present invention. N.B.
CC Sequences given in the disclosure of the present invention do not match
CC up with their corresponding SEQ ID NO: sequences given in the sequence
CC listing
XX
SQ Sequence 16310 BP; 3819 A; 4069 C; 4268 G; 4152 T; 0 U; 2 Other;

Query Match 2.5%; Score 154.8; DB 3; Length 16310;
Best Local Similarity 67.6%; Pred. No. 1.4e-26;
Matches 296; Conservative 0; Mismatches 127; Indels 15; Gaps 5;

```
Qy        252 CTTTAGTTCTCAGCATATTCGTGCACCTTCATTTATCTCTCCTTCTCTCTCAAAGCTGCA 311
              || || ||     | |   | |   ||||   |  | |      |||      |    | |
Db      10659 CTCTACTTGAGGACCTGCCCCTTTGCCTTTTATGCTTTACTAGTCTTAGCTATTATTTCT 10718

Qy        312 GTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTC 371
              |||||||   ||| | ||||| |  || ||   |||||||||||||     |||| ||||
Db      10719 GTAGGGGCAGAAAGGGGTGAT-CCCTTCCTGACCCATCATAAGGGTTATGGCCAATACTC 10777

Qy        372 CTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTT 431
              ||||| |||||||||| |||| ||||| ||| | ||| | ||||||||||||||||||||
Db      10778 CTATAATAAAAGACAGATTAACAAGAGAAAAGCATAATACATTTATTTAATCAAAGTTTT 10837

Qy        432 ACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTT 491
              |  |||||||||| |||||||||||||||||| ||||||||||||          |||
Db      10838 AGGTGACATGGGAGCCTTCAGAAATGAAGACCCAAGGACCCAGGGGAAA----ACTATTT 10893

Qy        492 TTTTGCTGAGGTTCGATGAAGAATGGATAGC-ATGTAGCCATGTAGATTAGACAAAAGGA 550
              || |||| || || |||||||||||  | |||  || ||  |||||  | |   ||| |
Db      10894 TTATGCTTAGATTTGATGAAGAATGAACAGCTGTGCAGAAATGTAATTGAACAAAAGGAG 10953

Qy        551 TATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCT 610
              ||| ||||| |||| | |||| |  |          | ||     | ||   || |||||
Db      10954 TATCATCTAATGGTCACAGACTGGGACTGGGGGGATCCCAGCAAGGCCTGGCCATATTCT 11013

Qy        611 TCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGA-CTC 669
              ||||| ||||||         |||| ||||||| |||||||    |||| |||||| | |||
Db      11014 TCTTGGTCTCTC--------TGTACAGCATTCCTTCCTCCCAGGTATAGGGCAGAACCTC 11065

Qy        670 TTCTTCAATGAGGGTCTT 687
              ||||  ||||||||||||
Db      11066 TTCTGGAATGAGGGTCTT 11083
```

RESULT 43
AAF21086
ID AAF21086 standard; DNA; 16310 BP.
XX
AC AAF21086;
XX
DT 14-MAR-2001 (first entry)
XX
DE Human low adenosine antisense oligonucleotide related sequence #2653.
XX
KW Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW human; airway disorder; bronchoconstriction; lung inflammation;
KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
KW respiratory obstruction; pulmonary obstruction; impeded respiration;
KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200062736-A2.
XX

PD 26-OCT-2000.
XX
PF 24-MAR-2000; 2000WO-US008020.
XX
PR 06-APR-1999; 99US-0127958P.
XX
PA (UYEC-) UNIV EAST CAROLINA.
PA (NYCE/) NYCE J W.
XX
PI Nyce JW;
XX
DR WPI; 2000-679539/66.
XX
PT Low adenosine (A) content antisense oligonucleotides which do not trigger
PT adenosine receptors during metabolism, useful e.g. for treating cancers
PT and respiratory obstructions.
XX
PS Disclosure; Page 887-891; 1592pp; English.
XX
CC The present invention describes low adenosine (A) content antisense
CC oligonucleotides and compositions (I) comprising them. In the antisense
CC oligonucleotides the A is replaced by a 'Universal' or alternative base.
CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC The antisense oligonucleotides and (I) can be used to down-regulate the
CC expression and or activity of target polypeptides associated with
CC lung/respiratory disorders and malignancies, such as stimulating and
CC activating peptide factors and transmitters, transcription factors,
CC immunoglobulins and antibodies, antibody receptors, cytokines and
CC chemokines, endogenously produced specific and non-specific enzymes,
CC binding proteins, adhesion molecules and their receptors, cytokine and
CC chemokine receptors, adenosine receptors, bradykinin receptors, central
CC nervous system (CNS) and peripheral nervous and non-nervous system
CC receptors, CNS and peripheral nervous and non-nervous system peptide
CC transmitters, defensins, growth factors, vasoactive peptides and
CC receptors, binding proteins and malignancy associated proteins. The
CC antisense oligonucleotides may be used in this way to treat disorders
CC including respiratory obstruction (especially pulmonary obstruction
CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies) and/or
CC surfactant hypoproduction which are associated with a disease or
CC condition selected from pulmonary vasoconstriction, inflammation,
CC allergies, asthma, impeded respiration, respiratory distress syndrome
CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC fragments and antisense oligonucleotides used in the exemplification of
CC the present invention
XX
SQ Sequence 16310 BP; 3819 A; 4069 C; 4268 G; 4152 T; 0 U; 2 Other;

Query Match 2.5%; Score 154.8; DB 3; Length 16310;
Best Local Similarity 67.6%; Pred. No. 1.4e-26;
Matches 296; Conservative 0; Mismatches 127; Indels 15; Gaps 5;

```
Qy      252 CTTTAGTTCTCAGCATATTCGTGCACCTTCATTTATCTCTCCTTCTCTCTCAAAGCTGCA 311
            || || ||     | |   | |   ||||   | | |     |||     |    | |
Db    10659 CTCTACTTGAGGACCTGCCCCTTTGCCTTTTATGCTTTACTAGTCTTAGCTATTATTTCT 10718

Qy      312 GTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAAACTC 371
            |||||||   ||| | ||||| |  || ||   |||||||||||||     |||| ||||
Db    10719 GTAGGGGCAGAAAGGGGTGAT-CCCTTCCTGACCCATCATAAGGGTTATGGCCAATACTC 10777

Qy      372 CTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTT 431
            ||||| |||||||||| |||| ||||| ||| | ||| | |||||||||||||||||||
Db    10778 CTATAATAAAAGACAGATTAACAAGAGAAAAGCATAATACATTTATTTAATCAAAGTTTT 10837

Qy      432 ACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTT 491
            |  |||||||||| |||||||||||||||||||| ||||||||||||        |||
Db    10838 AGGTGACATGGGAGCCTTCAGAAATGAAGACCCAAGGACCCAGGGGAAA----ACTATTT 10893

Qy      492 TTTTGCTGAGGTTCGATGAAGAATGGATAGC-ATGTAGCCATGTAGATTAGACAAAAGGA 550
            || |||| || || |||||||||| | |||  || ||  |||||  | |   ||| |
Db    10894 TTATGCTTAGATTTGATGAAGAATGAACAGCTGTGCAGAAATGTAATTGAACAAAAGGAG 10953

Qy      551 TATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCT 610
            ||| ||||| |||| |  |||| |   |          | ||    | ||   || |||||
Db    10954 TATCATCTAATGGTCACAGACTGGGACTGGGGGGATCCCAGCAAGGCCTGGCCATATTCT 11013

Qy      611 TCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGA-CTC 669
            ||||| ||||||        |||| ||||||| |||||||   |||| ||||||  | |||
Db    11014 TCTTGGTCTCTC--------TGTACAGCATTCCTTCCTCCCAGGTATAGGGCAGAACCTC 11065

Qy      670 TTCTTCAATGAGGGTCTT 687
            ||||  ||||||||||||
Db    11066 TTCTGGAATGAGGGTCTT 11083
```

RESULT 44
ABZ96780
ID ABZ96780 standard; DNA; 16310 BP.
XX
AC ABZ96780;
XX
DT 17-OCT-2003 (first entry)
XX
DE Human nucleic acid sequence.
XX
KW Human; antisense; lung dysfunction; nasal airway dysfunction;
KW antiinflammatory steroid; ubiquinone; antiinflammatory; antiallergic;
KW antiasthmatic; hypotensive; immunosuppressive; cytostatic; gene therapy;
KW antisense gene therapy; respiratory; lung; adenosine sensitivity;
KW adenosine receptor; bronchodilation; bronchoconstriction; lung allergy;
KW lung inflammation; respiratory disease; ds.
XX
OS Homo sapiens.
XX
PN WO200285308-A2.
XX
PD 31-OCT-2002.
XX
PF 23-APR-2002; 2002WO-US013135.
XX
PR 24-APR-2001; 2001US-0286137P.
XX
PA (EPIG-) EPIGENESIS PHARM INC.
XX
PI Nyce JW, Li Y, Sandrasagra A, Katz E, Pabalan J, Aguilar D;
PI Miller S, Tang L, Shahabuddin S;
XX
DR WPI; 2003-229219/22.
XX
PT Pharmaceutical composition for treating ailments associated with impaired
PT respiration, has oligo(s) antisense to specific gene(s) or its
PT corresponding RNAs, and glucocorticoid or non-glucocorticoid steroid or
PT ubiquinone.
XX
PS Disclosure; SEQ ID NO 12022; 872pp; English.
XX
CC The invention relates to a novel pharmaceutical composition, which has a
CC first active agent comprising an oligonucleotide antisense to the
CC initiation codon, coding region, 5' or 3' end genomic flanking regions,
CC 5' and 3' intron-exon junctions, or regions within 2-10 nucleotides of
CC junctions of genes encoding a polypeptide associated with lung and/or
CC nasal airway dysfunction and a second active agent comprising an
CC antiinflammatory steroid and ubiquinone. A composition of the invention
CC has antiinflammatory, antiallergic, antiasthmatic, hypotensive,
CC immunosuppressive, and cytostatic activity. The composition may have a
CC use in antisense gene therapy. The composition is useful for treating or
CC preventing a respiratory, lung or malignant disease or condition, also
CC for enhancing the prophylactic or therapeutic respiratory effect of an
CC antiinflammatory steroid in a subject, for reducing or depleting levels
CC of, or reducing sensitivity to adenosine, reducing levels of adenosine
CC receptor, producing bronchodilation, increasing levels of ubiquinone or
CC lung surfactant in a subject's tissue, or treating bronchoconstriction,
CC lung inflammation, lung allergies, or a respiratory disease or condition.

```
CC   Note: The sequence data for this patent is not represented in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 16310 BP; 3819 A; 4069 C; 4268 G; 4152 T; 0 U; 2 Other;

  Query Match             2.5%;  Score 154.8;  DB 10;  Length 16310;
  Best Local Similarity   67.6%;  Pred. No. 1.4e-26;
  Matches  296;  Conservative    0;  Mismatches  127;  Indels   15;  Gaps    5;

Qy        252 CTTTAGTTCTCAGCATATTCGTGCACCTTCATTTATCTCTCCTTCTCTCTCAAAGCTGCA 311
              || || ||     | |   | |   ||||   |  | |     |||     |    | |
Db      10659 CTCTACTTGAGGACCTGCCCCTTTGCCTTTTATGCTTTACTAGTCTTAGCTATTATTTCT 10718

Qy        312 GTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTC 371
              |||||||   ||| | ||||| |  || ||   ||||||||||||     |||| ||||
Db      10719 GTAGGGGCAGAAAGGGGTGAT-CCCTTCCTGACCCATCATAAGGGTTATGGCCAATACTC 10777

Qy        372 CTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTT 431
              ||||| |||||||||| |||| ||||| ||| | ||| | ||||||||||||||||||||
Db      10778 CTATAATAAAAGACAGATTAACAAGAGAAAAGCATAATACATTTATTTAATCAAAGTTTT 10837

Qy        432 ACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTT 491
              |  |||||||||| |||||||||||||||||||| ||||||||||||       |||
Db      10838 AGGTGACATGGGAGCCTTCAGAAATGAAGACCCAAGGACCCAGGGGAAA----ACTATTT 10893

Qy        492 TTTTGCTGAGGTTCGATGAAGAATGGATAGC-ATGTAGCCATGTAGATTAGACAAAAGGA 550
              || |||| || || |||||||||| | |||  || ||  |||||  | |   ||| |
Db      10894 TTATGCTTAGATTTGATGAAGAATGAACAGCTGTGCAGAAATGTAATTGAACAAAAGGAG 10953

Qy        551 TATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCT 610
              ||| ||||| |||| |  |||| |   |       | ||    | ||    || |||||
Db      10954 TATCATCTAATGGTCACAGACTGGGACTGGGGGGATCCCAGCAAGGCCTGGCCATATTCT 11013

Qy        611 TCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGA-CTC 669
              ||||| ||||||         |||| ||||||| |||||||   |||| |||||| | |||
Db      11014 TCTTGGTCTCTC--------TGTACAGCATTCCTTCCTCCCAGGTATAGGGCAGAACCTC 11065

Qy        670 TTCTTCAATGAGGGTCTT 687
              ||||  ||||||||||||
Db      11066 TTCTGGAATGAGGGTCTT 11083


RESULT 45
ABD20629
ID   ABD20629 standard; DNA; 16310 BP.
XX
AC   ABD20629;
XX
DT   29-JUL-2004  (first entry)
XX
DE   Human pulmonary and inflammatory target DNA #240.
XX
KW   Human; antisense; bronchoconstriction; allergy; hyposecretion; pain;
KW   respiratory tract inflammation; adenosine sensitivity; lung; cancer;
KW   surfactant depletion; antiallergic; antiinflammatory; antiasthmatic;
KW   analgesic; hypotensive; immunosuppressive; cytostatic; cystic fibrosis;
KW   beta-adrenergic agonist; respiratory disease; pulmonary vasoconstriction;
KW   respiratory distress syndrome; allergic rhinitis; pulmonary hypertension;
KW   emphysema; chronic obstructive pulmonary disease; cancer; bronchitis;
KW   pulmonary transplantation rejection; ds.
XX
OS   Homo sapiens.
XX
PN   WO200285309-A2.
XX
PD   31-OCT-2002.
XX
PF   23-APR-2002; 2002WO-US013143.
XX
PR   24-APR-2001; 2001US-0286036P.
XX
PA   (EPIG-) EPIGENESIS PHARM INC.
XX
PI   Nyce JW,  Li Y,  Sandrasagra A,  Katz E,  Pabalan J,  Aguilar D;
PI   Miller S,  Tang L,  Shahabuddin S;
XX
DR   WPI; 2003-093058/08.
XX
PT   Pharmaceutical composition for treating asthma, has antisense
PT   oligonucleotide containing less percentage of adenosine, targeted to
PT   nucleic acids associated with lung airway or lung dysfunction, and
PT   bronchodilating agent.
XX
PS   Claim 15; SEQ ID NO 12022; 763pp; English.
XX
CC   This invention describes a novel composition (a) a first active agent,
CC   comprising oligonucleotides, effective for alleviating
CC   bronchoconstriction, respiratory tract inflammation, allergies and
CC   reducing adenosine sensitivity, levels of adenosine (A) or (A) receptors,
CC   surfactant depletion or hyposecretion, when administered to a mammal. The
CC   oligonucleotides are derived from a gene encoding or regulating
CC   expression of a target polypeptide associated with lung airway or lung
CC   dysfunction or cancer and can be anti-sense to the corresponding mRNA.
CC   The invention also describes a kit, that comprises: (a) a delivery
CC   device, in separate containers, (b) the oligonucleotides, (c)
CC   instructions for adding a carrier and for use of the kit. The composition
CC   of the invention has antiallergic, antiinflammatory, antiasthmatic,
CC   analgesic, hypotensive, immunosuppressive and cytostatic activity, is a
CC   beta-adrenergic agonist. The composition is useful for preventing or
CC   treating a respiratory, lung or malignant disease. The administered
CC   composition comprises oligo and is administered to reduce the production
CC   or availability, or to increase the degradation of the target mRNA or to
CC   reduce the amount of target polypeptide present in the lungs. The
CC   pulmonary obstruction, and/or bronchoconstriction and/or lung
CC   inflammation, allergies and/or surfactant hypoproduction are associated
CC   with a disease or condition such as pulmonary vasoconstriction,
CC   inflammation, allergies, asthma, impeded respiration, respiratory
CC   distress syndrome, pain, cystic fibrosis, allergic rhinitis, pulmonary
CC   hypertension, emphysema, chronic obstructive pulmonary disease, pulmonary
CC   transplantation rejection, pulmonary infections, bronchitis or cancer.
CC   The reduced adenosine content of the anti-sense oligos corresponding to
CC   thymidines present in the target RNA serves to prevent the breakdown of
CC   the oligonucleotides into products that free adenosine into the system
CC   e.g., lung, brain, heart, kidney, etc, tissue environment and thereby, to
CC   prevent any unwanted effects due to it
XX
SQ   Sequence 16310 BP; 3819 A; 4070 C; 4267 G; 4152 T; 0 U; 2 Other;

  Query Match             2.5%;  Score 154.8;  DB 11;  Length 16310;
  Best Local Similarity   67.6%;  Pred. No. 1.4e-26;
  Matches  296;  Conservative    0;  Mismatches  127;  Indels   15;  Gaps    5;

Qy        252 CTTTAGTTCTCAGCATATTCGTGCACCTTCATTTATCTCTCCTTCTCTCTCAAAGCTGCA 311
              || || ||     | |   | |   ||||   |  | |     |||     |    | |
Db      10659 CTCTACTTGAGGACCTGCCCCTTTGCCTTTTATGCTTTACTAGTCTTAGCTATTATTTCT 10718

Qy        312 GTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTC 371
              |||||||   ||| | ||||| |  || ||   ||||||||||||     |||| ||||
Db      10719 GTAGGGGCAGAAAGGGGTGAT-CCCTTCCTGACCCATCATAAGGGTTATGGCCAATACTC 10777

Qy        372 CTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTT 431
              ||||| |||||||||| |||| ||||| ||| | ||| | ||||||||||||||||||||
Db      10778 CTATAATAAAAGACAGATTAACAAGAGAAAAGCATAATACATTTATTTAATCAAAGTTTT 10837

Qy        432 ACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTT 491
              |  |||||||||| |||||||||||||||||||| ||||||||||||       |||
Db      10838 AGGTGACATGGGAGCCTTCAGAAATGAAGACCCAAGGACCCAGGGGAAA----ACTATTT 10893

Qy        492 TTTTGCTGAGGTTCGATGAAGAATGGATAGC-ATGTAGCCATGTAGATTAGACAAAAGGA 550
              || |||| || || |||||||||| | |||  || ||  |||||  | |   ||| |
Db      10894 TTATGCTTAGATTTGATGAAGAATGAACAGCTGTGCAGAAATGTAATTGAACAAAAGGAG 10953

Qy        551 TATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCT 610
              ||| ||||| |||| |  |||| |   |       | ||    | ||    || |||||
Db      10954 TATCATCTAATGGTCACAGACTGGGACTGGGGGGATCCCAGCAAGGCCTGGCCATATTCT 11013
```

```
Qy        611 TCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGA-CTC 669
              ||||| ||||||       |||| ||||||| |||||||   |||| |||||| | |||
Db      11014 TCTTGGTCTCTC--------TGTACAGCATTCCTTCCTCCCAGGTATAGGGCAGAACCTC 11065

Qy        670 TTCTTCAATGAGGGTCTT 687
              ||||  ||||||||||||
Db      11066 TTCTGGAATGAGGGTCTT 11083
```

RESULT 46
AAA34965
ID AAA34965 standard; DNA; 17634 BP.
XX
AC AAA34965;
XX
DT 28-JUL-2000 (first entry)
XX
DE Human adenosine receptor related polynucleotide SEQ ID NO:2654.
XX
KW Human; adenosine receptor; low adenosine antisense oligonucleotide;
KW phosphorothioate; impaired respiration; inflammation; allergy;
KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX
OS Homo sapiens.
XX
PN WO200009525-A2.
XX
PD 24-FEB-2000.
XX
PF 03-AUG-1999; 99WO-US017712.
XX
PR 03-AUG-1998; 98US-0095212P.
XX
PA (UYEC-) UNIV EAST CAROLINA.
XX
PI Nyce JW;
XX
DR WPI; 2000-205971/18.
XX
PT New antisense oligonucleotides useful for treating e.g. pulmonary
PT vasoconstruction, inflammation, allergies, asthma, hypertension,
PT bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT cancers.
XX
PS Disclosure; Page 819-823; 1343pp; English.
XX
CC The present invention describes a new composition comprising an antisense
CC oligonucleotide (ON) with low adenosine (up to 15%), which targets
CC nucleic acids involved in bronchoconstriction, allergies, and/or
CC inflammation. The ON can have antiinflammatory, antiallergic,
CC antiasthmatic, cytostatic and analgesic activities. The compositions are
CC useful for the treatment of diseases associated with inflammation,
CC impaired airways, including lung disease and diseases whose secondary
CC effects afflict the lungs of a subject. They can be used for treating
CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies, asthma,
CC impeded respiration, respiratory distress syndrome, pain, cystic
CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
CC pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
CC carcinomas, and cancers which may metastasise to the lungs, including
CC breast and prostate cancer. The reduction of the adenosine content of the
CC ONs reduces side effects. The A-containing ONs break down with the
CC release of deoxyadenosine which activates adenosine receptors causing
CC bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
CC nucleotide sequences given in the sequence listing from the present
CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last 185
CC sequences are also called SEQ ID NO:1 to 185, but the sequences differ
CC from the previously named sequences. SEQ ID NO:11 to 1680 (AAA32323 to
CC AAA33992) are specifically claimed ONs from the present invention. N.B.
CC Sequences given in the disclosure of the present invention do not match
CC up with their corresponding SEQ ID NO: sequences given in the sequence
CC listing
XX
SQ Sequence 17634 BP; 4117 A; 4457 C; 4576 G; 4482 T; 0 U; 2 Other;

Query Match 2.5%; Score 154.8; DB 3; Length 17634;
Best Local Similarity 67.6%; Pred. No. 1.4e-26;
Matches 296; Conservative 0; Mismatches 127; Indels 15; Gaps 5;

```
Qy        252 CTTTAGTTCTCAGCATATTCGTGCACCTTCATTTATCTCTCCTTCTCTCTCAAAGCTGCA 311
              || || ||     | |   | |   ||||   |  | |     |||     |    | |
Db      11983 CTCTACTTGAGGACCTGCCCCTTTGCCTTTTATGCTTTACTAGTCTTAGCTATTATTTCT 12042

Qy        312 GTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTC 371
              |||||||   ||| | ||||| |  || ||   ||||||||||||     |||| ||||
Db      12043 GTAGGGGCAGAAAGGGGTGAT-CCCTTCCTGACCCATCATAAGGGTTATGGCCAATACTC 12101

Qy        372 CTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTT 431
              ||||| |||||||||| |||| ||||| ||| | ||| | ||||||||||||||||||||
Db      12102 CTATAATAAAAGACAGATTAACAAGAGAAAAGCATAATACATTTATTTAATCAAAGTTTT 12161

Qy        432 ACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTT 491
              |  |||||||||| |||||||||||||||||||| ||||||||||||       |||
Db      12162 AGGTGACATGGGAGCCTTCAGAAATGAAGACCCAAGGACCCAGGGGAAA----ACTATTT 12217

Qy        492 TTTTGCTGAGGTTCGATGAAGAATGGATAGC-ATGTAGCCATGTAGATTAGACAAAAGGA 550
              || |||| || || |||||||||||| | |||  || ||  |||||  | |   ||| |
Db      12218 TTATGCTTAGATTTGATGAAGAATGAACAGCTGTGCAGAAATGTAATTGAACAAAAGGAG 12277

Qy        551 TATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCT 610
              ||| ||||| |||| |  ||||  |   |         | ||    | ||   || |||||
Db      12278 TATCATCTAATGGTCACAGACTGGGACTGGGGGGATCCCAGCAAGGCCTGGCCATATTCT 12337

Qy        611 TCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGA-CTC 669
              ||||| ||||||       |||| ||||||| |||||||   |||| |||||| | |||
Db      12338 TCTTGGTCTCTC--------TGTACAGCATTCCTTCCTCCCAGGTATAGGGCAGAACCTC 12389

Qy        670 TTCTTCAATGAGGGTCTT 687
              ||||  ||||||||||||
Db      12390 TTCTGGAATGAGGGTCTT 12407
```

RESULT 47
AAF21087
ID AAF21087 standard; DNA; 17634 BP.
XX
AC AAF21087;
XX
DT 14-MAR-2001 (first entry)
XX
DE Human low adenosine antisense oligonucleotide related sequence #2654.
XX
KW Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW human; airway disorder; bronchoconstriction; lung inflammation;
KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
KW respiratory obstruction; pulmonary obstruction; impeded respiration;
KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200062736-A2.
XX
PD 26-OCT-2000.
XX
PF 24-MAR-2000; 2000WO-US008020.
XX

PR 06-APR-1999; 99US-0127958P.
XX
PA (UYEC-) UNIV EAST CAROLINA.
PA (NYCE/) NYCE J W.
XX
PI Nyce JW;
XX
DR WPI; 2000-679539/66.
XX
PT Low adenosine (A) content antisense oligonucleotides which do not trigger
PT adenosine receptors during metabolism, useful e.g. for treating cancers
PT and respiratory obstructions.
XX
PS Disclosure; Page 891-896; 1592pp; English.
XX
CC The present invention describes low adenosine (A) content antisense
CC oligonucleotides and compositions (I) comprising them. In the antisense
CC oligonucleotides the A is replaced by a 'Universal' or alternative base.
CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC The antisense oligonucleotides and (I) can be used to down-regulate the
CC expression and or activity of target polypeptides associated with
CC lung/respiratory disorders and malignancies, such as stimulating and
CC activating peptide factors and transmitters, transcription factors,
CC immunoglobulins and antibodies, antibody receptors, cytokines and
CC chemokines, endogenously produced specific and non-specific enzymes,
CC binding proteins, adhesion molecules and their receptors, cytokine and
CC chemokine receptors, adenosine receptors, bradykinin receptors, central
CC nervous system (CNS) and peripheral nervous and non-nervous system
CC receptors, CNS and peripheral nervous and non-nervous system peptide
CC transmitters, defensins, growth factors, vasoactive peptides and
CC receptors, binding proteins and malignancy associated proteins. The
CC antisense oligonucleotides may be used in this way to treat disorders
CC including respiratory obstruction (especially pulmonary obstruction
CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies) and/or
CC surfactant hypoproduction which are associated with a disease or
CC condition selected from pulmonary vasoconstriction, inflammation,
CC allergies, asthma, impeded respiration, respiratory distress syndrome
CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC fragments and antisense oligonucleotides used in the exemplification of
CC the present invention
XX
SQ Sequence 17634 BP; 4117 A; 4456 C; 4576 G; 4483 T; 0 U; 2 Other;

Query Match 2.5%; Score 154.8; DB 3; Length 17634;
Best Local Similarity 67.6%; Pred. No. 1.4e-26;
Matches 296; Conservative 0; Mismatches 127; Indels 15; Gaps 5;

```
Qy      252 CTTTAGTTCTCAGCATATTCGTGCACCTTCATTTATCTCTCCTTCTCTCTCAAAGCTGCA 311
            || || ||     | |   | |  ||||  |  | |  . ||     |    | |
Db    11983 CTCTACTTGAGGACCTGCCCCTTTGCCTTTTATGCTTTACTAGTCTTAGCTATTATTTCT 12042

Qy      312 GTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTC 371
            |||||||   ||| | ||||| |  || ||   ||||||||||||     |||| ||||
Db    12043 GTAGGGGCAGAAAGGGGTGAT-CCCTTCCTGACCCATCATAAGGGTTATGGCCAATACTC 12101

Qy      372 CTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTT 431
            ||||| |||||||||| |||| ||||| ||| | ||| | ||||||||||||||||||||
Db    12102 CTATAATAAAAGACAGATTAACAAGAGAAAAGCATAATACATTTATTTAATCAAAGTTTT 12161

Qy      432 ACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTT 491
            |  |||||||||| |||||||||||||||||||| ||||||||||||      |||
Db    12162 AGGTGACATGGGAGCCTTCAGAAATGAAGACCCAAGGACCCAGGGGAAA----ACTATTT 12217

Qy      492 TTTTGCTGAGGTTCGATGAAGAATGGATAGC-ATGTAGCCATGTAGATTAGACAAAAGGA 550
            || |||| || || |||||||||||| | |||  || ||  |||||  | |   ||| |
Db    12218 TTATGCTTAGATTTGATGAAGAATGAACAGCTGTGCAGAAATGTAATTGAACAAAAGGAG 12277

Qy      551 TATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCT 610
            ||| ||||| |||| |  ||||  |   |        | ||    | ||   || |||||
```

```
Db    12278 TATCATCTAATGGTCACAGACTGGGACTGGGGGGATCCCAGCAAGGCCTGGCCATATTCT 12337

Qy      611 TCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGA-CTC 669
            ||||| ||||||        |||| |||||||| |||||||   |||| |||||| | |||
Db    12338 TCTTGGTCTCTC--------TGTACAGCATTCCTTCCTCCCAGGTATAGGGCAGAACCTC 12389

Qy      670 TTCTTCAATGAGGGTCTT 687
            ||||  ||||||||||||
Db    12390 TTCTGGAATGAGGGTCTT 12407
```

RESULT 48
ABZ96781
ID ABZ96781 standard; DNA; 17634 BP.
XX
AC ABZ96781;
XX
DT 17-OCT-2003 (first entry)
XX
DE Human nucleic acid sequence.
XX
KW Human; antisense; lung dysfunction; nasal airway dysfunction;
KW antiinflammatory steroid; ubiquinone; antiinflammatory; antiallergic;
KW antiasthmatic; hypotensive; immunosuppressive; cytostatic; gene therapy;
KW antisense gene therapy; respiratory; lung; adenosine sensitivity;
KW adenosine receptor; bronchodilation; bronchoconstriction; lung allergy;
KW lung inflammation; respiratory disease; ds.
XX
OS Homo sapiens.
XX
PN WO200285308-A2.
XX
PD 31-OCT-2002.
XX
PF 23-APR-2002; 2002WO-US013135.
XX
PR 24-APR-2001; 2001US-0286137P.
XX
PA (EPIG-) EPIGENESIS PHARM INC.
XX
PI Nyce JW, Li Y, Sandrasagra A, Katz E, Pabalan J, Aguilar D;
PI Miller S, Tang L, Shahabuddin S;
XX
DR WPI; 2003-229219/22.
XX
PT Pharmaceutical composition for treating ailments associated with impaired
PT respiration, has oligo(s) antisense to specific gene(s) or its
PT corresponding RNAs, and glucocorticoid or non-glucocorticoid steroid or
PT ubiquinone.
XX
PS Disclosure; SEQ ID NO 12023; 872pp; English.
XX
CC The invention relates to a novel pharmaceutical composition, which has a
CC first active agent comprising an oligonucleotide antisense to the
CC initiation codon, coding region, 5' or 3' end genomic flanking regions,
CC 5' and 3' intron-exon junctions, or regions within 2-10 nucleotides of
CC junctions of genes encoding a polypeptide associated with lung and/or
CC nasal airway dysfunction and a second active agent comprising an
CC antiinflammatory steroid and ubiquinone. A composition of the invention
CC has antiinflammatory, antiallergic, antiasthmatic, hypotensive,
CC immunosuppressive, and cytostatic activity. The composition may have a
CC use in antisense gene therapy. The composition is useful for treating or
CC preventing a respiratory, lung or malignant disease or condition, also
CC for enhancing the prophylactic or therapeutic respiratory effect of an
CC antiinflammatory steroid in a subject, for reducing or depleting levels
CC of, or reducing sensitivity to adenosine, reducing levels of adenosine
CC receptor, producing bronchodilation, increasing levels of ubiquinone or
CC lung surfactant in a subject's tissue, or treating bronchoconstriction,
CC lung inflammation, lung allergies, or a respiratory disease or condition.
CC Note: The sequence data for this patent is not represented in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX

```
SQ   Sequence 17634 BP; 4117 A; 4456 C; 4576 G; 4483 T; 0 U; 2 Other;

  Query Match             2.5%;  Score 154.8;  DB 10;  Length 17634;
  Best Local Similarity   67.6%;  Pred. No. 1.4e-26;
  Matches  296;  Conservative    0;  Mismatches  127;  Indels   15;  Gaps    5;

Qy        252 CTTTAGTTCTCAGCATATTCGTGCACCTTCATTTATCTCTCCTTCTCTCTCAAAGCTGCA 311
              || || ||     | |   | |   ||||   |  | |     |||      |    | |
Db      11983 CTCTACTTGAGGACCTGCCCCTTTGCCTTTTATGCTTTACTAGTCTTAGCTATTATTTCT 12042

Qy        312 GTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTC 371
              |||||||   ||| | ||||| |  || ||   |||||||||||||     |||| ||||
Db      12043 GTAGGGGCAGAAAGGGGTGAT-CCCTTCCTGACCCATCATAAGGGTTATGGCCAATACTC 12101

Qy        372 CTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTT 431
              ||||| |||||||||| |||| ||||| ||| | ||| | ||||||||||||||||||||
Db      12102 CTATAATAAAAGACAGATTAACAAGAGAAAAGCATAATACATTTATTTAATCAAAGTTTT 12161

Qy        432 ACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTT 491
              |  ||||||||||| |||||||||||||||||||| ||||||||||||        |||
Db      12162 AGGTGACATGGGAGCCTTCAGAAATGAAGACCCAAGGACCCAGGGGAAA----ACTATTT 12217

Qy        492 TTTTGCTGAGGTTCGATGAAGAATGGATAGC-ATGTAGCCATGTAGATTAGACAAAAGGA 550
              || |||| || || |||||||||| | |||  || ||  |||||  | |   ||| |
Db      12218 TTATGCTTAGATTTGATGAAGAATGAACAGCTGTGCAGAAATGTAATTGAACAAAAGGAG 12277

Qy        551 TATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCT 610
              ||| ||||| |||| |  |||| |   |       | ||    | ||   || |||||
Db      12278 TATCATCTAATGGTCACAGACTGGGACTGGGGGGATCCCAGCAAGGCCTGGCCATATTCT 12337

Qy        611 TCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGA-CTC 669
              ||||| ||||||         |||| ||||||| |||||||   |||| ||||||  | |||
Db      12338 TCTTGGTCTCTC--------TGTACAGCATTCCTTCCTCCCAGGTATAGGGCAGAACCTC 12389

Qy        670 TTCTTCAATGAGGGTCTT 687
              ||||  ||||||||||||
Db      12390 TTCTGGAATGAGGGTCTT 12407


RESULT 49
ABD20630
ID   ABD20630 standard; DNA; 17634 BP.
XX
AC   ABD20630;
XX
DT   29-JUL-2004  (first entry)
XX
DE   Human pulmonary and inflammatory target DNA #241.
XX
KW   Human; antisense; bronchoconstriction; allergy; hyposecretion; pain;
KW   respiratory tract inflammation; adenosine sensitivity; lung; cancer;
KW   surfactant depletion; antiallergic; antiinflammatory; antiasthmatic;
KW   analgesic; hypotensive; immunosuppressive; cytostatic; cystic fibrosis;
KW   beta-adrenergic agonist; respiratory disease; pulmonary vasoconstriction;
KW   respiratory distress syndrome; allergic rhinitis; pulmonary hypertension;
KW   emphysema; chronic obstructive pulmonary disease; cancer; bronchitis;
KW   pulmonary transplantation rejection; ds.
XX
OS   Homo sapiens.
XX
PN   WO200285309-A2.
XX
PD   31-OCT-2002.
XX
PF   23-APR-2002; 2002WO-US013143.
XX
PR   24-APR-2001; 2001US-0286036P.
XX
PA   (EPIG-) EPIGENESIS PHARM INC.
XX
PI   Nyce JW,  Li Y,  Sandrasagra A,  Katz E,  Pabalan J,  Aguilar D;
PI   Miller S,  Tang L,  Shahabuddin S;
XX
DR   WPI; 2003-093058/08.
XX
PT   Pharmaceutical composition for treating asthma, has antisense
PT   oligonucleotide containing less percentage of adenosine, targeted to
PT   nucleic acids associated with lung airway or lung dysfunction, and
PT   bronchodilating agent.
XX
PS   Claim 15; SEQ ID NO 12023; 763pp; English.
XX
CC   This invention describes a novel composition (a) a first active agent,
CC   comprising oligonucleotides, effective for alleviating
CC   bronchoconstriction, respiratory tract inflammation, allergies and
CC   reducing adenosine sensitivity, levels of adenosine (A) or (A) receptors,
CC   surfactant depletion or hyposecretion, when administered to a mammal. The
CC   oligonucleotides are derived from a gene encoding or regulating
CC   expression of a target polypeptide associated with lung airway or lung
CC   dysfunction or cancer and can be anti-sense to the corresponding mRNA.
CC   The invention also describes a kit, that comprises: (a) a delivery
CC   device, in separate containers, (b) the oligonucleotides, (c)
CC   instructions for adding a carrier and for use of the kit. The composition
CC   of the invention has antiallergic, antiinflammatory, antiasthmatic,
CC   analgesic, hypotensive, immunosuppressive and cytostatic activity, is a
CC   beta-adrenergic agonist. The composition is useful for preventing or
CC   treating a respiratory, lung or malignant disease. The administered
CC   composition comprises oligo and is administered to reduce the production
CC   or availability, or to increase the degradation of the target mRNA or to
CC   reduce the amount of target polypeptide present in the lungs. The
CC   pulmonary obstruction, and/or bronchoconstriction and/or lung
CC   inflammation, allergies and/or surfactant hypoproduction are associated
CC   with a disease or condition such as pulmonary vasoconstriction,
CC   inflammation, allergies, asthma, impeded respiration, respiratory
CC   distress syndrome, pain, cystic fibrosis, allergic rhinitis, pulmonary
CC   hypertension, emphysema, chronic obstructive pulmonary disease, pulmonary
CC   transplantation rejection, pulmonary infections, bronchitis or cancer.
CC   The reduced adenosine content of the anti-sense oligos corresponding to
CC   thymidines present in the target RNA serves to prevent the breakdown of
CC   the oligonucleotides into products that free adenosine into the system
CC   e.g., lung, brain, heart, kidney, etc, tissue environment and thereby, to
CC   prevent any unwanted effects due to it
XX
SQ   Sequence 17634 BP; 4117 A; 4456 C; 4576 G; 4483 T; 0 U; 2 Other;

  Query Match             2.5%;  Score 154.8;  DB 11;  Length 17634;
  Best Local Similarity   67.6%;  Pred. No. 1.4e-26;
  Matches  296;  Conservative    0;  Mismatches  127;  Indels   15;  Gaps    5;

Qy        252 CTTTAGTTCTCAGCATATTCGTGCACCTTCATTTATCTCTCCTTCTCTCTCAAAGCTGCA 311
              || || ||     | |   | |   ||||   |  | |     |||      |    | |
Db      11983 CTCTACTTGAGGACCTGCCCCTTTGCCTTTTATGCTTTACTAGTCTTAGCTATTATTTCT 12042

Qy        312 GTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTC 371
              |||||||   ||| | ||||| |  || ||   |||||||||||||     |||| ||||
Db      12043 GTAGGGGCAGAAAGGGGTGAT-CCCTTCCTGACCCATCATAAGGGTTATGGCCAATACTC 12101

Qy        372 CTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTT 431
              ||||| |||||||||| |||| ||||| ||| | ||| | ||||||||||||||||||||
Db      12102 CTATAATAAAAGACAGATTAACAAGAGAAAAGCATAATACATTTATTTAATCAAAGTTTT 12161

Qy        432 ACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTT 491
              |  ||||||||||| |||||||||||||||||||| ||||||||||||        |||
Db      12162 AGGTGACATGGGAGCCTTCAGAAATGAAGACCCAAGGACCCAGGGGAAA----ACTATTT 12217

Qy        492 TTTTGCTGAGGTTCGATGAAGAATGGATAGC-ATGTAGCCATGTAGATTAGACAAAAGGA 550
              || |||| || || |||||||||| | |||  || ||  |||||  | |   ||| |
Db      12218 TTATGCTTAGATTTGATGAAGAATGAACAGCTGTGCAGAAATGTAATTGAACAAAAGGAG 12277

Qy        551 TATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCT 610
              ||| ||||| |||| |  |||| |   |       | ||    | ||   || |||||
Db      12278 TATCATCTAATGGTCACAGACTGGGACTGGGGGGATCCCAGCAAGGCCTGGCCATATTCT 12337

Qy        611 TCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGA-CTC 669
              ||||| ||||||         |||| ||||||| |||||||   |||| ||||||  | |||
Db      12338 TCTTGGTCTCTC--------TGTACAGCATTCCTTCCTCCCAGGTATAGGGCAGAACCTC 12389
```

```
Qy        670 TTCTTCAATGAGGGTCTT 687
              ||||  ||||||||||||
Db      12390 TTCTGGAATGAGGGTCTT 12407


RESULT 50
ACN37223/c
ID   ACN37223 standard; DNA; 20001 BP.
XX
AC   ACN37223;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Human periodontal disease related gene TNFA SEQ ID NO:133.
XX
KW   periodontal disease; polymorphism; ds; human; gene; SNP;
KW   single nucleotide polymorphism.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    287
FT                   /*tag= a
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide G,A"
FT   misc_feature    515
FT                   /*tag= b
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide T,C"
FT   misc_feature    889
FT                   /*tag= c
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide G,A"
FT   misc_feature    3127
FT                   /*tag= d
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide G,A"
FT   misc_feature    3699
FT                   /*tag= e
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide T,G"
FT   misc_feature    5145
FT                   /*tag= f
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide C,G"
FT   misc_feature    5178
FT                   /*tag= g
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide A,G"
FT   misc_feature    5397
FT                   /*tag= h
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide A,T"
FT   misc_feature    5398
FT                   /*tag= i
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide A,T"
FT   misc_feature    5910
FT                   /*tag= j
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide A,G"
FT   misc_feature    6335
FT                   /*tag= k
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide G,A"
FT   misc_feature    6545
FT                   /*tag= l
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide deletion,AG"
FT   misc_feature    6667
FT                   /*tag= m
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide T,C"
FT   misc_feature    6668
FT                   /*tag= n
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide A,G"
FT   misc_feature    6971
FT                   /*tag= o
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide A,G"
FT   misc_feature    7041
FT                   /*tag= p
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide C,A"
FT   misc_feature    7042
FT                   /*tag= q
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide A,G"
FT   misc_feature    7329
FT                   /*tag= r
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide C,G"
FT   misc_feature    7456
FT                   /*tag= s
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide C,T"
FT   misc_feature    7593
FT                   /*tag= t
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide C,G"
FT   misc_feature    7657
FT                   /*tag= u
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide A,C"
FT   misc_feature    7661
FT                   /*tag= v
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide C,T"
FT   misc_feature    7684
FT                   /*tag= w
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide A,C"
FT   misc_feature    8679
FT                   /*tag= x
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide A,G"
FT   misc_feature    8748
FT                   /*tag= y
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide A,T"
FT   misc_feature    9208
FT                   /*tag= z
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide T,C"
FT   misc_feature    9354
FT                   /*tag= aa
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide G,A"
FT   misc_feature    9376
FT                   /*tag= ab
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide A,C"
FT   misc_feature    9382
FT                   /*tag= ac
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide T,C"
FT   misc_feature    9553
FT                   /*tag= ad
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide C,A"
FT   misc_feature    9555
FT                   /*tag= ae
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide C,A"
```

```
FT   misc_feature    9591
FT                   /*tag= af
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide G,A"
FT   misc_feature    9664
FT                   /*tag= ag
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide A,G"
FT   misc_feature    9671
FT                   /*tag= ah
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide C,G"
FT   misc_feature    9720
FT                   /*tag= ai
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide C,T"
FT   misc_feature    9931
FT                   /*tag= aj
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide A,G"
FT   misc_feature    9995
FT                   /*tag= ak
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide A,G"
FT   misc_feature    10001
FT                   /*tag= al
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide A,G"
FT   misc_feature    10306
FT                   /*tag= am
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide C,G"
FT   misc_feature    10391
FT                   /*tag= an
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide T,C"
FT   misc_feature    10505
FT                   /*tag= ao
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide G,T"
FT   misc_feature    10572
FT                   /*tag= ap
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide A,C"
FT   misc_feature    10658
FT                   /*tag= aq
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide A,G"
FT   misc_feature    10727
FT                   /*tag= ar
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide A,G"
FT   misc_feature    10927
FT                   /*tag= as
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide deletion,G"
FT   misc_feature    11089
FT                   /*tag= at
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "Variable nucleotide A,G"
FT   misc_feature    11125
FT                   /*tag= au

  Query Match             2.5%;  Score 154.8;  DB 13;  Length 20001;
  Best Local Similarity   67.6%;  Pred. No. 1.5e-26;
  Matches  296;  Conservative    0;  Mismatches  127;  Indels   15;  Gaps    5;

Qy        252 CTTTAGTTCTCAGCATATTCGTGCACCTTCATTTATCTCTCCTTCTCTCTCAAAGCTGCA 311
              || || ||     | |   | |   ||||   |  | |     |||     |    | |
Db       4955 CTCTACTTGAGGACCTGCCCCTTTGCCTTTTATGCTTTACTAGTCTTAGCTATTATTTCT 4896

Qy        312 GTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTC 371
              |||||||   ||| | ||||| |  || ||   ||||||||||||     |||| ||||

Db       4895 GTAGGGGCAGAAAGGGGTGAT-CCCTTCCTGACCCATCATAAGGGTTATGGCCAATACTC 4837

Qy        372 CTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTT 431
              ||||| |||||||||| |||| ||||| ||| | ||| | ||||||||||||||||||||
Db       4836 CTATAATAAAAGACAGATTAACAAGAGAAAAGCATAATACATTTATTTAATCAAAGTTTT 4777

Qy        432 ACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTT 491
              |  |||||||||| |||||||||||||||||||| ||||||||||||        |||
Db       4776 AGGTGACATGGGAGCCTTCAGAAATGAAGACCCAAGGACCCAGGGGAAA----ACTATTT 4721

Qy        492 TTTTGCTGAGGTTCGATGAAGAATGGATAGC-ATGTAGCCATGTAGATTAGACAAAAGGA 550
              || |||| || || |||||||||| | |||  || ||  |||||  | |   ||| |
Db       4720 TTATGCTTAGATTTGATGAAGAATGAACAGCTGTGCAGAAATGTAATTGAACAAAAGGAG 4661

Qy        551 TATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCT 610
              ||| ||||| |||| |  ||||  |  |         | ||    | ||   || |||||
Db       4660 TATCATCTAATGGTCACAGACTGGGACTGGGGGGATCCCAGCAAGGCCTGGCCATATTCT 4601

Qy        611 TCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGA-CTC 669
              ||||| ||||||        |||| ||||||| |||||||   |||| ||||||  | |||
Db       4600 TCTTGGTCTCTC--------TGTACAGCATTCCTTCCTCCCAGGTATAGGGCAGAACCTC 4549

Qy        670 TTCTTCAATGAGGGTCTT 687
              ||||  ||||||||||||
Db       4548 TTCTGGAATGAGGGTCTT 4531


RESULT 51
ABK84756
ID   ABK84756 standard; cDNA; 81800 BP.
XX
AC   ABK84756;
XX
DT   14-AUG-2002  (first entry)
XX
DE   Human cDNA differentially expressed in granulocytic cells #1327.
XX
KW   Human; ss; granulocytic cell; DNA chip; bacterial infection;
KW   viral infection; parasitic infection; protozoal infection;
KW   fungal infection; sterile inflammatory disease; psoriasis;
KW   rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KW   cardiac reperfusion injury; renal reperfusion injury; ARDS;
KW   adult respiratory distress syndrome; inflammatory bowel disease;
KW   Crohn's disease; ulcerative colitis; periodontal disease;
KW   granulocyte activation; chronic inflammation; allergy.
XX
OS   Homo sapiens.
XX
PN   WO200228999-A2.
XX
PD   11-APR-2002.
XX
PF   03-OCT-2001; 2001WO-US030821.
XX
PR   03-OCT-2000; 2000US-0237189P.
XX
PA   (GENE-) GENE LOGIC INC.
XX
PI   Beazer-Barclay Y,  Weissman SM,  Yamaga S,  Vockley J;
XX
DR   WPI; 2002-435328/46.
XX
PT   Detecting granulocyte activation by detecting differential expression of
PT   genes associated with granulocyte activation, which serves as diagnostic
PT   markers that is useful for monitoring disease states and drug toxicity.
XX
PS   Claim 1; SEQ ID NO 1327; 114pp; English.
XX
CC   The invention relates to detecting (M1) granulocyte (GC) activation
CC   (GCA), by detecting the level of expression of gene(s) (Gs) identified by
CC   DNA chip analysis as given in the specification, and comparing the
CC   expression level to an expression level in an unactivated GC, where
CC   differential expression of Gs is indicative of GCA. Also included are
```

```
CC   modulating (M2) GA by contacting GC with an agent that alters the
CC   expression of at least one gene in Gs; (2) screening (M3) for an agent
CC   capable of modulating GCA or an inflammation (especially chronic) in a
CC   tissue, an allergic response in a subject, exposure of a subject to a
CC   pathogen or sterile inflammatory disease using the gene expression
CC   profile; (3) detecting (M4) an inflammation (especially chronic) in a
CC   tissue, an allergic response in a subject, exposure of a subject to a
CC   pathogen or sterile inflammatory disease, by detecting the level of
CC   expression in a sample of the tissue of gene(s) from Gs, where the level
CC   of expression of the gene is indicative of inflammation; (4) treating
CC   (M5) an inflammation (especially chronic) or in a tissue, an allergic
CC   response in a subject, exposure of a subject to a pathogen or sterile
CC   inflammatory disease, by contacting a tissue having inflammation with an
CC   agent that modulates the expression of gene(s) from Gs in the tissue. M1
CC   is useful for detecting GCA; M2 is useful for modulating GA; M3 is useful
CC   for screening an agent capable of modulating GCA preferably in an
CC   inflammation in a tissue; M4 is useful for detecting an inflammation
CC   (especially chronic) in a tissue, an allergic response in a subject,
CC   exposure of a subject to a pathogen or sterile inflammatory disease (e.g.
CC   psoriasis, rheumatoid arthritis, glomerulonephritis, asthma, thrombosis,
CC   cardiac reperfusion injury, renal reperfusion injury, ARDS, adult
CC   respiratory distress syndrome, inflammatory bowel disease, Crohn's
CC   disease, ulcerative colitis, periodontal disease; also bacterial
CC   infection, viral infection, parasitic infection, protozoal infection,
CC   fungal infection and M5 is useful for treating one of the above
CC   conditions. The present sequence represents a gene differentially
CC   expressed in granulocytes. Note: The sequence data for this patent did
CC   not form part of the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 81800 BP; 19886 A; 19955 C; 20623 G; 21336 T; 0 U; 0 Other;

  Query Match             2.5%;  Score 154.8;  DB 6;  Length 81800;
  Best Local Similarity   67.6%;  Pred. No. 2.9e-26;
  Matches  296;  Conservative    0;  Mismatches  127;  Indels   15;  Gaps    5;

Qy        252 CTTTAGTTCTCAGCATATTCGTGCACCTTCATTTATCTCTCCTTCTCTCTCAAAGCTGCA 311
              || || ||     | |   | |   ||||   |  | |     |||     |    | |
Db      49335 CTCTACTTGAGGACCTGCCCCTTTGCCTTTTATGCTTTACTAGTCTTAGCTATTATTTCT 49394

Qy        312 GTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCAACCAAAACTC 371
              |||||||   ||| | ||||| |  || ||   ||||||||||||     |||| ||||
Db      49395 GTAGGGGCAGAAAGGGGTGAT-CCCTTCCTGACCCATCATAAGGGTTATGGCCAATACTC 49453

Qy        372 CTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTAACAAATTTATTTAATCAAAGTTTT 431
              ||||| |||||||||| |||| ||||| ||| | ||| | ||||||||||||||||||||
Db      49454 CTATAATAAAAGACAGATTAACAAGAGAAAAGCATAATACATTTATTTAATCAAAGTTTT 49513

Qy        432 ACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAAACTGTCTGTTTT 491
              | ||||||||||| |||||||||||||||||||| ||||||||||||       |||
Db      49514 AGGTGACATGGGAGCCTTCAGAAATGAAGACCCAAGGACCCAGGGGAAA----ACTATTT 49569

Qy        492 TTTTGCTGAGGTTCGATGAAGAATGGATAGC-ATGTAGCCATGTAGATTAGACAAAAGGA 550
              || |||| || || |||||||||||| | |||  || ||  |||||  | |   ||| |
Db      49570 TTATGCTTAGATTTGATGAAGAATGAACAGCTGTGCAGAAATGTAATTGAACAAAAGGAG 49629

Qy        551 TATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCAGATTCT 610
              ||| ||||| |||| |  ||||  |    |        | ||    | ||    || |||||
Db      49630 TATCATCTAATGGTCACAGACTGGGACTGGGGGGATCCCAGCAAGGCCTGGCCATATTCT 49689

Qy        611 TCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAGGA-CTC 669
              ||||| ||||||        |||| ||||||| |||||||   |||| |||||| | |||
Db      49690 TCTTGGTCTCTC--------TGTACAGCATTCCTTCCTCCCAGGTATAGGGCAGAACCTC 49741

Qy        670 TTCTTCAATGAGGGTCTT 687
              ||||  ||||||||||||
Db      49742 TTCTGGAATGAGGGTCTT 49759


RESULT 52
ADS31041/c
ID   ADS31041 standard; DNA; 1042 BP.
XX
AC   ADS31041;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Human genome high complexity repeat found in the HIRA gene #74.
XX
KW   Human; ds;
KW   histone cell cycle regulation defective, S. cerevisiae homologue A; HIRA;
KW   high complexity repeat; in situ hybridisation; Southern blot;
KW   chromosome breakpoint; inherited genetic disease; neoplastic disorder;
KW   chromosome 22; DiGeorge syndrome; Velo-Cardio-facial syndrome.
XX
OS   Homo sapiens.
XX
PN   US2003224356-A1.
XX
PD   04-DEC-2003.
XX
PF   14-MAY-2001; 2001US-00854867.
XX
PR   16-MAY-2000; 2000US-00573080.
XX
PA   (KNOL/) KNOLL J H M.
PA   (ROGA/) ROGAN P K.
XX
PI   Knoll JHM,  Rogan PK;
XX
DR   WPI; 2002-062378/08.
XX
PT   Single copy genomic hybridization probes for detecting specific nucleic
PT   acid sequences in sample by in situ hybridization useful for detection of
PT   acquired or inherited genetic diseases.
XX
PS   Example 1; SEQ ID NO 74; 30pp; English.
XX
CC   The invention relates to a nucleic acid hybridisation probe comprising a
CC   labelled, single copy nucleic acids of at least 50 nucleotides, which
CC   will hybridise to a deduced single copy sequence interval in target
CC   nucleic acid (TNA) of known sequence. The single copy sequence is deduced
CC   by comparing the target nucleic acid (e.g. a disease causing gene) with a
CC   collection of high and low complexity repeat sequences as found in the
CC   genome of the organism from containing the target nucleic acid. The probe
CC   is generated by PCR on the target sequence. The probe is essentially free
CC   of blocking nucleic acid sequences which will hybridise to repeat
CC   sequences within the genome of which the TNA is a part, and is labelled
CC   with a label selected from fluorochrome-responsive labels, fluorochromes,
CC   calorimetric chemical, conjugated proteins, antibodies, antigens and
CC   their mixtures. The probe is useful in a hybridisation method, where the
CC   hybridisation method is from in situ hybridisation, Southern blot, and
CC   other methods in which nucleic acid is immobilised, where the method
CC   further comprises selecting a single copy nucleic acid which will
CC   hybridise to a duplicon or triplicon sequence domain. The probe is useful
CC   for determining the existence of previously unknown repeat sequence
CC   families in a genome. The method comprises reacting a labelled probe with
CC   the genome, causing the probe to hybridise and ascertaining if the probe
CC   hybridises to the genome at more than three preferably ten different
CC   locations as a determination of new repeat sequence family, where the
CC   determining step comprises selecting the single copy sequence from a
CC   duplicon or triplicon sequence domain. The probe is useful for
CC   determining a chromosome breakpoint and is useful in the fields for
CC   cytogenetics and molecular genetics for determining the presence of
CC   specific nucleic acid sequences in a sample of eukaryotic origin, e.g.
CC   the probes may be used to analyse specific chromosomal locations by in
CC   situ hybridisation as a detection of acquired or inherited genetic
CC   diseases especially for detection of genetic or neoplastic disorders.
CC   Unlike prior art techniques, the probe permits more precise chromosomal
CC   breakpoint determinations by in situ hybridisation. The genomic sequence
CC   comprising the human HIRA gene (histone cell cycle regulation defective,
CC   S. cerevisiae, homologue A) was analysed for single copy sequence
CC   intervals for use as probes of the invention. HIRA is located on
CC   chromosome 22 as a duplicate, deletions of 1 copy lead to DiGeorge and
CC   Velo-Cardio-facial syndromes. The present sequence is a high complexity
```

```
CC   repeat found within the human genome used to analyse the HIRA gene for
CC   repeat regions. Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic format
CC   directly from USPTO at seqdata.uspto.gov/sequence.html?DocID=20030224356.
XX
SQ   Sequence 1042 BP; 414 A; 149 C; 201 G; 252 T; 0 U; 26 Other;

  Query Match             2.5%;  Score 154.4;  DB 7;  Length 1042;
  Best Local Similarity   58.5%;  Pred. No. 4.8e-27;
  Matches  324;  Conservative    7;  Mismatches  198;  Indels   25;  Gaps    3;

Qy       1931 TTTTTTTTTAATTTTTAAAATTTGTATATATTTATGGGTTATA-AGTGAAATCTTTTTAG 1989
             |||||  || |    ||| | ||||| |||||||||||| || |  |||| || |    ||
Db       1042 TTTTTAATTGACACATAATAATTGTACATATTTATGGGGTACAGTGTGATATTTCGATAC 983

Qy       1990 ATGCATAGGTTGTATAGTGATAAAATCAGGGCTTTTAGGGTATTCATCACCTGAATGATG 2049
             ||| |||  :||| || |||| |||||||||    ||||  ||| |||||||| ||  ||
Db        982 ATGTATACAWTGTGTAATGATCAAATCAGGGTAATTAGCATATCCATCACCTCAAACATT 923

Qy       2050 TACATTGTACCCCTTAAGTAATTTCTCACCATCCGCTGACTTCTTGCCCCCTGGGTATTC 2109
             ||   | |      |   | |    ||| |||| ||      |                ||| |
Db        922 TATCATTTCTTTGTGTTGGGAACATTCAAAATCCTCTCTTCTAGCTATTTGAAAATATAC 863

Qy       2110 ATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCTGACTTCTT 2169
             |  |  | : || |    || ||  ||||   ||  |     ||  |          ||| |
Db        862 AATAAATTRTTGTTAACTATAGTCACCCTACGGTGCTACAGAACACTAGAACTTATTCCT 803

Qy       2170 GCCCCCTGGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACC 2229
              |   ||   | |   |                  |||||:|| ||||  ||  |||| |
Db        802 CCTATCTAACTGTAATT----------------TTGTAYCCWTTAACCAACCTCTCCCT 760

Qy       2230 ATCCGCTGACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTC 2289
             ||:|        |    |:||  | |||||  || ||| | |  |||  |||| || |||
Db        759 ATYC-------CCCCCTCYCCCTACCCTTCCCAGCCTCTAGTAACCACTATTCTACTCTC 707

Qy       2290 TACATCTATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGT 2349
             ||| |||||| |  ||  || |||||||  | || ||| |||||  |||||||  |||||||
Db        706 TACTTCTATGAGATCAACTTTTTTAGCTTCCACATATGAGTGAGAACATGCGGTATTTGT 647

Qy       2350 CTTTCTGTGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTG 2409
             |||||||||| ||| ||| ||| |||||  |||||| || |||| |  |||  | | ||:|:
Db        646 CTTTCTGTGCCTGGCTTATTTCACTTAGCATAATGTCCTCCAGGTTCATCCATGTTGYTR 587

Qy       2410 CAAAAGGCATGATTTCATTCTTTTTTATGGCTATGTTCTTTCCCAATTTAGATAAAGAAC 2469
             |||| | || ||||||||||||||||||||||||    |  | | ||| | |  ||| | | |
Db        586 CAAATGACAGGATTTCATTCTTTTTTATGGCTGAATAGTATTCCATTGTGTATATATACC 527

Qy       2470 ACTCGCACTTGCTC 2483
             ||     |||  ||
Db        526 ACATTTTCTTTATC 513


RESULT 53
ABL57909_0
WP Sequence split into 4 fragments  LOCUS ABL57909 Accession Abl57909
WP     Fragment Name      Begin     End
WP     ABL57909_0             1     110000
WP     ABL57909_1        100001     210000
WP     ABL57909_2        200001     310000
WP     ABL57909_3        300001     368004
ID   ABL57909 standard; DNA; 368004 BP.
XX
AC   ABL57909;
XX
DT   05-JUL-2002  (first entry)
XX
DE   Human transporter protein gene.
XX
KW   Human; transport protein; gene; gene therapy; glutamate receptor 4; SNP;
KW   brain; foetal brain; chromosome 11; single nucleotide polymorphism; ds.
XX
OS   Homo sapiens.
XX

FH   Key             Location/Qualifiers
FT   variation       replace(577,C)
FT                   /*tag=  r
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(1895,G)
FT                   /*tag=  s
FT                   /standard_name= "Single nucleotide polymorphism"
FT   CDS             2001. .366005
FT                   /*tag=  a
FT                   /product= "Human transporter protein"
FT   exon            2001. .2088
FT                   /*tag=  b
FT                   /number= 1
FT   variation       replace(2765,A)
FT                   /*tag=  t
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            3279. .3437
FT                   /*tag=  c
FT                   /number= 2
FT   variation       replace(10364,G)
FT                   /*tag=  u
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(11079,G)
FT                   /*tag=  v
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(11514,G)
FT                   /*tag=  w
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(20503,G)
FT                   /*tag=  x
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(20505,G)
FT                   /*tag=  y
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(23307,A)
FT                   /*tag=  z
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(27764,C)
FT                   /*tag=  aa
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(27766,C)
FT                   /*tag=  ab
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(28939,A)
FT                   /*tag=  ac
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(29021,C)
FT                   /*tag=  ad
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(29908,A)
FT                   /*tag=  ae
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(30398,T)
FT                   /*tag=  af
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(30440,T)
FT                   /*tag=  ag
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(31718,G)
FT                   /*tag=  ah
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(35206,C)
FT                   /*tag=  ai
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(35304,A)
FT                   /*tag=  aj
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(36362,A)
FT                   /*tag=  ak
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(36436,A)
FT                   /*tag=  al
FT                   /standard_name= "Single nucleotide polymorphism"
```

```
FT   variation       replace(43063,C)
FT                   /*tag= am
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(43202,C)
FT                   /*tag= an
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(44576,G)
FT                   /*tag= ao
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(44622,G)
FT                   /*tag= ap
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(45378,T)
FT                   /*tag= aq
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(45685,T)
FT                   /*tag= ar
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(45998,T)
FT                   /*tag= as
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(46173,C)
FT                   /*tag= at
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(47636,A)
FT                   /*tag= au
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(48011,G)
FT                   /*tag= av
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(48012,C)
FT                   /*tag= aw
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(48019,A)
FT                   /*tag= ax
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(50175,G)
FT                   /*tag= ay
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(50919,T)
FT                   /*tag= az
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(51730,T)
FT                   /*tag= ba
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(53975,G)
FT                   /*tag= bb
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(54377,C)
FT                   /*tag= bc
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(54388,G)
FT                   /*tag= bd
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(61997,G)
FT                   /*tag= be
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(62272,C)
FT                   /*tag= bf
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(63212,G)
FT                   /*tag= bg
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(65792,C)
FT                   /*tag= bh
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(68883,C)
FT                   /*tag= bi
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(69135,G)
FT                   /*tag= bj
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(69992,T)
FT                   /*tag= bk
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(70260,A)
FT                   /*tag= bl
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(73199,T)
FT                   /*tag= bm
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(73820,C)
FT                   /*tag= bn
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(74460,T)
FT                   /*tag= bo
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(76045,C)
FT                   /*tag= bp
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(79986,G)
FT                   /*tag= bq
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(81815,A)
FT                   /*tag= br
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(81861,A)
FT                   /*tag= bs
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(83741,C)
FT                   /*tag= bt
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(84120,A)
FT                   /*tag= bu
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(86862,A)
FT                   /*tag= bv
FT                   /standard_name= "Single nucleotide polymorphism"
```

```
  Query Match             2.5%;  Score 153.8;  DB 6;  Length 110000;
  Best Local Similarity   74.7%;  Pred. No. 5.9e-26;
  Matches  219;  Conservative    0;  Mismatches   72;  Indels    2;  Gaps    2;

Qy       2177 GGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCT 2236
              | ||||| | | |||||||| | |||||||| || | |||||||||||| ||| ||||| |
Db      13776 GAGTATGCTTTACCTGAATAACGTGCATTGCACTCATTAAGTAATTTATCATCATCCACC 13835

Qy       2237 GACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTC-TACATC 2295
              | |||  ||||| |||  ||| ||| || || |||| || | |||| ||||    ||
Db      13836 CTCCTCTAACCCCCCCATTATTCAGAGTCTACAATGTCTATTAGTCCAAACTCAGTTGTC 13895

Qy       2296 TATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTG-TCTTTC 2354
              ||| |||||||||  |||||||| |||||||| | |  ||||| |||||||| | ||||
Db      13896 CATGAGTACACATTGCTTAGCTCCCACTTATAAATAAGCACATGTAATATTTGATTTTTC 13955

Qy       2355 TGTGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAA 2414
              |||||||| ||||||| ||||| |||||||||| | ||||| |||| | ||||| |||
Db      13956 TGTGTCTGACTTGTTTCACTTAAAATAATGGCCTCTAGTTCCATCTGTGTTGCTGAAAAC 14015

Qy       2415 GGCATGATTTCATTCTTTTTTATGGCTATGTTCTTTCCCAATTTAGATAAAGA 2467
                ||  |||||||||||||||||| | ||  | | ||| |  | ||| | |
Db      14016 AACACAATTTCATTCTTTTTTATGGCCAAGTAGTATTCCATTGGATATATATA 14068
```

```
RESULT 54
ADS31325/c
ID   ADS31325 standard; DNA; 897 BP.
XX
AC   ADS31325;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Human genome high complexity repeat found in the HIRA gene #358.
XX
KW   Human; ds;
KW   histone cell cycle regulation defective, S. cerevisiae homologue A; HIRA;
```

KW high complexity repeat; in situ hybridisation; Southern blot;
KW chromosome breakpoint; inherited genetic disease; neoplastic disorder;
KW chromosome 22; DiGeorge syndrome; Velo-Cardio-facial syndrome.
XX
OS Homo sapiens.
XX
PN US2003224356-A1.
XX
PD 04-DEC-2003.
XX
PF 14-MAY-2001; 2001US-00854867.
XX
PR 16-MAY-2000; 2000US-00573080.
XX
PA (KNOL/) KNOLL J H M.
PA (ROGA/) ROGAN P K.
XX
PI Knoll JHM, Rogan PK;
XX
DR WPI; 2002-062378/08.
XX
PT Single copy genomic hybridization probes for detecting specific nucleic
PT acid sequences in sample by in situ hybridization useful for detection of
PT acquired or inherited genetic diseases.
XX
PS Example 1; SEQ ID NO 358; 30pp; English.
XX
CC The invention relates to a nucleic acid hybridisation probe comprising a
CC labelled, single copy nucleic acids of at least 50 nucleotides, which
CC will hybridise to a deduced single copy sequence interval in target
CC nucleic acid (TNA) of known sequence. The single copy sequence is deduced
CC by comparing the target nucleic acid (e.g. a disease causing gene) with a
CC collection of high and low complexity repeat sequences as found in the
CC genome of the organism from containing the target nucleic acid. The probe
CC is generated by PCR on the target sequence. The probe is essentially free
CC of blocking nucleic acid sequences which will hybridise to repeat
CC sequences within the genome of which the TNA is a part, and is labelled
CC with a label selected from fluorochrome-responsive labels, fluorochromes,
CC calorimetric chemical, conjugated proteins, antibodies, antigens and
CC their mixtures. The probe is useful in a hybridisation method, where the
CC hybridisation method is from in situ hybridisation, Southern blot, and
CC other methods in which nucleic acid is immobilised, where the method
CC further comprises selecting a single copy nucleic acid which will
CC hybridise to a duplicon or triplicon sequence domain. The probe is useful
CC for determining the existence of previously unknown repeat sequence
CC families in a genome. The method comprises reacting a labelled probe with
CC the genome, causing the probe to hybridise and ascertaining if the probe
CC hybridises to the genome at more than three preferably ten different
CC locations as a determination of new repeat sequence family, where the
CC determining step comprises selecting the single copy sequence from a
CC duplicon or triplicon sequence domain. The probe is useful for
CC determining a chromosome breakpoint and is useful in the fields for
CC cytogenetics and molecular genetics for determining the presence of
CC specific nucleic acid sequences in a sample of eukaryotic origin, e.g.
CC the probes may be used to analyse specific chromosomal locations by in
CC situ hybridisation as a detection of acquired or inherited genetic
CC diseases especially for detection of genetic or neoplastic disorders.
CC Unlike prior art techniques, the probe permits more precise chromosomal
CC breakpoint determinations by in situ hybridisation. The genomic sequence
CC comprising the human HIRA gene (histone cell cycle regulation defective,
CC S. cerevisiae, homologue A) was analysed for single copy sequence
CC intervals for use as probes of the invention. HIRA is located on
CC chromosome 22 as a duplicate, deletions of 1 copy lead to DiGeorge and
CC Velo-Cardio-facial syndromes. The present sequence is a high complexity
CC repeat found within the human genome used to analyse the HIRA gene for
CC repeat regions. Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from USPTO at seqdata.uspto.gov/sequence.html?DocID=20030224356.
XX
SQ Sequence 897 BP; 362 A; 152 C; 162 G; 191 T; 0 U; 30 Other;

Query Match 2.5%; Score 151.4; DB 7; Length 897;
Best Local Similarity 69.1%; Pred. No. 2.4e-26;

Matches 212; Conservative 6; Mismatches 87; Indels 2; Gaps 1;

```
Qy       2177 GGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCT 2236
             | |||  ||||||| ||||  ||: ||||||||||  || |||||||  ||| |  || |
Db        817 GTGTACCCATCACCCGAATAGTGWACATTGTACCCAATAGGTAATTTTTCATC--CCTCA 760

Qy       2237 GACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTCTACATCT 2296
               | |||    |||||  |||| |||| ||||| |||| || || |||| || ||   |
Db        759 CCCCTCTCCTACCCTCCACCTTTTGAGTCTCCARTGTCTATTATACCACTCTATATGCCC 700

Qy       2297 ATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGTCTTTCTG 2356
               | | ||| |||   |||||||| |||||||||||| ||||| ::|| ||||  ||||
Db        699 GTATATACCCATAGCTTAGCTCCCACTTATAAGTGAGAACATAYRATGTTTGGTTTTCCA 640

Qy       2357 TGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAAGG 2416
               |  |||  ||  || |||||  |||||:| | ||::|| ||| ||| |||||||||| |
Db        639 TTCCTGAGTTACTTCACTTAGAATAATRGTCTCCARYTCCATCCAGGTTGCTGCAAATGC 580

Qy       2417 CATGATTTCATTCTTTTTTATGGCTATGTTCTTTCCCAATTTAGATAAAGAACACTCGCA 2476
               ||| ||||||||||||||||||||  ||  | | |||   || ||| | | |||
Db        579 CATTATTTCATTCTTTTTTATGGCTGMGTAGTATTCCATGGTATATATATAYCACATTTT 520

Qy       2477 CTTGCTC 2483
               |||  ||
Db        519 CTTTATC 513
```

RESULT 55
ADC87622/c
ID ADC87622 standard; DNA; 222880 BP.
XX
AC ADC87622;
XX
DT 01-JAN-2004 (first entry)
XX
DE Human GPCR related polynucleotide SEQ ID NO:2075.
XX
KW ds; human; GPCR; guanosine triphosphate-binding protein coupled receptor;
KW gene therapy.
XX
OS Homo sapiens.
XX
PN EP1270724-A2.
XX
PD 02-JAN-2003.
XX
PF 18-JUN-2002; 2002EP-00013517.
XX
PR 18-JUN-2001; 2001JP-00246789.
XX
PA (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI Suwa M, Asai K, Akiyama Y, Aburatani H;
XX
DR WPI; 2003-315783/31.
XX
PT New polynucleotide, useful for preparing a composition for treating a
PT patient in need of increased or suppressed activity or expression of the
PT guanosine triphosphate-binding protein coupled receptor.
XX
PS Disclosure; SEQ ID NO 2075; 28pp; English.
XX
CC The invention relates to a novel polynucleotide encoding a guanosine
CC triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC the invention may have a use in gene therapy. The polynucleotide and
CC polypeptide are useful for preparing a composition for treating a patient
CC in need of increased or suppressed activity or expression of the
CC guanosine triphosphate-binding protein coupled receptor. The protein
CC sequences shown in ADC87618-ADC87623 represent polynucleotide sequences
CC related to the invention.
XX
SQ Sequence 222880 BP; 66561 A; 41704 C; 41435 G; 71579 T; 0 U; 1601 Other;

```
  Query Match             2.5%;  Score 150.2;  DB 10;  Length 222880;
  Best Local Similarity   72.5%;  Pred. No. 6.3e-25;
  Matches  208;  Conservative    0;  Mismatches   78;  Indels    1;  Gaps    1;

Qy       2185 ATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCTGACTTCTT 2244
              ||||||| ||| |||| | |||||||| |||||||| || |   | |  |  | || |
Db     155724 ATCACCTTAATAATGTACCTTGTACCCAATAAGTAATTACTTATTGT-CTATCCCCTCCT 155666

Qy       2245 GCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTCTACATCTATGTGTAC 2304
                ||| | | ||||   ||   |||||||| || ||||||||| |||| | |||||||||
Db     155665 CTCCCTTTACCCTTACCAGTTGCCATTGTCTATGATTCCACACCCTACGTTTATGTGTAC 155606

Qy       2305 ACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGTCTTTCTGTGTCTGTC 2364
              |||||||||||||||||| |||||||  ||||     | |||||||||   ||| ||
Db     155605 ACATTATTTAGCTCCTACTCATAAGTGTAAACACCTGGTGTTTGTCTTTTCATGTGTGAA 155546

Qy       2365 TTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAAGGCATGATTT 2424
              || |||||||| | |||||||||| ||| ||  ||| | | |||||||||| |||||||||
Db     155545 TTATTTTACTCAAGATAATGGCCTCCATTTACATCCATGTTGCTGCAAAATACATGATTT 155486

Qy       2425 CATTCTTTTTTATGGCTATGTTCTTTCCCAATTTAGATAAAGAACAC 2471
              |||||||||| |||||||  |  | | |   | |  ||  |   |||
Db     155485 CATTCTTTTTCATGGCTAAATAGTATTCGGCTGTTCATGTATCCCAC 155439


RESULT 56
ABD32923_7/c
Continuation (8 of 8) of ABD32923 from base 700001 (Human cancer-associated genomic DNA
WP Sequence split into 8 fragments  LOCUS ABD32923 Accession Abd32923
WP     Fragment Name      Begin     End
WP     ABD32923_0             1    110000
WP     ABD32923_1        100001    210000
WP     ABD32923_2        200001    310000
WP     ABD32923_3        300001    410000
WP     ABD32923_4        400001    510000
WP     ABD32923_5        500001    610000
WP     ABD32923_6        600001    710000
WP     ABD32923_7        700001    788759

  Query Match             2.5%;  Score 149.8;  DB 13;  Length 88759;
  Best Local Similarity   69.9%;  Pred. No. 5.1e-25;
  Matches  202;  Conservative    0;  Mismatches   87;  Indels    0;  Gaps    0;

Qy       2179 GTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCTGA 2238
              |||  |||||||  ||      |  |||||||| | | ||||||||||| |     |
Db      48991 GTAACCATCACCCAAACAGCCTATATTGTACCCTTCATGTAATTTCTCATCCCTTACCCC 48932

Qy       2239 CTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTCTACATCTAT 2298
              | ||   ||| | || | |||  || ||| | |||| || |||||||||| |  || |
Db      48931 CATCCCACCCTCCCACCTTTCCAAGTCTCTAATGTCTATTATTCCACACTCGATGTCCAG 48872

Qy       2299 GTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGTCTTTCTGTG 2358
              |||||||||||||||||||| ||||| |||||| ||||||   ||||||  |||||||
Db      48871 GTGTACACATTATTTAGCTCCCACTTAAAAGTGAGAACATGTGGTATTTGATTTTCTGTT 48812

Qy       2359 TCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAAGGCA 2418
              ||||  || ||| ||||| || | || || ||||||| ||| | | | |||||||||  ||
Db      48811 TCTGAATTATTTCACTTAAGAGAGTGACCTCCAGTTCCATCCACGATACTGCAAAAGACA 48752

Qy       2419 TGATTTCATTCTTTTTTATGGCTATGTTCTTTCCCAATTTAGATAAAGA 2467
              | |||| ||| || || ||||||   | | | ||  |  | | | | |
Db      48751 TTATTTTATTTTTGTTAATGGCTGAATAATATTCCTTTGCATAAATATA 48703


RESULT 57
ADQ59518
ID   ADQ59518 standard; DNA; 100762 BP.
XX
AC   ADQ59518;
XX
DT   07-OCT-2004  (first entry)
XX
DE   Human cancer-associated (CA) gene sequence SEQ ID NO:154.
XX
KW   human; cancer-associated gene; cancer-associated protein; cytostatic;
KW   gene therapy; vaccine; tyrosine kinase antagonist;
KW   G-protein coupled receptor antagonist; cancer; lymphoma; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO2004058288-A1.
XX
PD   15-JUL-2004.
XX
PF   15-DEC-2003; 2003WO-US040082.
XX
PR   17-DEC-2002; 2002US-00322696.
XX
PA   (SAGR-) SAGRES DISCOVERY INC.
XX
PI   Morris DW,  Malandro MS;
XX
DR   WPI; 2004-543349/52.
DR   P-PSDB; ADQ59520.
XX
PT   New cancer-associated nucleic acid for diagnosing, preventing or treating
PT   cancer (e.g. lymphoma) or for screening agents that may be used for
PT   treating or preventing cancer.
XX
PS   Claim 16; SEQ ID NO 154; 143pp; English.
XX
CC   The present invention describes human cancer-associated (CA) nucleotide
CC   sequences (I). Also described: (1) an expression vector comprising (I);
CC   (2) a host cell comprising (I) or the expression vector; (3) a microarray
CC   for detecting a CA nucleic acid; (4) an isolated polypeptide encoded
CC   within an open reading frame of a CA sequence; (5) an isolated antibody,
CC   or its antigen binding fragment, that binds to the above polypeptide; (6)
CC   a hybridoma that produces the monoclonal antibody described above; (7) a
CC   pharmaceutical composition comprising the antibody and a pharmaceutical
CC   excipient; (8) a kit for detecting or diagnosing cancer cells, comprising
CC   the above (monoclonal) antibody or polynucleotide that selectively
CC   hybridises to any of the polynucleotide sequences mentioned above; (9)
CC   methods for diagnosing cancer or for detecting the presence or absence of
CC   cancer cells in an individual; (10) a method for inhibiting growth of
CC   cancer cells in an individual; (11) a method for delivering a therapeutic
CC   agent to cancer cells in an individual; (12) an electronic library
CC   comprising the polynucleotide or polypeptide, or their fragments,
CC   mentioned above; (13) a method of screening for anticancer activity; (14)
CC   methods for detecting cancer associated with expression of a polypeptide
CC   or the presence of the antibody in a test cell or serum sample; (15) a
CC   method for screening for a bioactive agent capable of modulating the
CC   activity of a CA protein encoded by the above nucleic acid molecule; and
CC   (16) a method for treating cancers. (I) has cytostatic activity, and can
CC   be used in gene therapy, in vaccines, as a tyrosine kinase antagonist,
CC   and as a G-protein coupled receptor antagonist. The compositions and
CC   methods of the present invention can be used for diagnosing, preventing
CC   and treating cancer, especially lymphomas. They may also be used in
CC   screening for agents that may be used for treating or preventing cancer.
CC   The present sequence represents a human CA gene sequence, which is given
CC   in the exemplification of the present invention. Note: The sequence data
CC   for this patent did not form part of the printed specification, but was
CC   obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 100762 BP; 23383 A; 25739 C; 25440 G; 26180 T; 0 U; 20 Other;

  Query Match             2.5%;  Score 149.4;  DB 12;  Length 100762;
  Best Local Similarity   71.1%;  Pred. No. 6.8e-25;
  Matches  214;  Conservative    0;  Mismatches   81;  Indels    6;  Gaps    1;

Qy       2177 GGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCT 2236
              |||||| | |||||  |||     | |||||||||||||| ||||||||||   ||| |
Db      64406 GGGTATCCCTCACCCAAATAGCATACATTGTACCCCTTAAATAATTTCTCATGGTCCACC 64465

Qy       2237 GACTTCTTG------CCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTCT 2290
```

```
                 || | |       |||||||| |||||| || |||  ||| |  | |||||||| |||
Db      64466 TACCTATACCTCCCACCCCCTCACCCTTCTAAGTCTCTGTTGCCTTTTATTCCACTCTCC 64525

Qy       2291 ACATCTATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGTC 2350
              |  ||  |||| ||||| |||||  |||| ||||| |||||| |||| ||| |||||||||
Db      64526 AGCTCCGTGTGGACACAGTATTTCACTCCCACTTAGAAGTGAGAACAGGCAGTATTTGTC 64585

Qy       2351 TTTCTGTGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTGC 2410
              |||||||||||||  |   |  ||||| ||  ||||||  |||||| ||| | | ||||||
Db      64586 TTTCTGTGTCTGAATGTCTGCACTTAAGAATATGGCCTTCAGTTCCATCCATGTTGCTGC 64645

Qy       2411 AAAAGGCATGATTTCATTCTTTTTTATGGCTATGTTCTTTCCCAATTTAGATAAAGAACA 2470
              |||| | ||||||| ||| || | ||||||  ||  | | ||| | |  ||| |   ||
Db      64646 GAAAGACTTGATTTCTTTCCTTATGATGGCTGAGTAGTATTCCATTGTGTATAGATGCCA 64705

Qy       2471 C 2471
              |
Db      64706 C 64706
```

RESULT 58
ADO47193/c
ID ADO47193 standard; DNA; 200000 BP.
XX
AC ADO47193;
XX
DT 15-JUL-2004 (first entry)
XX
DE DNA sequence of a human immunoglobulin heavy chain variable region.
XX
KW immunoglobulin; B cell; germline; region V; region D; region J;
KW recombination-promoting protein; recombination activating gene 1; RAG-1;
KW RAG-2; ss; terminal deoxynucleotidyltransferase; TdT; V(D)J recombinase.
XX
OS Homo sapiens.
XX
PN WO2004029249-A1.
XX
PD 08-APR-2004.
XX
PF 30-SEP-2003; 2003WO-AU001286.
XX
PR 30-SEP-2002; 2002US-0415024P.
XX
PA (CENT-) CENTENARY INST CANCER MEDICINE & CELL BI.
XX
PI Jolly C;
XX
DR WPI; 2004-316126/29.
XX
PT New vector comprising one or more immunoglobulin regions selected from V,
PT D, and J regions, a 5' flanking region, and a 3' flanking region, useful
PT for reverting cell lines to a pro-B cell-like state or to a germline-like
PT state.
XX
PS Example 1; Page 234-283; 382pp; English.
XX
CC The specification describes a method for the preparation of
CC immunoglobulins. The method uses a vector for reverting cell lines to a
CC pro-B cell-like state or to a germline-like state, by replacing one or
CC more of the immunoglobulin regions V, D, and J of the rearranged
CC immunoglobulin gene with the germ-line or unrearranged V, D or J regions.
CC The vector can comprise a polynucleotide sequence encoding a
CC recombination-promoting protein, selected from recombination activating
CC gene 1 (RAG-1), RAG-2, terminal deoxynucleotidyltransferase (TdT). These
CC proteins collectively constitute a V(D)J recombinase. The method is
CC useful for the preparation of immunoglobulin heavy and/or light chains
CC and their fragments. The present sequence encodes a human immunoglobulin
CC heavy chain variable region, and is used as a template to construct
CC vectors for use in the method of the invention.
XX
SQ Sequence 200000 BP; 59140 A; 40391 C; 40764 G; 59705 T; 0 U; 0 Other;

```
  Query Match             2.4%;  Score 147.8;  DB 12;  Length 200000;
  Best Local Similarity   67.2%;  Pred. No. 2.3e-24;
  Matches  209;  Conservative    0;  Mismatches  102;  Indels    0;  Gaps    0;

Qy       2162 GACTTCTTGCCCCCTGGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAAT 2221
              | | ||||| |      |||| ||| | |  ||| |||| ||  ||||| |||||||||
Db     197566 GTCGTCTTGGCTTTAATGTATCCATGATCCAAATAATGTACATCATACCCTTTAAGTAAT 197507

Qy       2222 TTCTCACCATCCGCTGACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATT 2281
              |||||||||| | |  |  |  ||| | | |||  ||||||| || || ||||||||||
Db     197506 TTCTCACCATTCTCCCAACTTCTGCTCTCCCATGTTTCTGAGTCTGCAATGTCCATCATT 197447

Qy       2282 CCACACTCTACATCTATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGCA 2341
              ||   |||||      |||| ||| |||| |||||||| | ||||||||| ||  ||
Db     197446 CCTTTCTCTATGGGCCTGTGCACATATTACTTAGCTCCCATTTATAAGTGAAAAAGTGTG 197387

Qy       2342 ATATTTGTCTTTCTGTGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTA 2401
              | | ||   ||||   |||   || ||  |||||  |||||| |||  ||||| ||| |
Db     197386 GTGTGTGCTGTTCTACTTCTCGTTTATTAGACTTAAAATAATGACCCAGAGTTCCATCCA 197327

Qy       2402 GGCTGCTGCAAAAGGCATGATTTCATTCTTTTTTATGGCTATGTTCTTTCCCAATTTAGA 2461
              |  |||||||||  ||||||| ||||| ||  |||||||   |  | |   | | || |
Db     197326 TGTAGCTGCAAAATACATGATTCCATTCCTTAATATGGCTGGATAGTATTTAATTGTATA 197267

Qy       2462 TAAAGAACACT 2472
              || |    |||
Db     197266 TATATGTGACT 197256
```

RESULT 59
ABN83429/c
ID ABN83429 standard; DNA; 126512 BP.
XX
AC ABN83429;
XX
DT 21-AUG-2002 (first entry)
XX
DE Human transporter protein gene.
XX
KW Human; sodium/calcium exchanger; transporter; brain; heart; kidney; lung;
KW spleen; testis; leukocyte; foetal brain; chromosome 14; gene;
KW single nucleotide polymorphism; SNP; ds.
XX
OS Homo sapiens.
XX
FH Key   Location/Qualifiers
FT variation   replace(378,T)
FT   /*tag= a
FT   /standard_name= "Single nucleotide polymorphism"
FT variation   replace(741. .742,C-)
FT   /*tag= b
FT   /standard_name= "Single nucleotide polymorphism"
FT   /note= "This variation is an indel"
FT variation   replace(2002,T)
FT   /*tag= c
FT   /standard_name= "Single nucleotide polymorphism"
FT CDS   2010. .124505
FT   /*tag= d
FT   /product= "Human transporter"
FT   /note= "Contains 5 introns"
FT exon   2010. .3793
FT   /*tag= e
FT   /number= 1
FT variation   replace(2381,C)
FT   /*tag= f
FT   /standard_name= "Single nucleotide polymorphism"
FT intron   3794. .109509
FT   /*tag= g
FT   /number= 1
FT variation   replace(5165,T)
FT   /*tag= h
FT   /standard_name= "Single nucleotide polymorphism"
FT variation   replace(5402,G)

```
FT                   /*tag= i
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(6794,C)
FT                   /*tag= j
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(9883,G)
FT                   /*tag= k
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(10210,C)
FT                   /*tag= l
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(12220,G)
FT                   /*tag= m
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(13842,G)
FT                   /*tag= n
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(14200,A)
FT                   /*tag= o
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(15878,T)
FT                   /*tag= p
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(16030,G)
FT                   /*tag= q
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(16292,C)
FT                   /*tag= r
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(16506,G)
FT                   /*tag= s
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(17953,A)
FT                   /*tag= t
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(23832,G)
FT                   /*tag= u
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(25001,A)
FT                   /*tag= v
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(25141,G)
FT                   /*tag= w
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(25191,G)
FT                   /*tag= x
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(26147,A)
FT                   /*tag= y
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(27400,G)
FT                   /*tag= z
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(27401,T)
FT                   /*tag= aa
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(29278,C)
FT                   /*tag= ab
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(31437,G)
FT                   /*tag= ac
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(31857,G)
FT                   /*tag= ad
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(33155,A)
FT                   /*tag= ae
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(39487,C)
FT                   /*tag= af
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(41449,C)
FT                   /*tag= ag
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(42420,C)
FT                   /*tag= ah
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(43256,C)
FT                   /*tag= ai
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(43967,C)
FT                   /*tag= aj
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(48603: .48604,A-)
FT                   /*tag= ak
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   /note= "This variation is an indel"
FT   variation       replace(49560,T)
FT                   /*tag= al
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(52729,G)
FT                   /*tag= am
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(55031,G)
FT                   /*tag= an
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(55066,C)
FT                   /*tag= ao
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(56912,G)
FT                   /*tag= ap
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(58480,T)
FT                   /*tag= aq
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(61128,A)
FT                   /*tag= ar
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(61320,A)
FT                   /*tag= as
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(61444,C)
FT                   /*tag= at
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(62641,C)
FT                   /*tag= au
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(63023,G)
FT                   /*tag= av
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(63051,C)
FT                   /*tag= aw
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(64989,G)
FT                   /*tag= ax
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(65929,A)
FT                   /*tag= ay
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(66694,G)
FT                   /*tag= az
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(66755,A)
FT                   /*tag= ba
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(66879,C)
FT                   /*tag= bb
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(69156,T)
FT                   /*tag= bc
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(69280,T)
FT                   /*tag= bd
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(70647,T)
FT                   /*tag= be
```

```
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(71867,T)
FT                   /*tag=  bf
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(71900,T)
FT                   /*tag=  bg
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(71901,A)
FT                   /*tag=  bh
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(72369,T)
FT                   /*tag=  bi
```

```
  Query Match             2.4%;  Score 147.6;  DB 6;  Length 126512;
  Best Local Similarity   73.3%;  Pred. No. 2.1e-24;
  Matches  189;  Conservative    0;  Mismatches   69;  Indels    0;  Gaps    0;

Qy       2177 GGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCT 2236
              | ||||||||||||  ||| |||| | |||||| |.|||||||||| ||| |||   |
Db      45171 GTGTATTCATCACCCAAATAATGTACGTTGTACTCATTAAGTAATTTTTCATCATTGACC 45112

Qy       2237 GACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTCTACATCT 2296
                || |    |||| || ||||| ||| ||||| |||| || ||||||||||||    |
Db      45111 CTCTCCCATTCCCCCCACCCTTCCGAGTCTCCAATGTCTATTATTCCACACTCTGTGCCC 45052

Qy       2297 ATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGTCTTTCTG 2356
              |||  ||  |||||||||||||| | || ||||||| ||||||    ||||| |||||||
Db      45051 ATGCATATGCATTATTTAGCTCCCAATTGTAAGTGAGAACATGTGGAATTTGCCTTTCTG 44992

Qy       2357 TGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAAGG 2416
              | ||||  ||| || ||||| || |||||||  |||||||  || |   ||||||||||
Db      44991 TTTCTGAGTTGCTTCACTTAAGAGAATGGCTTCCAGTTCCTTCCATATTGCTGCAAAAGA 44932

Qy       2417 CATGATTTCATTCTTTTT 2434
              ||   |||  |  |||||
Db      44931 CAGACTTTGTTCTTTTTT 44914
```

```
RESULT 60
ABS55320_0/c
WP Sequence split into 5 fragments  LOCUS ABS55320 Accession Abs55320
WP     Fragment Name      Begin     End
WP     ABS55320_0             1    110000
WP     ABS55320_1        100001    210000
WP     ABS55320_2        200001    310000
WP     ABS55320_3        300001    410000
WP     ABS55320_4        400001    506286
ID   ABS55320 standard; DNA; 506286 BP.
XX
AC   ABS55320;
XX
DT   06-JAN-2003  (first entry)
XX
DE   Human transporter protein genomic DNA sequence.
XX
KW   Human; transporter protein; hypoglycaemia; antidiabetic; gene therapy;
KW   ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    341512. .343967
FT                   /*tag=  a
FT                   /note= "This section of sequence contains unallowable
FT                   bases which have been substituted by n nucleotides"
XX
PN   WO200268472-A2.
XX
PD   06-SEP-2002.
XX
PF   17-JAN-2002; 2002WO-US001202.
XX
PR   31-JAN-2001; 2001US-0265128P.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Yan C,  Li Z,  Ketchum KA,  Beasley EM;
XX
DR   WPI; 2002-723196/78.
XX
PT   New human transporter peptide, useful for identifying agents that can be
PT   used to treat a disease or condition mediated by a human transporter
PT   protein e.g. hypoglycemia.
XX
PS   Claim 4; Page 221-466; 466pp; English.
XX
CC   The present invention relates to a new peptide. The invention is useful
CC   for treating a disease or condition mediated by a human transporter
CC   protein e.g. hypoglycaemia. The present nucleic acid sequence represents
CC   the human transporter protein genomic DNA sequence. Note: This sequence
CC   is shorter that the printed sequence (507421bp) as the polymorphisms
CC   shown in the specification have been substituted by the corresponding
CC   degenerate codons
XX
SQ   Sequence 506286 BP; 151127A; 88917C; 95869G; 167602T; 0U; 27710ther;
```

```
  Query Match             2.4%;  Score 147.2;  DB 6;  Length 110000;
  Best Local Similarity   73.7%;  Pred. No. 2.5e-24;
  Matches  227;  Conservative    0;  Mismatches   78;  Indels    3;  Gaps    3;

Qy       2177 GGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCT 2236
              |||||     | |||||||| |  |  ||||||| | ||||| ||||||||| ||||| |
Db      55279 GGGTAACTGTGACCTGAATAACATATATTGTACTCATTAAG-AATTTCTCATCATCCACC 55221

Qy       2237 GACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTCTACATCT 2296
                | ||   | || ||| ||| |  || ||||| |||| |||||||||| |||||||||
Db      55220 CTC-TCCCACGCCTTCACCCTCCCAAGTCTCCAATGTCTATCATTCCACTCTCTACATCC 55162

Qy       2297 ATGTGTACACA-TTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGTCTTTCT 2355
              |||||||||||| || |||| | || |||||| ||||| |  ||   ||||||||||||||
Db      55161 ATGTGTACACATTTTTTTAACACCCACTTATGAGTGAGATTATTTGATATTTGTCTTTCT 55102

Qy       2356 GTGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAAG 2415
              |||||||  ||| || ||||| ||||| ||   ||| ||| ||| | | ||||||| |||
Db      55101 GTGTCTGGGTTGGTTCACTTAAGATAACGGATTCCAATTCCATCCATGTTGCTGCAGAAG 55042

Qy       2416 GCATGATTTCATTCTTTTTTATGGCTATGTTCTTTCCCAATTTAGATAAAGAACACTCGC 2475
              |||||||||||||||||||||||||  ||    ||||| | |  ||| |   |||  |
Db      55041 ACATGATTTCATTCTTTTTTATGGCTGAGTAGCATCCCATTGTGTATATATGCCACATGT 54982

Qy       2476 ACTTGCTC 2483
              |||  ||
Db      54981 TCTTTATC 54974
```

```
RESULT 61
ADQ59398_0/c
WP Sequence split into 4 fragments  LOCUS ADQ59398 Accession Adq59398
WP     Fragment Name      Begin     End
WP     ADQ59398_0             1    110000
WP     ADQ59398_1        100001    210000
WP     ADQ59398_2        200001    310000
WP     ADQ59398_3        300001    357652
ID   ADQ59398 standard; DNA; 357652 BP.
XX
AC   ADQ59398;
XX
DT   07-OCT-2004  (first entry)
XX
DE   Human cancer-associated (CA) gene sequence SEQ ID NO:34.
XX
KW   human; cancer-associated gene; cancer-associated protein; cytostatic;
KW   gene therapy; vaccine; tyrosine kinase antagonist;
KW   G-protein coupled receptor antagonist; cancer; lymphoma; gene; ds.
XX
OS   Homo sapiens.
```

```
XX
PN   WO2004058288-A1.
XX
PD   15-JUL-2004.
XX
PF   15-DEC-2003; 2003WO-US040082.
XX
PR   17-DEC-2002; 2002US-00322696.
XX
PA   (SAGR-) SAGRES DISCOVERY INC.
XX
PI   Morris DW,  Malandro MS;
XX
DR   WPI; 2004-543349/52.
DR   P-PSDB; ADQ59400.
XX
PT   New cancer-associated nucleic acid for diagnosing, preventing or treating
PT   cancer (e.g. lymphoma) or for screening agents that may be used for
PT   treating or preventing cancer.
XX
PS   Claim 16; SEQ ID NO 34; 143pp; English.
XX
CC   The present invention describes human cancer-associated (CA) nucleotide
CC   sequences (I). Also described: (1) an expression vector comprising (I);
CC   (2) a host cell comprising (I) or the expression vector; (3) a microarray
CC   for detecting a CA nucleic acid; (4) an isolated polypeptide encoded
CC   within an open reading frame of a CA sequence; (5) an isolated antibody,
CC   or its antigen binding fragment, that binds to the above polypeptide; (6)
CC   a hybridoma that produces the monoclonal antibody described above; (7) a
CC   pharmaceutical composition comprising the antibody and a pharmaceutical
CC   excipient; (8) a kit for detecting or diagnosing cancer cells, comprising
CC   the above (monoclonal) antibody or polynucleotide that selectively
CC   hybridises to any of the polynucleotide sequences mentioned above; (9)
CC   methods for diagnosing cancer or for detecting the presence or absence of
CC   cancer cells in an individual; (10) a method for inhibiting growth of
CC   cancer cells in an individual; (11) a method for delivering a therapeutic
CC   agent to cancer cells in an individual; (12) an electronic library
CC   comprising the polynucleotide or polypeptide, or their fragments,
CC   mentioned above; (13) a method of screening for anticancer activity; (14)
CC   methods for detecting cancer associated with expression of a polypeptide
CC   or the presence of the antibody in a test cell or serum sample; (15) a
CC   method for screening for a bioactive agent capable of modulating the
CC   activity of a CA protein encoded by the above nucleic acid molecule; and
CC   (16) a method for treating cancers. (I) has cytostatic activity, and can
CC   be used in gene therapy, in vaccines, as a tyrosine kinase antagonist,
CC   and as a G-protein coupled receptor antagonist. The compositions and
CC   methods of the present invention can be used for diagnosing, preventing
CC   and treating cancer, especially lymphomas. They may also be used in
CC   screening for agents that may be used for treating or preventing cancer.
CC   The present sequence represents a human CA gene sequence, which is given
CC   in the exemplification of the present invention. Note: The sequence data
CC   for this patent did not form part of the printed specification, but was
CC   obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 357652 BP; 107153 A; 63035 C; 68169 G; 119275 T; 0 U; 20 Other;

  Query Match             2.4%;  Score 147.2;  DB 12;  Length 110000;
  Best Local Similarity   73.7%;  Pred. No. 2.5e-24;
  Matches  227;  Conservative    0;  Mismatches   78;  Indels    3;  Gaps    3;

Qy       2177 GGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCT 2236
              |||||    | |||||||| |  |  ||||||| | ||||| |||||||||| ||||| |
Db      63266 GGGTAACTGTGACCTGAATAACATATATTGTACTCATTAAG-AATTTCTCATCATCCACC 63208

Qy       2237 GACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTCTACATCT 2296
                | ||   | || ||| ||| |  || ||||| |||| |||||||||| |||||||||
Db      63207 CTC-TCCCACGCCTTCACCCTCCCAAGTCTCCAATGTCTATCATTCCACTCTCTACATCC 63149

Qy       2297 ATGTGTACACA-TTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGTCTTTCT 2355
              |||||||||||  || |||| | || |||||| ||||| |  ||   |||||||||||||
Db      63148 ATGTGTACACATTTTTTTAACACCCACTTATGAGTGAGATTATTTGATATTTGTCTTTCT 63089

Qy       2356 GTGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAAG 2415
              |||||||  ||| || ||||| ||||| ||    ||| ||| ||| | | ||||||| |||
Db      63088 GTGTCTGGGTTGGTTCACTTAAGATAACGGATTCCAATTCCATCCATGTTGCTGCAGAAG 63029

Qy       2416 GCATGATTTCATTCTTTTTTATGGCTATGTTCTTTCCCAATTTAGATAAAGAACACTCGC 2475
              |||||||||||||||||||||||||  ||    ||||| | |  ||| |   |||  |
Db      63028 ACATGATTTCATTCTTTTTTATGGCTGAGTAGCATCCCATTGTGTATATATGCCACATGT 62969

Qy       2476 ACTTGCTC 2483
              |||  ||
Db      62968 TCTTTATC 62961


RESULT 62
ABS55200/c
ID   ABS55200 standard; DNA; 341511 BP.
XX
AC   ABS55200;
XX
DT   06-JAN-2003  (first entry)
XX
DE   Genomic DNA encoding human transporter protein.
XX
KW   Human; transporter protein; hypoglycaemia; antidiabetic; gene therapy;
KW   gene; ds; single nucleotide polymorphism; SNP.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             2001. .339512
FT                   /*tag=  a
FT                   /product= "Human transporter protein"
FT   exon            2001. .2073
FT                   /*tag=  b
FT                   /number= 1
FT   intron          2074. .30237
FT                   /*tag=  c
FT                   /number= 1
FT   variation       replace(2557,T)
FT                   /*tag=  w
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(5492,C)
FT                   /*tag=  x
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(5930,A)
FT                   /*tag=  y
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(7269,C)
FT                   /*tag=  z
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(9650,G/T)
FT                   /*tag=  aa
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(10664,T)
FT                   /*tag=  ab
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(11043,A/G/C)
FT                   /*tag=  ac
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(13109,T)
FT                   /*tag=  ad
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(13392,A)
FT                   /*tag=  ae
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(13395. .13397,GA)
FT                   /*tag=  af
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(13563,C)
FT                   /*tag=  ag
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(13694,A)
FT                   /*tag=  ah
```

```
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(14075,A)
FT                   /*tag= ai
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(14573,A)
FT                   /*tag= aj
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(15013,T)
FT                   /*tag= ak
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(15073. .15075,GT)
FT                   /*tag= al
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(15848,T)
FT                   /*tag= am
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(16346,T)
FT                   /*tag= an
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(16367,A)
FT                   /*tag= ao
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(16639,G)
FT                   /*tag= ap
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(18149,A/T)
FT                   /*tag= aq
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(18243,C/A)
FT                   /*tag= ar
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(18293,G/A)
FT                   /*tag= as
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(19185,C)
FT                   /*tag= at
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(24728. .24730,AA)
FT                   /*tag= au
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(24790,T)
FT                   /*tag= av
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(24950,T)
FT                   /*tag= aw
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(28608,G)
FT                   /*tag= ax
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(29244,C)
FT                   /*tag= ay
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(29651,T)
FT                   /*tag= az
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   exon            30238. .30361
FT                   /*tag= d
FT                   /number= 2
FT   intron          30362. .40719
FT                   /*tag= e
FT                   /number= 2
FT   variation       replace(34034,T)
FT                   /*tag= ba
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(36671,G)
FT                   /*tag= bb
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(36836,T)
FT                   /*tag= bc
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(37010,A)
FT                   /*tag= bd
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(38461,A/C)
FT                   /*tag= be
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(38588,A/C)
FT                   /*tag= bf
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(39673,A)
FT                   /*tag= bg
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(39854,C)
FT                   /*tag= bh
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   exon            40720. .40826
FT                   /*tag= f
FT                   /number= 3
FT   intron          40827. .45732
FT                   /*tag= g
FT                   /number= 3
FT   variation       replace(41325,A)
FT                   /*tag= bi
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(41412,A)
FT                   /*tag= bj
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(41475,G)
FT                   /*tag= bk
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(41541,G)
FT                   /*tag= bl
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(41649,G)
FT                   /*tag= bm
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(41887,T)
FT                   /*tag= bn
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(42790,A)
FT                   /*tag= bo
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   exon            45733. .45861
FT                   /*tag= h
FT                   /number= 4
FT   intron          45862. .59818
FT                   /*tag= i
FT                   /number= 4
FT   variation       replace(46178,T)
FT                   /*tag= bp
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(46282,C)
FT                   /*tag= bq
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(46382,C)
FT                   /*tag= br
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(46550,A)
FT                   /*tag= bs
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(49271. .49273,AT)
FT                   /*tag= bt
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(49586,A)
FT                   /*tag= bu
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(49835,G)
FT                   /*tag= bv
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(52080,G)
FT                   /*tag= bw
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"

  Query Match             2.4%;  Score 147.2;  DB 6;  Length 341511;
  Best Local Similarity   73.7%;  Pred. No. 4.2e-24;
  Matches  227;  Conservative    0;  Mismatches   78;  Indels    3;  Gaps    3;
```

```
Qy       2177 GGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCT 2236
             |||||    | |||||||| |  |  ||||||| | ||||| |||||||||| ||||| |
Db      55279 GGGTAACTGTGACCTGAATAACATATATTGTACTCATTAAG-AATTTCTCATCATCCACC 55221

Qy       2237 GACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTCTACATCT 2296
               | ||   | || ||| ||| |  || ||||| |||| |||||||||| |||||||||
Db      55220 CTC-TCCCACGCCTTCACCCTCCCAAGTCTCCAATGTCTATCATTCCACTCTCTACATCC 55162

Qy       2297 ATGTGTACACA-TTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGTCTTTCT 2355
             |||||||||| || |||| | || |||||| ||||| |  ||   |||||||||||||
Db      55161 ATGTGTACACATTTTTTTAACACCCACTTATGAGTGAGATTATTTGATATTTGTCTTTCT 55102

Qy       2356 GTGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAAG 2415
             |||||||  ||| || ||||| ||||| ||   ||| ||| ||| | | ||||||| |||
Db      55101 GTGTCTGGGTTGGTTCACTTAAGATAACGGATTCCAATTCCATCCATGTTGCTGCAGAAG 55042

Qy       2416 GCATGATTTCATTCTTTTTTATGGCTATGTTCTTTCCCAATTTAGATAAAGAACACTCGC 2475
             |||||||||||||||||||||||||  ||    ||||| | |  ||| |   |||  |
Db      55041 ACATGATTTCATTCTTTTTTATGGCTGAGTAGCATCCCATTGTGTATATATGCCACATGT 54982

Qy       2476 ACTTGCTC 2483
              |||  ||
Db      54981 TCTTTATC 54974


RESULT 63
ABQ74964_1/c
Continuation (2 of 8) of ABQ74964 from base 100001 (Human kinase protein genomic DNA seq
WP Sequence split into 8 fragments  LOCUS ABQ74964 Accession Abq74964
WP     Fragment Name      Begin     End
WP     ABQ74964_0             1    110000
WP     ABQ74964_1        100001    210000
WP     ABQ74964_2        200001    310000
WP     ABQ74964_3        300001    410000
WP     ABQ74964_4        400001    510000
WP     ABQ74964_5        500001    610000
WP     ABQ74964_6        600001    710000
WP     ABQ74964_7        700001    786431

  Query Match             2.4%;  Score 146.8;  DB 6;  Length 110000;
  Best Local Similarity   75.6%;  Pred. No. 3.1e-24;
  Matches  195;  Conservative    0;  Mismatches   62;  Indels    1;  Gaps    1;

Qy       2141 AGTAATTTCTCACCATCCGCTGACTTCTTGCCCCCTGGGTATTCATCACCTGAATGATGT 2200
             | |||| | |  |      | |||  ||  | |    | ||| |||||||| |||| ||||
Db      70144 ACTAATATATTGCATATTGGTGAAGTCAGGGCTTTAGTGTACTCATCACCGGAATAATGT 70085

Qy       2201 GCATTGTACCCCTTAAGTAATTTCTCACCATCCGCTGACTTCTTGCCCCCTCATCCTTCT 2260
              ||||| | || ||||||||||||||| ||||| |   | | || |||  ||| | |||
Db      70084 ACATTGCA-CCATTAAGTAATTTCTCATCATCCACCCCCCTTTTACCCTTTCACCTTTCC 70026

Qy       2261 GAGGCTCCATTGTCCATCATTCCACACTCTACATCTATGTGTACACATTATTTAGCTCCT 2320
              || |||||||||| |||||||||||  | |||||||||||||| ||||||||||||||||
Db      70025 AAGTCTCCATTGTCTATCATTCCACAGTGTACATCTATGTGTGCACATTATTTAGCTCCC 69966

Qy       2321 ACTTATAAGTGATAACATGCAATATTTGTCTTTCTGTGTCTGTCTTGTTTTACTTATGAT 2380
              |||||||| ||| |||||||| |||||||||||||| ||| |   |||||| | ||| |||
Db      69965 ACTTATAAATGAGAACATGCAGTATTTGTCTTTCTATGTGTCAGTTGTTTCATTTAAGAT 69906

Qy       2381 AATGGCCCCCAGTTCTAT 2398
              ||| |||  |||||| ||
Db      69905 AATAGCCTGCAGTTCCAT 69888


RESULT 64
ADC24763
ID   ADC24763 standard; DNA; 334462 BP.
XX
AC   ADC24763;
XX
DT   18-DEC-2003  (first entry)
XX
DE   Human wild type HNL4X genomic sequence.
XX
KW   ds; nootropic; neuroleptic; tranquilizer; gene therapy; synaptogenesis;
KW   mutation; neurological disease; mental disorder; psychiatric illness;
KW   autism; Asperger syndrome; schizophrenia;
KW   attention deficit hyperactivity disorder.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   exon            10670. .10828
FT                   /*tag= a
FT                   /number= 1b
FT   intron          10829. .83743
FT                   /*tag= b
FT                   /number= 1b
FT   exon            83744. .84513
FT                   /*tag= c
FT                   /number= 2
FT   intron          84514. .201675
FT                   /*tag= d
FT                   /number= 2
FT   exon            201676. .201828
FT                   /*tag= e
FT                   /number= 3
FT   intron          201829. .240358
FT                   /*tag= f
FT                   /number= 3
FT   exon            240359. .240469
FT                   /*tag= g
FT                   /number= 3b
FT   intron          240470. .312939
FT                   /*tag= h
FT                   /number= 3b
FT   exon            312940. .313125
FT                   /*tag= i
FT                   /number= 4
FT   intron          313126. .318312
FT                   /*tag= j
FT                   /number= 4
FT   exon            318313. .319102
FT                   /*tag= k
FT                   /number= 5
FT   intron          319103. .327017
FT                   /*tag= l
FT                   /number= 5
FT   exon            327018. .330406
FT                   /*tag= m
FT                   /number= 6
XX
PN   WO2003045998-A2.
XX
PD   05-JUN-2003.
XX
PF   02-DEC-2002; 2002WO-FR004134.
XX
PR   30-NOV-2001; 2001CA-02364106.
XX
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
PA   (INSP ) INST PASTEUR.
PA   (ASSI-) ASSISTANCE PUBLIQUE HOPITAUX PARIS.
XX
PI   Bourgeron T,  Jamain S,  Quach H,  Betancur C,  Leboyer M;
PI   Gillberg C;
XX
DR   WPI; 2003-493399/46.
XX
PT   New nucleic acid encoding mutant protein involved in synaptogenesis,
PT   useful for treatment and diagnosis of e.g. autism, Asperger syndrome, and
PT   schizophrenia.
XX
PS   Claim 10; SEQ ID NO 1; 416pp; French.
XX
```

```
CC   The invention relates to an isolated or purified polynucleotide encoding
CC   a polypeptide (the wild-type form of which is involved in synaptogenesis)
CC   that includes at least one mutation associated with development of
CC   neurological disease and/or a predisposition to development of mental
CC   disorders or psychiatric illness. The polypeptide are used to screen for
CC   agents that modulate their activity. Also nucleic acid, polypeptide,
CC   polypeptide-specific antibodies, vectors containing he nucleic acid and
CC   host cells containing the vector, are useful as pharmaceuticals for
CC   treating mental and neurological disorders, specifically autism, Asperger
CC   syndrome, schizophrenia and attention deficit hyperactivity disorder. The
CC   wild-type forms of the nucleic acid and polypeptide can be used
CC   similarly. Also detecting mutations in the nucleic acid and polypeptide,
CC   or measuring activity of the polypeptide, can be used to detect
CC   biochemical disorders that affect formation of synapses and to diagnose
CC   mental disease. This sequence corresponds to the genomic sequence of the
CC   human wild type HNL4X gene.
XX
SQ   Sequence 334462 BP; 97253 A; 64494 C; 66298 G; 106417 T; 0 U; 0 Other;

  Query Match             2.4%;  Score 146;  DB 10;  Length 334462;
  Best Local Similarity   72.9%;  Pred. No. 8.1e-24;
  Matches  188;  Conservative    0;  Mismatches   70;  Indels    0;  Gaps    0;

Qy       2177 GGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCT 2236
             | |||  || ||||  ||| |  | ||||||| || ||||||||||||||     |
Db      26206 GTGTAACCACCACCCAAATAACATACATTGTATCCATTAAGTAATTTCTCATTCCTCATC 26265

Qy       2237 GACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTCTACATCT 2296
               | |    ||| | || | || |||  |||||  ||| || |||||||| ||||   ||
Db      26266 ATCCTACCACCCTCCCACCTTTTTGATTCTCCAGGGTCTATTATTCCACTCTCTGTGTCC 26325

Qy       2297 ATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGTCTTTCTG 2356
              |||||||||||||||||||||||| |||||  |||||  |||||    | |||| |||||||
Db      26326 ATGTGTACACATTATTTAGCTCCCACTTAGGAGTGAGGACATGTGGTTTTTGACTTTCTG 26385

Qy       2357 TGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAAGG 2416
              | || |  |||||| ||||| | |||||||| ||||||| ||| | ||| |||||||||
Db      26386 TTTCCGAGTTGTTTCACTTAAGGTAATGGCCTCCAGTTCCATCCATGCTCCTGCAAAAGA 26445

Qy       2417 CATGATTTCATTCTTTTT 2434
                ||  |||| ||||||||
Db      26446 GATAGTTTCGTTCTTTTT 26463


RESULT 65
ACN44446
ID   ACN44446 standard; DNA; 135800 BP.
XX
AC   ACN44446;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Human genomic sequence hCG22243.
XX
KW   Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO2003073826-A2.
XX
PD   12-SEP-2003.
XX
PF   28-FEB-2003; 2003WO-US006235.
XX
PR   01-MAR-2002; 2002US-00087192.
XX
PA   (SAGR-) SAGRES DISCOVERY.
XX
PI   Morris DW;
XX
DR   WPI; 2003-328604/31.
XX
PT   Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
PT   comprises a nucleotide sequence.
XX
PS   Claim 1; SEQ ID NO 898; 0pp; English.
XX
CC   The present invention relates to novel DNA and protein sequences which
CC   are associated with carcinomas. The sequences are useful for: (i) for
CC   screening drug candidates; (ii) for screening of bioactive agent capable
CC   of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
CC   a bioactive agent capable of modulating the activity of CAP; (iv) for
CC   evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
CC   carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
CC   carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
CC   (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
CC   determining Carcinoma Associated (CA) gene copy number. In addition, the
CC   CA genes are useful as DNA vaccines and the CAP are useful as markers of
CC   carcinoma including lymphoma. The present sequence is one such CA coding
CC   sequence. Note: This patent is an equivalent to basic patent
CC   US2002182586A1, for which no sequence data was published
XX
SQ   Sequence 135800 BP; 31639 A; 32696 C; 34887 G; 36578 T; 0 U; 0 Other;

  Query Match             2.4%;  Score 145.8;  DB 11;  Length 135800;
  Best Local Similarity   69.1%;  Pred. No. 6e-24;
  Matches  295;  Conservative    0;  Mismatches  107;  Indels   25;  Gaps    6;

Qy        540 AGACAAAAGGATATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCT 599
              ||||     || ||||| | |||||||| | |||  || || | |    ||| ||||||  |
Db     133378 AGACCCGGGGGTATGACCCAGTGGTAATGAACT--GGAGGGATCTTGCCAAAGCCTGTTT 133435

Qy        600 ATTCAGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGG 659
              ||||||||||| ||  |||||         |||   ||||    ||  ||  | |||||
Db     133436 GTTCAGATTCTTTCTGGCCTCTC--------TGTGAGGCATGTCCTCGCCCCAAATATGG 133487

Qy        660 GGCAGGACTCTTCTTCAATGAGGGTCTT--------CAAGGGAGAAGGGAGAAAGTGGCC 711
              || ||||  ||| | |||||||||| |         |  | | | | ||       |
Db     133488 GGTAGGATCCTTTTGCAATGAGGGTCCTGTGGCCACCTTGAGGGGAAGGTAGGTCAGAGG 133547

Qy        712 TTTTTAGATTTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTG 771
              | | ||| |||||||||| |||| | || |||||||||||||||| |||||| || ||||
Db     133548 TCTCTAGGTTTTATGGCTGGCTTTGCGGGAGAGGAGTTCTAGTTTTTATGACACACCTTG 133607

Qy        772 GGGAAGAGGAATTCTGGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAG 831
              |||| ||| |||||| |||||| |||| ||| |||  | |||||  |||  |  ||  |
Db     133608 GGGAGGAGAAATTCTAGTTTCTATGACCTGC-TTCGGGGAGAAATGGGAAAAGAAGAAAT 133666

Qy        832 GAGGGCAGGAGATGGTCAGAAAGAGACTTGGCTGCTTCTGAGGGCTTCCGCTCTCCTTTA 891
              |||||||||||| ||||||| ||||||||   |||||||||| ||||   |||||||||
Db     133667 GAGGGCAGGAGAAGGTCAGAGAGAGACTT---TGCTTCTGAGGCCTTCTAATCTCCTTTA 133723

Qy        892 GTTCCAAGTACTTCTTAGCATACCAAAGCACTATACTTTGGCATATGGTTTTCTGAGCTC 951
              |||| |||||   || ||| | || |||  | | ||||  |  ||| ||||||||||| |
Db     133724 GTTCAAAGTA---CTCAGCCTGCCCAAGGTCCAGACTTCAGGGTATTGTTTTCTGAGCCC 133780

Qy        952 TAACACT 958
                |||||
Db     133781 CGACACT 133787


RESULT 66
ADQ97391
ID   ADQ97391 standard; DNA; 120447 BP.
XX
AC   ADQ97391;
XX
DT   07-OCT-2004  (first entry)
XX
DE   Human cancer associated sequence HD08-042, SEQ ID 368.
XX
KW   Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Human; ds.
XX
OS   Homo sapiens.
XX
PN   WO2004060304-A2.
```

```
XX
PD   22-JUL-2004.
XX
PF   22-DEC-2003; 2003WO-US041389.
XX
PR   27-DEC-2002; 2002US-00330773.
XX
PA   (SAGR-) SAGRES DISCOVERY INC.
XX
PI   Morris DW,  Malandro MS;
XX
DR   WPI; 2004-543781/52.
XX
PT   New isolated cancer associated nucleic acids comprising at least 10
PT   contiguous nucleotides, useful for diagnosing, preventing and/or treating
PT   cancers such as leukemia and lymphoma.
XX
PS   Claim 1; SEQ ID NO 368; 199pp; English.
XX
CC   The present invention relates to cancer associated sequences (ADQ97025-
CC   ADQ98004). The sequences are useful for the diagnosis, prevention and/or
CC   treatment of cancer, such as leukemia and lymphoma. Note: The sequence
CC   data for this patent did not form part of the printed specification, but
CC   was obtained in electronic formate directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 120447 BP; 33790 A; 24730 C; 24661 G; 33309 T; 0 U; 3957 Other;

  Query Match             2.4%;  Score 144.8;  DB 12;  Length 120447;
  Best Local Similarity   75.0%;  Pred. No. 9.9e-24;
  Matches  195;  Conservative    0;  Mismatches   62;  Indels    3;  Gaps    1;

Qy       2180 TATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCTGAC 2239
             ||| |||||||  ||| |  | ||| ||| || | |||||||||  |    |||| ||  |
Db      30493 TATACATCACCCAAATAACATACATGGTATCCATGAAGTAATTTGT---TATCCACTCCC 30549

Qy       2240 TTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTCTACATCTATG 2299
              ||    | || ||   | |  || |||||| ||| ||||||||| ||||||  || ||
Db      30550 CTCCCCTCTCCCCACTGTCCCAAGTCTCCATGGTCTATCATTCCATACTCTATGTCCATT 30609

Qy       2300 TGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGTCTTTCTGTGT 2359
             ||||||||||||||| | | |||  | ||||| |||||||| |||||||||||||||||
Db      30610 TGTACACATTATTTAGTTTCCACTCCTGAGTGAGAACATGCAGTATTTGTCTTTCTGTGT 30669

Qy       2360 CTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAAGGCAT 2419
              |||  |||||| ||||| |||||||||| ||||||| | | | ||||||||||| |||
Db      30670 CTGAGTTGTTTCACTTAAGATAATGGCCTCCAGTTCCACCCATGTTGCTGCAAAAGACAT 30729

Qy       2420 GATTTCATTCTTTTTTATGG 2439
             ||||| ||| ||| |  | |
Db      30730 GATTTTATTTTTTATGGTTG 30749


RESULT 67
ABK89296
ID   ABK89296 standard; DNA; 326014 BP.
XX
AC   ABK89296;
XX
DT   16-OCT-2002  (first entry)
XX
DE   Human gene for novel serine/threonine serine kinase.
XX
KW   Human; ds; gene; serine/threonine protein kinase; inflammation; cancer;
KW   arteriosclerosis; psoriasis; SNF kinase; transgenic; chromosome 3; SNP;
KW   single nucleotide polymorphism.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       replace(1139. .1140,GGT)
FT                   /*tag=  ai
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(1399,G)
FT                   /*tag=  aj
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(1871. .1873,TG)
FT                   /*tag=  ak
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(1875. .1877,TA)
FT                   /*tag=  al
FT                   /standard_name= "Single nucleotide polymorphism"
FT   CDS             3000. .323016
FT                   /*tag=  a
FT                   /product= "Serine/threonine protein kinase"
FT   exon            3000. .3169
FT                   /*tag=  b
FT                   /number= 1
FT   intron          3170. .53737
FT                   /*tag=  c
FT                   /number= 1
FT   variation       replace(3180,G)
FT                   /*tag=  am
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(7590,A)
FT                   /*tag=  an
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(9315,C)
FT                   /*tag=  ao
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(9572,A)
FT                   /*tag=  ap
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(11004,G)
FT                   /*tag=  aq
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(11379. .11380,AAA)
FT                   /*tag=  ar
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(12478,A)
FT                   /*tag=  as
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(12486. .12487,TTT)
FT                   /*tag=  at
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(13904,G)
FT                   /*tag=  au
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(25444. .25446,TT)
FT                   /*tag=  av
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(26805,A)
FT                   /*tag=  aw
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(28010,R)
FT                   /*tag=  ax
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(28107,Y)
FT                   /*tag=  ay
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(28178,R)
FT                   /*tag=  az
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(28356,T)
FT                   /*tag=  ba
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(29535,G)
FT                   /*tag=  bb
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(31808,T)
FT                   /*tag=  bc
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(32807,C)
FT                   /*tag=  bd
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(34531,G)
```

```
FT                   /*tag= be
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(35318,G)
FT                   /*tag= bf
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(35463,C)
FT                   /*tag= bg
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(35477,T)
FT                   /*tag= bh
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(37050,G)
FT                   /*tag= bi
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(37470,G)
FT                   /*tag= bj
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(37481,A)
FT                   /*tag= bk
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(37499,T)
FT                   /*tag= bl
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(40178. .40179,GRA)
FT                   /*tag= bm
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(40787,A)
FT                   /*tag= bn
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(44582,G)
FT                   /*tag= bo
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(45615,A)
FT                   /*tag= bp
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(46091,T)
FT                   /*tag= bq
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(47844,C)
FT                   /*tag= br
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(48013,T)
FT                   /*tag= bs
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(48072,A)
FT                   /*tag= bt
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(50802,T)
FT                   /*tag= bu
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(53305,A)
FT                   /*tag= bv
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(53378. .53380,GG)
FT                   /*tag= bw
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            53738. .53903
FT                   /*tag= d
FT                   /number= 2
FT   intron          53904. .83114
FT                   /*tag= e
FT                   /number= 2
FT   variation       replace(56233,G)
FT                   /*tag= bx
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(56453,C)
FT                   /*tag= by
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(59300,T)
FT                   /*tag= bz
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(60998,A)
FT                   /*tag= ca
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(63013,A)
FT                   /*tag= cb
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(63293,A)
FT                   /*tag= cc
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(64181,T)
FT                   /*tag= cd
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(64191,A)
FT                   /*tag= ce
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(68992,C)
FT                   /*tag= cf
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(76246,Y)
FT                   /*tag= cg
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(77368,G)
FT                   /*tag= ch
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(78043,A)
FT                   /*tag= ci
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(79395,T)
FT                   /*tag= cj
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(79474,G)
FT                   /*tag= ck
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(79992,T)
FT                   /*tag= cl
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(81065,C)
FT                   /*tag= cm
```

```
  Query Match             2.4%;  Score 144.8;  DB 6;  Length 326014;
  Best Local Similarity   77.3%;  Pred. No. 1.6e-23;
  Matches  201;  Conservative    0;  Mismatches   57;  Indels    2;  Gaps    2;

Qy       2177 GGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTC-ACCATCCGC 2235
              | ||| |||||||||||||   || |||| ||||| ||||||||||| |  | ||  | |
Db     218474 GTGTAATCATCACCTGAATATGGTACATTATACCCATTAAGTAATTTTTTTATCACTCAC 218533

Qy       2236 TGACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTCTACATC 2295
                ||    |  | ||  |||||||| |||||  |||| | | ||||||||| || |||
Db     218534 CTCTCTCCCAACTTCCCACTCTTCTGAGTCTCCAACGTCCGTTACTCCACACTGTATATC 218593

Qy       2296 TATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGTCTTTCT 2355
              || |||||||||||||||||||| |||||||||||| |||||    |||||| | ||||
Db     218594 CATATGTACACATTATTTAGCTCCCACTTATAAGTGAGAACAT-TGGTATTTGACATTCT 218652

Qy       2356 GTGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAAG 2415
              || ||||  |||||| ||||| |||||||||| |||||||| |||| | ||||||||||
Db     218653 GTTTCTGAGTTGTTTCACTTAAGATAATGGCCTCCAGTTCTTTCTATGTTGCTGCAAAAT 218712

Qy       2416 GCATGATTTCATTCTTTTTT 2435
              |||||||| ||||||||||
Db     218713 TCATGATTTTATTCTTTTTT 218732
```

```
RESULT 68
ADQ94981
ID   ADQ94981 standard; DNA; 326014 BP.
XX
AC   ADQ94981;
XX
DT   23-SEP-2004  (first entry)
XX
DE   Human kinase genomic DNA.
XX
KW   Kinase; diagnosis; cancer; tissue growth abnormality;
```

```
KW   drug screening assay; cytostatic; gene therapy; human; SNP;
KW   single nucleotide polymorphism; gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       replace(1138. .1140,GT)
FT                   /*tag= a
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(1399,G)
FT                   /*tag= b
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(1871. .1873,TG)
FT                   /*tag= c
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(1875. .1877,TA)
FT                   /*tag= d
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   CDS             3000. .23016
FT                   /*tag= f
FT                   /product= "Human kinase protein"
FT                   /transl_except= (pos:3024. .3026, aa:Arg)
FT                   /transl_except= (pos:142261. .142263, aa:Glu)
FT                   /transl_except= (pos:142261. .144102, aa:Glu-Ser)
FT   exon            3000. .3169
FT                   /*tag= e
FT   intron          3170. .53737
FT                   /*tag= g
FT   variation       replace(3180,A)
FT                   /*tag= h
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(7590,A)
FT                   /*tag= i
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(9315,C)
FT                   /*tag= j
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(9572,A)
FT                   /*tag= k
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(11004,G)
FT                   /*tag= l
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(11378. .11380,AA)
FT                   /*tag= m
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(12478,A)
FT                   /*tag= n
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(12485. .12487,TC)
FT                   /*tag= o
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(13904,G)
FT                   /*tag= p
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(25444. .25446,TT)
FT                   /*tag= q
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(26805,A)
FT                   /*tag= r
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(28010,R)
FT                   /*tag= s
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(28107,Y)
FT                   /*tag= t
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(28178,R)
FT                   /*tag= u
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(28356,T)
FT                   /*tag= v
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(29535,G)
FT                   /*tag= w
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(31808,T)
FT                   /*tag= x
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(32807,C)
FT                   /*tag= y
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(34531,G)
FT                   /*tag= z
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(35318,G)
FT                   /*tag= aa
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(35463,G)
FT                   /*tag= ab
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(35477,T)
FT                   /*tag= ac
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(37050,G)
FT                   /*tag= ad
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(37470,G)
FT                   /*tag= ae
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(37481,A)
FT                   /*tag= af
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(37499,T)
FT                   /*tag= ag
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(40177. .40179,AA)
FT                   /*tag= ah
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(40178,A)
FT                   /*tag= ai
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(40787,A)
FT                   /*tag= aj
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(44582,G)
FT                   /*tag= ak
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(45615,A)
FT                   /*tag= al
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(46091,T)
FT                   /*tag= am
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(47844,C)
FT                   /*tag= an
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(48013,T)
FT                   /*tag= ao
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(48072,A)
FT                   /*tag= ap
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(50802,T)
FT                   /*tag= aq
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(53305,A)
FT                   /*tag= ar
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(53378. .53380,GG)
FT                   /*tag= as
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   exon            53738. .53903
FT                   /*tag= at
FT   intron          53904. .83114
FT                   /*tag= au
```

```
FT   variation       replace(56233,G)
FT                   /*tag=  av
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(56453,C)
FT                   /*tag=  aw
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(59300,T)
FT                   /*tag=  ax
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(60998,A)
FT                   /*tag=  ay
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(63013,A)
FT                   /*tag=  az
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(63293,A)
FT                   /*tag=  ba
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(64181,T)
FT                   /*tag=  bb
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(64191,A)
FT                   /*tag=  bc
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(68992,C)
FT                   /*tag=  bd
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(76246,Y)
FT                   /*tag=  be
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(77368,G)
FT                   /*tag=  bf
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(78043,A)
FT                   /*tag=  bg
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(79395,T)
FT                   /*tag=  bh
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(79474,G)
FT                   /*tag=  bi
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"
FT   variation       replace(79992,T)
FT                   /*tag=  bj
FT                   /standard_name= "Single nucleotide polymorphism (SNP)"

  Query Match             2.4%;  Score 144.8;  DB 12;  Length 326014;
  Best Local Similarity   77.3%;  Pred. No. 1.6e-23;
  Matches  201;  Conservative    0;  Mismatches   57;  Indels    2;  Gaps    2;

Qy       2177 GGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTC-ACCATCCGC 2235
              | ||| |||||||||||||   || |||| ||||| ||||||||||| |  | ||  | |
Db     218474 GTGTAATCATCACCTGAATATGGTACATTATACCCATTAAGTAATTTTTTTATCACTCAC 218533

Qy       2236 TGACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTCTACATC 2295
                ||    |  | ||  |||||||| |||||  |||| | | |||||||| || |||
Db     218534 CTCTCTCCCAACTTCCCACTCTTCTGAGTCTCCAACGTCCGTTACTCCACACTGTATATC 218593

Qy       2296 TATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGTCTTTCT 2355
              || |||||||||||||||||||| |||||||||||| |||||    |||||| | ||||
Db     218594 CATATGTACACATTATTTAGCTCCCACTTATAAGTGAGAACAT-TGGTATTTGACATTCT 218652

Qy       2356 GTGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAAG 2415
              || ||||  |||||| ||||| |||||||||| |||||||| |||| | ||||||||||
Db     218653 GTTTCTGAGTTGTTTCACTTAAGATAATGGCCTCCAGTTCTTTCTATGTTGCTGCAAAAT 218712

Qy       2416 GCATGATTTCATTCTTTTTT 2435
              |||||||| ||||||||||
Db     218713 TCATGATTTTATTCTTTTTT 218732


RESULT 69
ABT10146
```

```
ID   ABT10146 standard; cDNA; 140167 BP.
XX
AC   ABT10146;
XX
DT   04-DEC-2002  (first entry)
XX
DE   Human breast cancer associated coding sequence SEQ ID NO: 280.
XX
KW   Human; breast specific gene; breast cancer; differential expression;
KW   cytostatic; gene therapy; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200259271-A2.
XX
PD   01-AUG-2002.
XX
PF   25-JAN-2002; 2002WO-US002176.
XX
PR   25-JAN-2001; 2001US-0263757P.
PR   25-APR-2001; 2001US-0286090P.
PR   23-MAY-2001; 2001US-0292517P.
XX
PA   (GENE-) GENE LOGIC INC.
XX
PI   Orr MS,  Nation M,  Diggans JC,  Zeng W;
XX
DR   WPI; 2002-674803/72.
XX
PT   Diagnosing breast cancer in a patient comprises detecting the level of
PT   gene expression in cell or tissue samples, where a differential gene
PT   expression is indicative of breast cancer.
XX
PS   Claim 1; SEQ ID NO 280; 260pp + Sequence Listing; English.
XX
CC   The present invention relates to methods of diagnosing breast cancer in a
CC   patient, which comprise detecting the level of expression in a tissue
CC   sample of two or more genes selected from those shown in ABT09867-
CC   ABT11112, where a differential expression of the genes indicates breast
CC   cancer. The methods are useful in diagnosing, treating, detecting the
CC   progression, and in monitoring treatment of breast cancer in patients.
CC   The methods are also useful as a screening tool for agents that modulate
CC   the onset or progression of breast cancer. The breast cancer genes may be
CC   used as diagnostic markers for the prediction or identification of the
CC   malignant state of breast tissue, for confirming the type and progression
CC   of cancer, and for drug screening and assays. The present sequence is a
CC   coding sequence of the invention. Note: The sequence data for this patent
CC   did not form part of the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub.published_pct_sequences
XX
SQ   Sequence 140167 BP; 45038 A; 27502 C; 26800 G; 40827 T; 0 U; 0 Other;

  Query Match             2.4%;  Score 144.4;  DB 6;  Length 140167;
  Best Local Similarity   75.7%;  Pred. No. 1.3e-23;
  Matches  206;  Conservative    0;  Mismatches   61;  Indels    5;  Gaps    2;

Qy       2177 GGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCT 2236
              | |||  ||||||    | | |||| ||| |||||| ||||||||||||||| ||||| |
Db     1991 GTGTACACATCACTGAAGTAATGTACATGGTACCCATTAAGTAATTTCTCATCATCCACC 2050

Qy       2237 GACTTCTTGC-CCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTCTACATC 2295
                ||||     ||||||| |||| ||  ||||| |    | |||||||| |||||| ||
Db     2051 TCCTTCCCATGCCCCTCACCCTTTAGATCCTCCACT----AACATTCCACTCTCTACTTC 2106

Qy       2296 TATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGTCTTTCT 2355
              | ||| |||||||||| ||||||  ||||||||| | || ||||| || ||||||||||
Db     2107 CAAGTGCACACATTATTCAGCTCCCTCTTATAAGTTAGAATATGCAGTACTTGTCTTTCT 2166

Qy       2356 GTGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAAG 2415
              |||||||  ||| || ||||| || |||||||  |  |||  ||| |  ||||||||||
Db     2167 GTGTCTGCGTTGCTTCACTTAAGAAAATGGCCTCTGGTTGCATCCACATTGCTGCAAAAG 2226
```

```
Qy       2416 GCATGATTTCATTCTTTTTTATGGCTATGTTC 2447
              ||||||||||||||| |||||||  ||| | |
Db       2227 ACATGATTTCATTCTCTTTTATGTTTATATAC 2258
```

RESULT 70
ADL13638/c
ID ADL13638 standard; DNA; 186528 BP.
XX
AC ADL13638;
XX
DT 06-MAY-2004 (first entry)
XX
DE Osteoarthritis-associated polymorphic nucleotide #170.
XX
KW ds; gene; osteopathic; antiinflammatory; antiarthritic; gene therapy;
KW joint space narrowing; osteophyte development; joint pain;
KW osteoarthritis; SNP; single nucleotide polymorphism.
XX
OS Homo sapiens.
XX
PN WO2003054166-A2.
XX
PD 03-JUL-2003.
XX
PF 19-DEC-2002; 2002WO-US041225.
XX
PR 20-DEC-2001; 2001US-0342603P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Jones KA, Schafer A;
XX
DR WPI; 2003-559141/52.
XX
PT Determining susceptibility of an individual to joint space narrowing,
PT osteophyte development and/or joint pain comprises identifying whether
PT the individual has at least one polymorphism in a polynucleotide encoding
PT a protein.
XX
PS Disclosure; SEQ ID NO 170; 297pp; English.
XX
CC The invention relates to a method of determining susceptibility of an
CC individual to joint space narrowing and/or osteophyte development and/or
CC joint pain comprising identifying whether the individual has at least one
CC polymorphism in a polynucleotide encoding at least one of the protein
CC listed in the specification. The methods, composition and agent are
CC useful for modulating the susceptibility of an individual to joint space
CC narrowing and/or osteophyte development and/or joint pain that is
CC associated with a disease, preferably osteoarthritis. The cell line and
CC the non-human animal are useful for screening for an agent for diagnosing
CC an individual having susceptibility to joint space narrowing and/or
CC osteophyte development and/or joint pain. This sequence corresponds to
CC the polynucleotide encoding a protein listed in the specification. (Note:
CC The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences).
XX
SQ Sequence 186528 BP; 49245 A; 38352 C; 38583 G; 46735 T; 0 U; 13613 Other;

```
  Query Match             2.4%;  Score 144.2;  DB 10;  Length 186528;
  Best Local Similarity   71.9%;  Pred. No. 1.7e-23;
  Matches  202;  Conservative    0;  Mismatches   78;  Indels    1;  Gaps    1;

Qy       2170 GCCCCCTGGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACC 2229
              | |  | | || | ||||||  ||   | |  |||||||||| | ||  ||  || || |
Db     102643 GTCTTCAGTGTGTCCATCACTGGAGCAAAGCACATTGTACCCATCAAACAAACTCCCATC 102584

Qy       2230 ATCCGCTGACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTC 2289
              ||| |   ||||   |||| ||| || ||  || | || ||||||||||||||||||||
Db     102583 CTCCACGCGCTTCCCTCCCCTTCACCCCTCCAAGTCCCCGTTGTCCATCATTCCACACTC 102524

Qy       2290 TACATCTATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGT 2349
              | | ||  |||||||||||| ||||||  || |||| ||||||| | ||||  |||||||
Db     102523 TTCTTCCCTGTGTACACATCATTTAGTGCCCACTT-TAAGTGACAGCATGTGATATTTGG 102465

Qy       2350 CTTTCTGTGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTG 2409
              ||||||||||   |||||| |||||  |||| |||| ||| ||| ||| | | ||||
Db     102464 TTTTCTGTGTCTCAGTTGTTTCACTTAGCATAACGGCCTCCATTTCCATCCATGTTGCTA 102405

Qy       2410 CAAAAGGCATGATTTCATTCTTTTTTATGGCTATGTTCTTT 2450
              |||| | |||||||||| |||| |||| |   ||| || | |
Db     102404 CAAAGGACATGATTTCCTTCTCTTTTTTCTTTATTTTTTAT 102364
```

RESULT 71
ADJ12604
ID ADJ12604 standard; DNA; 22913 BP.
XX
AC ADJ12604;
XX
DT 20-MAY-2004 (first entry)
XX
DE DNA fragment of a BAC clone that encodes a human secreted protein Seq458.
XX
KW human; secreted; cancer; haematopoietic disease; anaemia;
KW multiple myeloma; reproductive system disorder; prostatitis;
KW inguinal hernia; musculoskeletal disease; systemic lupus erythematosus;
KW gout; cardiovascular disease; arrhythmia; hypernatraemia; fetal disease;
KW fetal alcohol syndrome; Down's syndrome; excretory disease;
KW urinary incontinence; renal disorder; neural; sensory disease;
KW Alzheimer's disease; meningitis; respiratory disease; emphysema;
KW occupational lung disease; endocrine disease; diabetes;
KW glomerulonephritis; digestive disease; portal hypertension;
KW irritable bowel syndrome; epithelial disease; scleroderma;
KW epidermolysis bullosa; cytostatic; antianemic; antiarthritic;
KW antiasthmatic; anti-HIV; immunosuppressive; antiinflammatory;
KW antipsoriatic; antibacterial; osteopathic; dermatological; antigout;
KW immunomodulator; antiarrhythmic; cardiant; nootropic; antilipemic;
KW nephrotropic; uropathic; neuroprotective; antiparkinsonian; tranquilizer;
KW antidiabetic; anabolic; hypertensive; vulnerary; ds.
XX
OS Homo sapiens.
XX
PN US2004010132-A1.
XX
PD 15-JAN-2004.
XX
PF 30-OCT-2001; 2001US-00984429.
XX
PR 09-OCT-1997; 97US-0061463P.
PR 09-OCT-1997; 97US-0061527P.
PR 09-OCT-1997; 97US-0061529P.
PR 09-OCT-1997; 97US-0061532P.
PR 09-OCT-1997; 97US-0061536P.
PR 09-OCT-1997; 97US-0071498P.
PR 08-OCT-1998; 98WO-US021142.
PR 08-APR-1999; 99US-00288143.
PR 01-NOV-2000; 2000US-0244591P.
XX
PA (ROSE/) ROSEN C A.
PA (BREW/) BREWER L A.
PA (DUAN/) DUAN R D.
PA (RUBE/) RUBEN S M.
PA (FLOR/) FLORENCE K A.
PA (GREE/) GREENE J M.
PA (YOUN/) YOUNG P E.
PA (FERR/) FERRIE A M.
PA (YUGG/) YU G.
PA (FLOR/) FLORENCE C.
PA (EBNE/) EBNER R.
PA (OLSE/) OLSEN H.
XX
PI Rosen CA, Brewer LA, Duan RD, Ruben SM, Florence KA, Greene JM;
PI Young PE, Ferrie AM, Yu G, Florence C, Ebner R, Olsen H;
XX

```
DR   WPI; 2004-090518/09.
XX
PT   New isolated nucleic acids and polypeptides, useful for diagnosing,
PT   treating, preventing or ameliorating diseases or disorders e.g. cancer,
PT   anemia, arthritis, asthma, inflammatory bowel disease or Alzheimer's
PT   disease.
XX
PS   Disclosure; SEQ ID NO 458; 286pp; English.
XX
CC   This invention relates to novel polynucleotides encoding human secreted
CC   proteins. Specifically, it refers to the vectors, host cells, recombinant
CC   and synthetic methods for producing human polynucleotides, polypeptides
CC   and antibodies. Furthermore, it relates to screening methods to identify
CC   agonists and antagonists that can be used to inhibit or enhance the
CC   production and function of the secreted proteins. The present invention
CC   describes these compositions as useful for diagnosing, treating or
CC   preventing disorders such as cancer, haematopoietic diseases including
CC   anaemia and multiple myeloma, reproductive system disorders including
CC   prostatitis and inguinal hernia, musculoskeletal diseases including
CC   systemic lupus erythematosus and gout, cardiovascular disease including
CC   arrhythmia and hypernatraemia, mixed fetal diseases including fetal
CC   alcohol syndrome and Down's syndrome, excretory diseases including
CC   urinary incontinence and renal disorders, neural or sensory disease
CC   including Alzheimer's disease and meningitis, respiratory disease
CC   including emphysema and occupational lung disease, endocrine diseases
CC   including diabetes and glomerulonephritis, digestive diseases including
CC   portal hypertension and irritable bowel syndrome and connective tissue or
CC   epithelial diseases including scleroderma and epidermolysis bullosa. As
CC   such, there are various activites such as cytostatic, antianemic,
CC   antiarthritic, antiasthmatic, anti-HIV, immunosuppressive,
CC   antiinflammatory, antipsoriatic, antibacterial, osteopathic,
CC   dermatological, antigout, immunomodulator, antiarrhythmic, cardiant,
CC   nootropic, antilipemic, nephrotropic, uropathic, neuroprotective,
CC   antiparkinsonian, tranquilizer, antidiabetic, anabolic, hypertensive and
CC   vulnerary. This polynucleotide is a DNA fragment of a BAC clone that
CC   encodes a human secreted protein of the invention. NOTE: This sequence
CC   does not appear in the printed specification but has been obtained in
CC   electronic format from the US patent office at the following web site
CC   www.seqdata.uspto.gov/sequence.html; Document ID: 20040010132.
XX
SQ   Sequence 22913 BP; 7363 A; 5191 C; 4930 G; 5428 T; 0 U; 1 Other;

  Query Match             2.4%;  Score 143.6;  DB 12;  Length 22913;
  Best Local Similarity   70.7%;  Pred. No. 9e-24;
  Matches  191;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;

Qy       2179 GTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCTGA 2238
              |||| |||||||  ||    |  ||||||||||   | |||||||  ||| |   | ||
Db      10248 GTATCCATCACCCAAACAGCGAACATTGTACCCAAGAGGTAATTTTTCAACCCTCACTCC 10307

Qy       2239 CTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTCTACATCTAT 2298
              | || | ||  | || | ||   || ||||| ||||  | |||  || || ||  ||| |
Db      10308 CCTCCTACCTTCCCACCTTTTGTAGTCTCCAGTGTCTGTTATTTTACTCTGTATGTCTGT 10367

Qy       2299 GTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGTCTTTCTGTG 2358
              |||||| |||| ||||||||| |||||||| |||  ||||||  |||||| ||| ||||
Db      10368 GTGTACCCATTGTTTAGCTCCCACTTATAAATGAGCACATGCGGTATTTGACTTCCTGTT 10427

Qy       2359 TCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAAGGCA 2418
              ||||  || ||| || || ||||||||||  |||||||||| | | ||||||||||| ||
Db      10428 TCTGAGTTATTTCACATAGGATAATGGCCTTCAGTTCTATCCATGTTGCTGCAAAAGACA 10487

Qy       2419 TGATTTCATTCTTTTTTATGGCTATGTTCT 2448
              || ||||||| || | |      ||  |||
Db      10488 TGGTTTCATTTTTATATGCATTAATTATCT 10517


RESULT 72
AAK95240_03
Continuation (4 of 15) of AAK95240 from base 300001 (Human neuregulin-1 gene. )
WP Sequence split into 15 fragments  LOCUS AAK95240 Accession Aak95240
WP     Fragment Name      Begin     End
WP     AAK95240_00            1    110000
WP     AAK95240_01       100001    210000
WP     AAK95240_02       200001    310000
WP     AAK95240_03       300001    410000
WP     AAK95240_04       400001    510000
WP     AAK95240_05       500001    610000
WP     AAK95240_06       600001    710000
WP     AAK95240_07       700001    810000
WP     AAK95240_08       800001    910000
WP     AAK95240_09       900001   1010000
WP     AAK95240_10      1000001   1110000
WP     AAK95240_11      1100001   1210000
WP     AAK95240_12      1200001   1310000
WP     AAK95240_13      1300001   1410000
WP     AAK95240_14      1400001   1503900

  Query Match             2.3%;  Score 142.8;  DB 4;  Length 110000;
  Best Local Similarity   73.2%;  Pred. No. 2.9e-23;
  Matches  183;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy       2188 ACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCTGACTTCTTGCC 2247
              ||   ||  | ||  ||||| ||| |||| ||||||||||  ||    |    || |
Db      37824 ACAAAAAACAAGTATATTGTGCCCATTAAATAATTTCTCATTATTTATTCCTCTCCTATG 37883

Qy       2248 CCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTCTACATCTATGTGTACACA 2307
              || || |||||||||  || ||||||  ||| ||  | |||||   || |||||||  ||
Db      37884 TCCCCACCCTTCTGAGTTTCTATTGTCTATCTTTTTATACTCTGTGTCCATGTGTATGCA 37943

Qy       2308 TTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGTCTTTCTGTGTCTGTCTTG 2367
              |||||||||||| |||| ||| ||| ||||||||  |||| | ||||||||  ||| ||||
Db      37944 TTATTTAGCTCCCACTTGTAAATGAGAACATGCAGCATTTATGTTTCTGTGACTGACTTG 38003

Qy       2368 TTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAAGGCATGATTTCAT 2427
              ||| | ||| || ||||||| ||||||  ||| | | ||||||||||| ||||| || ||
Db      38004 TTTCAATTAAGACAATGGCCTCCAGTTTCATCCATGTTGCTGCAAAAGACATGACTTTAT 38063

Qy       2428 TCTTTTTTAT 2437
              ||| |||| |
Db      38064 TCTGTTTTTT 38073


RESULT 73
AAK96733_03
Continuation (4 of 15) of AAK96733 from base 300001 (Human neuregulin-1 gene. )
WP Sequence split into 15 fragments  LOCUS AAK96733 Accession Aak96733
WP     Fragment Name      Begin     End
WP     AAK96733_00            1    110000
WP     AAK96733_01       100001    210000
WP     AAK96733_02       200001    310000
WP     AAK96733_03       300001    410000
WP     AAK96733_04       400001    510000
WP     AAK96733_05       500001    610000
WP     AAK96733_06       600001    710000
WP     AAK96733_07       700001    810000
WP     AAK96733_08       800001    910000
WP     AAK96733_09       900001   1010000
WP     AAK96733_10      1000001   1110000
WP     AAK96733_11      1100001   1210000
WP     AAK96733_12      1200001   1310000
WP     AAK96733_13      1300001   1410000
WP     AAK96733_14      1400001   1503900

  Query Match             2.3%;  Score 142.8;  DB 4;  Length 110000;
  Best Local Similarity   73.2%;  Pred. No. 2.9e-23;
  Matches  183;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy       2188 ACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCTGACTTCTTGCC 2247
              ||   ||  | ||  ||||| ||| |||| ||||||||||  ||    |    || |
Db      37824 ACAAAAAACAAGTATATTGTGCCCATTAAATAATTTCTCATTATTTATTCCTCTCCTATG 37883

Qy       2248 CCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTCTACATCTATGTGTACACA 2307
              || || |||||||||  || ||||||  ||| ||  | |||||   || |||||||  ||
Db      37884 TCCCCACCCTTCTGAGTTTCTATTGTCTATCTTTTTATACTCTGTGTCCATGTGTATGCA 37943
```

```
Qy       2308 TTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGTCTTTCTGTGTCTGTCTTG 2367
             ||||||||||||| |||| ||| ||| |||||||| |||| | |||||||| ||| ||||
Db      37944 TTATTTAGCTCCCACTTGTAAATGAGAACATGCAGCATTTATGTTTCTGTGACTGACTTG 38003

Qy       2368 TTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAAGGCATGATTTCAT 2427
             ||| | ||| || ||||||| |||||| ||| | | |||||||||||| ||||| || ||
Db      38004 TTTCAATTAAGACAATGGCCTCCAGTTTCATCCATGTTGCTGCAAAAGACATGACTTTAT 38063

Qy       2428 TCTTTTTTAT 2437
             ||| |||| |
Db      38064 TCTGTTTTTT 38073


RESULT 74
ABT00010_03
Continuation (4 of 15) of ABT00010 from base 300001 (Human neuregulin 1 gene. )
WP Sequence split into 15 fragments  LOCUS ABT00010 Accession Abt00010
WP     Fragment Name      Begin     End
WP     ABT00010_00            1    110000
WP     ABT00010_01       100001    210000
WP     ABT00010_02       200001    310000
WP     ABT00010_03       300001    410000
WP     ABT00010_04       400001    510000
WP     ABT00010_05       500001    610000
WP     ABT00010_06       600001    710000
WP     ABT00010_07       700001    810000
WP     ABT00010_08       800001    910000
WP     ABT00010_09       900001   1010000
WP     ABT00010_10      1000001   1110000
WP     ABT00010_11      1100001   1210000
WP     ABT00010_12      1200001   1310000
WP     ABT00010_13      1300001   1410000
WP     ABT00010_14      1400001   1503841

  Query Match             2.3%;  Score 142.8;  DB 6;  Length 110000;
  Best Local Similarity   73.2%;  Pred. No. 2.9e-23;
  Matches  183;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy       2188 ACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCTGACTTCTTGCC 2247
             ||   ||  | ||  ||||| ||| |||| ||||||||||  ||    |    || |
Db      37824 ACAAAAAACAAGTATATTGTGCCCATTAAATAATTTCTCATTATTTATTCCTCTCCTATG 37883

Qy       2248 CCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTCTACATCTATGTGTACACA 2307
              || || |||||||||  || ||||||  ||| ||  | |||||   || |||||||  ||
Db      37884 TCCCCACCCTTCTGAGTTTCTATTGTCTATCTTTTTATACTCTGTGTCCATGTGTATGCA 37943

Qy       2308 TTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGTCTTTCTGTGTCTGTCTTG 2367
             ||||||||||||| |||| ||| ||| |||||||| |||| | |||||||| ||| ||||
Db      37944 TTATTTAGCTCCCACTTGTAAATGAGAACATGCAGCATTTATGTTTCTGTGACTGACTTG 38003

Qy       2368 TTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAAGGCATGATTTCAT 2427
             ||| | ||| || ||||||| |||||| ||| | | |||||||||||| ||||| || ||
Db      38004 TTTCAATTAAGACAATGGCCTCCAGTTTCATCCATGTTGCTGCAAAAGACATGACTTTAT 38063

Qy       2428 TCTTTTTTAT 2437
             ||| |||| |
Db      38064 TCTGTTTTTT 38073


RESULT 75
ABT01503_03
Continuation (4 of 15) of ABT01503 from base 300001 (Human neuregulin 1 gene. )
WP Sequence split into 15 fragments  LOCUS ABT01503 Accession Abt01503
WP     Fragment Name      Begin     End
WP     ABT01503_00            1    110000
WP     ABT01503_01       100001    210000
WP     ABT01503_02       200001    310000
WP     ABT01503_03       300001    410000
WP     ABT01503_04       400001    510000
WP     ABT01503_05       500001    610000
WP     ABT01503_06       600001    710000
WP     ABT01503_07       700001    810000
WP     ABT01503_08       800001    910000
WP     ABT01503_09       900001   1010000
WP     ABT01503_10      1000001   1110000
WP     ABT01503_11      1100001   1210000
WP     ABT01503_12      1200001   1310000
WP     ABT01503_13      1300001   1410000
WP     ABT01503_14      1400001   1503841

  Query Match             2.3%;  Score 142.8;  DB 6;  Length 110000;
  Best Local Similarity   73.2%;  Pred. No. 2.9e-23;
  Matches  183;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy       2188 ACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCTGACTTCTTGCC 2247
             ||   ||  | ||  ||||| ||| |||| ||||||||||  ||    |    || |
Db      37824 ACAAAAAACAAGTATATTGTGCCCATTAAATAATTTCTCATTATTTATTCCTCTCCTATG 37883

Qy       2248 CCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTCTACATCTATGTGTACACA 2307
              || || |||||||||  || ||||||  ||| ||  | |||||   || |||||||  ||
Db      37884 TCCCCACCCTTCTGAGTTTCTATTGTCTATCTTTTTATACTCTGTGTCCATGTGTATGCA 37943

Qy       2308 TTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGTCTTTCTGTGTCTGTCTTG 2367
             ||||||||||||| |||| ||| ||| |||||||| |||| | |||||||| ||| ||||
Db      37944 TTATTTAGCTCCCACTTGTAAATGAGAACATGCAGCATTTATGTTTCTGTGACTGACTTG 38003

Qy       2368 TTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAAGGCATGATTTCAT 2427
             ||| | ||| || ||||||| |||||| ||| | | |||||||||||| ||||| || ||
Db      38004 TTTCAATTAAGACAATGGCCTCCAGTTTCATCCATGTTGCTGCAAAAGACATGACTTTAT 38063

Qy       2428 TCTTTTTTAT 2437
             ||| |||| |
Db      38064 TCTGTTTTTT 38073


RESULT 76
ADH77486_03
Continuation (4 of 15) of ADH77486 from base 300001 (Human neuregulin gene. )
WP Sequence split into 15 fragments  LOCUS ADH77486 Accession Adh77486
WP     Fragment Name      Begin     End
WP     ADH77486_00            1    110000
WP     ADH77486_01       100001    210000
WP     ADH77486_02       200001    310000
WP     ADH77486_03       300001    410000
WP     ADH77486_04       400001    510000
WP     ADH77486_05       500001    610000
WP     ADH77486_06       600001    710000
WP     ADH77486_07       700001    810000
WP     ADH77486_08       800001    910000
WP     ADH77486_09       900001   1010000
WP     ADH77486_10      1000001   1110000
WP     ADH77486_11      1100001   1210000
WP     ADH77486_12      1200001   1310000
WP     ADH77486_13      1300001   1410000
WP     ADH77486_14      1400001   1503841

  Query Match             2.3%;  Score 142.8;  DB 12;  Length 110000;
  Best Local Similarity   73.2%;  Pred. No. 2.9e-23;
  Matches  183;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy       2188 ACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCTGACTTCTTGCC 2247
             ||   ||  | ||  ||||| ||| |||| ||||||||||  ||    |    || |
Db      37824 ACAAAAAACAAGTATATTGTGCCCATTAAATAATTTCTCATTATTTATTCCTCTCCTATG 37883

Qy       2248 CCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTCTACATCTATGTGTACACA 2307
              || || |||||||||  || ||||||  ||| ||  | |||||   || |||||||  ||
Db      37884 TCCCCACCCTTCTGAGTTTCTATTGTCTATCTTTTTATACTCTGTGTCCATGTGTATGCA 37943

Qy       2308 TTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGTCTTTCTGTGTCTGTCTTG 2367
             ||||||||||||| |||| ||| ||| |||||||| |||| | |||||||| ||| ||||
Db      37944 TTATTTAGCTCCCACTTGTAAATGAGAACATGCAGCATTTATGTTTCTGTGACTGACTTG 38003

Qy       2368 TTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAAGGCATGATTTCAT 2427
             ||| | ||| || ||||||| |||||| ||| | | |||||||||||| ||||| || ||
Db      38004 TTTCAATTAAGACAATGGCCTCCAGTTTCATCCATGTTGCTGCAAAAGACATGACTTTAT 38063
```

```
Qy       2428 TCTTTTTTAT 2437
              ||| |||| |
Db      38064 TCTGTTTTTT 38073


RESULT 77
AAS46545/c
ID   AAS46545 standard; DNA; 7163 BP.
XX
AC   AAS46545;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Tumour suppressor gene derived chemically modified sequence #267.
XX
KW   Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer;
KW   tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW   cytosine methylation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200168912-A2.
XX
PD   20-SEP-2001.
XX
PF   15-MAR-2001; 2001WO-EP002955.
XX
PR   15-MAR-2000; 2000DE-01013847.
PR   06-APR-2000; 2000DE-01019058.
PR   07-APR-2000; 2000DE-01019173.
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2001-602752/68.
XX
PT   Fragments of chemically modified genes associated with tumor suppressor
PT   genes and oncogenes, useful in designing primers and probes for analyzing
PT   diseases associated with cytosine methylation state e.g. cancer.
XX
PS   Claim 1; SEQ ID NO 267; 27pp; English.
XX
CC   The invention relates to a nucleic acid comprising a sequence of 18
CC   bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC   bisulphite, of genes associated with tumour suppression and oncogenes
CC   having a sequence taken from 536 (actually 533 since numbers 408, 458 and
CC   500 are missing from the sequence listing) sequences (Ss) and sequences
CC   complementary to (Ss). The nucleic acid may be a peptide nucleic acid-
CC   oligomer (PNA) of at least 9 nucleotides and may form part of a set of
CC   probes for detecting the cytosine methylation state and/or single
CC   nucleotide polymorphisms and also to be used in an array for analysing
CC   diseases associated with CpG dinucleotides e.g. cancers and tumours. The
CC   probes can also be used in a method for ascertaining genetic and/or
CC   epigenetic parameters for the diagnosis and/or therapy of existing
CC   diseases or the predisposition to specific diseases, by analysing
CC   cytosine methylations. The parameters may be compared to another set of
CC   genetic and/or epigenetic parameters, the differences serving as basis
CC   for diagnosis and/or prognosis events which are disadvantageous to
CC   patients. The present sequence is one of the 533 genomic sequences
CC   derived from tumour suppressor genes and oncogenes. Note: The sequence
CC   data for this patent did not form part of the printed specification, but
CC   was obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 7163 BP; 1854 A; 216 C; 2051 G; 3042 T; 0 U; 0 Other;

  Query Match             2.3%;  Score 142;  DB 4;  Length 7163;
  Best Local Similarity   58.9%;  Pred. No. 1.3e-23;
  Matches  291;  Conservative    0;  Mismatches  185;  Indels   18;  Gaps    2;

Qy       1971 ATAAGTGAAATCTTTTTAGATGCATAGGTTGTATAGTGATAAAATCAGGGCTTTTAGGGT 2030
              | || |  ||| |||||| ||  ||| ||| |||| | |||||||||    |||||  |
Db       2147 ACAAATATAATTTTTTTACATAAATACGTTATATAATAATAAAATCAAACATTTTACTAT 2088

Qy       2031 ATTCATCACCTGAATGATGTACATTGTACCCCTTAAGTAATTTCTCACCATCCGCTGACT 2090
              || ||||| || ||| || ||||||    | ||  || || |        ||  |    ||
Db       2087 ATCCATCAACTAAATAATATACATTAACCTCCCAAAATACTAAAATTACAAACATAAACC 2028

Qy       2091 TCTTGCCCCCTGGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCT 2150
              |    ||       |   || |  | | |          |       | |||  ||
Db       2027 ACAATACCAAACCACAAATATAATTTTATTACACGAATACATTATATAATAATAAAATCA 1968

Qy       2151 CACCATCCGCTGACTTCTTGCCCCCTGGGTATTCATCACCTGAATGATGTGCATTGTACC 2210
              ||  |   ||                  ||| ||| |||| ||| ||   |||| |||
Db       1967 AACATTTTACTA----------------TATCCATAACCTAAATAATACACATTATACT 1925

Qy       2211 CCTTAAGTAATTT-CTCACCATCCGCTGACTTCTTGCCCCCTCATCCTTCTGAGGCTCCA 2269
              | |||| |||||| |||| ||||  ||  ||||   || || || |||||  |   || |
Db       1924 CGTTAAATAATTTCCTCATCATCTACTTCCTTCCCACCTCCCCACCCTTCCAAATTTCTA 1865

Qy       2270 TTGTCCATCATTCCACACTCTACATCTATGTGTACACATTATTTAGCTCCTACTTATAAG 2329
              || |  || |||||||  ||| ||| || | ||||||||||||| |||| ||||| ||
Db       1864 TTATTTATAATTCCACATACTATATCCATATATACACATTATTTAACTCCCACTTACAAA 1805

Qy       2330 TGATAACATGCAATATTTGTCTTTCTGTGTCTGTCTTGTTTTACTTATGATAATGGCCCC 2389
              | | || || |||||||| ||||||| | |||   || ||| ||| |   |||  ||
Db       1804 TAAAAAAATTCAATATTTATCTTTCTATTTCTAAATTATTTCACTCAAATTAATAACCTT 1745

Qy       2390 CAGTTCTATCTAGGCTGCTGCAAAAGGCATGATTTCATTCTTTTTTATGGCTATGTTCTT 2449
              || ||  ||  |   | || |||||  ||| ||||||||| |||||||  |    | |
Db       1744 CAATTTCATTCATATTACTACAAAAAACATAATTTCATTATTTTTTATAACCGAATAATA 1685

Qy       2450 TCCCAATTTAGATA 2463
              | ||| | || |||
Db       1684 TTCCATTATATATA 1671


RESULT 78
ADC86642/c
ID   ADC86642 standard; DNA; 349881 BP.
XX
AC   ADC86642;
XX
DT   01-JAN-2004  (first entry)
XX
DE   Human GPCR gene SEQ ID NO:1095.
XX
KW   ds; gene; human; GPCR;
KW   guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1270724-A2.
XX
PD   02-JAN-2003.
XX
PF   18-JUN-2002; 2002EP-00013517.
XX
PR   18-JUN-2001; 2001JP-00246789.
XX
PA   (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA   (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI   Suwa M,  Asai K,  Akiyama Y,  Aburatani H;
XX
DR   WPI; 2003-315783/31.
DR   P-PSDB; ADC86643.
XX
PT   New polynucleotide, useful for preparing a composition for treating a
PT   patient in need of increased or suppressed activity or expression of the
PT   guanosine triphosphate-binding protein coupled receptor.
XX
PS   Claim 1; SEQ ID NO 1095; 28pp; English.
```

```
XX
CC   The invention relates to a novel polynucleotide encoding a guanosine
CC   triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC   the invention may have a use in gene therapy. The polynucleotide and
CC   polypeptide are useful for preparing a composition for treating a patient
CC   in need of increased or suppressed activity or expression of the
CC   guanosine triphosphate-binding protein coupled receptor. The
CC   polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC   invention.
XX
SQ   Sequence 349881 BP; 96804 A; 74573 C; 73713 G; 103046 T; 0 U; 1745 Other;

  Query Match             2.3%;  Score 141.2;  DB 10;  Length 349881;
  Best Local Similarity   69.9%;  Pred. No. 1.2e-22;
  Matches  207;  Conservative    0;  Mismatches   83;  Indels    6;  Gaps    1;

Qy        545 AAAGGATATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAGCAAGGCCTGTCTATTCA 604
              | ||  | ||||||| |||| |  | || | |||| ||| | |||||||||||||| | ||
Db     106652 ACAGAGTGTGATCTAATGGTGACAGGCTGAAGGGGGAACCCTGCAAGGCCTGTCTGTCCA 106593

Qy        605 GATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCCTCCTGAGTATGGGGCAG 664
              ||||| |||||  ||||||   |      |    | | |      |     |||    |||
Db     106592 GATTCATCTTGGCCTCTCTGGGTAGCATTGCGTCCTCC------CAGATATATATTCCAG 106539

Qy        665 GACTCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAGTGGCCTTTTTAGATTTTA 724
              ||| | |||  |||||| |||||| |||||||||||||  ||| ||||| ||| || |
Db     106538 GACCCCTCTGGAATGAGGATCTTCACGGGAGAAGGGAGAGGGTGACCTTTCTAGGTTCTG 106479

Qy        725 TGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTATGACCCATCTTGGGGAAGAGGAATT 784
              |||||||||| |||| ||| |  ||| ||||||||||||||| ||||||||||||||||
Db     106478 TGGCTTGCTTTGGGGGAGAAGGATTCCAGTTTCTATGACCCACCTTGGGGAAGAGGAATT 106419

Qy        785 CTGGTTTCTGTGACTTGCTTTCATGAAGAAAGAGGAGTAAGAGGCAGGAGGGCAGG 840
              |||||||||  ||| ||||||     |||  |  | |||  |||  |   || |||
Db     106418 CTGGTTTCTACGACCTGCTTTGGGAGAGATGGGAGGGTAGAAGGAGGTCAGGGAGG 106363


RESULT 79
AAA28150
ID   AAA28150 standard; DNA; 41684 BP.
XX
AC   AAA28150;
XX
DT   29-JAN-2001  (first entry)
XX
DE   Human purH gene genomic DNA sequence.
XX
KW   Human; purH; biallelic marker; single nucleotide polymorphism; SNP;
KW   diagnosis; prevention; treatment; prostate cancer; tumour;
KW   chromosome 1q34-q35; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   5'UTR           1. .2000
FT                   /*tag= a
FT   exon            2001. .2096
FT                   /*tag= b
FT                   /number= 1
FT   intron          2097. .2432
FT                   /*tag= c
FT                   /number= 1
FT   exon            2433. .2559
FT                   /*tag= d
FT                   /number= 2
FT   intron          2560. .8091
FT                   /*tag= e
FT                   /number= 2
FT   primer_bind     6137. .6157
FT                   /*tag= az
FT   primer_bind     6472. .6490
FT                   /*tag= cd
FT   misc_binding    6479. .6503
FT                   /*tag= dp
FT                   /bound_moiety= "Probe 99-32284-107"
FT   variation       replace(6491,T)
FT                   /*tag= ag
FT   primer_bind     complement(6492. .6510)
FT                   /*tag= ce
FT   primer_bind     complement(6577. .6597)
FT                   /*tag= ba
FT   exon            8092. .8168
FT                   /*tag= f
FT                   /number= 3
FT   intron          8169. .9599
FT                   /*tag= g
FT                   /number= 3
FT   exon            9600. .9666
FT                   /*tag= h
FT                   /number= 4
FT   intron          9667. .15177
FT                   /*tag= i
FT                   /number= 4
FT   primer_bind     14864. .14882
FT                   /*tag= bb
FT   exon            15178. .15266
FT                   /*tag= j
FT                   /number= 5
FT   primer_bind     15215. .15233
FT                   /*tag= cf
FT   misc_binding    15222. .15246
FT                   /*tag= dq
FT                   /bound_moiety= "Probe 99-5602-372"
FT   variation       replace(15234,G)
FT                   /*tag= ah
FT   primer_bind     complement(15235. .15253)
FT                   /*tag= cg
FT   intron          15267. .15923
FT                   /*tag= k
FT                   /number= 5
FT   primer_bind     complement(15292. .15312)
FT                   /*tag= bc
FT   primer_bind     15837. .15855
FT                   /*tag= bd
FT   primer_bind     15849. .15867
FT                   /*tag= ch
FT   misc_binding    15856. .15880
FT                   /*tag= dr
FT                   /bound_moiety= "Probe 5-290-32"
FT   variation       replace(15868,T)
FT                   /*tag= ai
FT   primer_bind     complement(15869. .15887)
FT                   /*tag= ci
FT   exon            15924. .16075
FT                   /*tag= l
FT                   /number= 6
FT   intron          16076. .16758
FT                   /*tag= m
FT                   /number= 6
FT   primer_bind     complement(16429. .16266)
FT                   /*tag= be
FT   primer_bind     16599. .16617
FT                   /*tag= bf
FT   primer_bind     16710. .16728
FT                   /*tag= cj
FT   misc_binding    16717. .16741
FT                   /*tag= ds
FT                   /bound_moiety= "Probe 99-22573-321"
FT   variation       replace(16729,T)
FT                   /*tag= aj
FT   primer_bind     complement(16730. .16748)
FT                   /*tag= ck
FT   exon            16759. .16915
FT                   /*tag= n
FT                   /number= 7
FT   intron          16916. .22308
```

```
FT                   /*tag=  o
FT                   /number= 7
FT   primer_bind     complement(17030. .17049)
FT                   /*tag=  bg
FT   primer_bind     18131. .18150
FT                   /*tag=  bh
FT   primer_bind     18292. .18310
FT                   /*tag=  cl
FT   misc_binding    18299. .18323
FT                   /*tag=  dt
FT                   /bound_moiety= "Probe 99-22586-300"
FT   variation       replace(18311,C)
FT                   /*tag=  ak
FT   primer_bind     complement(18312. .18330)
FT                   /*tag=  cm
FT   primer_bind     18553. .18571
FT                   /*tag=  cn
FT   misc_binding    18560. .18584
FT                   /*tag=  du
FT                   /bound_moiety= "Probe 99-22586-39"
FT   variation       replace(18572,T)
FT                   /*tag=  al
FT   primer_bind     complement(18573. .18591)
FT                   /*tag=  co
FT   primer_bind     complement(18592. .18610)
FT                   /*tag=  bi
FT   exon            22309. .22434
FT                   /*tag=  p
FT                   /number= 8
FT   intron          22435. .23276
FT                   /*tag=  q
FT                   /number= 8
FT   primer_bind     22710. .22727
FT                   /*tag=  bj
FT   primer_bind     22887. .22905
FT                   /*tag=  cp
FT   misc_binding    22894. .22918
FT                   /*tag=  dv
FT                   /bound_moiety= "Probe 99-5596-197"
FT   variation       replace(22906,A)
FT                   /*tag=  am
FT   primer_bind     complement(22907. .22925)
FT                   /*tag=  cq
FT   primer_bind     23100. .23118
FT                   /*tag=  bk
FT   primer_bind     complement(23130. .23149)
FT                   /*tag=  bl
FT   primer_bind     23156. .23174
FT                   /*tag=  cr
FT   misc_binding    23163. .23187
FT                   /*tag=  dw
FT                   /bound_moiety= "Probe 5-293-76"
FT   variation       replace(23175,T)
FT                   /*tag=  an
FT   primer_bind     complement(23176. .23194)
FT                   /*tag=  cs
FT   primer_bind     23234. .23252
FT                   /*tag=  ct
FT   misc_binding    23241. .23265
FT                   /*tag=  dx
FT                   /bound_moiety= "Probe 5-293-155"
FT   variation       replace(23253,G)
FT                   /*tag=  ao
FT   exon            23277. .23384
FT                   /*tag=  r
FT                   /number= 9
FT   intron          23385. .24840
FT                   /*tag=  s
FT                   /number= 9
FT   primer_bind     complement(23512. .23530)
FT                   /*tag=  bm
FT   exon            24841. .24926
FT                   /*tag=  t
```

```
FT                   /number= 10
FT   intron          24927. .25956
FT                   /*tag=  u
FT                   /number= 10
FT   primer_bind     25822. .25840
FT                   /*tag=  bn
FT   exon            25957. .26046
FT                   /*tag=  v
FT                   /number= 11
FT   intron          26047. .28699
FT                   /*tag=  w
FT                   /number= 11
FT   primer_bind     26087. .26105
FT                   /*tag=  cv
FT   misc_binding    26094. .26118
FT                   /*tag=  dy
FT                   /bound_moiety= "Probe 5-294-285"
FT   variation       replace(26106,C)
FT                   /*tag=  ap
FT   primer_bind     complement(26107. .26125)
```

```
  Query Match             2.3%;  Score 141;  DB 3;  Length 41684;
  Best Local Similarity   63.7%;  Pred. No. 5.2e-23;
  Matches  375;  Conservative    0;  Mismatches  175;  Indels   39;  Gaps    9;

Qy       302 CAAAGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCATCATAAGGGTCGCA 361
             | ||| |   |||||||| ||| ||||||| |  |||    |||||| |||  ||| |
Db      4727 CCAAGATTTTGTAGGGGTGGAAAGGTGTGATCC--TTTGCTCTCCATCGTAAACGTCACG 4784

Qy       362 ACCAAAACTCCTATAGTAAAAGACAGGT--TAATAAGAGCAAAACCTAACAAATTTATTT 419
             |||| | |  ||||  | |||||||||  ||||||||| ||| | ||||||||||||||
Db      4785 GCCAATATTTTTATAAGAGAAGACAGGTTATAATAAGAGAAAAGCATAACAAATTTATTT 4844

Qy       420 AATCAAAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAAGACCCAGGGGAA 479
             || |||||||||||||||||| ||| |||||| |||||||||||||||||  |||  ||
Db      4845 AA-CAAAGTTTTACATGACATGAGAGCCTTCAG-AATGAAGACCCAAAGACAGAGGAAAA 4902

Qy       480 ACTGTCTGTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTAGCCATGTAGATT 539
             ||  ||   ||||| || | |||||| |  ||||||||| || | |  |  |    ||||
Db      4903 ACCATC-CATTTTTATGTTTAGGTTCAACAAAGAATGGACAGAAAGGTGGAAGTATGATT 4961

Qy       540 AGACAAAAGGATATGATCTAGTGGTAAAGGACTCAG--GGGGAAACACAGCAAGGCCTGT 597
             ||||  | |          || ||   |||| ||  || ||||  ||| || ||||||
Db      4962 GGACAGCAAGGATATGGTGTATGCTAGTAGACTGAGGTGGAGAAAGCCAGGAAAGCCTGT 5021

Qy       598 CTATTCAGATTCTTCTTGATCTCTCTCTCTCTATGT---------ATAGCATTCTTTCCT 648
             || | ||||||||||||||   |||||      | | |           ||    |||
Db      5022 CTGTCCAGATTCTTCTTGGCTTCTCTAAAATTCTTTCTCCACCCCCTAAGGATCTCCTGA 5081

Qy       649 CCTGAGTATGGGGCAGGACTCTTCTTCAATGAGGGTCTTCAAGGGAGAAGGGAGAAAG-- 706
             |||    |||||  ||||          | |     | || ||     ||  ||||||
Db      5082 CCTACTAATGGGCAAGGAGATGAGAGGATTTCTTTACGTCCAGCTCCTAGACAGAAAGCC 5141

Qy       707 ---------------TGGCCTTTTTAGATTTTATGGCTTGCTTCGGGGAAGAGGAGTTCT 751
                           |     |||| || ||||| || |||| |||||| | |||||
Db      5142 AGTGGAAAGTTAGAATCATAAGTTTAAATCTTATGACTGGCTTTGGGGAAAAAGAGTTTT 5201

Qy       752 AGTTTCTATGACCCATCTTGGGGAAGAGGAATTCTGGTTTCTGTGACTTGCTTTCATGAA 811
             ||||||| | | |   | |||||||||| ||||||  ||||||||||||        |||
Db      5202 AGTTTCTGTAAACTGCCCTGGGGAAGAGAAATTCTCATTTCTGTGACTTCAGG----GAA 5257

Qy       812 GAAAGAGGAGTAAGAGGCAGGAGGGCAGGAGATGGTCAGAAAGAGACTT 860
             ||| || | || ||||||| |||||||||||  |  | | | | | ||
Db      5258 GAATGAAGGGTGAGAGGCAAGAGGGCAGGAGAAAGCTATATATATATTT 5306
```

```
RESULT 80
ADD50650
ID   ADD50650 standard; DNA; 175590 BP.
XX
AC   ADD50650;
XX
DT   15-JAN-2004  (first entry)
```

XX
DE BAC sequence #1 containing hCHT DNA.
XX
KW Human; high-affinity choline transporter; hCHT; chromosome 2q12;
KW cholinergic function; Parkinson's disease; Huntington's disease;
KW Alzheimer's disease; schizophrenia; dysautonomia; myasthenia gravis;
KW brain; cholinergic signalling; antiparkinsonian; anticonvulsant;
KW nootropic; neuroprotective; neuroleptic; bacterial artificial chromosome;
KW BAC; ds.
XX
OS Homo sapiens.
XX
PN US2003114399-A1.
XX
PD 19-JUN-2003.
XX
PF 23-JUL-2001; 2001US-00911077.
XX
PR 23-JUL-2001; 2001US-00911077.
XX
PA (BLAK/) BLAKELY R D.
PA (APPA/) APPARSUNDARAM S.
PA (FERG/) FERGUSON S.
XX
PI Blakely RD, Apparsundaram S, Ferguson S;
XX
DR WPI; 2003-810914/76.
XX
PT Novel isolated polynucleotide encoding human or mouse high affinity
PT choline transporter polypeptide, useful in gene therapy to increase
PT cholinergic function in a cell of a patient suffering from Alzheimer's
PT disease.
XX
PS Example 1; SEQ ID NO 13; 74pp; English.
XX
CC The present invention relates to the isolation of polynucleotide
CC sequences encoding human and mouse high-affinity choline transporter
CC (hCHT and mCHT respectively), and the proteins they encode. The gene
CC encoding hCHT is located on chromosome 2q12. The polynucleotide sequence
CC encoding hCHT, is useful for expressing hCHT recombinantly. The hCHT
CC polynucleotide sequence when delivered to a cell, increases cholinergic
CC function in the cell that is in a patient having Parkinson's disease,
CC Huntington's disease, Alzheimer's disease, schizophrenia, dysautonomia or
CC myasthenia gravis. The hCHT antibody is useful for controlling
CC transporter CHT proteins to the brain, and for treating the above
CC mentioned diseases. The antibody is also useful for diagnosing the above
CC mentioned disorders and to detect the influence of cholinergic
CC signalling. The present sequence represents a bacterial artificial
CC chromosome (BAC) sequence containing hCHT DNA. Note: The sequence data
CC for this patent was obtained in electronic format directly from the USPTO
CC web site at seqdata.uspto.gov.
XX
SQ Sequence 175590 BP; 55029 A; 33099 C; 32863 G; 53896 T; 0 U; 703 Other;

Query Match 2.3%; Score 140.4; DB 10; Length 175590;
Best Local Similarity 71.6%; Pred. No. 1.4e-22;
Matches 212; Conservative 0; Mismatches 81; Indels 3; Gaps 2;

```
Qy       2177 GGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCT 2236
              || ||| |||||||  ||| | || | |||||||| |||||| |||||  |  ||||
Db      31651 GGATATCCATCACCCAAATAACGTACTTTGTACCCATTAAGTGATTTCCAATTATCCATC 31710

Qy       2237 GACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTCTACATCT 2296
                ||||   | || |||  |||||||| |||||||||| |||||| | | ||||||   |
Db      31711 CCCTTCCA-CTCCTTCACTCTTCTGAGTCTCCATTGTCTATCATTTCCCTCTCTAC--TT 31767

Qy       2297 ATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGTCTTTCTG 2356
              |  | ||||| || | | || |  | |||||||||| |||||||||||||| | ||||||
Db      31768 ACATATACACCTTTTCTGGCACTCATTTATAAGTGAGAACATGCAATATTTATATTTCTG 31827

Qy       2357 TGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAAGG 2416
              |||||| ||||||| | |||||| |||| || || |||| ||| | | | |||||||||
Db      31828 TGTCTGGCTTGTTTCATTTATGACAATGCCCTCCTGTTCCATCCAAGTTCCTGCAAAAGA 31887

Qy       2417 CATGATTTCATTCTTTTTTATGGCTATGTTCTTTCCCAATTTAGATAAAGAACACT 2472
              |||||||||||  |||| |      ||  |  ||     |||| ||| |  ||| |
Db      31888 CATGATTTCATGTTTTTATGGCTGAATAGTACTTTATTCTTTATATATACCACATT 31943
```

RESULT 81
ADS31318/c
ID ADS31318 standard; DNA; 904 BP.
XX
AC ADS31318;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human genome high complexity repeat found in the HIRA gene #351.
XX
KW Human; ds;
KW histone cell cycle regulation defective, S. cerevisiae homologue A; HIRA;
KW high complexity repeat; in situ hybridisation; Southern blot;
KW chromosome breakpoint; inherited genetic disease; neoplastic disorder;
KW chromosome 22; DiGeorge syndrome; Velo-Cardio-facial syndrome.
XX
OS Homo sapiens.
XX
PN US2003224356-A1.
XX
PD 04-DEC-2003.
XX
PF 14-MAY-2001; 2001US-00854867.
XX
PR 16-MAY-2000; 2000US-00573080.
XX
PA (KNOL/) KNOLL J H M.
PA (ROGA/) ROGAN P K.
XX
PI Knoll JHM, Rogan PK;
XX
DR WPI; 2002-062378/08.
XX
PT Single copy genomic hybridization probes for detecting specific nucleic
PT acid sequences in sample by in situ hybridization useful for detection of
PT acquired or inherited genetic diseases.
XX
PS Example 1; SEQ ID NO 351; 30pp; English.
XX
CC The invention relates to a nucleic acid hybridisation probe comprising a
CC labelled, single copy nucleic acids of at least 50 nucleotides, which
CC will hybridise to a deduced single copy sequence interval in target
CC nucleic acid (TNA) of known sequence. The single copy sequence is deduced
CC by comparing the target nucleic acid (e.g. a disease causing gene) with a
CC collection of high and low complexity repeat sequences as found in the
CC genome of the organism from containing the target nucleic acid. The probe
CC is generated by PCR on the target sequence. The probe is essentially free
CC of blocking nucleic acid sequences which will hybridise to repeat
CC sequences within the genome of which the TNA is a part, and is labelled
CC with a label selected from fluorochrome-responsive labels, fluorochromes,
CC calorimetric chemical, conjugated proteins, antibodies, antigens and
CC their mixtures. The probe is useful in a hybridisation method, where the
CC hybridisation method is from in situ hybridisation, Southern blot, and
CC other methods in which nucleic acid is immobilised, where the method
CC further comprises selecting a single copy nucleic acid which will
CC hybridise to a duplicon or triplicon sequence domain. The probe is useful
CC for determining the existence of previously unknown repeat sequence
CC families in a genome. The method comprises reacting a labelled probe with
CC the genome, causing the probe to hybridise and ascertaining if the probe
CC hybridises to the genome at more than three preferably ten different
CC locations as a determination of new repeat sequence family, where the
CC determining step comprises selecting the single copy sequence from a
CC duplicon or triplicon sequence domain. The probe is useful for
CC determining a chromosome breakpoint and is useful in the fields for
CC cytogenetics and molecular genetics for determining the presence of
CC specific nucleic acid sequences in a sample of eukaryotic origin, e.g.
CC the probes may be used to analyse specific chromosomal locations by in

```
CC   situ hybridisation as a detection of acquired or inherited genetic
CC   diseases especially for detection of genetic or neoplastic disorders.
CC   Unlike prior art techniques, the probe permits more precise chromosomal
CC   breakpoint determinations by in situ hybridisation. The genomic sequence
CC   comprising the human HIRA gene (histone cell cycle regulation defective,
CC   S. cerevisiae, homologue A) was analysed for single copy sequence
CC   intervals for use as probes of the invention. HIRA is located on
CC   chromosome 22 as a duplicate, deletions of 1 copy lead to DiGeorge and
CC   Velo-Cardio-facial syndromes. The present sequence is a high complexity
CC   repeat found within the human genome used to analyse the HIRA gene for
CC   repeat regions. Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic format
CC   directly from USPTO at seqdata.uspto.gov/sequence.html?DocID=20030224356.
XX
SQ   Sequence 904 BP; 356 A; 171 C; 173 G; 191 T; 0 U; 13 Other;

  Query Match             2.3%;  Score 140.2;  DB 7;  Length 904;
  Best Local Similarity   65.1%;  Pred. No. 1.4e-23;
  Matches  205;  Conservative    1;  Mismatches  109;  Indels    0;  Gaps    0;

Qy       2182 TTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCTGACTT 2241
              |||||||||    | ||  |||| ||||||  || ||| ||| |       | |   |||
Db        814 TTCATCACCCAGGTAATAAGCATAGTACCCGATAGGTAGTTTTTTGATCCTCTCCCTCTT 755

Qy       2242 CTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTCTACATCTATGTG 2301
              |   |||  |     |       |||| ||| ||||  |  ||||   || ||  || |||||
Db        754 CCCACCCTCCACCCTCAAGTAGGCCCCAGTGTCTGTTGTTCCCTTCTTTATGTCCATGTG 695

Qy       2302 TACACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGTCTTTCTGTGTCT 2361
              ||| || |:||||||||| |||||||||||| |||||||  ||||||  |||||||  ||
Db        694 TACTCAATRTTTAGCTCCCACTTATAAGTGAGAACATGCGGTATTTGGTTTTCTGTTCCT 635

Qy       2362 GTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAAGGCATGA 2421
              |  ||  ||| || | |||||||||| |||| || ||| | | | ||||||| | |||||
Db        634 GCRTTAGTTTGCTAAGGATAATGGCCTCCAGCTCCATCCATGTTCCTGCAAAGGACATGA 575

Qy       2422 TTTCATTCTTTTTTATGGCTATGTTCTTTCCCAATTTAGATAAAGAACACTCGCACTTGC 2481
              | ||||||||||||||||||   |  | | |||   |  |||   | |||     |||
Db        574 TCTCATTCTTTTTTATGGCTGCATAGTATTCCATGGTGTATATGTACCACATTTTCTTTA 515

Qy       2482 TCTTACTTCTATTTG 2496
              ||     | |  |||
Db        514 TCCAGTCTATCATTG 500


RESULT 82
ADO34927_0
WP Sequence split into 4 fragments  LOCUS ADO34927 Accession Ado34927
WP     Fragment Name      Begin     End
WP     ADO34927_0             1     110000
WP     ADO34927_1        100001     210000
WP     ADO34927_2        200001     310000
WP     ADO34927_3        300001     382256
ID   ADO34927 standard; DNA; 382256 BP.
XX
AC   ADO34927;
XX
DT   12-AUG-2004  (first entry)
XX
DE   Human voltage-gated potassium channel-interacting protein 1 DNA sequence.
XX
KW   ds; diabetes; polymorphism;
KW   voltage-gated potassium channel-interacting protein 1; KChIP1; diagnosis;
KW   chromosome 5.
XX
OS   Homo sapiens.
XX
PN   WO2004041193-A2.
XX
PD   21-MAY-2004.
XX
PF   31-OCT-2003; 2003WO-US034681.
XX
PR   01-NOV-2002; 2002US-0423545P.
PR   25-FEB-2003; 2003US-0449945P.
PR   09-JUN-2003; 2003US-0477111P.
XX
PA   (DECO-) DECODE GENETICS EHF.
XX
PI   Reynisdottir I,  Gulcher JR,  Grant SF,  Thorleifsson G;
XX
DR   WPI; 2004-390509/36.
XX
PT   Diagnosing susceptibility of individual to type II diabetes, involves
PT   detecting polymorphism in voltage-gated potassium channel-interacting
PT   protein 1 nucleic acid.
XX
PS   Disclosure; SEQ ID NO 1; 370pp; English.
XX
CC   The invention relates to a method of diagnosing (M1) a susceptibility to
CC   type II diabetes in an individual, by: (a) detecting a polymorphism in a
CC   voltage-gated potassium (Kv) channel-interacting protein 1 (KChIP1)
CC   nucleic acid; (b) detecting an alteration in the expression or
CC   composition of a polypeptide encoded by KChIP1 nucleic acid in a test
CC   sample; (c) determining the presence of a haplotype (H) comprising a
CC   haplotype designated A1-A6 and B1-B5 and haplotype of Icelandic
CC   population designated D2-D6 and Danish population designated D2-D5 as
CC   given in the specification at the 5q35 loci; (d) obtaining a nucleic acid
CC   sample, and analyzing for the presence of (H) at the 5q35 loci comprising
CC   a KChIP1 gene; (e) determining the presence of haplotype (H1) comprising
CC   one or more markers and/or single nucleotide polymorphisms (SNPs) of 142
CC   markers and/or SNPs (PM) as given in the specification in the locus on
CC   chromosome 5q35; or (f) screening for an at-risk haplotype in the KChIP1
CC   nucleic acid that is more frequently present in an individual susceptible
CC   to type II diabetes compared to an individual who is not susceptible to
CC   type II diabetes. (M1) is useful for diagnosing susceptibility to type II
CC   diabetes in an individual. A host cell is useful for producing a
CC   polypeptide encoded by an isolated nucleic acid molecule having
CC   polymorphism. An agent is useful for altering expression of a KChIP1
CC   nucleic acid. The type II diabetes therapeutic agent is useful for the
CC   treatment of a disease or condition associated with KChIP1 in an
CC   individual. This sequence corresponds to the voltage-gated potassium (Kv)
CC   channel-interacting protein 1 (KChIP1) genomic sequence.
XX
SQ   Sequence 382256 BP; 104735 A; 89702 C; 87238 G; 100581 T; 0 U; 0 Other;

  Query Match             2.3%;  Score 140.2;  DB 12;  Length 110000;
  Best Local Similarity   73.9%;  Pred. No. 1.3e-22;
  Matches  178;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy       2177 GGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCT 2236
              |||||| | || ||  ||||| ||  ||||||||| | || |||||||||| ||||| ||
Db      75900 GGGTATCCGTCCCCACAATGACGTATATTGTACCCATCAACTAATTTCTCATCATCCACT 75959

Qy       2237 GACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTCTACATCT 2296
                 |   | ||||||  |||||  || ||||||||||  || || ||||  |||||||||
Db      75960 CCTCCCCCACACCCTCACCCTTCCAAGCCTCCATTGTCTATTATCCCACTCTCTACATCC 76019

Qy       2297 ATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGTCTTTCTG 2356
              |||| ||||  ||||||| |||| | ||||  |||  ||||||  ||||||| ||||||
Db      76020 ATGTATACAGGTTATTTATCTCCCATTTATGGGTGGGAACATGTGATATTTGACTTTCTT 76079

Qy       2357 TGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAAGG 2416
                 ||| ||||||| ||||| |||||||    ||||||  ||    | ||||||||||
Db      76080 CACCTGGCTTGTTTCACTTAAGATAATGAATTCCAGTTTCATTCCAGTTGCTGCAAAAGA 76139

Qy       2417 C 2417
                |
Db      76140 C 76140


RESULT 83
ADQ18925
ID   ADQ18925 standard; DNA; 148497 BP.
XX
AC   ADQ18925;
```

XX
DT 26-AUG-2004 (first entry)
XX
DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 1744.
XX
KW soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
KW ds.
XX
OS Homo sapiens.
XX
PN WO2004048938-A2.
XX
PD 10-JUN-2004.
XX
PF 26-NOV-2003; 2003WO-US038193.
XX
PR 26-NOV-2002; 2002US-0429739P.
XX
PA (PROT-) PROTEIN DESIGN LABS INC.
XX
PI Aziz N, Ginsburg WM, Zlotnik A;
XX
DR WPI; 2004-441208/41.
XX
PT Early detection of soft tissue sarcoma comprises determining expression
PT of a gene in a first soft tissue sample and a normal soft tissue sample
PT and comparing the gene expression, also useful in treating soft tissue
PT sarcoma.
XX
PS Example 2; SEQ ID NO 1744; 210pp; English.
XX
CC The invention relates to a novel method for detecting soft tissue sarcoma
CC which comprises obtaining a first soft tissue sample from an individual
CC and a normal soft tissue sample from the same or different individual,
CC determining the expression of a gene in both samples and comparing the
CC expression of the gene in both soft tissue samples, where a higher level
CC of protein expression in the first soft tissue sample indicates the
CC presence of soft tissue sarcoma. The method of the invention has
CC cytostatic applications and may be useful for detecting soft tissue
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC acid sequences may be useful in diagnostic and screening applications.
CC The current sequence is that of a human soft tissue sarcoma-upregulated
CC DNA of the invention. The current sequence is not shown within the
CC specification per se but was submitted in CD format by the inventor.
XX
SQ Sequence 148497 BP; 40996 A; 32911 C; 32421 G; 42169 T; 0 U; 0 Other;

Query Match 2.3%; Score 140.2; DB 12; Length 148497;
Best Local Similarity 69.5%; Pred. No. 1.5e-22;
Matches 205; Conservative 0; Mismatches 88; Indels 2; Gaps 1;

```
Qy       2177 GGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCT 2236
Db      33930 GTGTAACCATCACCCGAATAATGTAAATGGTATCCAACAGGTACTTTTTCATCCCTCAAC 33989

Qy       2237 GACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTCTACATCT 2296
Db      33990 CCCCTTCCACCCTCCCACCTTGTGGAGTCTCCAGTGTCTATTATTCCACTCTGTATGTCC 34049

Qy       2297 ATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGTCTTTCTG 2356
Db      34050 ATGTGTACACATTGTTTAGCTCCCACTTATAAGTGAGA--ATGTGGTATTTGACTCTCTG 34107

Qy       2357 TGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAAGG 2416
Db      34108 TTTCTGAATTATTTTAGGTAGGATCATGGCCTCCAGTTCCATCCATATTGGTGCAAAAGA 34167

Qy       2417 CATGATTTCATTCTTTTTTATGGCTATGTTCTTTCCCAATTTAGATAAAGAACAC 2471
Db      34168 CATTCTTTCTTTCTTTTTTATGGTTGAGTAGTATTCCATGGTGTATGTATACCAC 34222
```

RESULT 84

ADO47192/c
ID ADO47192 standard; DNA; 200000 BP.
XX
AC ADO47192;
XX
DT 15-JUL-2004 (first entry)
XX
DE DNA sequence of a human immunoglobulin heavy chain variable region.
XX
KW immunoglobulin; B cell; germline; region V; region D; region J;
KW recombination-promoting protein; recombination activating gene 1; RAG-1;
KW RAG-2; ss; terminal deoxynucleotidyltransferase; TdT; V(D)J recombinase.
XX
OS Homo sapiens.
XX
PN WO2004029249-A1.
XX
PD 08-APR-2004.
XX
PF 30-SEP-2003; 2003WO-AU001286.
XX
PR 30-SEP-2002; 2002US-0415024P.
XX
PA (CENT-) CENTENARY INST CANCER MEDICINE & CELL BI.
XX
PI Jolly C;
XX
DR WPI; 2004-316126/29.
XX
PT New vector comprising one or more immunoglobulin regions selected from V,
PT D, and J regions, a 5' flanking region, and a 3' flanking region, useful
PT for reverting cell lines to a pro-B cell-like state or to a germline-like
PT state.
XX
PS Example 1; Page 185-234; 382pp; English.
XX
CC The specification describes a method for the preparation of
CC immunoglobulins. The method uses a vector for reverting cell lines to a
CC pro-B cell-like state or to a germline-like state, by replacing one or
CC more of the immunoglobulin regions V, D, and J of the rearranged
CC immunoglobulin gene with the germ-line or unrearranged V, D or J regions.
CC The vector can comprise a polynucleotide sequence encoding a
CC recombination-promoting protein, selected from recombination activating
CC gene 1 (RAG-1), RAG-2, terminal deoxynucleotidyltransferase (TdT). These
CC proteins collectively constitute a V(D)J recombinase. The method is
CC useful for the preparation of immunoglobulin heavy and/or light chains
CC and their fragments. The present sequence encodes a human immunoglobulin
CC heavy chain variable region, and is used as a template to construct
CC vectors for use in the method of the invention.
XX
SQ Sequence 200000 BP; 58364 A; 41011 C; 41621 G; 59004 T; 0 U; 0 Other;

Query Match 2.3%; Score 139.6; DB 12; Length 200000;
Best Local Similarity 67.7%; Pred. No. 2.4e-22;
Matches 210; Conservative 0; Mismatches 99; Indels 1; Gaps 1;

```
Qy       2156 TCCGCTGACTTCTTGCCCCCTGGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTA 2215
Db     136381 TACACAGTGGTATTGGCTTTAGTGTAACCAACACCCAAGTAACGTACATTGTACCCATTA 136322

Qy       2216 AGTAATTTCTCACCATCCGCTGACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCC 2275
Db     136321 GGTGATTTTTCATCATGCTGCCAACTCCTACCCTCCCATTCTGCTAAGTCTCCAATGTCC 136262

Qy       2276 ATCATTCCACACTCTACATCTATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAA 2335
Db     136261 ATAATCCCTCTCTCCATATCT-TGTGTACACACTGTTTTCCTCCCACTTATAAGTAATAA 136203

Qy       2336 CATGCAATATTTGTCTTTCTGTGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTC 2395
Db     136202 TGTGTGATATGTGGCTTTCTGTTTGTGAGTTAGTTCACTAATTATAATGTCCGCCAGTTC 136143

Qy       2396 TATCTAGGCTGCTGCAAAAGGCATGATTTCATTCTTTTTTATGGCTATGTTCTTTCCCAA 2455
```

```
              ||   |   |||||||||||| ||   |||||||| || || |||||   |  | |   |
Db     136142 TAGGCATCTTGCTGCAAAAGACACAGTTTCATTCCTTATTGTGGCTGACTAGTATTAAAT 136083

Qy       2456 TTTAGATAAA 2465
              | |  ||| |
Db     136082 TGTGCATATA 136073


RESULT 85
ADO34435_3
Continuation (4 of 7) of ADO34435 from base 300001 (Human SLIT-3 genomic sequence. )
WP Sequence split into 7 fragments  LOCUS ADO34435 Accession Ado34435
WP      Fragment Name      Begin     End
WP      ADO34435_0             1    110000
WP      ADO34435_1        100001    210000
WP      ADO34435_2        200001    310000
WP      ADO34435_3        300001    410000
WP      ADO34435_4        400001    510000
WP      ADO34435_5        500001    610000
WP      ADO34435_6        600001    634886

  Query Match             2.3%;  Score 139;  DB 12;  Length 110000;
  Best Local Similarity   74.2%;  Pred. No. 2.5e-22;
  Matches  219;  Conservative    0;  Mismatches   65;  Indels   11;  Gaps    3;

Qy       2177 GGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCT 2236
              || ||| | |||||  | | | || |||||||||| |||||||||||||| |||||
Db     100031 GGATATCCGTCACCCCAGTAACGTACATTGTACCCGTTAAGTAATTTCTCATCATCCAGC 100090

Qy       2237 GACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTCTACATCT 2296
              || |  |  |||||||  |||||||||||||||||||  ||        |||||  ||
Db     100091 CACCTTCTA-CCCCTCACGCTTCTGAGGCTCCATTGTCATTC------TCCTCTATGTCC 100143

Qy       2297 ATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGTCTTTCTG 2356
              |||||||||||| ||||||| | |||| ||||||| ||| ||  | |||||    ||
Db     100144 ATGTGTACACATTGTTTAGCTTCCACTTCTAAGTGAGAACCTGTGAAATTTG----TCCA 100199

Qy       2357 TGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAAGG 2416
              |||||| ||||||| |||||  |||||| || ||||||| ||| | | ||||||||||
Db     100200 TGTCTGACTTGTTTCACTTACAATAATGACCTCCAGTTCCATCCATGTTGCTGCAAAAGA 100259

Qy       2417 CATGATTTCATTCTTTTTTATGGCTATGTTCTTTCCCAATTTAGATAAAGAACAC 2471
              |||||||||||| ||||||||||    |  | | ||| | |  ||| | | |||
Db     100260 CATGATTTCATTCGTTTTTATGGCTGAATAGTATTCCATTGTGTATATATACCAC 100314


RESULT 86
ADO34435_4
Continuation (5 of 7) of ADO34435 from base 400001 (Human SLIT-3 genomic sequence. )
WP Sequence split into 7 fragments  LOCUS ADO34435 Accession Ado34435
WP      Fragment Name      Begin     End
WP      ADO34435_0             1    110000
WP      ADO34435_1        100001    210000
WP      ADO34435_2        200001    310000
WP      ADO34435_3        300001    410000
WP      ADO34435_4        400001    510000
WP      ADO34435_5        500001    610000
WP      ADO34435_6        600001    634886

  Query Match             2.3%;  Score 139;  DB 12;  Length 110000;
  Best Local Similarity   74.2%;  Pred. No. 2.5e-22;
  Matches  219;  Conservative    0;  Mismatches   65;  Indels   11;  Gaps    3;

Qy       2177 GGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCT 2236
              || ||| | |||||  | | | || |||||||||| |||||||||||||| |||||
Db          31 GGATATCCGTCACCCCAGTAACGTACATTGTACCCGTTAAGTAATTTCTCATCATCCAGC 90

Qy       2237 GACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTCTACATCT 2296
              || |  |  |||||||  |||||||||||||||||||  ||        |||||  ||
Db          91 CACCTTCTA-CCCCTCACGCTTCTGAGGCTCCATTGTCATTC------TCCTCTATGTCC 143

Qy       2297 ATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGTCTTTCTG 2356
              |||||||||||| ||||||| | |||| ||||||| ||| ||  | |||||    ||
Db         144 ATGTGTACACATTGTTTAGCTTCCACTTCTAAGTGAGAACCTGTGAAATTTG----TCCA 199

Qy       2357 TGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAAGG 2416
              |||||| ||||||| |||||  |||||| || ||||||| ||| | | ||||||||||
Db         200 TGTCTGACTTGTTTCACTTACAATAATGACCTCCAGTTCCATCCATGTTGCTGCAAAAGA 259

Qy       2417 CATGATTTCATTCTTTTTTATGGCTATGTTCTTTCCCAATTTAGATAAAGAACAC 2471
              |||||||||||| ||||||||||    |  | | ||| | |  ||| | | |||
Db         260 CATGATTTCATTCGTTTTTATGGCTGAATAGTATTCCATTGTGTATATATACCAC 314


RESULT 87
ADR07913
ID   ADR07913 standard; cDNA; 5666 BP.
XX
AC   ADR07913;
XX
DT   04-NOV-2004  (first entry)
XX
DE   Full length human cDNA useful for treating neurological disease Seq 1419.
XX
KW   gene; ss; human; oligo-capping method; diagnostic marker; gene therapy;
KW   osteoporosis; neurological disease; Alzheimer's disease;
KW   Parkinson's disease; dementia; short memory; cancer;
KW   sense or motor function; emotional reaction; fear response; panic;
KW   osteopathic; neuroprotective; nootropic; antiparkinsonian; cytostatic;
KW   tranquiliser.
XX
OS   Homo sapiens.
XX
PN   EP1447413-A2.
XX
PD   18-AUG-2004.
XX
PF   12-FEB-2004; 2004EP-00003145.
XX
PR   14-FEB-2003; 2003JP-00102207.
PR   09-MAY-2003; 2003JP-00131452.
XX
PA   (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI   Isogai T,  Yamamoto J,  Nishikawa T,  Isonó Y,  Sugiyama T,  Otsuki T;
PI   Wakamatsu A,  Ishii S,  Nagai K,  Irie R;
XX
DR   WPI; 2004-583265/57.
DR   P-PSDB; ADR09869.
XX
PT   New 1995 cDNA, useful for treating osteoporosis, neurological diseases,
PT   Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.
XX
PS   Claim 1; SEQ ID NO 1419; 2686pp; English.
XX
CC   This invention relates to novel, isolated full length human cDNA
CC   molecules and the encoded proteins thereof. Specifically, it refers to
CC   cDNA clones obtained by an oligo-capping method, where none of these
CC   clones are identical to any known human mRNAs. The present invention
CC   describes an immunoassay to identify agonists and antagonists, as well as
CC   antibodies, antisense molecules and siRNAs that can all be used to bind
CC   to and modulate expression of the cDNA molecules. As such, these
CC   molecules are useful for diagnostic markers or therapeutic targets for
CC   the various diseases or morbid states. In particular, they are useful in
CC   gene therapy for treating osteoporosis, neurological disease, Alzheimer's
CC   disease, Parkinson's disease, dementia, short memory and various cancers,
CC   as well as for maintaining equilibrium of sense or motor function, and
CC   for treating emotional reaction, fear response and panic. Accordingly,
CC   they exhibit osteopathic, neuroprotective, nootropic, antiparkinsonian,
CC   cytostatic and tranquiliser activities. This polynucleotide is a full
CC   length human cDNA sequence of the invention. NOTE: This sequence is not
CC   given in the sequence listing of the specification but can be obtained on
CC   CD-ROM from the European Patent Office, Vienna Sub-office.
XX
SQ   Sequence 5666 BP; 1684 A; 1261 C; 1249 G; 1472 T; 0 U; 0 Other;
```

```
  Query Match             2.3%;  Score 137.4;  DB 13;  Length 5666;
  Best Local Similarity   65.1%;  Pred. No. 1.5e-22;
  Matches  218;  Conservative    0;  Mismatches  116;  Indels    1;  Gaps    1;

Qy       2177 GGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCT 2236
             | ||| |||| | |  |||  | | |||||| ||| ||| ||||||||||||| |   |
Db       4199 GTGTAGTCATGATCCAAATAGTATACATTGTGCCCATTAGGTAATTTCTCACCCT-TACC 4257

Qy       2237 GACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTCTACATCT 2296
              || ||   | | | ||  ||||  || ||||| |||| || ||| | |  || |  ||
Db       4258 CACCTCCCACTCTCGCACTCTTCCAAGTCTCCAGTGTCTATTATTTCGCGGTCAATGTCC 4317

Qy       2297 ATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGTCTTTCTG 2356
              | ||||||| ||||||||||||  || | ||| ||| |||||||||| ||| |||||
Db       4318 AGGTGTACATATTATTTAGCTCTCACGTGTAACTGAGAACATGCAATACTTGACTTTCCA 4377

Qy       2357 TGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAAGG 2416
              | ||||  || ||| ||||| |||||||| | || |||| ||| | | ||||||||||
Db       4378 TTTCTGAGTTATTTCACTTAAGATAATGGTCTCCGGTTCCATCCATGTTGCTGCAAAAGA 4437

Qy       2417 CATGATTTCATTCTTTTTTATGGCTATGTTCTTTCCCAATTTAGATAAAGAACACTCGCA 2476
              |||||||| ||  |||| |     | |  | || |    | || ||  |    ||   |
Db       4438 CATGATTTTATCTTTTTATGACTGTTTAGTATTACATTGTGTATATGTATCCTACATTCT 4497

Qy       2477 CTTGCTCTTACTTCTATTTGGAATACTAATTCCTA 2511
              |||  |    |  | |||      |   ||||||
Db       4498 CTTTATTCAGATGATCTTTTATTGCCAGCTTCCTA 4532


RESULT 88
ADA02900
ID   ADA02900 standard; DNA; 31236 BP.
XX
AC   ADA02900;
XX
DT   06-NOV-2003  (first entry)
XX
DE   Human PTP4A2 carcinoma associated gene, SEQ ID NO:1418.
XX
KW   Human; carcinoma associated; oncogene; carcinoma; cancer; breast;
KW   prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO2003057146-A2.
XX
PD   17-JUL-2003.
XX
PF   26-DEC-2002; 2002WO-US041414.
XX
PR   26-DEC-2001; 2001US-00035832.
XX
PA   (SAGR-) SAGRES DISCOVERY.
XX
PI   Morris DW;
XX
DR   WPI; 2003-587068/55.
XX
PT   New recombinant nucleic acid encoding carcinoma associated protein,
PT   useful for preparing compositions for treating carcinomas.
XX
PS   Claim 1; SEQ ID NO 1418; 245pp; English.
XX
CC   The invention relates to recombinant carcinoma associated (CA) nucleic
CC   acid sequences from mouse and human (ADA01482-ADA03094), and to
CC   recombinant carcinoma associated proteins (CAP) encoded by them. The
CC   invention also encompasses expression vectors and host cells comprising a
CC   CA nucleic acid, a polypeptide (especially an antibody) that specifically
CC   binds to the protein, and a biochip comprising CA nucleic acid or
CC   fragments thereof. The sequences of the invention were identified using
CC   oncogenic retroviruses, which insert into the genome of the host organism
CC   at random. Many of these do not carry transduced host oncogenes or
CC   pathogenic trans-acting viral genes, meaning that cancer incidence is a
CC   direct consequence of the effects of proviral integration into host
CC   protooncogenes. The CA nucleic acid sequences can be used to diagnose
CC   carcinoma (especially breast cancer, prostate cancer, lymphoma or
CC   leukaemia) or a propensity to carcinoma by determination of the sequence
CC   of a CA gene, or by determination of CA gene expression in particular
CC   tissues. CA nucleic acids, proteins and antibodies are also useful as
CC   therapeutic agents and in screening and evaluating drug candidates. The
CC   present sequence represents a specifically claimed human CA nucleic acid
CC   sequence of the invention. Note: The complete sequence data for this
CC   patent did not form part of the printed specification, but was obtained
CC   in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 31236 BP; 8595 A; 6081 C; 6658 G; 9902 T; 0 U; 0 Other;

  Query Match             2.3%;  Score 137;  DB 9;  Length 31236;
  Best Local Similarity   69.8%;  Pred. No. 4.3e-22;
  Matches  185;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy       2201 GCATTGTACCCCTTAAGTAATTTCTCACCATCCGCTGACTTCTTGCCCCCTCATCCTTCT 2260
              |||| |||||   |||||| ||| |||    || ||  | |||  ||| |    |
Db       2082 GCATAGTACCTGATAAGTAGTTTTTCAAGCCCCACTCTCCTCTCACCCTCCACCCTCAAG 2141

Qy       2261 GAGGCTCCATTGTCCATCATTCCACACTCTACATCTATGTGTACACATTATTTAGCTCCT 2320
                | | ||| ||||  |  |||| | || |   |||||||||| || | |||||||||
Db       2142 TGGTCCCCAGTGTCTGTTGTTCCCCTCTTTGAGTCTATGTGTACTCAGTGTTTAGCTCCC 2201

Qy       2321 ACTTATAAGTGATAACATGCAATATTTGTCTTTCTGTGTCTGTCTTGTTTTACTTATGAT 2380
              |||||||||||| |||||||  ||||||  |||||||  |||  ||  ||  |||| |||
Db       2202 ACTTATAAGTGAGAACATGCGGTATTTGGTTTTCTGTTCCTGCGTTAATTCGCTTAGGAT 2261

Qy       2381 AATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAAGGCATGATTTCATTCTTTTTTATGGC 2440
              ||||||| |||| |  ||| | | |||||||||| | ||||||||||| ||||||||||
Db       2262 AATGGCCTCCAGCTGCATCCATGTTGCTGCAAAGGACATGATTTCATTATTTTTTATGGC 2321

Qy       2441 TATGTTCTTTCCCAATTTAGATAAA 2465
              | ||   | || |  || ||  |||
Db       2322 TTTGAGTTGTCACTCTTGAGTGAAA 2346


RESULT 89
ADB72638
ID   ADB72638 standard; DNA; 31236 BP.
XX
AC   ADB72638;
XX
DT   04-DEC-2003  (first entry)
XX
DE   Human PTP4A2 gene.
XX
KW   human; ds; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;
KW   cancer; neoplasm; adenocarcinoma; sarcoma; gene.
XX
OS   Homo sapiens.
XX
PN   WO2003008583-A2.
XX
PD   30-JAN-2003.
XX
PF   26-DEC-2001; 2001WO-US051291.
XX
PR   02-MAR-2001; 2001US-00798586.
PR   23-OCT-2001; 2001US-00004113.
PR   08-NOV-2001; 2001US-00052482.
PR   30-NOV-2001; 2001US-00997722.
PR   20-DEC-2001; 2001US-00034650.
XX
PA   (SAGR-) SAGRES DISCOVERY.
XX
PI   Morris DW,  Engelhard EK;
XX
DR   WPI; 2003-239337/23.
```

XX
PT New recombinant nucleic acid, useful for treating carcinomas, lymphomas,
PT cancers, neoplasm, adenocarcinoma, or sarcomas.
XX
PS Claim 1; SEQ ID NO 466; 2304pp; English.
XX
CC The invention relates to a novel recombinant nucleic acid comprising a
CC nucleotide sequence selected from any of the 660 sequences fully defined
CC in the specification. A polynucleotide of the invention has cytostatic
CC activity, and may have a use in gene therapy, or in a vaccine. The
CC recombinant nucleic acids and polypeptides are useful for treating
CC carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and
CC sarcomas. The present sequence represents a human gene of the invention.
XX
SQ Sequence 31236 BP; 8595 A; 6081 C; 6658 G; 9902 T; 0 U; 0 Other;

Query Match 2.3%; Score 137; DB 10; Length 31236;
Best Local Similarity 69.8%; Pred. No. 4.3e-22;
Matches 185; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

Qy 2201 GCATTGTACCCCTTAAGTAATTTCTCACCATCCGCTGACTTCTTGCCCCCTCATCCTTCT 2260
Db 2082 GCATAGTACCTGATAAGTAGTTTTTCAAGCCCCACTCTCCTCTCACCCTCCACCCTCAAG 2141

Qy 2261 GAGGCTCCATTGTCCATCATTCCACACTCTACATCTATGTGTACACATTATTTAGCTCCT 2320
Db 2142 TGGTCCCCAGTGTCTGTTGTTCCCCTCTTTGAGTCTATGTGTACTCAGTGTTTAGCTCCC 2201

Qy 2321 ACTTATAAGTGATAACATGCAATATTTGTCTTTCTGTGTCTGTCTTGTTTTACTTATGAT 2380
Db 2202 ACTTATAAGTGAGAACATGCGGTATTTGGTTTTCTGTTCCTGCGTTAATTCGCTTAGGAT 2261

Qy 2381 AATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAAGGCATGATTTCATTCTTTTTTATGGC 2440
Db 2262 AATGGCCTCCAGCTGCATCCATGTTGCTGCAAAGGACATGATTTCATTATTTTTTATGGC 2321

Qy 2441 TATGTTCTTTCCCAATTTAGATAAA 2465
Db 2322 TTTGAGTTGTCACTCTTGAGTGAAA 2346

RESULT 90
ADC85379
ID ADC85379 standard; DNA; 31236 BP.
XX
AC ADC85379;
XX
DT 01-JAN-2004 (first entry)
XX
DE Mouse Ptp4a2 coding sequence.
XX
KW Cytostatic; gene therapy; vaccine; cancer; carcinoma-associated gene; CA;
KW secreted; transmembrane; intracellular; ds.
XX
OS Mus sp.
XX
PN WO2003045230-A2.
XX
PD 05-JUN-2003.
XX
PF 02-DEC-2002; 2002WO-US038582.
XX
PR 30-NOV-2001; 2001US-00997722.
XX
PA (SAGR-) SAGRES DISCOVERY.
XX
PI Morris DW, Engelhard EK;
XX
DR WPI; 2003-513603/48.
XX
PT New recombinant nucleic acid comprising a nucleotide sequence of any of
PT the carcinoma-associated (CA) genes, useful for screening for drug
PT candidates for diagnosing or treating carcinomas.

XX
PS Claim 1; SEQ ID NO 165; 983pp; English.
XX
CC The invention relates to a recombinant nucleic acid comprising a
CC nucleotide sequence selected from any of the fully defined carcinoma-
CC associated (CA) genes from the 50 tables given in the specification. The
CC CA proteins are secreted, transmembrane or intracellular proteins. The
CC recombinant nucleic acids are useful for screening for drug candidates
CC for diagnosing or treating carcinomas. Sequences given in ADC85215-
CC ADC85514 represent CA genes of the invention.
XX
SQ Sequence 31236 BP; 8595 A; 6081 C; 6658 G; 9902 T; 0 U; 0 Other;

Query Match 2.3%; Score 137; DB 10; Length 31236;
Best Local Similarity 69.8%; Pred. No. 4.3e-22;
Matches 185; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

Qy 2201 GCATTGTACCCCTTAAGTAATTTCTCACCATCCGCTGACTTCTTGCCCCCTCATCCTTCT 2260
Db 2082 GCATAGTACCTGATAAGTAGTTTTTCAAGCCCCACTCTCCTCTCACCCTCCACCCTCAAG 2141

Qy 2261 GAGGCTCCATTGTCCATCATTCCACACTCTACATCTATGTGTACACATTATTTAGCTCCT 2320
Db 2142 TGGTCCCCAGTGTCTGTTGTTCCCCTCTTTGAGTCTATGTGTACTCAGTGTTTAGCTCCC 2201

Qy 2321 ACTTATAAGTGATAACATGCAATATTTGTCTTTCTGTGTCTGTCTTGTTTTACTTATGAT 2380
Db 2202 ACTTATAAGTGAGAACATGCGGTATTTGGTTTTCTGTTCCTGCGTTAATTCGCTTAGGAT 2261

Qy 2381 AATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAAGGCATGATTTCATTCTTTTTTATGGC 2440
Db 2262 AATGGCCTCCAGCTGCATCCATGTTGCTGCAAAGGACATGATTTCATTATTTTTTATGGC 2321

Qy 2441 TATGTTCTTTCCCAATTTAGATAAA 2465
Db 2322 TTTGAGTTGTCACTCTTGAGTGAAA 2346

RESULT 91
ADM74495
ID ADM74495 standard; DNA; 31236 BP.
XX
AC ADM74495;
XX
DT 01-JUL-2004 (first entry)
XX
DE Human carcinoma associated (CA) nucleic acid #82.
XX
KW Human; carcinoma associated nucleic acid; CA nucleic acid; gene; ds;
KW carcinoma associated protein; CAP; carcinoma; leukaemia; lymphoma;
KW cytostatic.
XX
OS Homo sapiens.
XX
PN US2004072154-A1.
XX
PD 15-APR-2004.
XX
PF 30-NOV-2001; 2001US-00997722.
XX
PR 22-DEC-2000; 2000US-00747377.
PR 02-MAR-2001; 2001US-00798586.
XX
PA (MORR/) MORRIS D W.
PA (ENGE/) ENGELHARD E K.
XX
PI Morris DW, Engelhard EK;
XX
DR WPI; 2004-328562/30.
XX
PT New carcinoma associated gene or protein, useful for preparing a
PT composition for diagnosing or treating carcinoma e.g., leukemia or
PT lymphoma.

XX
PS Claim 1; SEQ ID NO 166; 29pp; English.
XX
CC The invention relates to new recombinant nucleic acids. The invention
CC also relates to a host cell comprising a recombinant nucleic acid or
CC expression vector, an expression vector comprising a recombinant nucleic
CC acid, a recombinant protein, a method of screening for drug candidates, a
CC method of screening for a bioactive agent capable of binding to a
CC carcinoma associated protein (CAP) encoded by a nucleotide sequence, a
CC method of screening for a bioactive agent capable of modulating the
CC activity of a CAP, a method of evaluating the effect of a candidate
CC carcinoma drug, a method of diagnosing carcinoma, a method for inhibiting
CC the activity of a CAP, a method of treating carcinomas, a method of
CC neutralising the effect of a CAP and a method of diagnosing carcinoma or
CC propensity to carcinoma. A method of evaluating the effect of a candidate
CC carcinoma drug comprises administering the drug to a patient, removing a
CC cell sample from the patient and determining alterations in the
CC expression or activation of a gene comprising the nucleotide sequence. A
CC method of diagnosing carcinoma comprises determining the expression of
CC one or more genes comprising the nucleic acid sequence in a first tissue
CC type of a first individual and comparing the expression of the gene from
CC a second normal tissue type from the first individual or a second
CC unaffected individual, where a difference in the expression indicates
CC that the first individual has carcinoma. A method of inhibiting the
CC activity of a CAP comprises binding an inhibitor to the CAP. Treating
CC carcinomas comprises administering to a patient an inhibitor of CAP.
CC Neutralising the effect of a CAP comprises contacting an agent specific
CC for the CAP. The polypeptide specifically binds to the protein encoded by
CC the nucleic acid. It comprises an antibody that specifically binds to the
CC protein encoded by the nucleic acid. The nucleic acids are useful for
CC preparing a composition for diagnosing or treating carcinoma e.g.,
CC leukaemia or lymphoma. This sequence represents a human carcinoma
CC associated (CA) nucleic acid of the invention. Note: The sequence data
CC for this patent did not form part of the printed specification but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 31236 BP; 8595 A; 6081 C; 6658 G; 9902 T; 0 U; 0 Other;

Query Match 2.3%; Score 137; DB 12; Length 31236;
Best Local Similarity 69.8%; Pred. No. 4.3e-22;
Matches 185; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

```
Qy       2201 GCATTGTACCCCTTAAGTAATTTCTCACCATCCGCTGACTTCTTGCCCCCTCATCCTTCT 2260
             |||| |||||   |||||| ||| |||    || ||  | |||  ||| |    |
Db       2082 GCATAGTACCTGATAAGTAGTTTTTCAAGCCCCACTCTCCTCTCACCCTCCACCCTCAAG 2141

Qy       2261 GAGGCTCCATTGTCCATCATTCCACACTCTACATCTATGTGTACACATTATTTAGCTCCT 2320
               | ||| ||||  |  |||| | || |    ||||||||||| || | ||||||||||
Db       2142 TGGTCCCCAGTGTCTGTTGTTCCCCTCTTTGAGTCTATGTGTACTCAGTGTTTAGCTCCC 2201

Qy       2321 ACTTATAAGTGATAACATGCAATATTTGTCTTTCTGTGTCTGTCTTGTTTTACTTATGAT 2380
             ||||||||||| |||||||  ||||||  |||||||  |||  ||  ||  |||| |||
Db       2202 ACTTATAAGTGAGAACATGCGGTATTTGGTTTTCTGTTCCTGCGTTAATTCGCTTAGGAT 2261

Qy       2381 AATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAAGGCATGATTTCATTCTTTTTTATGGC 2440
             |||||||  |||| |  ||| | | |||||||||| | ||||||||||| ||||||||||
Db       2262 AATGGCCTCCAGCTGCATCCATGTTGCTGCAAAGGACATGATTTCATTATTTTTTATGGC 2321

Qy       2441 TATGTTCTTTCCCAATTTAGATAAA 2465
              | ||   | || |  || ||  |||
Db       2322 TTTGAGTTGTCACTCTTGAGTGAAA 2346
```

RESULT 92
AAL52246_1
Continuation (2 of 4) of AAL52246 from base 100001 (Human genomic DNA for the gene encod
WP Sequence split into 4 fragments LOCUS AAL52246 Accession Aal52246
WP Fragment Name Begin End
WP AAL52246_0 1 110000
WP AAL52246_1 100001 210000
WP AAL52246_2 200001 310000
WP AAL52246_3 300001 378361

Query Match 2.2%; Score 136.4; DB 10; Length 110000;
Best Local Similarity 69.5%; Pred. No. 1.1e-21;
Matches 214; Conservative 0; Mismatches 91; Indels 3; Gaps 2;

```
Qy       2177 GGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCT 2236
              | ||   |||||||||||  ||  |||  |||||  || |||||| |||  | || |
Db      17535 GTGTGCCCATCACCTGAATACTGAACATGATACCCAGTAGGTAATTTTTCA--ACCCTCA 17592

Qy       2237 GACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTCTACATCT 2296
                |  |||   | ||   | ||   |  | ||| | ||  | ||| |    | ||| ||
Db      17593 CGCCCCTTCTACTCTTCCCTTTTGAACTCCCCACTATCTGTTATTTCCTTTTGTACGTCC 17652

Qy       2297 ATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGTCTTTCTG 2356
              |||||||| |||| ||||||||||||||||||||| |||||||| ||||||  ||||||
Db      17653 ATGTGTACCCATTGTTTAGCTCCTACTTATAAGTGAGAACATGCAGTATTTGATTTTCTG 17712

Qy       2357 TGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTA-GGCTGCTGCAAAAG 2415
              | ||||  || ||||| ||| | |||||||| |||| || ||||| |||||||  |||||
Db      17713 TTTCTGAGTTATTTTAGTTAGGTTAATGGCCTCCAGCTCCATCTATGGCTGCTTCAAAAG 17772

Qy       2416 GCATGATTTCATTCTTTTTTATGGCTATGTTCTTTCCCAATTTAGATAAAGAACACTCGC 2475
                |||||||||||||||||  | | |||  | | |||   |  ||| |   ||| |
Db      17773 ATGTGATTTCATTCTTTTTTACAGTTGTGTAGTATTCCATGATGTATATATGCCACACTT 17832

Qy       2476 ACTTGCTC 2483
                |||  ||
Db      17833 TCTTTATC 17840
```

RESULT 93
ACN44874/c
ID ACN44874 standard; DNA; 34161 BP.
XX
AC ACN44874;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human genomic sequence hCG21464.
XX
KW Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.
XX
OS Homo sapiens.
XX
PN WO2003073826-A2.
XX
PD 12-SEP-2003.
XX
PF 28-FEB-2003; 2003WO-US006235.
XX
PR 01-MAR-2002; 2002US-00087192.
XX
PA (SAGR-) SAGRES DISCOVERY.
XX
PI Morris DW;
XX
DR WPI; 2003-328604/31.
XX
PT Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
PT comprises a nucleotide sequence.
XX
PS Claim 1; SEQ ID NO 1540; 0pp; English.
XX
CC The present invention relates to novel DNA and protein sequences which
CC are associated with carcinomas. The sequences are useful for: (i) for
CC screening drug candidates; (ii) for screening of bioactive agent capable
CC of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
CC a bioactive agent capable of modulating the activity of CAP; (iv) for
CC evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
CC carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
CC carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
CC (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
CC determining Carcinoma Associated (CA) gene copy number. In addition, the

CC CA genes are useful as DNA vaccines and the CAP are useful as markers of
CC carcinoma including lymphoma. The present sequence is one such CA coding
CC sequence. Note: This patent is an equivalent to basic patent
CC US2002182586A1, for which no sequence data was published
XX
SQ Sequence 34161 BP; 9518 A; 7175 C; 7133 G; 10335 T; 0 U; 0 Other;

Query Match 2.2%; Score 135.6; DB 11; Length 34161;
Best Local Similarity 66.3%; Pred. No. 9.9e-22;
Matches 195; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

Qy 2170 GCCCCCTGGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACC 2229
Db 7504 GCTTCCATTGATTCCCTCACCCAAATAGTGAACATTGTACCCTATAGGTAGTTTTTCAAA 7445

Qy 2230 ATCCGCTGACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTC 2289
Db 7444 CCTTGCCCCCCTCCCTGCCTCTCCCTTTTTGGGATCTGCAGTGTTCATTGTTCCCTTCTT 7385

Qy 2290 TACATCTATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGT 2349
Db 7384 TATGTCCATGTATACCCCATGTTTAGCTCTCACTTATAAGTGAGAACATGCAGTATTGGG 7325

Qy 2350 CTTTCTGTGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTG 2409
Db 7324 CTTTCTGTTTCTGTGTTAATTTGCTTAGGATAATGGCCTCTAGCTGCATCCATGTTGCTG 7265

Qy 2410 CAAAAGGCATGATTTCATTCTTTTTTATGGCTATGTTCTTTCCCAATTTAGATA 2463
Db 7264 CAAGGGACATGATTTTGTTCTTTTTTATGGCTGCATAGTATTCCTTGCTGTATA 7211

RESULT 94
ABD33349/c
ID ABD33349 standard; DNA; 34161 BP.
XX
AC ABD33349;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human cancer-associated (CA) gene HD07-063.
XX
KW Human; cancer-associated protein; CAP; cancer-associated gene; CA; gene;
KW ds; cancer; cytostatic.
XX
OS Homo sapiens.
XX
PN WO2004058146-A2.
XX
PD 15-JUL-2004.
XX
PF 15-DEC-2003; 2003WO-US040081.
XX
PR 17-DEC-2002; 2002US-00322281.
XX
PA (SAGR-) SAGRES DISCOVERY INC.
XX
PI Morris DW, Malandro MS;
XX
DR WPI; 2004-499109/47.
XX
PT Novel human cancer associated protein encoded within open reading frame
PT of cancer associated gene, useful as targets for diagnosing cancer.
XX
PS Claim 16; SEQ ID NO 436; 182pp; English.
XX
CC The invention relates to cancer-associated proteins (CAP) and the cancer-
CC associated (CA) nucleic acids encoding them. The invention also relates
CC to a method for treating cancers involving administering to a patient an
CC inhibitor of CAP, and a method of screening for anticancer activity in a
CC potential drug involving providing a cell that expresses a CA gene,
CC contacting a tissue sample derived from a cancer cell with an anticancer
CC drug candidate and monitoring the effect of the anticancer drug candidate
CC on expression of the CA gene. The CAP proteins are useful for detecting
CC cancer associated with expression of a CAP protein in a test cell sample
CC and for screening for a bioactive agent capable of modulating the
CC activity of a CAP protein. The CA nucleic acids are useful for diagnosing
CC cancer, involving determining the expression of a CA nucleic acid in a
CC tissue. This sequence represents a human CA gene of the invention. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 34161 BP; 9518 A; 7175 C; 7133 G; 10335 T; 0 U; 0 Other;

Query Match 2.2%; Score 135.6; DB 13; Length 34161;
Best Local Similarity 66.3%; Pred. No. 9.9e-22;
Matches 195; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

Qy 2170 GCCCCCTGGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACC 2229
Db 7504 GCTTCCATTGATTCCCTCACCCAAATAGTGAACATTGTACCCTATAGGTAGTTTTTCAAA 7445

Qy 2230 ATCCGCTGACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTC 2289
Db 7444 CCTTGCCCCCCTCCCTGCCTCTCCCTTTTTGGGATCTGCAGTGTTCATTGTTCCCTTCTT 7385

Qy 2290 TACATCTATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGT 2349
Db 7384 TATGTCCATGTATACCCCATGTTTAGCTCTCACTTATAAGTGAGAACATGCAGTATTGGG 7325

Qy 2350 CTTTCTGTGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTG 2409
Db 7324 CTTTCTGTTTCTGTGTTAATTTGCTTAGGATAATGGCCTCTAGCTGCATCCATGTTGCTG 7265

Qy 2410 CAAAAGGCATGATTTCATTCTTTTTTATGGCTATGTTCTTTCCCAATTTAGATA 2463
Db 7264 CAAGGGACATGATTTTGTTCTTTTTTATGGCTGCATAGTATTCCTTGCTGTATA 7211

RESULT 95
ADG20558
ID ADG20558 standard; DNA; 15620 BP.
XX
AC ADG20558;
XX
DT 11-MAR-2004 (first entry)
XX
DE Human RNASEL nucleotide sequence SEQ ID NO:2.
XX
KW prostate cancer; RNASEL; cytostatic; human; gene; ds; chromosome 1.
XX
OS Homo sapiens.
XX
PN WO2003106617-A2.
XX
PD 24-DEC-2003.
XX
PF 12-JUN-2003; 2003WO-IL000507.
XX
PR 12-JUN-2002; 2002US-0387589P.
XX
PA (TELA-) TEL AVIV MEDICAL CENT RES DEV FUND.
XX
PI Orr-Urtreger A, Rennert H, Bercovich D, Bar-Shira A, Yaron Y;
XX
DR WPI; 2004-082170/08.
XX
PT Determining predisposition of individual of Ashkenazi descent to prostate
PT cancer, involves determining presence or absence of one or more nucleic
PT acid sequence alteration in one or more allele of RNASEL gene of
PT individual.
XX
PS Claim 2; SEQ ID NO 2; 72pp; English.
XX
CC The present invention describes a method (M1) for determining

CC predisposition of subject, preferably Ashkenazi descent to prostate
CC cancer, comprising determining presence or absence of one or more nucleic
CC acid sequence alteration in one or more allele of RNASEL gene of the
CC individual, where presence of one or more nucleic acid sequence
CC alteration indicates predisposition to prostate cancer in the individual.
CC Also described: (1) oligonucleotide (I) specifically binding to the
CC nucleic acid alternation being chosen from deletion spanning nucleotide
CC 471-474 of ADG20557, C to T substitution at nucleotides 354 of ADG20557,
CC and deletion at nucleotide 11338427 of ADG20558; (2) a kit (II) for
CC diagnosing prostate cancer or predisposition to prostate cancer in a
CC subject, the kit comprising (I) and one or more reagent for detecting the
CC hybridization of (I) with nucleic acid isolated from subject; (3)
CC determining sensitivity of a subject to prospective interferon therapy,
CC comprising determining the presence or absence of one or more nucleic
CC acid alternation being chosen from deletion spanning nucleotide 471-474
CC of ADG20557, C to T substitution at nucleotides 354 of ADG20557, and
CC deletion at nucleotide 11338427 of ADG20558; (4) an antibody or its
CC fragment being capable of specifically binding to one or more portion of
CC amino acid residues 1-55 of RNASEL polypeptide; (5) an antibody (III) or
CC its fragment being capable of specifically binding a polypeptide
CC including amino acid sequence: Asn-Lys-Thr-Cys-Phe-His-Ser-Thr-Lys-Leu-
CC Thr-Val-Lys; and (6) a kit (IV) for diagnosing prostate cancer or a
CC predisposition to prostate cancer in a subject, the kit comprising the
CC antibody or its fragment and one or more reagent for detecting binding of
CC (III) to the polypeptide from the subject. RNASEL has cytostatic
CC activity. (M1) is useful for treating a subject having for predisposition
CC to prostate cancer, which involves specifically down regulating in the
CC subject expression of mutated RNASEL transcript having one or more
CC nucleic acid alternation being chosen from deletion spanning nucleotide
CC 471-474 of ADG20557, C to T substitution at nucleotides 354 of ADG20557,
CC and deletion at nucleotide 11338427 of ADG20558, and thereby preventing
CC the formation or halting the progression of prostate cancer in the
CC subject. The human RNASEL gene is located on chromosome 1, more
CC specifically to region 1q25. The present sequence is used in the
CC exemplification of the present invention.
XX
SQ Sequence 15620 BP; 4431 A; 3136 C; 3351 G; 4702 T; 0 U; 0 Other;

Query Match 2.2%; Score 135.4; DB 12; Length 15620;
Best Local Similarity 75.1%; Pred. No. 7.7e-22;
Matches 196; Conservative 0; Mismatches 61; Indels 4; Gaps 2;

```
Qy       2177 GGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCT 2236
             | |||  ||| ||||||| | |  |||||||||| |||||||||||||| |   | ||
Db      10105 GTGTAACCATTACCTGAATAACGCACATTGTACCCATTAAGTAATTTCTCATCCCTCTC- 10163

Qy       2237 GACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTCTACATCT 2296
               | ||   ||| | || | |||  || ||||| | |  || ||||||| || ||||||
Db      10164 --CCTCCAACCCTCCCACCTTTCCAAGTCTCCAAT-TATATTATTCCACTCTGTACATCC 10220

Qy       2297 ATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGTCTTTCTG 2356
              |||||||||||||||| |||||| ||||||||| || ||||||  ||||||| |||||||
Db      10221 ATGTGTACACATTATTGAGCTCCCACTTATAAGAGAGAACATGTGATATTTGACTTTCTG 10280

Qy       2357 TGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAAGG 2416
              | | |   |  ||| ||||| |||||||| ||||||| || | |  ||  ||||||
Db      10281 TTTGTAAGTCATTTCACTTAAAATAATGGCCTCCAGTTCCGTCCATGTGGCCACAAAAGA 10340

Qy       2417 CATGATTTCATTCTTTTTTAT 2437
              |||||||||||||| ||| |
Db      10341 CATGATTTCATTCTTCTTTTT 10361
```

RESULT 96
AAK85177
ID AAK85177 standard; DNA; 7268 BP.
XX
AC AAK85177;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:39989.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001354.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.

PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.

PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS Disclosure; SEQ ID NO 39989; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention
XX
SQ Sequence 7268 BP; 1732 A; 1341 C; 2069 G; 2126 T; 0 U; 0 Other;

Query Match 2.2%; Score 134.8; DB 4; Length 7268;
Best Local Similarity 73.9%; Pred. No. 7.5e-22;
Matches 198; Conservative 0; Mismatches 67; Indels 3; Gaps 2;

```
Qy       2179 GTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCTGA 2238
             |||| |||| ||||||| |||| ||||||| || |||||||||| ||| ||  |
Db       1162 GTATCCATCGCCTGAATAATGTACATTGTATCCATTAAGTAATTTTTCATCACTCATCTT 1221

Qy       2239 CTTCTTGCCCCCTCATCCTTCT-GAGGCTCCATTGTCCATCATTCCACACTCTACATCTA 2297
             | |||   ||    | |  | | ||| ||||| |||   ||||||||||||||   || |
Db       1222 CGTCTCATCCTTCTACCTGTTTGGAGTCTCCAATGTTTGTCATTCCACACTCTTTGTCCA 1281

Qy       2298 TGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGTCTTTCTGT 2357
             |||||||||||||| |||| | |||| ||||||| |||||    |||||| ||||||||
Db       1282 TGTGTACACATTATTCAGCTTCCACTTGTAAGTGAGAACATTTGGTATTTGACTTTCTGT 1341

Qy       2358 GTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAAGGC 2417
              ||||  |||||| | |||  ||||||||| ||||||| ||| |   |||| |||| |
Db       1342 TTCTGAGTTGTTTCATTTA--ATAATGGCCTCCAGTTCCATCCACATTGCTGTAAAAGAC 1399

Qy       2418 ATGATTTCATTCTTTTTTATGGCTATGT 2445
              ||  |||||| ||||||| |   | |||
Db       1400 ATAGTTTCATGCTTTTTTTTTTTTTGT 1427
```

RESULT 97
ADQ18153/c
ID ADQ18153 standard; DNA; 112604 BP.
XX
AC ADQ18153;
XX
DT 26-AUG-2004 (first entry)
XX
DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 970.
XX
KW soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
KW ds.
XX
OS Homo sapiens.
XX

PN WO2004048938-A2.
XX
PD 10-JUN-2004.
XX
PF 26-NOV-2003; 2003WO-US038193.
XX
PR 26-NOV-2002; 2002US-0429739P.
XX
PA (PROT-) PROTEIN DESIGN LABS INC.
XX
PI Aziz N, Ginsburg WM, Zlotnik A;
XX
DR WPI; 2004-441208/41.
XX
PT Early detection of soft tissue sarcoma comprises determining expression
PT of a gene in a first soft tissue sample and a normal soft tissue sample
PT and comparing the gene expression, also useful in treating soft tissue
PT sarcoma.
XX
PS Example 2; SEQ ID NO 970; 210pp; English.
XX
CC The invention relates to a novel method for detecting soft tissue sarcoma
CC which comprises obtaining a first soft tissue sample from an individual
CC and a normal soft tissue sample from the same or different individual,
CC determining the expression of a gene in both samples and comparing the
CC expression of the gene in both soft tissue samples, where a higher level
CC of protein expression in the first soft tissue sample indicates the
CC presence of soft tissue sarcoma. The method of the invention has
CC cytostatic applications and may be useful for detecting soft tissue
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC acid sequences may be useful in diagnostic and screening applications.
CC The current sequence is that of a human soft tissue sarcoma-upregulated
CC DNA of the invention. The current sequence is not shown within the
CC specification per se but was submitted in CD format by the inventor.
XX
SQ Sequence 112604 BP; 37437 A; 21989 C; 19831 G; 33347 T; 0 U; 0 Other;

Query Match 2.2%; Score 134.4; DB 12; Length 112604;
Best Local Similarity 62.1%; Pred. No. 3.4e-21;
Matches 365; Conservative 0; Mismatches 156; Indels 67; Gaps 7;

```
Qy      289 TCTCCTTCTCTCTCAAAGCTGCAGTAGGGGTGAAAACGTGTGATACATTTTCTCTTCCAT 348
            ||| |  |||  | || || |   | ||||   ||| || |||||| ||||||      |
Db    79367 TCTACACCTCATTGAATGCGGTTTTGGGGGCAGAAAAGTTTGATACCTTTTCTTACTGGT 79308

Qy      349 CATAAGGGTCGCAAC-CAAAACTCCTATAGTAAAAGACAGGTTAATAAGAGCAAAACCTA 407
            ||||||  ||  || |  |   |||||||   |||||||||||||| ||||| |||  | ||
Db    79307 GATAAGGCTCTCATCTGACCGCTCCTATTACAAAAGACAGGTTAACAAGAGAAAAGCATA 79248

Qy      408 ACAAATTTATTTAATCAAAGTTTTACATGACATGGGAGTCTTCAGAAATGAAGACCCAAA 467
            |||||  |||||||||  ||||||||  ||||  |||||||||||  |||||||||  ||||||
Db    79247 GCAAATGTATTTAATC-AAGTTTTATGTGACCTGGGAGTCTTCTGAAATGAAGGCCCAAA 79189

Qy      468 GACCCAGGGGAAACTGTCTGTTTTTTTTGCTGAGGTTCGATGAAGAATGGATAGCATGTA 527
            ||| |||||  |||||  |     |||||  |  ||  |||  |  ||||||||||||  |
Db    79188 GACGCAGGGAAAACTCT---ATTTTTATTCTTAGGATTGATGAAGAATGGACACTCCCCC 79132

Qy      528 GCCATGTAGATTAGACAAAAGGATATGATCTAGTGGTAAAGGACTCAGGGGGAAACACAG 587
              ||    | |   |  | |  ||    | | |||
Db    79131 ATCACTCTGTTACTATTAGATCATTCCCTTTTGTG------------------------- 79097

Qy      588 CAAGGCCTGTCTATTCAGATTCTTCTTGATCTCTCTCTCTCTATGTATAGCATTCTTTCC 647
                        ||  |   | || ||   |||    || | | ||| |||  ||||
Db    79096 -------------AACACCTGTCTGTTCATTCTTCTTGGCCTTTCTGTAGTATTGCTTCC 79050

Qy      648 TCCTGAGTATGGGGCAGGACTC-TTCTTCAATGAGGGTCTTCA-------AGGGAGAAGG 699
            || || |||||  || |||| |   ||  |||||| |||| |       | ||||||||
Db    79049 TCTTGGGTATGCTGCTGGACCCGCCCTGGAATGAGGCTCTTTAGGGAAGGATGGAGAAGG 78990

Qy      700 GAGAAAGTGGCCTTTTTAGATTTTATGGCTTGCTTCGGGGAAGAGGAGTTCTAGTTTCTA 759
            ||||  ||| ||||| | | ||||||||| || || || | |||||||||
Db    78989 GAGAGGGTGACCTTTCTGGGTTTTATGGCCTGTTTTGGAGGAGAGGAGTT---------- 78940
```

Qy 760 TGACCCATCTTGGGGAAGAGGAATTCTGGTTTCTGTGACTTGCTTTCATGAAGAAAGAGG 819
|||||||| | |||| || ||||| |||||||||| | ||| |||||
Db 78939 ------GCCTTGGGGAAAACAAATTTTGTTTTCTATGACTTGCTTCAGGGGAGATAGAGG 78886

Qy 820 AGTAAGAGGCAGGAGGGCAGGAGATGGTCAGAAAGAGACTTGGCTGCT 867
|| ||| ||| ||||||||||| ||||||| || ||| ||| ||
Db 78885 GATAGGAGACAGAAGGGCAGGAGAAAGTCAGAATAAGGCTTTGCTTCT 78838

RESULT 98
ACH32342
ID ACH32342 standard; cDNA; 493 BP.
XX
AC ACH32342;
XX
DT 13-OCT-2003 (first entry)
XX
DE Human endothelial cell cDNA #475.
XX
KW Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;
KW genome mapping; biodiversity; genetic disorder.
XX
OS Homo sapiens.
XX
PN US2003073623-A1.
XX
PD 17-APR-2003.
XX
PF 30-JUL-2001; 2001US-00918995.
XX
PR 30-JUL-2001; 2001US-00918995.
XX
PA (DRMA/) DRMANAC R T.
PA (LABA/) LABAT I.
PA (STAC/) STACHE-CRAIN B.
PA (DICK/) DICKSON M C.
PA (JONE/) JONES L W.
XX
PI Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;
XX
DR WPI; 2003-615964/58.
XX
PT New polynucleotide sequences obtained from various cDNA libraries, useful
PT as hybridization probes, as oligomers for PCR, for chromosome and gene
PT mapping, in the recombinant production of protein, or in generating
PT antisense DNA or RNA.
XX
PS Claim 1; SEQ ID NO 19554; 44pp; English.
XX
CC The invention relates to an isolated polynucleotide comprising any one of
CC 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
CC determined by the technique of SBH (sequencing by hybridisation). Also
CC included is a purified polypeptide comprising a sequence corresponding to
CC a reading frame of the novel polynucleotide. The nucleic acid sequences
CC are useful in diagnostics as expressed sequence tags (EST) for
CC identifying expressed genes or for physical mapping of the human genome,
CC in forensics, in assessing biodiversities, or in identifying mutations
CC responsible for genetic disorders and other traits. The nucleotide
CC sequences are also useful as hybridisation probes, as oligomers for PCR,
CC for chromosome and gene mapping, in the recombinant production of
CC protein, or in generating antisense DNA or RNA. The purified polypeptide
CC is useful for generating antibodies specific for it. The present sequence
CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=20030073623
XX
SQ Sequence 493 BP; 129 A; 101 C; 86 G; 175 T; 0 U; 2 Other;

Query Match 2.2%; Score 134; DB 9; Length 493;
Best Local Similarity 63.6%; Pred. No. 3.4e-22;
Matches 203; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

```
Qy       2177 GGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCTCACCATCCGCT 2236
Db        167 GTGAACCCATCACCCAAATAGTGAACATTGTACTTAACAGGGAATTTTTCAACCGTCACT 226

Qy       2237 GACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCACACTCTACATCT 2296
Db        227 CTTCTCCCACTCTCCCACCTTTTGGAGTCTCCAATGTCTATTTTTCCACTCTGTATGTCT 286

Qy       2297 ATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATATTTGTCTTTCTG 2356
Db        287 GTGTGTGCCCATTGCTTAGCTCCCACTTATAAATGAGAATATGTGGTATTTGACTTTCTA 346

Qy       2357 TGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCTGCTGCAAAAGG 2416
Db        347 GTTCAAGGTTATTTCCCTTAAGACAATGGCATTCGGTTCCATCCATGTTGCTGCAAAAGA 406

Qy       2417 CATGATTTCATTCTTTTTTATGGCTATGTTCTTTCCCAATTTAGATAAAGAACACTCGCA 2476
Db        407 CATGATTTCATTNTTTTTTTTAAGACTGAGCAGTAATATATGATTTAATCACCACTTTGA 466

Qy       2477 CTTGCTCTTACTTCTATTT 2495
Db        467 GAATATGTTATCTCATATT 485
```

RESULT 99
ABT10748
ID ABT10748 standard; cDNA; 121600 BP.
XX
AC ABT10748;
XX
DT 04-DEC-2002 (first entry)
XX
DE Human breast cancer associated coding sequence SEQ ID NO: 882.
XX
KW Human; breast specific gene; breast cancer; differential expression;
KW cytostatic; gene therapy; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200259271-A2.
XX
PD 01-AUG-2002.
XX
PF 25-JAN-2002; 2002WO-US002176.
XX
PR 25-JAN-2001; 2001US-0263757P.
PR 25-APR-2001; 2001US-0286090P.
PR 23-MAY-2001; 2001US-0292517P.
XX
PA (GENE-) GENE LOGIC INC.
XX
PI Orr MS, Nation M, Diggans JC, Zeng W;
XX
DR WPI; 2002-674803/72.
XX
PT Diagnosing breast cancer in a patient comprises detecting the level of
PT gene expression in cell or tissue samples, where a differential gene
PT expression is indicative of breast cancer.
XX
PS Claim 1; SEQ ID NO 882; 260pp + Sequence Listing; English.
XX
CC The present invention relates to methods of diagnosing breast cancer in a
CC patient, which comprise detecting the level of expression in a tissue
CC sample of two or more genes selected from those shown in ABT09867-
CC ABT11112, where a differential expression of the genes indicates breast
CC cancer. The methods are useful in diagnosing, treating, detecting the
CC progression, and in monitoring treatment of breast cancer in patients.
CC The methods are also useful as a screening tool for agents that modulate
CC the onset or progression of breast cancer. The breast cancer genes may be
CC used as diagnostic markers for the prediction or identification of the
CC malignant state of breast tissue, for confirming the type and progression
CC of cancer, and for drug screening and assays. The present sequence is a
CC coding sequence of the invention. Note: The sequence data for this patent
CC did not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub.published_pct_sequences
XX
SQ Sequence 121600 BP; 38372 A; 24652 C; 24048 G; 34528 T; 0 U; 0 Other;

  Query Match   2.2%; Score 134; DB 6; Length 121600;
  Best Local Similarity   65.2%; Pred. No. 4.4e-21;
  Matches 197; Conservative   0; Mismatches 105; Indels   0; Gaps   0;

```
Qy       2166 TCTTGCCCCCTGGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCT 2225
Db       4257 TTTGGCTTCTAGTGGACCCATAAACCAAATATTGAACATTGTACCCAATAGGTAATTTTT 4316

Qy       2226 CACCATCCGCTGACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCAC 2285
Db       4317 TAACCCTCATTCCTCTCCCACCCTACCCTGTTTTGGAGTCCCCTGTGTCTACTATTTCCA 4376

Qy       2286 ACTCTACATCTATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATAT 2345
Db       4377 TCTTTATGCCCATGTGTTCCCATTGTCTAGCTCCCACTTATAAGTAAGAATATGTGATAT 4436

Qy       2346 TTGTCTTTCTGTGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCT 2405
Db       4437 TTGATTTTCTTTTTCTGAGTTAGTTCACTTAGGATAATGGCCTCCAGCTCCATCCATGTT 4496

Qy       2406 GCTGCAAAAGGCATGATTTCATTCTTTTTTATGGCTATGTTCTTTCCCAATTTAGATAAA 2465
Db       4497 GCTGCAAAGGACATGATTTCATTCCTTCTTATGGCTGCATAATATTCCATGGGGTATATA 4556

Qy       2466 GA 2467
Db       4557 TA 4558
```

RESULT 100
ADQ18307
ID ADQ18307 standard; DNA; 121600 BP.
XX
AC ADQ18307;
XX
DT 26-AUG-2004 (first entry)
XX
DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 1125.
XX
KW soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
KW ds.
XX
OS Homo sapiens.
XX
PN WO2004048938-A2.
XX
PD 10-JUN-2004.
XX
PF 26-NOV-2003; 2003WO-US038193.
XX
PR 26-NOV-2002; 2002US-0429739P.
XX
PA (PROT-) PROTEIN DESIGN LABS INC.
XX
PI Aziz N, Ginsburg WM, Zlotnik A;
XX
DR WPI; 2004-441208/41.
XX
PT Early detection of soft tissue sarcoma comprises determining expression
PT of a gene in a first soft tissue sample and a normal soft tissue sample
PT and comparing the gene expression, also useful in treating soft tissue
PT sarcoma.
XX
PS Example 2; SEQ ID NO 1125; 210pp; English.
XX
CC The invention relates to a novel method for detecting soft tissue sarcoma

CC which comprises obtaining a first soft tissue sample from an individual
CC and a normal soft tissue sample from the same or different individual,
CC determining the expression of a gene in both samples and comparing the
CC expression of the gene in both soft tissue samples, where a higher level
CC of protein expression in the first soft tissue sample indicates the
CC presence of soft tissue sarcoma. The method of the invention has
CC cytostatic applications and may be useful for detecting soft tissue
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC acid sequences may be useful in diagnostic and screening applications.
CC The current sequence is that of a human soft tissue sarcoma-upregulated
CC DNA of the invention. The current sequence is not shown within the
CC specification per se but was submitted in CD format by the inventor.
XX
SQ Sequence 121600 BP; 38372 A; 24652 C; 24048 G; 34528 T; 0 U; 0 Other;

Query Match 2.2%; Score 134; DB 12; Length 121600;
Best Local Similarity 65.2%; Pred. No. 4.4e-21;
Matches 197; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

```
Qy       2166 TCTTGCCCCCTGGGTATTCATCACCTGAATGATGTGCATTGTACCCCTTAAGTAATTTCT 2225
             | | ||  |  | | |  ||| | |  |||  ||  ||||||||||  || ||||||| |
Db       4257 TTTGGCTTCTAGTGGACCCATAAACCAAATATTGAACATTGTACCCAATAGGTAATTTTT 4316

Qy       2226 CACCATCCGCTGACTTCTTGCCCCCTCATCCTTCTGAGGCTCCATTGTCCATCATTCCAC 2285
              | |   |  |    ||   |||   | |  ||  ||| | ||  |||| |  ||| |
Db       4317 TAACCCTCATTCCTCTCCCACCCTACCCTGTTTTGGAGTCCCCTGTGTCTACTATTTCCA 4376

Qy       2286 ACTCTACATCTATGTGTACACATTATTTAGCTCCTACTTATAAGTGATAACATGCAATAT 2345
              || ||   | |||||| | |||| | ||||||| |||||||||| | || |||  ||||
Db       4377 TCTTTATGCCCATGTGTTCCCATTGTCTAGCTCCCACTTATAAGTAAGAATATGTGATAT 4436

Qy       2346 TTGTCTTTCTGTGTCTGTCTTGTTTTACTTATGATAATGGCCCCCAGTTCTATCTAGGCT 2405
              |||  ||||| | ||||  ||  || ||||| |||||||||| |||| || ||| | | |
Db       4437 TTGATTTTCTTTTTCTGAGTTAGTTCACTTAGGATAATGGCCTCCAGCTCCATCCATGTT 4496

Qy       2406 GCTGCAAAAGGCATGATTTCATTCTTTTTTATGGCTATGTTCTTTCCCAATTTAGATAAA 2465
              |||||||| | |||||||||||||| || ||||||||    |  | | |||      ||| |
Db       4497 GCTGCAAAGGACATGATTTCATTCCTTCTTATGGCTGCATAATATTCCATGGGGTATATA 4556

Qy       2466 GA 2467
              |
Db       4557 TA 4558
```

Search completed: April 21, 2005, 00:49:04
Job time : 2020.96 secs

This Page Blank (uspto)

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 20, 2005, 22:49:10 ; Search time 13904.5 Seconds
(without alignments)
17319.707 Million cell updates/sec

Title: US-09-939-209A-3_COPY_1_4970
Perfect score: 4970
Sequence: 1 agttcaagaccagcctgagc.....gtggctgaatgtctgccccc 4970

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	4970	100.0	20300	6	AX451337	AX451337 Sequence
	2	4954.8	99.7	165329	9	AL583850	AL583850 Human DNA
c	3	4559.4	91.7	191699	2	AC031977	AC031977 Homo sapi
	4	411.2	8.3	203216	2	AC097328	AC097328 Pan trogl
	5	405.4	8.2	104321	2	CR759943	CR759943 Homo sapi
c	6	405.2	8.2	181946	2	AC097330	AC097330 Pan trogl
	7	402.2	8.1	109081	9	AL358074	AL358074 Human DNA
	8	396.8	8.0	164844	9	AC009058	AC009058 Homo sapi
	9	395.2	8.0	130349	9	AC011593	AC011593 Homo sapi
c	10	394.8	7.9	196413	2	AC097269	AC097269 Pan trogl
	11	390.8	7.9	189662	9	AC015468	AC015468 Homo sapi
c	12	387.4	7.8	125685	9	AC073310	AC073310 Homo sapi
	13	385.8	7.8	139118	9	AL133477	AL133477 Human DNA
	14	382.6	7.7	216021	9	HUAC004787	AC004787 Homo sapi
	15	379.8	7.6	92564	9	AY007685	AY007685 Homo sapi
c	16	379.4	7.6	161252	2	AC114955	AC114955 Homo sapi
c	17	378	7.6	202305	9	AC114291	AC114291 Homo sapi
c	18	373.6	7.5	192139	9	AC021955	AC021955 Homo sapi
c	19	372	7.5	99596	9	AC106873	AC106873 Homo sapi
	20	371.4	7.5	38096	9	BX901949	BX901949 Human DNA
c	21	371.4	7.5	110000	2	BX324168_0	BX324168 Homo sapi
	22	370.4	7.5	252512	2	AC097326	AC097326 Pan trogl
c	23	368.2	7.4	170646	2	AC123545	AC123545 Pan trogl
	24	367	7.4	112762	9	HS323B6	Z83841 Human DNA s
c	25	365.4	7.4	163976	9	AC108483	AC108483 Homo sapi
	26	365.4	7.4	163998	9	AC022166	AC022166 Homo sapi
c	27	364.2	7.3	153201	2	AC092863	AC092863 Homo sapi
c	28	363.6	7.3	151549	9	AC092694	AC092694 Homo sapi
	29	357.4	7.2	217985	9	AC021193	AC021193 Homo sapi
c	30	357.2	7.2	151424	2	CR812471	CR812471 Danio rer
c	31	356.6	7.2	121600	9	HS141H5	AL049176 Human DNA
c	32	356.6	7.2	127172	2	CR788239	CR788239 Danio rer
	33	355.4	7.2	282240	2	CR753865	CR753865 Danio rer
c	34	355.2	7.1	149969	2	AC024180	AC024180 Homo sapi
c	35	355.2	7.1	151047	2	CR759776	CR759776 Danio rer
c	36	355.2	7.1	162996	9	AC006441	AC006441 Homo sapi
c	37	354.8	7.1	182430	9	AL590381	AL590381 Human DNA
c	38	354	7.1	252450	2	CR753861	CR753861 Danio rer
	39	353.6	7.1	163651	2	CR790381	CR790381 Danio rer
c	40	353.6	7.1	164994	2	AC141288	AC141288 Homo sapi
c	41	353	7.1	165979	2	CR376741	CR376741 Danio rer
c	42	353	7.1	248167	2	CR627491	CR627491 Danio rer
	43	352.4	7.1	178254	9	AL731541	AL731541 Human DNA
	44	352.4	7.1	204486	2	CR752655	CR752655 Danio rer
c	45	352.2	7.1	178427	2	CR392006	CR392006 Danio rer

ALIGNMENTS

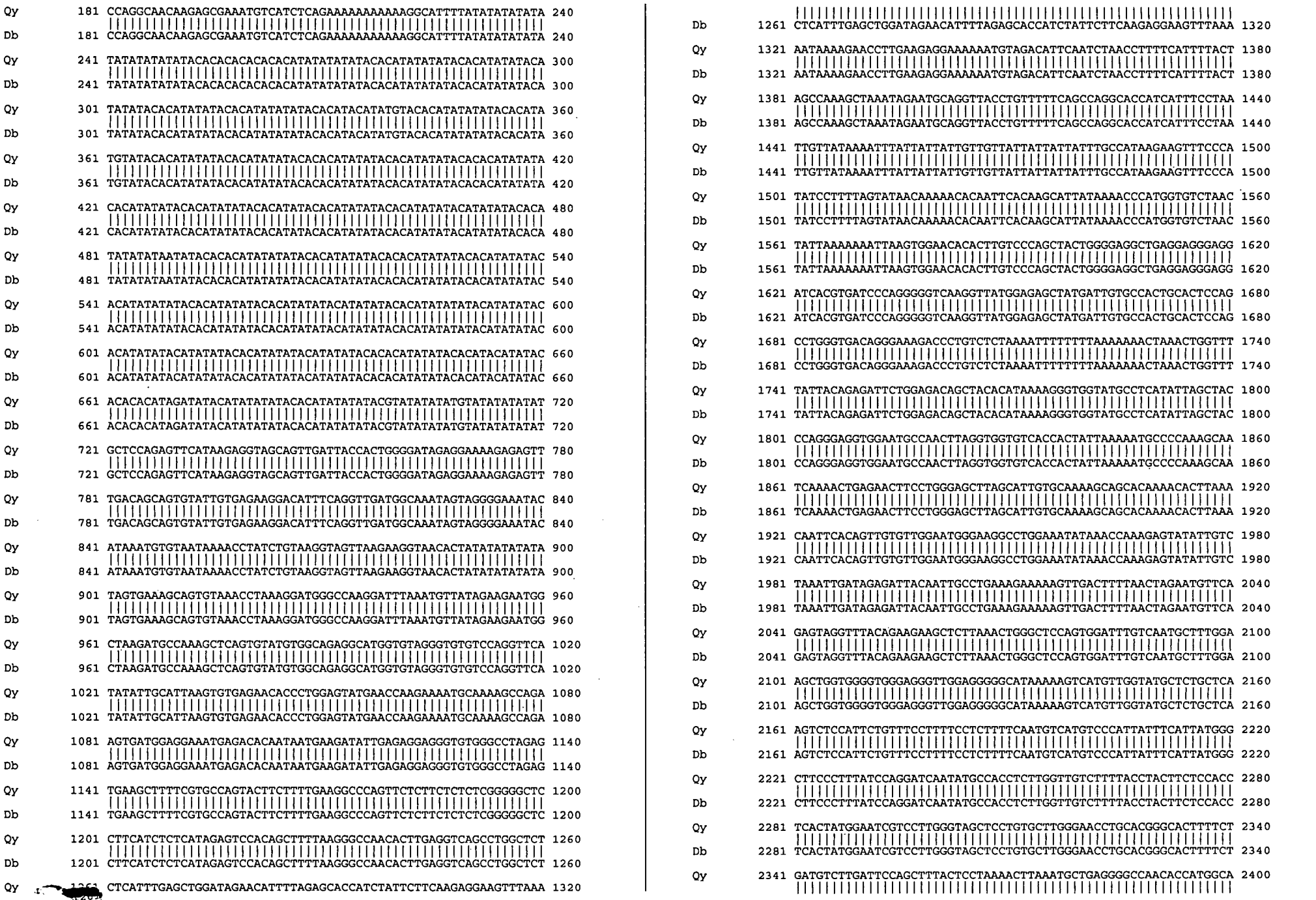
RESULT 1
AX451337
LOCUS AX451337 20300 bp DNA linear PAT 03-JUL-2002
DEFINITION Sequence 3 from Patent WO0216653.
ACCESSION AX451337
VERSION AX451337.1 GI:21698388
KEYWORDS .
SOURCE synthetic construct
ORGANISM synthetic construct
other sequences; artificial sequences.
REFERENCE 1
AUTHORS Levitt,P.R., Mirnics,K., Kodavali,V.C. and Nimgaonkar,V.L.
TITLE Methods and systems for facilitating the diagnosis and treatment of schizophrenia
JOURNAL Patent: WO 0216653-A 3 28-FEB-2002;
University of Pittsburgh (US)
FEATURES Location/Qualifiers
source 1. .20300
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="A genomic sequence containing RGS4 nucleic acid sequence and sequ ences upstream and downstream to the RGS4 nucleic acid sequence"
ORIGIN

Query Match 100.0%; Score 4970; DB 6; Length 20300;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4970; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAATTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAATTA 60

Qy         61 GACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAAT 120

Qy        121 CACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAGC 180
```


Qy 181 CCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATA 240
Db 181 CCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATA 240
Qy 241 TATATATATATACACACACACACACATATATATATACACATATATATACACATATATACA 300
Db 241 TATATATATATACACACACACACACATATATATATACACATATATATACACATATATACA 300
Qy 301 TATATACACATATATACACATATATATACACATACATATGTACACATATATATACACATA 360
Db 301 TATATACACATATATACACATATATATACACATACATATGTACACATATATATACACATA 360
Qy 361 TGTATACACATATATACACATATATACACACATATATACACATATATACACACATATATA 420
Db 361 TGTATACACATATATACACATATATACACACATATATACACATATATACACACATATATA 420
Qy 421 CACATATATACACATATATACACATATACACATATATACACATATATACATATATACACA 480
Db 421 CACATATATACACATATATACACATATACACATATATACACATATATACATATATACACA 480
Qy 481 TATATATAATATACACACATATATATACACATATATACACACATATATACACATATATAC 540
Db 481 TATATATAATATACACACATATATATACACATATATACACACATATATACACATATATAC 540
Qy 541 ACATATATATACACATATATACACATATATACATATATACACATATATATACATATATAC 600
Db 541 ACATATATATACACATATATACACATATATACATATATACACATATATATACATATATAC 600
Qy 601 ACATATATACATATATACACATATATACATATATACACACATATATACACATACATATAC 660
Db 601 ACATATATACATATATACACATATATACATATATACACACATATATACACATACATATAC 660
Qy 661 ACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATAT 720
Db 661 ACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATAT 720
Qy 721 GCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTT 780
Db 721 GCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTT 780
Qy 781 TGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATAC 840
Db 781 TGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATAC 840
Qy 841 ATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATATATATA 900
Db 841 ATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATATATATA 900
Qy 901 TAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTATAGAAGAATGG 960
Db 901 TAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTATAGAAGAATGG 960
Qy 961 CTAAGATGCCAAAGCTCAGTGTATGTGGCAGAGGCATGGTGTAGGGTGTGTCCAGGTTCA 1020
Db 961 CTAAGATGCCAAAGCTCAGTGTATGTGGCAGAGGCATGGTGTAGGGTGTGTCCAGGTTCA 1020
Qy 1021 TATATTGCATTAAGTGTGAGAACACCCTGGAGTATGAACCAAGAAAATGCAAAAGCCAGA 1080
Db 1021 TATATTGCATTAAGTGTGAGAACACCCTGGAGTATGAACCAAGAAAATGCAAAAGCCAGA 1080
Qy 1081 AGTGATGGAGGAAATGAGACACAATAATGAAGATATTGAGAGGAGGGTGTGGGCCTAGAG 1140
Db 1081 AGTGATGGAGGAAATGAGACACAATAATGAAGATATTGAGAGGAGGGTGTGGGCCTAGAG 1140
Qy 1141 TGAAGCTTTTCGTGCCAGTACTTCTTTTGAAGGCCCAGTTCTCTTCTCTCTCGGGGGCTC 1200
Db 1141 TGAAGCTTTTCGTGCCAGTACTTCTTTTGAAGGCCCAGTTCTCTTCTCTCTCGGGGGCTC 1200
Qy 1201 CTTCATCTCTCATAGAGTCCACAGCTTTTAAGGGCCAACACTTGAGGTCAGCCTGGCTCT 1260
Db 1201 CTTCATCTCTCATAGAGTCCACAGCTTTTAAGGGCCAACACTTGAGGTCAGCCTGGCTCT 1260
Qy 1261 CTCATTTGAGCTGGATAGAACATTTTAGAGCACCATCTATTCTTCAAGAGGAAGTTTAAA 1320
Db 1261 CTCATTTGAGCTGGATAGAACATTTTAGAGCACCATCTATTCTTCAAGAGGAAGTTTAAA 1320
Qy 1321 AATAAAAGAACCTTGAAGAGGAAAAAATGTAGACATTCAATCTAACCTTTTCATTTTACT 1380
Db 1321 AATAAAAGAACCTTGAAGAGGAAAAAATGTAGACATTCAATCTAACCTTTTCATTTTACT 1380
Qy 1381 AGCCAAAGCTAAATAGAATGCAGGTTACCTGTTTTTCAGCCAGGCACCATCATTTCCTAA 1440
Db 1381 AGCCAAAGCTAAATAGAATGCAGGTTACCTGTTTTTCAGCCAGGCACCATCATTTCCTAA 1440
Qy 1441 TTGTTATAAAATTTATTATTATTGTTGTTATTATTATTATTTGCCATAAGAAGTTTCCCA 1500
Db 1441 TTGTTATAAAATTTATTATTATTGTTGTTATTATTATTATTTGCCATAAGAAGTTTCCCA 1500
Qy 1501 TATCCTTTTAGTATAACAAAAACACAATTCACAAGCATTATAAAACCCATGGTGTCTAAC 1560
Db 1501 TATCCTTTTAGTATAACAAAAACACAATTCACAAGCATTATAAAACCCATGGTGTCTAAC 1560
Qy 1561 TATTAAAAAAAATTAAGTGGAACACACTTGTCCCAGCTACTGGGGAGGCTGAGGAGGGAGG 1620
Db 1561 TATTAAAAAAAATTAAGTGGAACACACTTGTCCCAGCTACTGGGGAGGCTGAGGAGGGAGG 1620
Qy 1621 ATCACGTGATCCCAGGGGGTCAAGGTTATGGAGAGCTATGATTGTGCCACTGCACTCCAG 1680
Db 1621 ATCACGTGATCCCAGGGGGTCAAGGTTATGGAGAGCTATGATTGTGCCACTGCACTCCAG 1680
Qy 1681 CCTGGGTGACAGGGAAAGACCCTGTCTCTAAAATTTTTTTTAAAAAAACTAAACTGGTTT 1740
Db 1681 CCTGGGTGACAGGGAAAGACCCTGTCTCTAAAATTTTTTTTAAAAAAACTAAACTGGTTT 1740
Qy 1741 TATTACAGAGATTCTGGAGACAGCTACACATAAAAGGGTGGTATGCCTCATATTAGCTAC 1800
Db 1741 TATTACAGAGATTCTGGAGACAGCTACACATAAAAGGGTGGTATGCCTCATATTAGCTAC 1800
Qy 1801 CCAGGGAGGTGGAATGCCAACTTAGGTGGTGTCACCACTATTAAAAATGCCCCAAAGCAA 1860
Db 1801 CCAGGGAGGTGGAATGCCAACTTAGGTGGTGTCACCACTATTAAAAATGCCCCAAAGCAA 1860
Qy 1861 TCAAAACTGAGAACTTCCTGGGAGCTTAGCATTGTGCAAAAGCAGCACAAAACACTTAAA 1920
Db 1861 TCAAAACTGAGAACTTCCTGGGAGCTTAGCATTGTGCAAAAGCAGCACAAAACACTTAAA 1920
Qy 1921 CAATTCACAGTTGTGTTGGAATGGGAAGGCCTGGAAATATAAACCAAAGAGTATATTGTC 1980
Db 1921 CAATTCACAGTTGTGTTGGAATGGGAAGGCCTGGAAATATAAACCAAAGAGTATATTGTC 1980
Qy 1981 TAAATTGATAGAGATTACAATTGCCTGAAAGAAAAAGTTGACTTTTAACTAGAATGTTCA 2040
Db 1981 TAAATTGATAGAGATTACAATTGCCTGAAAGAAAAAGTTGACTTTTAACTAGAATGTTCA 2040
Qy 2041 GAGTAGGTTTACAGAAGAAGCTCTTAAACTGGGCTCCAGTGGATTTGTCAATGCTTTGGA 2100
Db 2041 GAGTAGGTTTACAGAAGAAGCTCTTAAACTGGGCTCCAGTGGATTTGTCAATGCTTTGGA 2100
Qy 2101 AGCTGGTGGGGTGGGAGGGTTGGAGGGGGCATAAAAAGTCATGTTGGTATGCTCTGCTCA 2160
Db 2101 AGCTGGTGGGGTGGGAGGGTTGGAGGGGGCATAAAAAGTCATGTTGGTATGCTCTGCTCA 2160
Qy 2161 AGTCTCCATTCTGTTTCCTTTTCCTCTTTTCAATGTCATGTCCCATTATTTCATTATGGG 2220
Db 2161 AGTCTCCATTCTGTTTCCTTTTCCTCTTTTCAATGTCATGTCCCATTATTTCATTATGGG 2220
Qy 2221 CTTCCCTTTATCCAGGATCAATATGCCACCTCTTGGTTGTCTTTTACCTACTTCTCCACC 2280
Db 2221 CTTCCCTTTATCCAGGATCAATATGCCACCTCTTGGTTGTCTTTTACCTACTTCTCCACC 2280
Qy 2281 TCACTATGGAATCGTCCTTGGGTAGCTCCTGTGCTTGGGAACCTGCACGGGCACTTTTCT 2340
Db 2281 TCACTATGGAATCGTCCTTGGGTAGCTCCTGTGCTTGGGAACCTGCACGGGCACTTTTCT 2340
Qy 2341 GATGTCTTGATTCCAGCTTTACTCCTAAAACTTAAATGCTGAGGGGCCAACACCATGGCA 2400

```
Db      2341 GATGTCTTGATTCCAGCTTTACTCCTAAAACTTAAATGCTGAGGGGCCAACACCATGGCA 2400
Qy      2401 GTGGTAGGGATGGGAATGGGGGTCTTGTAACACACTACATAAACTACACGAAATAAACTA 2460
Db      2401 GTGGTAGGGATGGGAATGGGGGTCTTGTAACACACTACATAAACTACACGAAATAAACTA 2460
Qy      2461 CATGAAACTCAACATGTTTGCAAGACTCAGTTCACATCCATGAGGAGCTCATGCTTCTCC 2520
Db      2461 CATGAAACTCAACATGTTTGCAAGACTCAGTTCACATCCATGAGGAGCTCATGCTTCTCC 2520
Qy      2521 CTCCTGCTCCCCTAGCACACATGATTATCTCTATTTGGAAATGTTTGGCATTTTTGGTGA 2580
Db      2521 CTCCTGCTCCCCTAGCACACATGATTATCTCTATTTGGAAATGTTTGGCATTTTTGGTGA 2580
Qy      2581 AGTGAATGGTTCAATAACTTTCTCCACCATCAGAACAAAAGCTCTTTAAGGTTAGGGATG 2640
Db      2581 AGTGAATGGTTCAATAACTTTCTCCACCATCAGAACAAAAGCTCTTTAAGGTTAGGGATG 2640
Qy      2641 GGATCATACACACTTCCCTTGTCCAAGTCCCCATCACCCCTTATCTAGACAATTGCTACA 2700
Db      2641 GGATCATACACACTTCCCTTGTCCAAGTCCCCATCACCCCTTATCTAGACAATTGCTACA 2700
Qy      2701 GTTTCCTACACACTCTTCTAACCTCTTGCAGTCTATTTTCATAAAACAGCTAGAGAACTT 2760
Db      2701 GTTTCCTACACACTCTTCTAACCTCTTGCAGTCTATTTTCATAAAACAGCTAGAGAACTT 2760
Qy      2761 TGAGATGTAAGTCAAAAAATAGAACATGTCGCTCTTTCCCATTGTTTTTGAAATAAAGTT 2820
Db      2761 TGAGATGTAAGTCAAAAAATAGAACATGTCGCTCTTTCCCATTGTTTTTGAAATAAAGTT 2820
Qy      2821 CAACCCCCTTACCAGGGTCAACAAGGCCCTGCAATGATTTGGTCCTGTTAAAAATTCTTT 2880
Db      2821 CAACCCCCTTACCAGGGTCAACAAGGCCCTGCAATGATTTGGTCCTGTTAAAAATTCTTT 2880
Qy      2881 AGCCTTAACTCATGCTGTTCTTCCTTACACTCACTGCATTCTAGCCATTGAGGTTTCTAT 2940
Db      2881 AGCCTTAACTCATGCTGTTCTTCCTTACACTCACTGCATTCTAGCCATTGAGGTTTCTAT 2940
Qy      2941 GCATCAAACTTTTTTTGGTCCCAGCACTGTGCACATCCTTCTGGGTAGAATGCCCCTTGA 3000
Db      2941 GCATCAAACTTTTTTTGGTCCCAGCACTGTGCACATCCTTCTGGGTAGAATGCCCCTTGA 3000
Qy      3001 TTTGTATAATTAGCACCTCCTTCATCATTTAGGTCTTAGTATAACTACTACCTTCTTAGA 3060
Db      3001 TTTGTATAATTAGCACCTCCTTCATCATTTAGGTCTTAGTATAACTACTACCTTCTTAGA 3060
Qy      3061 GAAGCTCTGCTTCTTCATCCTATAAAAAAGTAAAATTCCTTACCCTGTTATTTTTTAAGT 3120
Db      3061 GAAGCTCTGCTTCTTCATCCTATAAAAAAGTAAAATTCCTTACCCTGTTATTTTTTAAGT 3120
Qy      3121 CATCCGTGTTTCATTCTGTTAAAGTTCTTATCACAATTTATCATTATTTTATTTACAGTC 3180
Db      3121 CATCCGTGTTTCATTCTGTTAAAGTTCTTATCACAATTTATCATTATTTTATTTACAGTC 3180
Qy      3181 ATGTGCCACATAACAATGTTTCAGTCAGGGATAGAACACAAATGTATCTGGCCCCATAAT 3240
Db      3181 ATGTGCCACATAACAATGTTTCAGTCAGGGATAGAACACAAATGTATCTGGCCCCATAAT 3240
Qy      3241 ATTATAAGCTGAGAAATTTCTATTAACTAGTGATATCGCAGCCATCATAAGTGTAATGCA 3300
Db      3241 ATTATAAGCTGAGAAATTTCTATTAACTAGTGATATCGCAGCCATCATAAGTGTAATGCA 3300
Qy      3301 GGACATTACCTTTTCTATGTTTAGATATGTTAGATACACAAATATATTTCATTGTGTTAT 3360
Db      3301 GGACATTACCTTTTCTATGTTTAGATATGTTAGATACACAAATATATTTCATTGTGTTAT 3360
Qy      3361 AATTTCCTACAGTATTCAGTACAGTAACATGCTGTACAGGTTTGTAACCTAGGAGTAATA 3420
Db      3361 AATTTCCTACAGTATTCAGTACAGTAACATGCTGTACAGGTTTGTAACCTAGGAGTAATA 3420
Qy      3421 GGCTATACCATACAGCTTAGGTGTGTAGTAGGCTATAACCATCTAGGTTTGTGTAAGTAC 3480
Db      3421 GGCTATACCATACAGCTTAGGTGTGTAGTAGGCTATAACCATCTAGGTTTGTGTAAGTAC 3480
Qy      3481 ATTCTATGATATTCCCACAATGATGAAATCACCTAACTACACATTTCTCAGAATGTTTCA 3540
Db      3481 ATTCTATGATATTCCCACAATGATGAAATCACCTAACTACACATTTCTCAGAATGTTTCA 3540
Qy      3541 CTGTTGTGAAGTGACCCATGACTATATTTTCCTATATACTTGATATTTTTGTGCATCTGC 3600
Db      3541 CTGTTGTGAAGTGACCCATGACTATATTTTCCTATATACTTGATATTTTTGTGCATCTGC 3600
Qy      3601 CCATGAGAATGTAGTGTAAGATCAAAGGATGCAAGAATGGGTTCTATCCAGTATAGTACC 3660
Db      3601 CCATGAGAATGTAGTGTAAGATCAAAGGATGCAAGAATGGGTTCTATCCAGTATAGTACC 3660
Qy      3661 CACTACACTGGTGGATGTCAATATGTATTTGTTAGATTAATATCTCAAGAATGAGCACCT 3720
Db      3661 CACTACACTGGTGGATGTCAATATGTATTTGTTAGATTAATATCTCAAGAATGAGCACCT 3720
Qy      3721 TTCTCAGACACATAAAAGATGCTCAATATAAAAGTTTGTTGAACTGAACGTTATTGGCAA 3780
Db      3721 TTCTCAGACACATAAAAGATGCTCAATATAAAAGTTTGTTGAACTGAACGTTATTGGCAA 3780
Qy      3781 ATGTAACATGATCGGATTTAAAGAGGAGCGAAACAGAGGTCTGGCTCAAACACCATACTT 3840
Db      3781 ATGTAACATGATCGGATTTAAAGAGGAGCGAAACAGAGGTCTGGCTCAAACACCATACTT 3840
Qy      3841 CTAGAGTGCATAAGAGGTAGCAGTTGATTACCACTGGCGACAGGAGAAAAAAGAGCTTGA 3900
Db      3841 CTAGAGTGCATAAGAGGTAGCAGTTGATTACCACTGGCGACAGGAGAAAAAAGAGCTTGA 3900
Qy      3901 CCGCAGGGTACTGTGAAGACATTTCAGGTTGATGGCACAGAACAGGGGAAATACATAAAT 3960
Db      3901 CCGCAGGGTACTGTGAAGACATTTCAGGTTGATGGCACAGAACAGGGGAAATACATAAAT 3960
Qy      3961 GTGTGGGAATATTCAGTGGTCTGGGATGACTACATAGTAGAATATAATGAAGAAAAGAGT 4020
Db      3961 GTGTGGGAATATTCAGTGGTCTGGGATGACTACATAGTAGAATATAATGAAGAAAAGAGT 4020
Qy      4021 GGAAGGGAAAGATGAAAAGTTGGAATGGGGATGAATTATGAAAGTACCAGAATGTTATGC 4080
Db      4021 GGAAGGGAAAGATGAAAAGTTGGAATGGGGATGAATTATGAAAGTACCAGAATGTTATGC 4080
Qy      4081 TAAGGAATCTAGATTTTAAAATGTGAGGGCAAATTGAAGTCCTGGGCACGTTACAAAACT 4140
Db      4081 TAAGGAATCTAGATTTTAAAATGTGAGGGCAAATTGAAGTCCTGGGCACGTTACAAAACT 4140
Qy      4141 AGAGGTCATAAAGTTTACCCTAATTTACCAAGATTTCCTAGAGGATCTATAATTGGAATC 4200
Db      4141 AGAGGTCATAAAGTTTACCCTAATTTACCAAGATTTCCTAGAGGATCTATAATTGGAATC 4200
Qy      4201 CAGATCTGCCTCTCTGTAAAGTTCAAGCACTTTCCATGACACCATACTGTTTCTTTCCAC 4260
Db      4201 CAGATCTGCCTCTCTGTAAAGTTCAAGCACTTTCCATGACACCATACTGTTTCTTTCCAC 4260
Qy      4261 CTGCACAATGCAAATGAACTCTTATGAAACTGCTGTTTCTATCCTGGGCTAAATGTTGCA 4320
Db      4261 CTGCACAATGCAAATGAACTCTTATGAAACTGCTGTTTCTATCCTGGGCTAAATGTTGCA 4320
Qy      4321 GAAAAAAGATTTAATCTTTGGGATAAGGCTATTTTGGGTTTTCTCCTACTTCTTGGGAAA 4380
Db      4321 GAAAAAAGATTTAATCTTTGGGATAAGGCTATTTTGGGTTTTCTCCTACTTCTTGGGAAA 4380
Qy      4381 CAAGGTTTTCTTCCCCTGGCTAATTAAGTGTGGTATTGTTCTTCCAGGGAAATCAGTGAT 4440
Db      4381 CAAGGTTTTCTTCCCCTGGCTAATTAAGTGTGGTATTGTTCTTCCAGGGAAATCAGTGAT 4440
Qy      4441 GCATCACCTGCTGCTATCAAATGTCAGGGTTGGAGTTCCTGATTTATTGCATGTGCCCAC 4500
Db      4441 GCATCACCTGCTGCTATCAAATGTCAGGGTTGGAGTTCCTGATTTATTGCATGTGCCCAC 4500
Qy      4501 AAAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCT 4560
Db      4501 AAAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCT 4560
```

```
Qy       4561 GTTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAA 4620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4561 GTTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAA 4620

Qy       4621 ATCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGG 4680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4621 ATCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGG 4680

Qy       4681 AGGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCT 4740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4681 AGGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCT 4740

Qy       4741 GAGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAACATCTGGTAGA 4800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4741 GAGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAACATCTGGTAGA 4800

Qy       4801 ACTAAAGAACGCAACGCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCC 4860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4801 ACTAAAGAACGCAACGCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCC 4860

Qy       4861 TTGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCT 4920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4861 TTGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCT 4920

Qy       4921 AGAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCC 4970
             |||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4921 AGAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCC 4970
```

RESULT 2
AL583850

```
LOCUS       AL583850              165329 bp    DNA     linear   PRI 15-NOV-2001
DEFINITION  Human DNA sequence from clone RP11-430G6 on chromosome 1, complete
            sequence.
ACCESSION   AL583850
VERSION     AL583850.5  GI:16973044
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 165329)
  AUTHORS   Tracey,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Nov 16, 2001 this sequence version replaced gi:15020514.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            RP11-430G6 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-430G6 It may be shorter because we sequence overlapping
            sections only once, except for a short overlap.
            The true right end of clone RP11-430G6 is at 165329 in this
            sequence. The true right end of clone RP11-331H2 is at 2000 in this
            sequence.
FEATURES             Location/Qualifiers
     source          1. .165329
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-430G6"
                     /clone_lib="RPCI-11.2"
ORIGIN
```

Query Match 99.7%; Score 4954.8; DB 9; Length 165329;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 4967; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

```
Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAATTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71683 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAATTA 71742

Qy         61 GACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71743 GACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAAT 71802

Qy        121 CACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71803 CACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAGC 71862

Qy        181 CCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71863 CCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATA 71922

Qy        241 TATATATATATACACACACACACATATATATATACACATATATATACACATATATACA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71923 TATATATATATACACACACACACATATATATATACACATATATATACACATATATACA 71982

Qy        301 TATATACACATATATACACATATATATACACATACATATGTACACATATATATACACATA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71983 TATATACACATATATACACATATATATACACATACATATGTACACATATATATACACATA 72042

Qy        361 TGTATACACATATATACACATATATACACACATATATACACATATATACACACATATATA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72043 TGTATACACATATATACACATATATACACACATATATACACATATATACACACATATATA 72102

Qy        421 CACATATATACACATATATACACATATACACATATATACACATATATACATATATACACA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72103 CACATATATACACATATATACACATATACACATATATACACATATATACATATATACACA 72162

Qy        481 TATATATAATATACACACATATATATACACATATATACACACATATATACACATATATAC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72163 TATATATAATATACACACATATATATACACATATATACACACATATATACACATATATAC 72222

Qy        541 ACATATATATACACATATATACACATATATACATATATACACATATATATACATATATAC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72223 ACATATATATACACATATATACACATATATACATATATACACATATATATACATATATAC 72282

Qy        601 ACATATATACATATATACACATATATACATATATACACACATATATACACATACATATAC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72283 ACATATATACATATATACACATATATACATATATACACACATATATACACATACATATAC 72342

Qy        661 ACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72343 ACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATAT 72402

Qy        721 GCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72403 GCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTT 72462
```

```
Qy       781 TGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATAC 840
Db     72463 TGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATAC 72522

Qy       841 ATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATATATATA 900
Db     72523 ATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATATATATA 72582

Qy       901 TAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTATAGAAGAATGG 960
Db     72583 TAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTATAGAAGAATGG 72642

Qy       961 CTAAGATGCCAAAGCTCAGTGTATGTGGCAGAGGCATGGTGTAGGGTGTGTCCAGGTTCA 1020
Db     72643 CTAAGATGCCAAAGCTCAGTGTATGTGGCAGAGGCATGGTGGAGGGTGTGTCCAGGTTCA 72702

Qy      1021 TATATTGCATTAAGTGTGAGAACACCCTGGAGTATGAACCAAGAAAATGCAAAAGCCAGA 1080
Db     72703 TATATTGCATTAAGTGTGAGAA-ACCCTGGAGTATGAACCAAGAAAATGCAAAAGCCAGA 72761

Qy      1081 AGTGATGGAGGAAATGAGACACAATAATGAAGATATTGAGAGGAGGGTGTGGGCCTAGAG 1140
Db     72762 AGTGATGGAGGAAATGAGACACAATAATGAAGATATTGAGAGGAGGGTGTGGGCCTAGAG 72821

Qy      1141 TGAAGCTTTTCGTGCCAGTACTTCTTTTGAAGGCCCAGTTCTCTTCTCTCTCGGGGGCTC 1200
Db     72822 TGAAGCTTTTCGTGCCAGTACTTCTTTTGAAGGCCCAGTTCTCTTCTCTCTCGGGGGCTC 72881

Qy      1201 CTTCATCTCTCATAGAGTCCACAGCTTTTAAGGGCCAACACTTGAGGTCAGCCTGGCTCT 1260
Db     72882 CTTCATCTCTCATAGAGTCCACAGCTTTTAAGGGCCAACACTTGAGGTCAGCCTGGCTCT 72941

Qy      1261 CTCATTTGAGCTGGATAGAACATTTTAGAGCACCATCTATTCTTCAAGAGGAAGTTTAAA 1320
Db     72942 CTCATTTGAGCTGGATAGAACATTTTAGAGCACCATCTATTCTTCAAGAGGAAGTTTAAA 73001

Qy      1321 AATAAAAGAACCTTGAAGAGGAAAAAATGTAGACATTCAATCTAACCTTTTCATTTTACT 1380
Db     73002 AATAAAAGAACCTTGAAGAGGAAAAAATGTAGACATTCAATCTAACCTTTTCATTTTACT 73061

Qy      1381 AGCCAAAGCTAAATAGAATGCAGGTTACCTGTTTTTCAGCCAGGCACCATCATTTCCTAA 1440
Db     73062 AGCCAAAGCTAAATAGAATGCAGGTTACCTGTTTTTCAGCCAGGCACCATCATTTCCTAA 73121

Qy      1441 TTGTTATAAAATTTATTATTATTGTTGTTATTATTATTATTTGCCATAAGAAGTTTCCCA 1500
Db     73122 TTGTTATAAAATTTATTATTATTGTTGTTATTATTATTATTTGCCATAAGAAGTTTCCCA 73181

Qy      1501 TATCCTTTTAGTATAACAAAAACACAATTCACAAGCATTATAAAACCCATGGTGTCTAAC 1560
Db     73182 TATCCTTTTAGTATAACAAAAACACAATTCACAAGCATTATAAAACCCATGGTGTCTAAC 73241

Qy      1561 TATTAAAAAAAATTAAGTGGAACACACTTGTCCCAGCTACTGGGGAGGCTGAGGAGGGAGG 1620
Db     73242 TATTAAAAAAAATTAAGTGGAACACACTTGTCCCAGCTACTGGGGAGGCTGAGGAGGGAGG 73301

Qy      1621 ATCACGTGATCCCAGGGGGTCAAGGTTATGGAGAGCTATGATTGTGCCACTGCACTCCAG 1680
Db     73302 ATCACGTGATCCCAGGGGGTCAAGGTTATGGAGAGCTATGATTGTGCCACTGCACTCCAG 73361

Qy      1681 CCTGGGTGACAGGGAAAGACCCTGTCTCTAAAATTTTTTTTAAAAAAACTAAACTGGTTT 1740
Db     73362 CCTGGGTGACAGGGAAAGACCCTGTCTCTAAAATTTTTTTTAAAAAAACTAAACTGGTTT 73421

Qy      1741 TATTACAGAGATTCTGGAGACAGCTACACATAAAAGGGTGGTATGCCTCATATTAGCTAC 1800
Db     73422 TATTACAGAGATTCTGGAGACAGCTACACATAAAAGGGTGGTATGCCTCATATTAGCTAC 73481

Qy      1801 CCAGGGAGGTGGAATGCCAACTTAGGTGGTGTCACCACTATTAAAAATGCCCCAAAGCAA 1860
Db     73482 CCAGGGAGGTGGAATGCCAACTTAGGTGGTGTCACCACTATTAAAAATGCCCCAAAGCAA 73541

Qy      1861 TCAAAACTGAGAACTTCCTGGGAGCTTAGCATTGTGCAAAAGCAGCACAAAACACTTAAA 1920
Db     73542 TCAAAACTGAGAACTTCCTGGGAGCTTAGCATTGTGCAAAAGCAGCACAAAACACTTAAA 73601

Qy      1921 CAATTCACAGTTGTGTTGGAATGGGAAGGCCTGGAAATATAAACCAAAGAGTATATTGTC 1980
Db     73602 CAATTCACAGTTGTGTTGGAATGGGAAGGCCTGGAAATATAAACCAAAGAGTATATTGTC 73661

Qy      1981 TAAATTGATAGAGATTACAATTGCCTGAAAGAAAAAGTTGACTTTTAACTAGAATGTTCA 2040
Db     73662 TAAATTGATAGAGATTACAATTGCCTGAAAGAAAAAGTTGACTTTTAACTAGAATGTTCA 73721

Qy      2041 GAGTAGGTTTACAGAAGAAGCTCTTAAACTGGGCTCCAGTGGATTTGTCAATGCTTTGGA 2100
Db     73722 GAGTAGGTTTACAGAAGAAGCTCTTAAACTGGGCTCCAGTGGATTTGTCAATGCTTTGGA 73781

Qy      2101 AGCTGGTGGGGTGGGAGGGTTGGAGGGGGCATAAAAAGTCATGTTGGTATGCTCTGCTCA 2160
Db     73782 AGCTGGTGGGGTGGGAGGGTTGGAGGGGGCATAAAAAGTCATGTTGGTATGCTCTGCTCA 73841

Qy      2161 AGTCTCCATTCTGTTTCCTTTTCCTCTTTTCAATGTCATGTCCCATTATTTCATTATGGG 2220
Db     73842 AGTCTCCATTCTGTTTCCTTTTCCTCTTTTCAATGTCATGTCCCATTATTTCATTATGGG 73901

Qy      2221 CTTCCCTTTATCCAGGATCAATATGCCACCTCTTGGTTGTCTTTTACCTACTTCTCCACC 2280
Db     73902 CTTCCCTTTATCCAGGATCAATATGCCACCTCTTGGTTGTCTTTTACCTACTTCTCCACC 73961

Qy      2281 TCACTATGGAATCGTCCTTGGGTAGCTCCTGTGCTTGGGAACCTGCACGGGCACTTTTCT 2340
Db     73962 TCACTATGGAATCGTCCTTGGGTAGCTCCTGTGCTTGGGAACCTGCACGGGCACTTTTCT 74021

Qy      2341 GATGTCTTGATTCCAGCTTTACTCCTAAAACTTAAATGCTGAGGGGCCAACACCATGGCA 2400
Db     74022 GATGTCTTGATTCCAGCTTTACTCCTAAAACTTAAATGCTGAGGGGCCAACACCATGGCA 74081

Qy      2401 GTGGTAGGGATGGGAATGGGGGTCTTGTAACACACTACATAAACTACACGAAATAAACTA 2460
Db     74082 GTGGTAGGGATGGGAATGGGGGTCTTGTAACACACTACATAAACTACACGAAATAAACTA 74141

Qy      2461 CATGAAACTCAACATGTTTGCAAGACTCAGTTCACATCCATGAGGAGCTCATGCTTCTCC 2520
Db     74142 CATGAAACTCAACATGTTTGCAAGACTCAGTTCACATCCATGAGGAGCTCATGCTTCTCC 74201

Qy      2521 CTCCTGCTCCCCTAGCACACATGATTATCTCTATTTGGAAATGTTTGGCATTTTTGGTGA 2580
Db     74202 CTCCTGCTCCCCTAGCACACATGATTATCTCTATTTGGAAATGTTTGGCATTTTTGGTGA 74261

Qy      2581 AGTGAATGGTTCAATAACTTTCTCCACCATCAGAACAAAAGCTCTTTAAGGTTAGGGATG 2640
Db     74262 AGTGAATGGTTCAATAACTTTCTCCACCATCAGAACAAAAGCTCTTTAAGGTTAGGGATG 74321

Qy      2641 GGATCATACACACTTCCCTTGTCCAAGTCCCCATCACCCCTTATCTAGACAATTGCTACA 2700
Db     74322 GGATCATACACACTTCCCTTGTCCAAGTCCCCATCACCCCTTATCTAGACAATTGCTACA 74381

Qy      2701 GTTTCCTACACACTCTTCTAACCTCTTGCAGTCTATTTTCATAAAACAGCTAGAGAACTT 2760
Db     74382 GTTTCCTACACACTCTTCTAACCTCTTGCAGTCTATTTTCATAAAACAGCTAGAGAACTT 74441

Qy      2761 TGAGATGTAAGTCAAAAAATAGAACATGTCGCTCTTTCCCATTGTTTTTGAAATAAAGTT 2820
Db     74442 TGAGATGTAAGTCAAAAAATAGAACATGTCGCTCTTTCCCATTGTTTTTGAAATAAAGTT 74501

Qy      2821 CAACCCCCTTACCAGGGTCAACAAGGCCCTGCAATGATTTGGTCCTGTTAAAAATTCTTT 2880
Db     74502 CAACCCCCTTACCAGGGTCAACAAGGCCCTGCAATGATTTGGTCCTGTTAAAAATTCTTT 74561

Qy      2881 AGCCTTAACTCATGCTGTTCTTCCTTACACTCACTGCATTCTAGCCATTGAGGTTTCTAT 2940
Db     74562 AGCCTTAACTCATGCTGTTCTTCCTTACACTCACTGCATTCTAGCCATTGAGGTTTCTAT 74621

Qy      2941 GCATCAAACTTTTTTTGGTCCCAGCACTGTGCACATCCTTCTGGGTAGAATGCCCCTTGA 3000
```

```
Db      74622 GCATCAAACTTTTTTTGGTCCCAGCACTGTGCACATCCTTCTGGGTAGAATGCCCCTTGA 74681

Qy       3001 TTTGTATAATTAGCACCTCCTTCATCATTTAGGTCTTAGTATAACTACTACCTTCTTAGA 3060
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74682 TTTGTATAATTAGCACCTCCTTCATCATTTAGGTCTTAGTATAACTACTACCTTCTTAGA 74741

Qy       3061 GAAGCTCTGCTTCTTCATCCTATAAAAAAGTAAAATTCCTTACCCTGTTATTTTTTAAGT 3120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74742 GAAGCTCTGCTTCTTCATCCTATAAAAAAGTAAAATTCCTTACCCTGTTATTTTTTAAGT 74801

Qy       3121 CATCCGTGTTTCATTCTGTTAAAGTTCTTATCACAATTTATCATTATTTTATTTACAGTC 3180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74802 CATCCGTGTTTCATTCTGTTAAAGTTCTTATCACAATTTATCATTATTTTATTTACAGTC 74861

Qy       3181 ATGTGCCACATAACAATGTTTCAGTCAGGGATAGAACACAAATGTATCTGGCCCCATAAT 3240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74862 ATGTGCCACATAACAATGTTTCAGTCAGGGATAGAACACAAATGTATCTGGCCCCATAAT 74921

Qy       3241 ATTATAAGCTGAGAAATTTCTATTAACTAGTGATATCGCAGCCATCATAAGTGTAATGCA 3300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74922 ATTATAAGCTGAGAAATTTCTATTAACTAGTGATATCGCAGCCATCATAAGTGTAATGCA 74981

Qy       3301 GGACATTACCTTTTCTATGTTTAGATATGTTAGATACACAAATATATTTCATTGTGTTAT 3360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74982 GGACATTACCTTTTCTATGTTTAGATATGTTAGATACACAAATATATTTCATTGTGTTAT 75041

Qy       3361 AATTTCCTACAGTATTCAGTACAGTAACATGCTGTACAGGTTTGTAACCTAGGAGTAATA 3420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75042 AATTTCCTACAGTATTCAGTACAGTAACATGCTGTACAGGTTTGTAACCTAGGAGTAATA 75101

Qy       3421 GGCTATACCATACAGCTTAGGTGTGTAGTAGGCTATAACCATCTAGGTTTGTGTAAGTAC 3480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75102 GGCTATACCATACAGCTTAGGTGTGTAGTAGGCTATAACCATCTAGGTTTGTGTAAGTAC 75161

Qy       3481 ATTCTATGATATTCCCACAATGATGAAATCACCTAACTACACATTTCTCAGAATGTTTCA 3540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75162 ATTCTATGATATTCCCACAATGATGAAATCACCTAACTACACATTTCTCAGAATGTTTCA 75221

Qy       3541 CTGTTGTGAAGTGACCCATGACTATATTTTCCTATATACTTGATATTTTTGTGCATCTGC 3600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75222 CTGTTGTGAAGTGACCCATGACTATATTTTCCTATATACTTGATATTTTTGTGCATCTGC 75281

Qy       3601 CCATGAGAATGTAGTGTAAGATCAAAGGATGCAAGAATGGGTTCTATCCAGTATAGTACC 3660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75282 CCATGAGAATGTAGTGTAAGATCAAAGGATGCAAGAATGGGTTCTATCCAGTATAGTACC 75341

Qy       3661 CACTACACTGGTGGATGTCAATATGTATTTGTTAGATTAATATCTCAAGAATGAGCACCT 3720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75342 CACTACACTGGTGGATGTCAATATGTATTTGTTAGATTAATATCTCAAGAATGAGCACCT 75401

Qy       3721 TTCTCAGACACATAAAAGATGCTCAATATAAAAGTTTGTTGAACTGAACGTTATTGGCAA 3780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75402 TTCTCAGACACATAAAAGATGCTCAATATAAAAGTTTGTTGAACTGAACGTTATTGGCAA 75461

Qy       3781 ATGTAACATGATCGGATTTAAAGAGGAGCGAAACAGAGGTCTGGCTCAAACACCATACTT 3840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75462 ATGTAACATGATCGGATTTAAAGAGGAGCGAAACAGAGGTCTGGCTCAAACACCATACTT 75521

Qy       3841 CTAGAGTGCATAAGAGGTAGCAGTTGATTACCACTGGCGACAGGAGAAAAAAGAGCTTGA 3900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75522 CTAGAGTGCATAAGAGGTAGCAGTTGATTACCACTGGCGACAGGAGAAAAAAGAGCTTGA 75581

Qy       3901 CCGCAGGGTACTGTGAAGACATTTCAGGTTGATGGCACAGAACAGGGGAAATACATAAAT 3960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75582 CCGCAGGGTACTGTGAAGACATTTCAGGTTGATGGCACAGAACAGGGGAAATACATAAAT 75641

Qy       3961 GTGTGGGAATATTCAGTGGTCTGGGATGACTACATAGTAGAATATAATGAAGAAAAGAGT 4020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75642 GTGTGGGAATATTCAGTGGTCTGGGATGACTACATAGTAGAATATAATGAAGAAAAGAGT 75701

Qy       4021 GGAAGGGAAAGATGAAAAGTTGGAATGGGGATGAATTATGAAAGTACCAGAATGTTATGC 4080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75702 GGAAGGGAAAGATGAAAAGTTGGAATGGGGATGAATTATGAAAGTACCAGAATGTTATGC 75761

Qy       4081 TAAGGAATCTAGATTTTAAAATGTGAGGGCAAATTGAAGTCCTGGGCACGTTACAAAACT 4140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75762 TAAGGAATCTAGATTTTAAAATGTGAGGGCAAATTGAAGTCCTGGGCACGTTACAAAACT 75821

Qy       4141 AGAGGTCATAAAGTTTACCCTAATTTACCAAGATTTCCTAGAGGATCTATAATTGGAATC 4200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75822 AGAGGTCATAAAGTTTACCCTAATTTACCAAGATTTCCTAGAGGATCTAGAATTGGAATC 75881

Qy       4201 CAGATCTGCCTCTCTGTAAAGTTCAAGCACTTTCCATGACACCATACTGTTTCTTTCCAC 4260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75882 CAGATCTGCCTCTCTGTAAAGTTCAAGCACTTTCCATGACACCATACTGTTTCTTTCCAC 75941

Qy       4261 CTGCACAATGCAAATGAACTCTTATGAAACTGCTGTTTCTATCCTGGGCTAAATGTTGCA 4320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75942 CTGCACAATGCAAATGAACTCTTATGAAACTGCTGTTTCTATCCTGGGCTAAATGTTGCA 76001

Qy       4321 GAAAAAAGATTTAATCTTTGGGATAAGGCTATTTTGGGTTTTCTCCTACTTCTTGGGAAA 4380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76002 GAAAAAAGATTTAATCTTTGGGATAAGGCTATTTTGGGTTTTCTCCTACTTCTTGGGAAA 76061

Qy       4381 CAAGGTTTTCTTCCCCTGGCTAATTAAGTGTGGTATTGTTCTTCCAGGGAAATCAGTGAT 4440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76062 CAAGGTTTTCTTCCCCTGGCTAATTAAGTGTGGTATTGTTCTTCCAGGGAAATCAGTGAT 76121

Qy       4441 GCATCACCTGCTGCTATCAAATGTCAGGGTTGGAGTTCCTGATTTATTGCATGTGCCCAC 4500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76122 GCATCACCTGCTGCTATCAAATGTCAGGGTTGGAGTTCCTGATTTATTGCATGTGCCCAC 76181

Qy       4501 AAAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCT 4560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76182 AAAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCT 76241

Qy       4561 GTTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAA 4620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76242 GTTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAA 76301

Qy       4621 ATCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGG 4680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76302 ATCAATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGG 76361

Qy       4681 AGGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCT 4740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76362 AGGCTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCT 76421

Qy       4741 GAGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAACATCTGGTAGA 4800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76422 GAGCCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAACATCTGGTAGA 76481

Qy       4801 ACTAAAGAACGCAACGCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCC 4860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76482 ACTAAAGAACGCAACGCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCC 76541

Qy       4861 TTGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCT 4920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76542 TTGAGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCT 76601

Qy       4921 AGAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCC 4970
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76602 AGAGACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCC 76651
```

RESULT 3
AC031977/c

```
LOCUS       AC031977              191699 bp    DNA     linear   HTG 12-APR-2001
DEFINITION  Homo sapiens chromosome 1 clone RP11-288O18, WORKING DRAFT
            SEQUENCE, 3 unordered pieces.
ACCESSION   AC031977
VERSION     AC031977.7  GI:13194952
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1  (bases 1 to 191699)
  AUTHORS Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
          Federspiel,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,
          Mao,J., Komp,C., Kottler,S., Lam,B., Marathe,R., Miranda,M.,
          Morehouse,A.J., Nguyen,M., Oefner,P., Palm,C.J., Ramirez,D.,
          Southwick,A.M., Webb,C., Wilhelmy,J., Yu,S. and Davis,R.W.
  JOURNAL Unpublished
REFERENCE 2  (bases 1 to 191699)
  AUTHORS Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
          Federspiel,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,
          Mao,J., Marathe,R., Morehouse,A.J., Oefner,P., Palm,C.J.,
          Ramirez,D., Wilhelmy,J., Yu,S. and Davis,R.W.
  TITLE   Direct Submission
  JOURNAL Submitted (03-APR-2000) DNA Sequencing and Technology Center,
          Stanford University, 855 California Avenue, Palo Alto, CA 94304,
          USA
COMMENT   On Mar 4, 2001 this sequence version replaced gi:9665085.
          -------------- Genome Center
               Center: Stanford DNA Sequencing and Technology Development
          Center
               Center code: SDSTDC
               Web site: http://sequence-www.stanford.edu/group/human/
               Contact: hum-info@sequence.stanford.edu
          -------------- Project Information
               Center project name: 880
               Center clone name: RP11-288O18
          -------------- Summary Statistics
               Sequencing Vector: M13mp18; X02513; 100% of reads
               Sequencing Vector: plasmid; plasmid_accession; 0% of reads
               Chemistry: Dye-primer; 1% of reads
               Chemistry: Dye-terminator Big Dye; 99% of reads
               Assembly program: Phrap; version 0.990319
               Consensus quality: 190680 bases at least Q40
               Consensus quality: 191287 bases at least Q30
               Consensus quality: 191336 bases at least Q20
               Insert size: 195548; agarose-fp
               Insert size: 191499; sum-of-contigs
               Quality coverage:  7.9x in Q20 bases; agarose-fp
               Quality coverage:  8.1x in Q20 bases; sum-of-contigs.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 3 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1    12646: contig of 12646 bp in length
          *    12647    12746: gap of unknown length
          *    12747    94961: contig of 82215 bp in length
          *    94962    95061: gap of unknown length
          *    95062   191699: contig of 96638 bp in length.
FEATURES             Location/Qualifiers
     source          1. .191699
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-288O18"
                     /clone_lib="RPCI human BAC library 11"
     misc_feature    1. .12646
                     /note="assembly_name:Contig5"
     misc_feature    12747. .94961
                     /note="assembly_name:Contig6
                     clone_end:T7"
     misc_feature    95062. .191699
                     /note="assembly_name:Contig7
                     clone_end:SP6"
ORIGIN

  Query Match             91.7%;  Score 4559.4;  DB 2;  Length 191699;
  Best Local Similarity   96.8%;  Pred. No. 0;

  Matches 4636;  Conservative    0;  Mismatches  136;  Indels   15;  Gaps    4;

Qy       194 GCGAAATGTCATCTCAGAAAAAAAAAAAGGCATTTTATATATATATATATATATATATAC 253
                ||| 
Db     12749 GCGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 12690

Qy       254 ACACACACACATATATATATACACATATATATACACATATATACATATATACACATAT 313
                                                     |||| ||||| | | |
Db     12689 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTACACATATATATACAC 12630

Qy       314 ATACACATATATATACACATACATATGTACACATATATATACACATATGTATACACATAT 373
                |||   ||| | ||| | | |||||| |||||| |||||||||||||| || ||| ||||
Db     12629 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACATATAT 12570

Qy       374 ATACACATATATACACACATATATACACATATATACACACATATATACACATATATACAC 433
                | |||||||||||||||||||||||||||||||    |||||||||||||||||||||
Db     12569 ACACACATATATACACACATATATACACATAT----ACACATATATACACATATATACAC 12514

Qy       434 ATATATACACATATACACATATATACACATATATACATATATACACATATATATAATATA 493
                |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db     12513 ATATATACACATATATACATATATACACATATATACATATATACACATATATATAATATA 12454

Qy       494 CACACATATATATACACATATATACACACATATATACACATATATACACATAT------A 547
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||      |
Db     12453 CACACATATATATACACATATATACACACATATATACACATATATACACATATATACACA 12394

Qy       548 TATACACATATATACACATATATACATATATACACATATATATACATATATACACATATA 607
                |||| |||||||||||||||||| | ||||| ||| ||||||||||||||||||||||||
Db     12393 TATATACATATATACACATATATATACATATATACACATATATACATATATACACATATA 12334

Qy       608 TACATATATACACATATATACATATATACACACATATATACACATACATATACACACACA 667
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     12333 TACATATATACACATATATACATATATACACACATATATACACATACATATACACACACA 12274

Qy       668 TAGATATACATATATATACACATATATATACGTATATATATGTATATATATATGCTCCAG 727
                |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db     12273 TAGATATACATATATATACACATATATATGTATATATATGTATATATATATGCTCCAG 12214

Qy       728 AGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGC 787
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     12213 AGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGC 12154

Qy       788 AGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATG 847
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     12153 AGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATG 12094

Qy       848 TGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACAC----TATATATATATATAG 903
                |||||||||||||||||||||||||||||||||||||||    |||||||||||||||
Db     12093 TGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATATATATATATAG 12034

Qy       904 TGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTATAGAAGAATGGCTA 963
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     12033 TGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTATAGAAGAATGGCTA 11974

Qy       964 AGATGCCAAAGCTCAGTGTATGTGGCAGAGGCATGGTGTAGGGTGTGTCCAGGTTCATAT 1023
                |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db     11973 AGATGCCAAAGCTCAGTGTATGTGGCAGAGGCATGGTGGAGGGTGTGTCCAGGTTCATAT 11914

Qy      1024 ATTGCATTAAGTGTGAGAACACCCTGGAGTATGAACCAAGAAAATGCAAAAGCCAGAAGT 1083
                |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db     11913 ATTGCATTAAGTGTGAGAA-ACCCTGGAGTATGAACCAAGAAAATGCAAAAGCCAGAAGT 11855

Qy      1084 GATGGAGGAAATGAGACACAATAATGAAGATATTGAGAGGAGGGTGTGGGCCTAGAGTGA 1143
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     11854 GATGGAGGAAATGAGACACAATAATGAAGATATTGAGAGGAGGGTGTGGGCCTAGAGTGA 11795

Qy      1144 AGCTTTTCGTGCCAGTACTTCTTTTGAAGGCCCAGTTCTCTTCTCTCTCGGGGGCTCCTT 1203
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     11794 AGCTTTTCGTGCCAGTACTTCTTTTGAAGGCCCAGTTCTCTTCTCTCTCGGGGGCTCCTT 11735

Qy      1204 CATCTCTCATAGAGTCCACAGCTTTTAAGGGCCAACACTTGAGGTCAGCCTGGCTCTCTC 1263
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     11734 CATCTCTCATAGAGTCCACAGCTTTTAAGGGCCAACACTTGAGGTCAGCCTGGCTCTCTC 11675
```

```
Qy       1264 ATTTGAGCTGGATAGAACATTTTAGAGCACCATCTATTCTTCAAGAGGAAGTTTAAAAAT 1323
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11674 ATTTGAGCTGGATAGAACATTTTAGAGCACCATCTATTCTTCAAGAGGAAGTTTAAAAAT 11615

Qy       1324 AAAAGAACCTTGAAGAGGAAAAAATGTAGACATTCAATCTAACCTTTTCATTTTACTAGC 1383
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11614 AAAAGAACCTTGAAGAGGAAAAAATGTAGACATTCAATCTAACCTTTTCATTTTACTAGC 11555

Qy       1384 CAAAGCTAAATAGAATGCAGGTTACCTGTTTTTCAGCCAGGCACCATCATTTCCTAATTG 1443
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11554 CAAAGCTAAATAGAATGCAGGTTACCTGTTTTTCAGCCAGGCACCATCATTTCCTAATTG 11495

Qy       1444 TTATAAAATTTATTATTATTGTTGTTATTATTATTATTTGCCATAAGAAGTTTCCCATAT 1503
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11494 TTATAAAATTTATTATTATTGTTGTTATTATTATTATTTGCCATAAGAAGTTTCCCATAT 11435

Qy       1504 CCTTTTAGTATAACAAAAACACAATTCACAAGCATTATAAAACCCATGGTGTCTAACTAT 1563
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11434 CCTTTTAGTATAACAAAAACACAATTCACAAGCATTATAAAACCCATGGTGTCTAACTAT 11375

Qy       1564 TAAAAAAATTAAGTGGAACACACTTGTCCCAGCTACTGGGGAGGCTGAGGAGGGAGGATC 1623
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11374 TAAAAAAATTAAGTGGAACACACTTGTCCCAGCTACTGGGGAGGCCGAGGAGGGAGGATC 11315

Qy       1624 ACGTGATCCCAGGGGGTCAAGGTTATGGAGAGCTATGATTGTGCCACTGCACTCCAGCCT 1683
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11314 ACGTGATCCCAGGGGGTCAAGGTTATGGAGAGCTATGATTGTGCCACTGCACTCCAGCCT 11255

Qy       1684 GGGTGACAGGGAAAGACCCTGTCTCTAAAATTTTTTTTAAAAAAACTAAACTGGTTTTAT 1743
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11254 GGGTGACAGGGAAAGACCCTGTCTCTAAAATTTTTTTTAAAAAAACTAAACTGGTTTTAT 11195

Qy       1744 TACAGAGATTCTGGAGACAGCTACACATAAAAGGGTGGTATGCCTCATATTAGCTACCCA 1803
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11194 TACAGAGATTCTGGAGACAGCTACACATAAAAGGGTGGTATGCCTCATATTAGCTACCCA 11135

Qy       1804 GGGAGGTGGAATGCCAACTTAGGTGGTGTCACCACTATTAAAAATGCCCCAAAGCAATCA 1863
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11134 GGGAGGTGGAATGCCAACTTAGGTGGTGTCACCACTATTAAAAATGCCCCAAAGCAATCA 11075

Qy       1864 AAACTGAGAACTTCCTGGGAGCTTAGCATTGTGCAAAAGCAGCACAAAACACTTAAACAA 1923
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11074 AAACTGAGAACTTCCTGGGAGCTTAGCATTGTGCAAAAGCAGCACAAAACACTTAAACAA 11015

Qy       1924 TTCACAGTTGTGTTGGAATGGGAAGGCCTGGAAATATAAACCAAAGAGTATATTGTCTAA 1983
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11014 TTCACAGTTGTGTTGGAATGGGAAGGCCTGGAAATATAAACCAAAGAGTATATTGTCTAA 10955

Qy       1984 ATTGATAGAGATTACAATTGCCTGAAAGAAAAAGTTGACTTTTAACTAGAATGTTCAGAG 2043
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10954 ATTGATAGAGATTACAATTGCCTGAAAGAAAAAGTTGACTTTTAACTAGAATGTTCAGAG 10895

Qy       2044 TAGGTTTACAGAAGAAGCTCTTAAACTGGGCTCCAGTGGATTTGTCAATGCTTTGGAAGC 2103
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10894 TAGGTTTACAGAAGAAGCTCTTAAACTGGGCTCCAGTGGATTTGTCAATGCTTTGGAAGC 10835

Qy       2104 TGGTGGGGTGGGAGGGTTGGAGGGGGCATAAAAAGTCATGTTGGTATGCTCTGCTCAAGT 2163
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10834 TGGTGGGGTGGGAGGGTTGGAGGGGGCATAAAAAGTCATGTTGGTATGCTCTGCTCAAGT 10775

Qy       2164 CTCCATTCTGTTTCCTTTTCCTCTTTTCAATGTCATGTCCCATTATTTCATTATGGGCTT 2223
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10774 CTCCATTCTGTTTCCTTTTCCTCTTTTCAATGTCATGTCCCATTATTTCATTATGGGCTT 10715

Qy       2224 CCCTTTATCCAGGATCAATATGCCACCTCTTGGTTGTCTTTTACCTACTTCTCCACCTCA 2283
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10714 CCCTTTATCCAGGATCAATATGCCACCTCTTGGTTGTCTTTTACCTACTTCTCCACCTCA 10655

Qy       2284 CTATGGAATCGTCCTTGGGTAGCTCCTGTGCTTGGGAACCTGCACGGGCACTTTTCTGAT 2343
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10654 CTATGGAATCGTCCTTGGGTAGCTCCTGTGCTTGGGAACCTGCACGGGCACTTTTCTGAT 10595
```

```
Qy       2344 GTCTTGATTCCAGCTTTACTCCTAAAACTTAAATGCTGAGGGGCCAACACCATGGCAGTG 2403
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10594 GTCTTGATTCCAGCTTTACTCCTAAAACTTAAATGCTGAGGGGCCAACACCATGGCAGTG 10535

Qy       2404 GTAGGGATGGGAATGGGGGTCTTGTAACACACTACATAAACTACACGAAATAAACTACAT 2463
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10534 GTAGGGATGGGAATGGGGGTCTTGTAACACACTACATAAACTACACGAAATAAACTACAT 10475

Qy       2464 GAAACTCAACATGTTTGCAAGACTCAGTTCACATCCATGAGGAGCTCATGCTTCTCCCTC 2523
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10474 GAAACTCAACATGTTTGCAAGACTCAGTTCACATCCATGAGGAGCTCATGCTTCTCCCTC 10415

Qy       2524 CTGCTCCCCTAGCACACATGATTATCTCTATTTGGAAATGTTTGGCATTTTTGGTGAAGT 2583
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10414 CTGCTCCCCTAGCACACATGATTATCTCTATTTGGAAATGTTTGGCATTTTTGGTGAAGT 10355

Qy       2584 GAATGGTTCAATAACTTTCTCCACCATCAGAACAAAAGCTCTTTAAGGTTAGGGATGGGA 2643
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10354 GAATGGTTCAATAACTTTCTCCACCATCAGAACAAAAGCTCTTTAAGGTTAGGGATGGGA 10295

Qy       2644 TCATACACACTTCCCTTGTCCAAGTCCCCATCACCCCTTATCTAGACAATTGCTACAGTT 2703
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10294 TCATACACACTTCCCTTGTCCAAGTCCCCATCACCCCTTATCTAGACAATTGCTACAGTT 10235

Qy       2704 TCCTACACACTCTTCTAACCTCTTGCAGTCTATTTTCATAAAACAGCTAGAGAACTTTGA 2763
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10234 TCCTACACACTCTTCTAACCTCTTGCAGTCTATTTTCATAAAACAGCTAGAGAACTTTGA 10175

Qy       2764 GATGTAAGTCAAAAAATAGAACATGTCGCTCTTTCCCATTGTTTTTGAAATAAAGTTCAA 2823
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10174 GATGTAAGTCAAAAAATAGAACATGTCGCTCTTTCCCATTGTTTTTGAAATAAAGTTCAA 10115

Qy       2824 CCCCCTTACCAGGGTCAACAAGGCCCTGCAATGATTTGGTCCTGTTAAAAATTCTTTAGC 2883
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10114 CCCCCTTACCAGGGTCAACAAGGCCCTGCAATGATCTGGTCCTGTTAAAAATTCTTTAGC 10055

Qy       2884 CTTAACTCATGCTGTTCTTCCTTACACTCACTGCATTCTAGCCATTGAGGTTTCTATGCA 2943
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10054 CTTAACTCATGCTGTTTTTCCTTACACTCACTGCATTCTAGCCATTGAGGTTTCTATGCA 9995

Qy       2944 TCAAACTTTTTTTGGTCCCAGCACTGTGCACATCCTTCTGGGTAGAATGCCCCTTGATTT 3003
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9994 TCAAACTTTTTTTGGTCCCAGCACTGTGCACATCCTTCTGGGTAGAATGCCCCTTGATTT 9935

Qy       3004 GTATAATTAGCACCTCCTTCATCATTTAGGTCTTAGTATAACTACTACCTTCTTAGAGAA 3063
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9934 GTATAATTAGCACCTCCTTCATCATTTAGGTCTTAGTATAACTACTACCTTCTTAGAGAA 9875

Qy       3064 GCTCTGCTTCTTCATCCTATAAAAAAGTAAAATTCCTTACCCTGTTATTTTTTAAGTCAT 3123
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9874 GCTCTGCTTCTTCATCCTATAAAAAAGTAAAATTCCTTACCCTGTTATTTTTTAAGTCAT 9815

Qy       3124 CCGTGTTTCATTCTGTTAAAGTTCTTATCACAATTTATCATTATTTTATTTACAGTCATG 3183
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9814 CCGTGTTTCATTCTGTTAAAGTTCTTATCACAATTTATCATTATTTTATTTACAGTCATG 9755

Qy       3184 TGCCACATAACAATGTTTCAGTCAGGGATAGAACACAAATGTATCTGGCCCCATAATATT 3243
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9754 TGCCACATAACAATGTTTCAGTCAGGGATAGAACACAAATGTATCTGGCCCCATAATATT 9695

Qy       3244 ATAAGCTGAGAAATTTCTATTAACTAGTGATATCGCAGCCATCATAAGTGTAATGCAGGA 3303
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9694 ATAAGCTGAGAAATTTCTATTAACTAGTGATATCGCAGCCATCATAAGTGTAATGCAGGA 9635

Qy       3304 CATTACCTTTTCTATGTTTAGATATGTTAGATACACAAATATATTTCATTGTGTTATAAT 3363
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9634 CATTACCTTTTCTATGTTTAGATATGTTAGATACACAAATATATTTCATTGTGTTATAAT 9575

Qy       3364 TTCCTACAGTATTCAGTACAGTAACATGCTGTACAGGTTTGTAACCTAGGAGTAATAGGC 3423
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9574 TTCCTACAGTATTCAGTACAGTAACATGCTGTACAGGTTTGTAACCTAGGAGTAATAGGC 9515

Qy       3424 TATACCATACAGCTTAGGTGTGTAGTAGGCTATAACCATCTAGGTTTGTGTAAGTACATT 3483
```

```
Db       9514 TATACCATACAGCTTAGGTGTGTAGTAGGCTATAACCATCTAGGTTTGTGTAAGTACATT 9455

Qy       3484 CTATGATATTCCCACAATGATGAAATCACCTAACTACACATTTCTCAGAATGTTTCACTG 3543
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9454 CTATGATATTCCCACAATGATGAAATCACCTAACTACACATTTCTCAGAATGTTTCACTG 9395

Qy       3544 TTGTGAAGTGACCCATGACTATATTTTCCTATATACTTGATATTTTTGTGCATCTGCCCA 3603
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9394 TTGTGAAGTGACCCATGACTATATTTTCCTATATACTTGATATTTTTGTGCATCTGCCCA 9335

Qy       3604 TGAGAATGTAGTGTAAGATCAAAGGATGCAAGAATGGGTTCTATCCAGTATAGTACCCAC 3663
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9334 TGAGAATGTAGTGTAAGATCAAAGGATGCAAGAATGGGTTCTATCCAGTATAGTACCCAC 9275

Qy       3664 TACACTGGTGGATGTCAATATGTATTTGTTAGATTAATATCTCAAGAATGAGCACCTTTC 3723
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9274 TACACTGGTGGATGTCAATATGTATTTGTTAGATTAATATCTCAAGAATGAGCACCTTTC 9215

Qy       3724 TCAGACACATAAAAGATGCTCAATATAAAAGTTTGTTGAACTGAACGTTATTGGCAAATG 3783
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9214 TCAGACACATAAAAGATGCTCAATATAAAAGTTTGTTGAACTGAACGTTATTGGCAAATG 9155

Qy       3784 TAACATGATCGGATTTAAAGAGGAGCGAAACAGAGGTCTGGCTCAAACACCATACTTCTA 3843
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9154 TAACATGATCGGATTTAAAGAGGAGCGAAACAGAGGTCTGGCTCAAACACCATACTTCTA 9095

Qy       3844 GAGTGCATAAGAGGTAGCAGTTGATTACCACTGGCGACAGGAGAAAAAAGAGCTTGACCG 3903
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9094 GAGTGCATAAGAGGTAGCAGTTGATTACCACTGGCGACAGGAGAAAAAAGAGCTTGACCG 9035

Qy       3904 CAGGGTACTGTGAAGACATTTCAGGTTGATGGCACAGAACAGGGGAAATACATAAATGTG 3963
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9034 CAGGGTACTGTGAAGACATTTCAGGTTGATGGCACAGAACAGGGGAAATACATAAATGTG 8975

Qy       3964 TGGGAATATTCAGTGGTCTGGGATGACTACATAGTAGAATATAATGAAGAAAAGAGTGGA 4023
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8974 TGGGAATATTCAGTGGTCTGGGATGACTACATAGTAGAATATAATGAAGAAAAGAGTGGA 8915

Qy       4024 AGGGAAAGATGAAAAGTTGGAATGGGGATGAATTATGAAAGTACCAGAATGTTATGCTAA 4083
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8914 AGGGAAAGATGAAAAGTTGGAATGGGGATGAATTATGAAAGTACCAGAATGTTATGCTAA 8855

Qy       4084 GGAATCTAGATTTTAAAATGTGAGGGCAAATTGAAGTCCTGGGCACGTTACAAAACTAGA 4143
             |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db       8854 GGAATCTAGATTTTAAAATGTGAGGGCAAATTGAAGTCCAGGGCACGTTACAAAACTAGA 8795

Qy       4144 GGTCATAAAGTTTACCCTAATTTACCAAGATTTCCTAGAGGATCTATAATTGGAATCCAG 4203
             |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db       8794 GGTCATAAAGTTTACCCTAATTTACCAAGATTTCCTAGAGGATCTAGAATTGGAATCCAG 8735

Qy       4204 ATCTGCCTCTCTGTAAAGTTCAAGCACTTTCCATGACACCATACTGTTTCTTTCCACCTG 4263
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8734 ATCTGCCTCTCTGTAAAGTTCAAGCACTTTCCATGACACCATACTGTTTCTTTCCACCTG 8675

Qy       4264 CACAATGCAAATGAACTCTTATGAAACTGCTGTTTCTATCCTGGGCTAAATGTTGCAGAA 4323
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8674 CACAATGCAAATGAACTCTTATGAAACTGCTGTTTCTATCCTGGGCTAAATGTTGCAGAA 8615

Qy       4324 AAAAGATTTAATCTTTGGGATAAGGCTATTTTGGGTTTTCTCCTACTTCTTGGGAAACAA 4383
             |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db       8614 AAAAGATTTAATCTTTGGGATAAGGCTATTTTGGGTTTTCTCCTCCTTCTTGGGAAACAA 8555

Qy       4384 GGTTTTCTTCCCCTGGCTAATTAAGTGTGGTATTGTTCTTCCAGGGAAATCAGTGATGCA 4443
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8554 GGTTTTCTTCCCCTGGCTAATTAAGTGTGGTATTGTTCTTCCAGGGAAATCAGTGATGCA 8495

Qy       4444 TCACCTGCTGCTATCAAATGTCAGGGTTGGAGTTCCTGATTTATTGCATGTGCCCACAAA 4503
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8494 TCACCTGCTGCTATCAAATGTCAGGGTTGGAGTTCCTGATTTATTGCATGTGCCCACAAA 8435

Qy       4504 GCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTGTT 4563
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8434 GCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTGTT 8375

Qy       4564 TACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAAATC 4623
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db       8374 TACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAACTC 8315

Qy       4624 AATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGAGG 4683
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8314 AATAACTGAACTCTCATTGATGTCTAGGCCATTGTAGTAAACAATAAAGAAGGAGGGAGG 8255

Qy       4684 CTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTGAG 4743
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8254 CTTCTGACAACTGAGAGGAAATTGTCATCTGAAGTGGTGCAAGCACAGCCTGGGGCTGAG 8195

Qy       4744 CCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAACATCTGGTAGAACT 4803
             |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db       8194 CCTTGGCCTACATCCTGCCCAAGTGGAGGATCAGTGCCCCATTTAATATCTGGTAGAACT 8135

Qy       4804 AAAGAACGCAACGCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCTTG 4863
             |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db       8134 AAAGAACGCAACTCCTGCCACAATGACTTATTTCCCTGCATTTGATACCGTCAATCCTTG 8075

Qy       4864 AGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTAGA 4923
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8074 AGAAATGTTTTCTTTTGTTCTCCCTGAGCAAAGGTTGGAAAAATTTGAAATTTACCTAGA 8015

Qy       4924 GACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCC 4970
             |||||||||||||||||||||||||||||||||||||||||||||||
Db       8014 GACCACACATAGTTCACATCCTGCTGTGTGGCTGAATGTCTGCCCCC 7968
```

RESULT 4
AC097328

```
LOCUS       AC097328              203216 bp    DNA     linear   HTG 20-OCT-2001
DEFINITION  Pan troglodytes clone RP43-174K14, WORKING DRAFT SEQUENCE, 21
            unordered pieces.
ACCESSION   AC097328
VERSION     AC097328.1  GI:16118151
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 203216)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
```

```
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 203216)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-OCT-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     -------------- Genome Center
              Center: Baylor College of Medicine
              Center code: BCM
              Web site: http://www.hgsc.bcm.tmc.edu/
              Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
              Center project name: ZUAM
              Center clone name: RP43-174K14
            -------------- Summary Statistics
              Sequencing vector: Plasmid; M77789
              Chemistry: Dye-terminator Big Dye: 100% of reads
              Assembly program: Phrap; version 0.990329
              Consensus quality: 196812 bases at least Q40
              Consensus quality: 208422 bases at least Q30
              Consensus quality: 218080 bases at least Q20
              Estimated insert size: 228347; sum-of-contigs estimation
              Quality coverage: 0x in Q20 bases; agarose-fp estimation
              Quality coverage: 7.1x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 21 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    30770: contig of 30770 bp in length
            *    30771    30870: gap of unknown length
            *    30871    52356: contig of 21486 bp in length
            *    52357    52456: gap of unknown length
            *    52457    72697: contig of 20241 bp in length
            *    72698    72797: gap of unknown length
            *    72798    90396: contig of 17599 bp in length
            *    90397    90496: gap of unknown length
            *    90497    96134: contig of 5638 bp in length
            *    96135    96234: gap of unknown length
            *    96235   114433: contig of 18199 bp in length
            *   114434   114533: gap of unknown length
            *   114534   125184: contig of 10651 bp in length
            *   125185   125284: gap of unknown length
            *   125285   135194: contig of 9910 bp in length
            *   135195   135294: gap of unknown length
            *   135295   140273: contig of 4979 bp in length
            *   140274   140373: gap of unknown length
            *   140374   148490: contig of 8117 bp in length
            *   148491   148590: gap of unknown length
            *   148591   153435: contig of 4845 bp in length
            *   153436   153535: gap of unknown length
            *   153536   162229: contig of 8694 bp in length
            *   162230   162329: gap of unknown length
            *   162330   167843: contig of 5514 bp in length
            *   167844   167943: gap of unknown length
            *   167944   175297: contig of 7354 bp in length
            *   175298   175397: gap of unknown length
            *   175398   179409: contig of 4012 bp in length
            *   179410   179509: gap of unknown length
            *   179510   185931: contig of 6422 bp in length
            *   185932   186031: gap of unknown length
            *   186032   190690: contig of 4659 bp in length
            *   190691   190790: gap of unknown length
            *   190791   194808: contig of 4018 bp in length
            *   194809   194908: gap of unknown length
            *   194909   198788: contig of 3880 bp in length
            *   198789   198888: gap of unknown length
            *   198889   201055: contig of 2167 bp in length
            *   201056   201155: gap of unknown length
            *   201156   203216: contig of 2061 bp in length.
FEATURES             Location/Qualifiers
     source          1. .203216
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="RP43-174K14"
ORIGIN

  Query Match             8.3%;  Score 411.2;  DB 2;  Length 203216;
  Best Local Similarity   76.7%;  Pred. No. 1.1e-72;
  Matches  554;  Conservative    0;  Mismatches  163;  Indels    5;  Gaps    4;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCA-TCTCTACTAAAAATACAAAATT 59
              ||||||| |||||||||| |||||||||||||||||  |||||||||||| ||||||| 
Db     114617 AGTTCAAAACCAGCCTGACCAACATGGTGAAACCCCCGTCTCTACTAAAAACACAAAATT 114676

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   ||||||||| | || ||||||| ||||||||||   ||| ||  ||||||||||
Db     114677 AGCTGGGCATGGTGGTGCATGCCTGTAGTCCCAGCTACTCAGGAAGCTAAGGCAGGAGAA 114736

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGC-AAGATCATGCCATTGCACTCCA 178
              ||||||||||  | || || ||||||||| | || |  |||||||||| |||||||||
Db     114737 TCACTTGAACC--CAGGAGGCGGAGGTTGCAGTGAACTGAGATCATGCCACTGCACTCCA 114794

Qy        179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATA 238
              |||   | |||| |||         |  |    | | |    |       | | |||||||||
Db     114795 GCCTGAGTAACAAAAGCATCTCAAAAAATATATATATATGTATATATGTGTGTATATATA 114854

Qy        239 TATATATATATATACACACACACACATATATATATACACATATATATACACATATATA 298
              | | |||||||||       | | | |  | ||||||||    | | |||| |  | ||||
Db     114855 TGTGTATATATATGTGTGTATATATATGTGTATATATATGTGTGTGTATATATGTGTATA 114914

Qy        299 CATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACA 358
              |||||||  | |||||||  | |||||||||  | |||  |||| ||  | |||||||||  | |
Db     114915 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 114974

Qy        359 TATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATA 418
              |||  ||||  | |||||||  | |||||||  | |  |||||||  | |||||||  | |  |||||
Db     114975 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 115034

Qy        419 TACACATATATACACATATATACACATATACACATATATACACATATATACATATATACA 478
              ||  | |||||||  | |||||||  | |||||  | |||||||  | |||||||  |||||||  |
Db     115035 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 115094

Qy        479 CATATATATAATATACACACATATATATACACATATATACACACATATATACACATATAT 538
              ||||||||  |||||  | |  |||||||||  | |||||||  | |  | |||||||  | ||||||
Db     115095 TATATATAT-ATATATATATATATATATATATATATATATATATATATATATATATAT 115153

Qy        539 ACACATATATATACACATATATACACATATATACATATATACACATATATATACATATAT 598
              | |  |||||||||  | |||||||  |||||||  |||||||  | |||||||||  ||||||
Db     115154 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 115213

Qy        599 ACACATATATACATATATACACATATATACATATATACACACATATATACACATACATAT 658
              | |  |||||||  |||||||  | |||||||  |||||||  | |  |||||||  | |||  ||||
Db     115214 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 115273

Qy        659 ACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATAT 718
              | | |   |||  |||||  |||||||||  | |||||||||   |||||||||  |||||||||
```

```
Db     115274 ATATATCTATATATATATATATATATATATATATATATATATATATATATATATATAT 115333

Qy        719 AT 720
              ||
Db     115334 AT 115335


RESULT 5
CR759943
LOCUS       CR759943              104321 bp    DNA     linear   HTG 01-OCT-2004
DEFINITION  Homo sapiens chromosome 6 clone DAAP-21A6, WORKING DRAFT SEQUENCE.
ACCESSION   CR759943
VERSION     CR759943.2  GI:53748680
KEYWORDS    HTG; HTGS_PHASE2; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 104321)
  AUTHORS   Phillimore,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 1, 2004 this sequence version replaced gi:51965768.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bAP21A6
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 103081 bases at least Q40
            Consensus quality: 103313 bases at least Q30
            Consensus quality: 103455 bases at least Q20
            Insert size: 104321; sum-of-contigs
            Insert size: 110232; 0.4% error; agarose-fp
            Quality coverage: 7.35x in Q20 bases; sum-of-contigs Quality
            coverage: 7.39x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   104321: contig of 104321 bp in length.
FEATURES            Location/Qualifiers
     source         1. .104321
                    /organism="Homo sapiens"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:9606"
                    /chromosome="6"
                    /clone="DAAP-21A6"
                    /clone_lib="APD"
     misc_feature   1. .104321
                    /note="assembly_fragment:01274"
ORIGIN

  Query Match             8.2%;  Score 405.4;  DB 2;  Length 104321;
  Best Local Similarity   75.3%;  Pred. No. 1.8e-71;
  Matches  558;  Conservative    0;  Mismatches  176;  Indels    7;  Gaps    4;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC---AAAA 57
              ||||| ||||||||||  || ||||||||||||||||||||||||||||||    ||||
Db      96311 AGTTCGAGACCAGCCTGGCCAGCATGGTGAAACCCCATCTCTACTAAAAATACAAAAAAA 96370

Qy         58 TTAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAG 117
              |||| ||||  ||||     |  | || || ||||| |||| | |||||| | |||||| |
Db      96371 TTAGCCAGGTGTGGTAGCGCGTGTCTCTAGTCCCACCTACCTGGGAGGCTGTGGCAGGCG 96430

Qy        118 AATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGC-AAGATCATGCCATTGCACTC 176
              |||  ||| ||||  |  | || |||||||||| | ||||  |||||| |||| |||||||
Db      96431 AATTGCTTCAACC--CGTGAGGCGGAGGTTGCAGTGAGCTGAGATCACGCCACTGCACTC 96488

Qy        177 CAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATA 236
              |||||  |||||| ||||| | |      ||||| |||||||| |       | ||||||||
Db      96489 CAGCCTGGGCAAC-AGAGCAAGAATCAGTCTCAAAAAAAAAATATATATATATATATATA 96547

Qy        237 TATATATATATATATACACACACACACATATATATATACACATATATATACACATATA 296
              |||||||||||||||| | | | | | ||||||||||| | |||||||| | |||||
Db      96548 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 96607

Qy        297 TACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACA 356
              || ||||||| | ||||||| | |||||||| | ||| |||| || | |||||||| |
Db      96608 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 96667

Qy        357 CATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATA 416
              |||| |||| | ||||||| | ||||||| | | ||||||| | ||||||| | | |||
Db      96668 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 96727

Qy        417 TATACACATATATACACATATATACACATATACACATATATACACATATATACATATATA 476
              |||| | ||||||| | ||||||| | ||||| | ||||||| | ||||||| ||| |||
Db      96728 TATATATATATATATATATATATATATATATATATATATATATATATATATATAAATA 96787

Qy        477 CACATATATATAATATACACACATATATATACACATATATACACACATATATACACATAT 536
              | ||||| |||  |                ||| | ||||||| | | ||||||| | ||||
Db      96788 TAAATATAAATATAAATATATATATAATATATATATATATATATATATATATATAT 96847

Qy        537 ATACACATATATATACACATATATACACATATATACATATATACACATATATATACATAT 596
              ||| | |||||||| | ||||||| | ||||||| ||||||| | |||||||| ||||
Db      96848 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 96907

Qy        597 ATACACATATATACATATATACACATATATACATATATACACACATATATACACATACAT 656
              ||| | ||||||| ||||||| | ||||||| ||||||| | | ||||||| | ||| ||
Db      96908 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 96967

Qy        657 ATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATAT 716
              ||| | | | ||| ||||| |||| |||| | |||| | ||  ||  | ||||| |||||
Db      96968 ATATATATATATATATATATATATGTATATATATATGTGTATATACGTGTATGTGTATAT 97027

Qy        717 ATATGCTCCAGAGTTCATAAG 737
              |||       | ||  ||| |
Db      97028 ATACATGTATGTGTATATATG 97048


RESULT 6
AC097330/c
LOCUS       AC097330              181946 bp    DNA     linear   HTG 20-OCT-2001
DEFINITION  Pan troglodytes clone RP43-54O15, WORKING DRAFT SEQUENCE, 16
            unordered pieces.
ACCESSION   AC097330
VERSION     AC097330.1  GI:16118153
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 181946)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
```

```
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 181946)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-OCT-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: ZUAP
                Center clone name: RP43-54O15
            -------------- Summary Statistics
                Sequencing vector: Plasmid; M77789
                Sequencing vector: M13; L08821
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 169928 bases at least Q40
                Consensus quality: 176100 bases at least Q30
                Consensus quality: 180968 bases at least Q20
                Estimated insert size: 183838; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 6.9x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 16 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    17715: contig of 17715 bp in length
            *    17716    17815: gap of unknown length
            *    17816    44117: contig of 26302 bp in length
            *    44118    44217: gap of unknown length
            *    44218    67185: contig of 22968 bp in length
            *    67186    67285: gap of unknown length
            *    67286    84740: contig of 17455 bp in length
            *    84741    84840: gap of unknown length
            *    84841   100199: contig of 15359 bp in length
            *   100200   100299: gap of unknown length
            *   100300   112190: contig of 11891 bp in length
            *   112191   112290: gap of unknown length
            *   112291   124548: contig of 12258 bp in length
            *   124549   124648: gap of unknown length
            *   124649   134134: contig of 9486 bp in length
            *   134135   134234: gap of unknown length
            *   134235   143821: contig of 9587 bp in length
            *   143822   143921: gap of unknown length
            *   143922   152987: contig of 9066 bp in length
            *   152988   153087: gap of unknown length
            *   153088   159063: contig of 5976 bp in length
            *   159064   159163: gap of unknown length
            *   159164   166547: contig of 7384 bp in length
            *   166548   166647: gap of unknown length
            *   166648   171663: contig of 5016 bp in length
            *   171664   171763: gap of unknown length
            *   171764   177471: contig of 5708 bp in length
            *   177472   177571: gap of unknown length
            *   177572   179822: contig of 2251 bp in length
            *   179823   179922: gap of unknown length
            *   179923   181946: contig of 2024 bp in length.
FEATURES             Location/Qualifiers
     source          1. .181946
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="RP43-54O15"
ORIGIN

  Query Match             8.2%;  Score 405.2;  DB 2;  Length 181946;
  Best Local Similarity   68.3%;  Pred. No. 1.8e-71;
  Matches  634;  Conservative    0;  Mismatches  288;  Indels    6;  Gaps    5;

Qy        1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
            ||||| |||||| ||||  | | ||||||||||||||| |||||||||| |||||| ||||||
Db   155131 AGTTCGAGACCAACCTGGACTATATGGTGAAACCCCGTCTCTACTAGAAATACAAAAATT 155072

Qy       60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
            ||   |||||||||  ||||||||||||||||||| |||||| ||||||  ||||||||||
Db   155071 AGCTGGGCATGGTGGCACACGCCTGTAATCCCAACTACTTGGGAGGCTAAGGCAGGAGAA 155012

Qy      120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCCA 178
            |||||||||   | || || ||||||| | ||| | |||||   ||| |||||||||
Db   155011 TCACTTGAAC--TCGGGAGGCAGAGGTTGCAGTGAGCCCAGATCGCACCACTGCACTCCA 154954

Qy      179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATA 238
            ||| |||| |       |   || |||     | |||| |||| |     |       |  | |
Db   154953 GCCTAGGCGAGTGAGACTCCATCTCAAAAAAAAAAAGAAAAGAAAATTGAGGGCTGGACA 154894

Qy      239 TATATATATATATACACACACACACATATATATATACACATATATATACACATATATA 298
            |      || |      | | ||     | |  ||||| | ||||||||| | |||||||
Db   154893 CAGTGGCTTACACCTGTAATCCCAGCACTTTGGATATATATATATATATATATATATA 154834

Qy      299 CATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACA 358
            ||||||| | ||||||| | ||||||||| | ||| |||| || | ||||||||| | |
Db   154833 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 154774

Qy      359 TATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATA 418
            ||| |||| | ||||||| | ||||||| | | ||||||| | ||||||| | | |||||
Db   154773 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 154714

Qy      419 TACACATATATACACATATATACACATATACACATATATACACATATATACATATATACA 478
            || | ||||||| | ||||||| | ||||| | ||||||| | ||||||| ||||||| |
Db   154713 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 154654

Qy      479 CATATATATA-ATATACACACATATATATACACATATATACACACATATATACACATATA 537
            |||||||||  ||||| | | ||||||||| | ||||||| | | ||||||| | |||||
Db   154653 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 154594

Qy      538 TACACATATATATACACATATATACACATATATACATATATACACATATATATACATATA 597
```

```
Db     154593 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 154534

Qy        598 TACACATATATACATATATACACATATATACATATATACACACATATATACACATACATA 657
Db     154533 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 154474

Qy        658 TACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATA 717
Db     154473 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 154414

Qy        718 TATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGA 777
Db     154413 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 154354

Qy        778 GTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAA 837
Db     154353 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 154294

Qy        838 TACATAAATGTGTAATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTAACACTATATATA 896
Db     154293 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 154234

Qy        897 TATATAGTGAAAGCAGTGTAAACCTAAA 924
Db     154233 TATATATATATATATATATATATATA 154206


RESULT 7
AL358074
LOCUS       AL358074              109081 bp    DNA     linear   PRI 21-NOV-2001
DEFINITION  Human DNA sequence from clone RP11-423C15 on chromosome 9, complete
            sequence.
ACCESSION   AL358074
VERSION     AL358074.22  GI:17048358
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 109081)
  AUTHORS   Hammond,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Nov 22, 2001 this sequence version replaced gi:17043779.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr9
            RP11-423C15 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-423C15 It may be shorter because we sequence overlapping
            sections only once, except for a short overlap.
            The true right end of clone RP11-423C15 is at 109081 in this
            sequence. The true left end of clone RP11-770B5 is at 78953 in this
            sequence. The true right end of clone RP11-12A16 is at 2000 in this
            sequence.
FEATURES             Location/Qualifiers
     source          1. .109081
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /clone="RP11-423C15"
                     /clone_lib="RPCI-11.2"
     misc_feature    complement(65197. .65223)
                     /note="Single clone region. Sequence from reads from a
                     short insert library derived from a single pUC clone.
                     Restriction digest data confirm the assembly."
     misc_feature    65533. .65757
                     /note="Single clone region. Sequence from reads from a
                     short insert library derived from a single pUC clone.
                     Restriction digest data confirm the assembly."
     misc_feature    98714. .98750
                     /note="Single clone region. Sequence from reads from a
                     short insert library derived from a single pUC clone.
                     Restriction digest data confirm the assembly."
     misc_feature    98834. .98842
                     /note="Single clone region. Sequence from reads from a
                     short insert library derived from a single pUC clone.
                     Restriction digest data confirm the assembly."
     misc_feature    99111. .99161
                     /note="Single clone region. Sequence from reads from a
                     short insert library derived from a single pUC clone.
                     Restriction digest data confirm the assembly"
     misc_feature    99163. .99191
                     /note="Single clone region. Sequence from reads from a
                     short insert library derived from a single pUC clone.
                     Restriction digest data confirm the assembly"
ORIGIN

  Query Match             8.1%;  Score 402.2;  DB 9;  Length 109081;
  Best Local Similarity   78.2%;  Pred. No. 7.7e-71;
  Matches  564;  Conservative    0;  Mismatches  138;  Indels   19;  Gaps    6;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC--AAAAT 58
Db      11620 AGTTTGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAAAAAC 11679

Qy         59 TAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGA 118
Db      11680 TAGCCAAGCGTGGTGGCACGCACCTGTAATCCCAGCTACTCGGGCAGCTGAGGCAGGAGA 11739

Qy        119 ATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCC 177
Db      11740 ATCGCTTGAACC--CAAGAGGCAGAGGTTGCAGTGAGCCAAGATCATGCCATTGCACTCC 11797

Qy        178 AGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATAT 237
Db      11798 AGCTCAAGTAACAAGAGCGAAATTCCGCCTCAAAAAAAAAAAA---------ATATATAT 11848

Qy        238 ATATATATATATACACACACACACATATATATATACACATATATATACACATATAT 297
Db      11849 ATATATATACACACACACACATACATGTATATATACACACACATATACACATATATAC 11908

Qy        298 ACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACAC 357
Db      11909 ACACATATATACACATACACACACATATATACACATATACACACACATATATACATATAC 11968

Qy        358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417
Db      11969 ACACACATATATACATACACACACATATATACACGTATACACACACACATACACATATAT 12028
```

```
Qy        418 ATACACATATATACACATATATACACATATACACATATATACACATATATACATATATAC 477
              | |||||||||||||||||| ||||| ||| |||  |||||| | | | |||||||| ||
Db      12029 ACACACATATATACACATATACACACACATATACACGTATACATACACACACATATACAC 12088

Qy        478 ACATATATATA-ATATACACACATATATATACACATATATACACACATATATACACACATAT 536
              ||||||| |  |||| | |||  |||||| ||| ||||| |||| ||||||||||||||
Db      12089 GTATATATACACGTATATATACACGTATATATACACATATATACACGTATATACACACATAT 12148

Qy        537 ATACACATATATATACACATATATACACATATATACATATATACAC----ATATATATAC 592
              |||||| |||||||||||||||||||||| |||||||| ||||| ||    ||||||||||
Db      12149 ATACACGTATATATACACATATATACACGTATATACACATATATACACGTATATATATAC 12208

Qy        593 ATATATACACATATATACATATATACACATATATACATATATACACACATATATACACAT 652
              | | ||| | ||| | ||| |||||||||| | | ||||||||| ||||| ||| |||||
Db      12209 ACACATATATATACACACACATATACACACACACACATATATATACACACATACACACAC 12268

Qy        653 ACATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTAT 712
              ||| | ||||||||| | ||||| |||||  |   |    |||   |   | |  ||||
Db      12269 ACACACACACACACACACATATATATATAGTTGGCCGGGCATAGGGGCTCACACCTGTAA 12328

Qy        713 A 713
              |
Db      12329 A 12329
```

```
RESULT 8
AC009058
LOCUS       AC009058              164844 bp    DNA     linear   PRI 13-MAR-2003
DEFINITION  Homo sapiens chromosome 16 clone RP11-292B23, complete sequence.
ACCESSION   AC009058
VERSION     AC009058.10  GI:28933531
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 164844)
  AUTHORS   DOE Joint Genome Institute, Stanford Human Genome Center and Los
            Alamos National Laboratory.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 164844)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 164844)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-DEC-2001) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   4  (bases 1 to 164844)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-2001) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   5  (bases 1 to 164844)
  AUTHORS   DOE Joint Genome Institute, Stanford Human Genome Center and Los
            Alamos National Laboratory.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAR-2003) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Mar 13, 2003 this sequence version replaced gi:17432456.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center and Los Alamos
            National Laboratory
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 100% of Sequence;
            Estimated Total Number of Errors is 0.
FEATURES             Location/Qualifiers
     source          1. .164844
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone="RP11-292B23"
ORIGIN
```

```
  Query Match             8.0%;  Score 396.8;  DB 9;  Length 164844;
  Best Local Similarity   75.4%;  Pred. No. 8.6e-70;
  Matches  559;  Conservative    0;  Mismatches  173;  Indels    9;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAATTA 60
              ||||| | |||| ||||  |||||| ||||||||||||||||||||||||     |||||||
Db     140343 AGTTCGAAACCACCCTGGCCAACATAGTGAAACCCCATCTCTACTAAAATACAAAAATTA 140402

Qy         61 GACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAAT 120
              | ||||| |||||  | |||||||||||||||||||||  |||||| |||||||||| ||
Db     140403 GCCAGGCGTGGTGGCAGACGCCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGAGGAT 140462

Qy        121 CACTTGAACCTGCTGGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCCAG 179
              |||||||||  | || ||  |||||||| | ||| | ||||  ||||| || |||||||
Db     140463 TGCTTGAACC--CAGGAGGCAGAGGTTGCAGTGAGCCCAGATTGTGCCACTGGACTCCAG 140520

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  ||  |||        | |         |  | | | | |          |||||||||
Db     140521 CCTGGGAGACAGAGTAAGACTCCGTCTCAAATACATATATACATACACACATATATATAC 140580

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACAC----ATAT 295
              | ||||| ||| ||||| | | | ||| ||||| ||||||| ||| ||||||     ||||
Db     140581 ACATATACATACACACATATATATACACATATACATACACACATACATACACACATATAT 140640

Qy        296 ATACATATATACACATATATACACATATATATACACATACATATGTACACATATATAT-A 354
              | ||| |||||||| ||||||||||||||||||||||| ||||| || || ||| ||||||| |
Db     140641 ACACACATATACATATATATACACATATATATACATATACACATATATACACATATATAA 140700

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              | | |   ||| | ||| ||| | | | ||| |||||||||||||| ||||| ||||| |
Db     140701 TATACACACATATATATACATATATACACATATACACATATATACATATATACACACATA 140760

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              |||| |||||||||||||||||||||| ||| |||||||||||| ||| | |||||||
Db     140761 TATACACACATATATACACATATATACATATACACACATATATACATATACACACATATA 140820

Qy        475 TACACATATATATAATATACACACATATATATACACATATATACACACATATATACACAT 534
              ||||||| | |  | |||||||||| | ||| ||||||| | ||||| | | | ||||||
Db     140821 CACACATACACA-CACATACACACACACATACACACATACACACACACATACACACACAT 140879

Qy        535 ATATACACATATATATACACATATATACACATATATACATATATACACATATATATACAT 594
              | | | ||| | ||| ||||| |||||||| ||| |  || | | ||| | |||||||
Db     140880 ACACATACACACATACACACACATATACACACATACATGTACACATACACACATATACAC 140939

Qy        595 ATATACACATATATACATATATACACATATATACATATATACACACATATATACACATAC 654
              ||||||||| |||||||||||||||| ||||||||||||||||||||||||| |||||
Db     140940 ATATACACACATATACATATATACACACATATACATATATACACACATATATATACATAT 140999

Qy        655 ATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATAT 714
              ||| ||| | | | ||||||| |||| |||| ||||| |||  || | | |  |||||
Db     141000 ATACACATATATACATATATACACATATGTACATATATACATATATACACACACATATAT 141059

Qy        715 ATATATGCTCCAGAGTTCATA 735
              | |||   |     | ||||
Db     141060 GTGTATATACACACATACATA 141080
```

```
RESULT 9
AC011593
LOCUS       AC011593              130349 bp    DNA     linear   PRI 07-NOV-2001
DEFINITION  Homo sapiens BAC clone RP11-437E9 from 2, complete sequence.
ACCESSION   AC011593
VERSION     AC011593.8  GI:14190772
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 130349)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
   PUBMED   9847074
REFERENCE   2  (bases 1 to 130349)
  AUTHORS   Pearman,C. and Cotton,M.
  TITLE     The sequence of Homo sapiens BAC clone RP11-437E9
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 130349)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 130349)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-MAY-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 130349)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-NOV-2001) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On May 23, 2001 this sequence version replaced gi:13518261.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0437E09
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. John D.
            McPherson, Department of Genetics, Washington University, St. Louis
            MO.  For additional information about the map position of this
            sequence, see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries.  Genomics 51:1-8.  The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at the Roswell Park Cancer Institute
            (http://bacpac.med.buffalo.edu)
            VECTOR:  pBACe3.6
            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is RP11-289J14, 200 bp overlap; the
            clone sequenced to the right is AC061984.  Actual start of this
            clone is at base position 150577 of RP11-289J14; actual end is at
            base position 130349 of RP11-437E9.
FEATURES              Location/Qualifiers
     source           1. .130349
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:9606"
                      /chromosome="2"
                      /map="2"
                      /clone="RP11-437E9"
                      /clone_lib="RPCI-11"
     repeat_region    1813. .2297
                      /rpt_family="MaLR"
     repeat_region    2618. .3136
                      /rpt_family="MaLR"
     repeat_region    3200. .3370
                      /rpt_family="MaLR"
     repeat_region    3522. .3856
                      /rpt_family="Alu"
     repeat_region    3866. .4290
                      /rpt_family="MaLR"
     repeat_region    4325. .4805
                      /rpt_family="ERV1"
     repeat_region    4910. .5014
                      /rpt_family="MER1_type"
     repeat_region    5258. .5600
                      /rpt_family="L1"
     repeat_region    5601. .5772
                      /rpt_family="Alu"
     repeat_region    5773. .6644
                      /rpt_family="L1"
     repeat_region    6773. .7356
                      /rpt_family="L1"
     repeat_region    7389. .7516
                      /rpt_family="L1"
     repeat_region    7529. .8188
                      /rpt_family="L1"
     repeat_region    8189. .8490
                      /rpt_family="Alu"
     repeat_region    8491. .8590
                      /rpt_family="L1"
     repeat_region    8605. .8943
                      /rpt_family="L1"
     repeat_region    9017. .9042
                      /rpt_family="(CA)n"
     repeat_region    9062. .9104
                      /rpt_family="MER53"
     repeat_region    9105. .9394
                      /rpt_family="Alu"
     repeat_region    9395. .9508
                      /rpt_family="MER53"
     repeat_region    9786. .9966
                      /rpt_family="C-rich"
     repeat_region    9946. .10036
                      /rpt_family="(TCC)n"
     repeat_region    10080. .10374
                      /rpt_family="Alu"
     repeat_region    12491. .12521
                      /rpt_family="(CA)n"
     repeat_region    13138. .13321
                      /rpt_family="MIR"
     repeat_region    13414. .13442
                      /rpt_family="(TAA)n"
     repeat_region    13615. .13925
                      /rpt_family="Alu"
     repeat_region    16183. .16266
                      /rpt_family="MIR"
     repeat_region    16629. .16997
                      /rpt_family="MaLR"
     repeat_region    17643. .17821
                      /rpt_family="Tc2"
     repeat_region    18157. .18200
                      /rpt_family="(TG)n"
     repeat_region    19591. .19619
                      /rpt_family="AT_rich"
     repeat_region    21204. .21375
```

```
                /rpt_family="MaLR"
     repeat_region   21376. .21940
                /rpt_family="ERVL"
     repeat_region   21962. .22901
                /rpt_family="L1"
     repeat_region   22909. .23443
                /rpt_family="L1"
     repeat_region   23492. .23987
                /rpt_family="L1"
     repeat_region   23988. .24016
                /rpt_family="(TTTTG)n"
     repeat_region   24017. .24634
                /rpt_family="L1"
     repeat_region   24636. .25726
                /rpt_family="L1"
     repeat_region   25727. .26040
                /rpt_family="Alu"
     repeat_region   26041. .27537
                /rpt_family="L1"
     repeat_region   27607. .28123
                /rpt_family="ERVL"
     repeat_region   28282. .28808
                /rpt_family="ERVL"
     repeat_region   28809. .29192
                /rpt_family="MaLR"
     repeat_region   29193. .29577
                /rpt_family="ERVL"
     repeat_region   29628. .29895
                /rpt_family="Alu"
     repeat_region   29960. .30264
                /rpt_family="ERVL"
     repeat_region   30265. .30573
                /rpt_family="Alu"
     repeat_region   30574. .30753
                /rpt_family="ERVL"
     repeat_region   30774. .32056
                /rpt_family="ERVL"
     repeat_region   32100. .32242
                /rpt_family="ERVL"
     repeat_region   32477. .32750
                /rpt_family="MaLR"
     repeat_region   32751. .32847
                /rpt_family="L1"
     repeat_region   32848. .33164
                /rpt_family="Alu"
     repeat_region   33165. .33302

  Query Match            8.0%;  Score 395.2;  DB 9;  Length 130349;
  Best Local Similarity  67.8%;  Pred. No. 1.9e-69;
  Matches  645;  Conservative    0;  Mismatches  293;  Indels   14;  Gaps    6;

Qy          4 TCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC--AAAATTAG 61
              ||||||||| ||||   |||||||||| |||||||||||||||||||||  ||||||||
Db      39934 TCAAGACCATCCTGGCTAACATGGTGAATCCCCATCTCTACTAAAAATACAAAAAATTAG 39993

Qy         62 ACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAATC 121
               ||||| |||||       ||||||| |||||||||||  |||||| ||||||||||||
Db      39994 CCAGGCGTGGTGGCGGGTGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATG 40053

Qy        122 ACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAGCC 181
               | ||||||  | || ||||||| ||||   |||| |||||   ||||||||||||||||
Db      40054 GCGTGAACC--CGGGAGGTGGAGCTTGCAATGAGCCAGATCGCACCATTGCACTCCAGCC 40111

Qy        182 CAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTT---TATATATATA 238
               | ||||  |||| | |   ||||| | ||| | | | |    || |   || ||||||| 
Db      40112 TGGACAAC-GGAGCAAGACTCCATCTAAAAAATATATATAAATATATACATAAATATATA 40170

Qy        239 TATATATATATATACACACACACACATATATATATACACATATATATACACATATATA 298
              |||||| ||||| ||| | | | | | ||||| ||||| ||||| ||||| |   ||| |
Db      40171 TATATAAATATAGACATAAATATATATATATAAATATAGACATAAATATATATGCATAAA 40230

Qy        299 CATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACA 358
              |  |||||||  |||| ||| | ||| | ||| | ||| |||| |||| | ||||||| | |

Db      40231 TATATATATGCATAAATATATATAAAAATATATATAAATATATACATAAATATATATAAA 40290

Qy        359 TATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATA 418
              ||| || | | ||||||| | ||||||| | | ||||||| | |||||||||||| |||||
Db      40291 TATATACAAAAATATATATAAATATATAAAAAAATATATAAATATATATACACATATATA 40350

Qy        419 TACACATATATACACATATATACACATATACACATATATACACATATATACATATATACA 478
              | | ||||| | ||||||||| ||||||| ||||| ||| | || |||| ||||||| |
Db      40351 AATATATATACATACATATATAAACATATATACATAAATATATATGTATAAATATATATA 40410

Qy        479 CATATATATA---ATATACACACATATATATACACATATATACACACATATATACACATA 535
              || ||| |||    |||  | | ||||| ||| | ||||||  | | |||||||  |||||
Db      40411 CACATAAATATATGTATGAATATATATACATAAATATATATGTATAAATATATATACATA 40470

Qy        536 TATACACATATATATACACATATATACACATATATACATATATACACATATATATACATA 595
              ||| | | | ||||| ||||| ||| | ||| ||| ||||| | | |||||||||  |||
Db      40471 AATATATAAAGATATATACATAAATATATATAAATATATATACATAAATATATATAAATA 40530

Qy        596 TATACACATATATACATATATACACATATATACATATATACACACATATATACACATACA 655
              |||| | ||| ||| ||| ||| | ||||| | ||||| | | | ||||| | | ||| |
Db      40531 TATATAAATAGATATATAAATATATATATAAATATATAAATATATATATAAATATATAAA 40590

Qy        656 TATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATA 715
              |||| | | | | ||||| |||||||| | ||||||| |  ||||||||  ||||||
Db      40591 TATATAAAAATAGATATATAAATATATATATAAATATATAAATATATATATAAATATATA 40650

Qy        716 TATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGA 775
              || ||     |   |  |||  |      |      ||  | |    ||| |   | | |
Db      40651 TAAATATATAAATATATATATAAATATATATAAATATATAAATATATATAAATATATAAA 40710

Qy        776 GAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGA 835
              |  |  | |    | |||     | |   | || |      |     |||||       |
Db      40711 TATATATATAAATATATATAAATATATAAATATATATAAATATATATAAATATATAAATA 40770

Qy        836 AATACATAAATGTGTAATAAAACCTATCTGTA---AGGTAGTTAAGAAGGTAACACTATA 892
              ||| |||||| | ||  || |  ||  | ||   |   |  ||  ||  ||  | ||||
Db      40771 TATATATAAATATATATAAATATATAAATATATATATAAATATATAAATATATAAATATA 40830

Qy        893 TATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAA 944
              |||||| |||   |||    | ||||  || |  |      ||  || || |
Db      40831 TATATAAATATATAAATATATATAAATATATATAAATATATAAATATATATA 40882


RESULT 10
AC097269/c
LOCUS       AC097269               196413 bp    DNA     linear   HTG 20-OCT-2001
DEFINITION  Pan troglodytes clone RP43-11J15, WORKING DRAFT SEQUENCE, 18
            unordered pieces.
ACCESSION   AC097269
VERSION     AC097269.1  GI:16117532
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 196413)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
```

```
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 196413)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-OCT-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: ZUAW
                Center clone name: RP43-11J15
            -------------- Summary Statistics
                Sequencing vector: Plasmid; M77789
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 190674 bases at least Q40
                Consensus quality: 194102 bases at least Q30
                Consensus quality: 196834 bases at least Q20
                Estimated insert size: 194298; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 8.1x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 18 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    17120: contig of 17120 bp in length
            *    17121    17220: gap of unknown length
            *    17221    36306: contig of 19086 bp in length
            *    36307    36406: gap of unknown length
            *    36407    58958: contig of 22552 bp in length
            *    58959    59058: gap of unknown length
            *    59059    82373: contig of 23315 bp in length
            *    82374    82473: gap of unknown length
            *    82474    96867: contig of 14394 bp in length
            *    96868    96967: gap of unknown length
            *    96968   109418: contig of 12451 bp in length
            *   109419   109518: gap of unknown length
            *   109519   120755: contig of 11237 bp in length
            *   120756   120855: gap of unknown length
            *   120856   134377: contig of 13522 bp in length
            *   134378   134477: gap of unknown length
            *   134478   144432: contig of 9955 bp in length
            *   144433   144532: gap of unknown length
            *   144533   155013: contig of 10481 bp in length
            *   155014   155113: gap of unknown length
            *   155114   161933: contig of 6820 bp in length
            *   161934   162033: gap of unknown length
            *   162034   168622: contig of 6589 bp in length
            *   168623   168722: gap of unknown length
            *   168723   175724: contig of 7002 bp in length
            *   175725   175824: gap of unknown length
            *   175825   180213: contig of 4389 bp in length
            *   180214   180313: gap of unknown length
            *   180314   186640: contig of 6327 bp in length
            *   186641   186740: gap of unknown length
            *   186741   190642: contig of 3902 bp in length
            *   190643   190742: gap of unknown length
            *   190743   194248: contig of 3506 bp in length
            *   194249   194348: gap of unknown length
            *   194349   196413: contig of 2065 bp in length.
FEATURES             Location/Qualifiers
     source          1. .196413
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="RP43-11J15"
ORIGIN

  Query Match             7.9%;  Score 394.8;  DB 2;  Length 196413;
  Best Local Similarity   75.8%;  Pred. No. 2.1e-69;
  Matches  547;  Conservative    0;  Mismatches  157;  Indels   18;  Gaps    4;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC------- 53
              |||||||||||||||||| ||||||||| |||||||||||| |||||| |
Db     196321 AGTTCAAGACCAGCCTGACCAACATGGTGGAACCCCATCTCTAGTAAAAAAAAAAAAAAA 196262

Qy         54 -------AAAATTAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGC 106
                    |||||||| |||| |||||   || ||||| ||||| | |||||| ||||||
Db     196261 AAAAAAAAAAATTAGCCAGGTGTGGTGGGGCATGCCTGAAATCCTAACTACTTGGGAGGC 196202

Qy        107 CGAGGCAGGAGAATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGC-AAGATCATG 165
              ||||||||| ||| |||||| |   |  | ||||||||||||   ||||  | |||| |
Db     196201 TGAGGCAGGAAAATTGCTTGAAGC--CAAGAGGTGGAGGTTGCAATGAGCTGATATCACG 196144

Qy        166 CCATTGCACTCCAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAGGCA 225
              || |||||||||||||  |||||||| |  ||||   ||||||| |||||| |   | |
Db     196143 CCGTTGCACTCCAGCCTGGGCAACAAAAATGAAACTCCATCTCAAAAAAAAGAGGGGCCG 196084

Qy        226 TTTTATATATATATATATATATATACACACACACACACATATATATATACACATATAT 285
                |  |  ||  |  || | || || | | | | | | |||||||||||| | ||||||
Db     196083 GGAGAGGTGGATCAAGCTAGAAATCTATATATATATATATATATATATATATATATAT 196024

Qy        286 ATACACATATATACATATATACACATATATACACATATATATACACATACATATGTACAC 345
              ||| | ||||||| ||||||| | ||||||| | |||||||| | ||| |||| || |
Db     196023 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 195964

Qy        346 ATATATATACACATATGTATACACATATATACACATATATACACACATATATACACATAT 405
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | ||||
Db     195963 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 195904

Qy        406 ATACACACATATATACACATATATACACATATATACACATATACACATATATACACATAT 465
              ||| | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | ||||
Db     195903 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 195844

Qy        466 ATACATATATACACATATATATAATATACACACATATATATACACATATATACACACATA 525
              ||| ||||||| | |||||||| ||||| | | |||||||||| | ||||||| | | |||
Db     195843 ATATATATATATATATATATAT-ATATATATATATATATATATATATATATATATATATA 195785

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | |||||||| | ||||||| | ||||||| ||||||| | |||
Db     195784 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 195725
```

```
Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db     195724 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 195665

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | ||||||||   |||  |
Db     195664 TATATATATATATATATATATATATATATATATATATATATATATATATATCTATATCCA 195605

Qy        706 TA 707
              ||
Db     195604 CA 195603
```

RESULT 11
AC015468
LOCUS AC015468 189662 bp DNA linear PRI 01-MAY-2001
DEFINITION Homo sapiens chromosome 8, clone RP11-369E15, complete sequence.
ACCESSION AC015468
VERSION AC015468.5 GI:13899433
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 189662)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 8, clone RP11-369E15
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 189662)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B., Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 189662)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P., Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (01-MAY-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On May 1, 2001 this sequence version replaced gi:12313808.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
---------- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
---------- Project Information
Center project name: L2466
Center clone name: 369_E_15
----------.

```
FEATURES             Location/Qualifiers
     source          1. .189662
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="RP11-369E15"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   1. .100
                     /rpt_family="MLT1A1"
     repeat_region   1594. .1741
                     /rpt_family="GA-rich"
     repeat_region   complement(2782. .2957)
                     /rpt_family="MIR"
     repeat_region   3311. .3471
                     /rpt_family="MIR"
     repeat_region   3964. .4067
                     /rpt_family="MIR"
     repeat_region   complement(4528. .4771)
                     /rpt_family="L2"
     repeat_region   4776. .4871
                     /rpt_family="MER4A"
     repeat_region   4869. .5223
                     /rpt_family="MER4A"
     repeat_region   complement(5255. .5765)
                     /rpt_family="L2"
     repeat_region   complement(5805. .6003)
                     /rpt_family="THE1A"
     repeat_region   6004. .6035
                     /rpt_family="(TTA)n"
     repeat_region   complement(6151. .6319)
                     /rpt_family="SVA"
     repeat_region   complement(6320. .6468)
                     /rpt_family="THE1A"
     repeat_region   8722. .8751
                     /rpt_family="(TTTTA)n"
     repeat_region   9155. .9553
                     /rpt_family="MSTB"
     repeat_region   9599. .9982
                     /rpt_family="THE1C"
     repeat_region   complement(10113. .10320)
                     /rpt_family="AluY"
     repeat_region   10822. .10970
                     /rpt_family="MIR3"
     unsure          11210. .11215
                     /note="<30 qual SNGL region"
     repeat_region   11794. .12080
                     /rpt_family="AluSx"
     repeat_region   12974. .13010
                     /rpt_family="(TG)n"
     repeat_region   complement(13091. .13304)
                     /rpt_family="MER106A"
     repeat_region   13484. .13858
                     /rpt_family="MLT1C"
     repeat_region   14025. .14260
                     /rpt_family="L2"
     repeat_region   14584. .14679
                     /rpt_family="MIR"
     repeat_region   14722. .14825
                     /rpt_family="L2"
     repeat_region   15167. .15332
```

```
                     /rpt_family="MER33"
  repeat_region      15436. .15788
                     /rpt_family="L2"
  repeat_region      complement(16078. .16165)
                     /rpt_family="AluJ/FRAM"
  repeat_region      16532. .17094
                     /rpt_family="MLT1F1"
  repeat_region      17653. .17674
                     /rpt_family="(TAA)n"
  repeat_region      complement(17675. .17892)
                     /rpt_family="MER7C"
  repeat_region      complement(17897. .18650)
                     /rpt_family="LTR17"
  repeat_region      complement(18651. .18996)
                     /rpt_family="MER7C"
  repeat_region      18999. .19103
                     /rpt_family="(TAA)n"
  repeat_region      complement(19203. .19283)
                     /rpt_family="MIR"
  repeat_region      20464. .20523
                     /rpt_family="A-rich"
  repeat_region      complement(21279. .21902)
                     /rpt_family="MER41B"
  repeat_region      21922. .22092
                     /rpt_family="MIR"
  repeat_region      22249. .22373
                     /rpt_family="MIR"
  repeat_region      23956. .24058
                     /rpt_family="L2"
  repeat_region      complement(24069. .24198)
                     /rpt_family="MIR"
  repeat_region      25149. .25213
                     /rpt_family="L2"
  repeat_region      complement(25783. .26130)
                     /rpt_family="MLT1I"
  repeat_region      complement(26174. .26539)
                     /rpt_family="MLT1A1"
  repeat_region      complement(28187. .28528)
                     /rpt_family="AluY"
  repeat_region      complement(28792. .29263)
                     /rpt_family="Charlie8"
  repeat_region      complement(30652. .30730)
                     /rpt_family="MER53"
  repeat_region      complement(30986. .31117)
                     /rpt_family="L1ME1"
  repeat_region      31113. .31556
                     /rpt_family="L1ME2"
  repeat_region      31590. .32174
                     /rpt_family="L1ME2"
  repeat_region      32162. .32588
                     /rpt_family="L1ME3"
  repeat_region      33242. .33490
                     /rpt_family="MIR"
  repeat_region      complement(34416. .34803)
                     /rpt_family="MSTB"
  repeat_region      complement(35044. .35350)
                     /rpt_family="AluSg"
  repeat_region      36407. .36695
                     /rpt_family="AluSx"
  repeat_region      complement(37670. .37790)
                     /rpt_family="FLAM_A"
  repeat_region      complement(38065. .38132)
                     /rpt_family="L2"
  repeat_region      complement(38504. .38590)
                     /rpt_family="MIR"
  repeat_region      39044. .39228
                     /rpt_family="MIR3"
  repeat_region      complement(40317. .40473)
                     /rpt_family="MIR"

Query Match             7.9%;  Score 390.8;  DB 9;  Length 189662;
Best Local Similarity   65.9%;  Pred. No. 1.3e-68;
Matches  643;  Conservative    0;  Mismatches  327;  Indels    6;  Gaps    5;
```

```
Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAAT-ACAAAATT 59
              ||||| |||||  ||||  ||||||| ||||||||||||||||||||||| | ||||||
Db      65793 AGTTCGAGACCCACCTGGCCAACATGATGAAACCCCATCTCTACTAAAAATGAAAAAATT 65852

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGA-GGCCGAGGCAGGAGA 118
              ||   ||  |||||  |   || | |||||||||||||  ||| ||| ||||||| |||
Db      65853 AGCTGGGTGTGGTGGCAGGTGCATCTAATCCCAGCTACTCTGGAGGGCTGAGGCAGAAGA 65912

Qy        119 ATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCC 177
              ||  ||||||||  | || |||||  |||||  | |||  | |||||  ||||||||||||
Db      65913 ATTGCTTGAACC--CAGGAGGTGGACGTTGCAGTGAGCCGAGATCGCGCCATTGCACTCC 65970

Qy        178 AGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATAT 237
              ||||  ||||||||||||| |||      |||||  |||||||||          |||||||||
Db      65971 AGCCTGGGCAACAAGAGCAAAACTCTGTCTCAAAAAAAAAAATTTTATAAATATATATAT 66030

Qy        238 ATATATATATATATACACACACACACATATATATATACACATATATATACACATATAT 297
              | ||||||||||| | | | | | | |||||||  ||| | |||  ||||| | | ||||
Db      66031 AAATATATATATAAATATATATAAATATATATATAAATATATATAAATATATATAAATAT 66090

Qy        298 ACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACAC 357
              | ||||| | | ||||| | | ||||| ||| | ||| |||| || | |||||||||  |
Db      66091 ATATATAAATATATATAAATATATATAAATATATATAAATATATATAAATATATATATAA 66150

Qy        358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417
              |||| |||| | ||||| | | ||||||| | | ||||||| | ||||||| | | ||||
Db      66151 ATATATATAAATATATAAAAATATATATATAAATATATATAAATATATATAAATATATAT 66210

Qy        418 ATACACATATATACACATATATACACATATACACATATATACACATATATACATATATAC 477
              | | | ||||||| | ||||||| | ||||| | ||||||| | | | ||| |||||||
Db      66211 AAATATATATATAAATATATATAAATATATATAAATATATATATAAAAATATATATATAA 66270

Qy        478 ACATATATATAATATACACACATATATATACACATATATACACACATATATACACATATA 537
              | ||||||| |||||| | | | ||||||| | ||||||| | | ||||||| | || ||
Db      66271 ATATATATA-AATATATATATAAATATATATAAATATATATATAAATATATAAATATGTA 66329

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATACATATA 597
              || | ||||||| | |  |||||| | || |||| | ||||| | |||| |||| | |||
Db      66330 TATAAATATATAAATATGTATATAAATATGTATATAAATATATAAATATGTATATAAATA 66389

Qy        598 TACACATATATACATATATACACATATATACATATATACACACATATATACACATACATA 657
              || | ||| ||| ||| |||    ||||| | ||||||| | | ||||||| | ||  |||
Db      66390 TATATATAAATATATAAATATGTATATAAATATATATATAAATATATATAAATATGTATA 66449

Qy        658 TACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATA 717
              || | | | ||| ||||  ||||| ||| | ||||||| |  ||||||  | |||||||||
Db      66450 TAAATATATATAAATATGTATATAAATATATATATATAAATATATATAAAAATATATATA 66509

Qy        718 TATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGA 777
              ||        | |  |||         |  | | ||     |    ||| |   ||| | |
Db      66510 TAAATATATAAATATATATATAAATATATATAAATATATATAAATATATATATAAATATA 66569

Qy        778 GTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAA 837
                  | |   | ||   | |  | || |        |    |||||       | |
Db      66570 TAAATATATATATATAAATATATATATAAATATATATAAAAATATAAATATATATAAATA 66629

Qy        838 TACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATATAT 897
              || ||| |  | ||  | |  |      |   ||  |   |    |||| ||||
Db      66630 TATATATAAATATATATATAAAAAATATATAAAAATATATATATAAATATATATAAATAT 66689

Qy        898 ATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTATAGAAGAA 957
              |||||   | |       || |  || ||  |   ||  | | |||| || |     ||
Db      66690 ATATAAATATATATAAATATATATATAGTAGTGAAGAATCACTAAAATATTTTTATGAAA 66749

Qy        958 TGGCTAAGATGCCAAA 973
                ||  ||  | | |
Db      66750 AAAGTATGAATCTAGA 66765
```

RESULT 12
AC073310/c
LOCUS AC073310 125685 bp DNA linear PRI 08-OCT-2003

```
DEFINITION  Homo sapiens BAC clone RP11-49G5 from 7, complete sequence.
ACCESSION   AC073310
VERSION     AC073310.7  GI:13176604
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 125685)
  AUTHORS   Sulston,J.E. and Wilson,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
   PUBMED   9847074
REFERENCE   2  (bases 1 to 125685)
  AUTHORS   Haakenson,W., Hannah,C. and Kang,K.
  TITLE     The sequence of Homo sapiens BAC clone RP11-49G5
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 125685)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUN-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 125685)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 125685)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2001) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   6  (bases 1 to 125685)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2001) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   7  (bases 1 to 125685)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-APR-2003) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   8  (bases 1 to 125685)
  AUTHORS   Wilson,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-OCT-2003) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Mar 1, 2001 this sequence version replaced gi:11024926.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0049G05
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            The sequence of this clone was established as part of a mapping and
            sequencing collaboration between the NHGRI Chromosome 7 Mapping
            Project (Eric D. Green, Director), John D. McPherson in the
            Department of Genetics (Washington University), and the Washington
            University Genome Sequencing Center. For additional information
            about the map position of this sequence, see
            http://www.nhgri.nih.gov/DIR/GTB/CHR7 , send
            mailto:egreen@nhgri.nih.gov , or see http://genome.wustl.edu

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries. Genomics 51:1-8. The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at http://www.chori.org
            VECTOR:  pBACe3.6

            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is RP11-409M7, 200 bp overlap the
            clone sequenced to the right is RP11-374N8.  Actual start of this
            clone is at base position 14572 of RP11-409M7 actual end is at base
            position 125685 of RP11-49G5.
FEATURES             Location/Qualifiers
     source          1. .125685
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7"
                     /clone="RP11-49G5"
                     /clone_lib="RPCI-11"
     repeat_region   1013. .1151
                     /rpt_family="CR1"
     repeat_region   2238. .2394
                     /rpt_family="MIR"
     repeat_region   2548. .2668
                     /rpt_family="L2"
     repeat_region   3206. .5231
                     /rpt_family="L1"
     repeat_region   5694. .5715
                     /rpt_family="AT_rich"
     repeat_region   7142. .7353
                     /rpt_family="MER1_type"
     repeat_region   7490. .7666
                     /rpt_family="ERVL"
     repeat_region   7671. .7771
                     /rpt_family="L1"
     repeat_region   7951. .7985
                     /rpt_family="AT_rich"
     repeat_region   9572. .9809
                     /rpt_family="L1"
     repeat_region   9811. .9831
                     /rpt_family="AT_rich"
     repeat_region   9895. .10017
                     /rpt_family="MaLR"
     repeat_region   10320. .10375
                     /rpt_family="MIR"
     repeat_region   10395. .10732
                     /rpt_family="L1"
     repeat_region   10911. .11118
                     /rpt_family="(TA)n"
     unsure          10914. .11684
                     /note="unresolved simple sequence repeat"
     repeat_region   11130. .11196
                     /rpt_family="L1"
     repeat_region   11217. .11663
                     /rpt_family="(TA)n"
     repeat_region   11667. .11918
                     /rpt_family="Alu"
     repeat_region   11919. .11970
```

```
                     /rpt_family="L1"
     repeat_region   12356. .12508
                     /rpt_family="L1"
     repeat_region   12646. .12813
                     /rpt_family="L1"
     repeat_region   12863. .13111
                     /rpt_family="L1"
     repeat_region   13280. .13592
                     /rpt_family="L1"
     repeat_region   13768. .14032
                     /rpt_family="L1"
     repeat_region   14048. .14489
                     /rpt_family="L1"
     repeat_region   15108. .15378
                     /rpt_family="Alu"
     repeat_region   15563. .15584
                     /rpt_family="AT_rich"
     repeat_region   17280. .17442
                     /rpt_family="MIR"
     repeat_region   18313. .18467
                     /rpt_family="L1"
     repeat_region   18468. .18676
                     /rpt_family="L1"
     repeat_region   18677. .18701
                     /rpt_family="AT_rich"
     repeat_region   18797. .19155
                     /rpt_family="MaLR"
     repeat_region   19594. .19911
                     /rpt_family="Alu"
     repeat_region   21451. .21639
                     /rpt_family="MER53"
     repeat_region   22809. .22986
                     /rpt_family="L2"
     repeat_region   23094. .23371
                     /rpt_family="L1"
     repeat_region   23372. .23678
                     /rpt_family="Alu"
     repeat_region   23679. .23806
                     /rpt_family="L1"
     repeat_region   23820. .23999
                     /rpt_family="L1"
     repeat_region   24445. .24470
                     /rpt_family="AT_rich"
     repeat_region   24596. .24649
                     /rpt_family="(CA)n"
     repeat_region   26292. .26765
                     /rpt_family="L1"
     repeat_region   26766. .27123
                     /rpt_family="MaLR"
     repeat_region   27124. .27202
                     /rpt_family="L1"
     repeat_region   27203. .27556
                     /rpt_family="MaLR"

  Query Match             7.8%;  Score 387.4;  DB 9;  Length 125685;
  Best Local Similarity   67.7%;  Pred. No. 6.8e-68;
  Matches  647;  Conservative    0;  Mismatches  296;  Indels   12;  Gaps    7;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              |||||||||||||||||   ||||||||||||||||||| | | ||||||||||||| ||||||
Db      11886 AGTTCAAGACCAGCCTAGCCAACATGGTGAAACCCCGTTTTTACTAAAAATACAAAAATT 11827

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              || ||||| ||||| |    ||||||||||||||| ||    ||||||  ||||||||||
Db      11826 AGCCAGGCGTGGTGGT---TGCCTGTAATCCCAGCCACACGGGAGGCTAAGGCAGGAGAA 11770

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGC-AAGATCATGCCATTGCACTCCA 178
              |  |||||| |||  || |  ||||||||| | ||||  |||||| |||| |||||||||
Db      11769 TTGCTTGAATCTG--GGAGACGGAGGTTGCAGTGAGCTGAGATCACGCCACTGCACTCCA 11712

Qy        179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATA 238
              |||  ||  || |||| || |   ||||||| ||||||||||||     |  ||||||||||
Db      11711 GCCTGGGTGAC-AGAGTGAGACTCCATCTCAAAAAAAAAAAAAAAAAGTCATATATATATA 11653

Qy        239 TATATATATATATACACACACACACATATATATATACACATATATATACACATATATA 298
              |||||||||||||| | | | | | | | ||||||| | | ||||||| | | |||||
Db      11652 TATATATATATATATATATATATATAAAAATATATATATAAATATATATAAATATA 11593

Qy        299 CATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACA 358
              ||||||| | |||| || | |||||||||| | ||| || | || | ||||||| | | |
Db      11592 TATATATAAATATATTTATAAATATATATATAAATATATTTATAAATATATATAAATATA 11533

Qy        359 TATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATA 418
              ||| |||| | ||||||| | ||||||  | |-| ||||| | ||| ||| | | |||||
Db      11532 TATATATAAATATATATATAAATATATGTATAAAAATATATATATAAATATAAATATATA 11473

Qy        419 TACACATATATACACATATATACACATATACACATATATACACATATATACATATATACA 478
              || | ||||||| | | | ||| | ||||| | ||| ||| | ||||| | ||||||| |
Db      11472 TATAAATATATATATAAAAATATATATATAAATATAAATATATATATAAATATATATATA 11413

Qy        479 CATATATATA---ATATACACACATATATATACACATATATACACACATATATACACATA 535
              |||||||||    ||||| | | ||||| ||| | ||||||| | | ||||| | |  ||
Db      11412 AATATATATATATATATAAATATATATAAATATATATATATATATATATATAAATATCTA 11353

Qy        536 TATACACATATATATACACATATATACACATATATACATATATACACATATATATACATA 595
              |||| |||||||||| | ||||||| | | ||| |||||||| | |||||||||| | |
Db      11352 TATAAACATATATATATAAATATATATATAAACATATATATATAAATATATATATAAACA 11293

Qy        596 TATACACATATATACATATATACACATATATACATATATACACACATATATACACATACA 655
              |||| | ||| ||| ||||||| |||||||| ||| ||| | | ||||||| | ||| |
Db      11292 TATATATATAAATATATATATAAACATATATATATAAATATATATATATATATATATA 11233

Qy        656 TATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATA 715
              |||| | | ||| | | | |||||||| || |||||||||    | |||||| |||| |
Db      11232 TATATATATATATATATAAACATATATATATTCAAATATATATATAAACATATATATATAA 11173

Qy        716 TATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGA 775
              |||||    |  | |||          |  |   ||   |      |  | | |
Db      11172 TATATATATAAACATATATATATAAATATATATATAAACATATATAATATAACATATATA 11113

Qy        776 GAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGA 835
              |  |  | |    | ||     | |   | || |     | |    | |||       |
Db      11112 TATATATATATATATATAAATATATATATAAATATATATATAAATATATATATAAATA 11053

Qy        836 AATACATAAATGTGTAATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTAACACTATATA 894
              ||| |||||| | ||   | |  ||| | || |  ||  ||   |   |  ||||||
Db      11052 TATATATAAATATATATATATATATATATATAAATATATATATATAAATATATATATA 10993

Qy        895 TATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTA 949
              || |||||   | |    | || |  ||||         ||  || || || | |
Db      10992 TAAATATATATATATATATATATATAAATATATATATAAATATATATATATAA 10938


RESULT 13
AL133477
LOCUS       AL133477              139118 bp    DNA     linear   PRI 02-FEB-2001
DEFINITION  Human DNA sequence from clone RP11-172F4 on chromosome 9q22.2-31.1,
            complete sequence.
ACCESSION   AL133477
VERSION     AL133477.16  GI:12666198
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 139118)
  AUTHORS   Lloyd,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Feb 5, 2001 this sequence version replaced gi:12584335.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
```

```
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows.  An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated repeat sequence elements.  Where the sequence is
ambiguous, there is an annotation using the 'unsure' feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
Group.  Further information can be found at
http://www.sanger.ac.uk/HGP/Chr9
RP11-172F4 is from the library RPCI-11.1 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
IMPORTANT: This sequence is not the entire insert of clone
RP11-172F4 It may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true left end of clone RP11-172F4 is at 1 in this sequence. The
true left end of clone RP11-240L7 is at 139019 in this sequence.
The true right end of clone RP11-392G7 is at 43511 in this
sequence.
FEATURES             Location/Qualifiers
     source          1. .139118
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /map="q22.2-31.1"
                     /clone="RP11-172F4"
                     /clone_lib="RPCI-11.1"
     repeat_region   45. .78
                     /note="L1MB2 repeat: matches 6125. .6158 of consensus"
     repeat_region   195. .361
                     /note="MER20 repeat: matches 39. .218 of consensus"
     repeat_region   488. .688
                     /note="MLT1B repeat: matches 205. .390 of consensus"
     repeat_region   689. .986
                     /note="AluY repeat: matches 1. .297 of consensus"
     repeat_region   987. .1144
                     /note="MLT1B repeat: matches 5. .205 of consensus"
     repeat_region   1149. .1500
                     /note="MST-INTERNAL repeat: matches 1289. .1651 of
                     consensus"
     repeat_region   1812. .2130
                     /note="MLT1-INTERNAL repeat: matches 404. .728 of
                     consensus"
     repeat_region   2175. .2472
                     /note="MLT2FB repeat: matches 101. .414 of consensus"
     repeat_region   2729. .2832
                     /note="MER5B repeat: matches 2. .80 of consensus"
     repeat_region   2833. .3146
                     /note="AluSx repeat: matches 1. .308 of consensus"
     repeat_region   3147. .3236
                     /note="MER5B repeat: matches 80. .177 of consensus"
     repeat_region   5694. .6008
                     /note="AluSx repeat: matches 1. .306 of consensus"
     repeat_region   6183. .6318
                     /note="MER5A repeat: matches 10. .189 of consensus"
     repeat_region   6556. .6717
                     /note="FRAM repeat: matches -5. .161 of consensus"
     repeat_region   7313. .7619
                     /note="AluSx repeat: matches 1. .311 of consensus"
     repeat_region   7907. .8217
                     /note="AluY repeat: matches 1. .311 of consensus"
     repeat_region   8545. .8837
                     /note="AluSg repeat: matches 1. .293 of consensus"
     repeat_region   9101. .9389
                     /note="AluJo repeat: matches 1. .282 of consensus"
     repeat_region   9399. .9701
                     /note="AluY repeat: matches 1. .311 of consensus"
     repeat_region   9702. .10014
                     /note="AluSx repeat: matches 1. .311 of consensus"
     repeat_region   10050. .10292
                     /note="AluSp repeat: matches 66. .310 of consensus"
     repeat_region   10300. .10615
                     /note="AluSx repeat: matches 4. .312 of consensus"
     repeat_region   14460. .14754
                     /note="AluJo repeat: matches 13. .308 of consensus"
     repeat_region   14846. .15042
                     /note="AluJb repeat: matches 96. .296 of consensus"
     repeat_region   15166. .15281
                     /note="AluJb repeat: matches 9. .96 of consensus"
     repeat_region   15506. .15781
                     /note="AluSg repeat: matches 1. .276 of consensus"
     repeat_region   16165. .16466
                     /note="AluSx repeat: matches 1. .300 of consensus"
     repeat_region   18757. .19057
                     /note="AluSx repeat: matches 1. .301 of consensus"
     repeat_region   19097. .19144
                     /note="12 copies 4 mer tata 77% conserved"
     repeat_region   19359. .19662
                     /note="AluSg repeat: matches 1. .304 of consensus"
     repeat_region   19743. .20060
                     /note="AluSx repeat: matches 1. .305 of consensus"
     repeat_region   21872. .22179
                     /note="AluSx repeat: matches 1. .299 of consensus"
     repeat_region   27049. .27427
                     /note="MLT1B repeat: matches 7. .390 of consensus"
     repeat_region   29141. .29384
                     /note="MIR repeat: matches 2. .260 of consensus"
     misc_feature    29833. .30559
                     /note="CpG island"
                     /evidence=not_experimental
     repeat_region   31569. .31653
                     /note="MER91A repeat: matches 3. .63 of consensus"
     repeat_region   31939. .32227
                     /note="AluY repeat: matches 1. .291 of consensus"
     repeat_region   32228. .32323
                     /note="48 copies 2 mer gg 67% conserved"
     repeat_region   32248. .32297
                     /note="10 copies 5 mer gacgg 88% conserved"
     repeat_region   32324. .32378
                     /note="11 copies 5 mer aagaa 78% conserved"
     repeat_region   33239. .33518
                     /note="AluJo repeat: matches 1. .289 of consensus"
     repeat_region   34313. .34610
                     /note="AluSq repeat: matches 1. .298 of consensus"
     repeat_region   35793. .35899
                     /note="MIR repeat: matches 1. .116 of consensus"
     repeat_region   37385. .37604
                     /note="L2 repeat: matches 2436. .2659 of consensus"
     repeat_region   38070. .38181
                     /note="L2 repeat: matches 2563. .2729 of consensus"
     repeat_region   40862. .41123
                     /note="AluSq repeat: matches 39. .302 of consensus"
     repeat_region   41722. .42016
                     /note="AluSp repeat: matches 1. .297 of consensus"
     repeat_region   43048. .43521
                     /note="L1MB5 repeat: matches 4783. .5239 of consensus"
     repeat_region   43522. .43695
                     /note="FAM repeat: matches 1. .174 of consensus"
     repeat_region   43696. .44552
                     /note="L1MB5 repeat: matches 5239. .6138 of consensus"
     repeat_region   44598. .44834
                     /note="L1MB4 repeat: matches 5814. .6065 of consensus"
     repeat_region   44835. .45133
                     /note="AluSg repeat: matches 1. .295 of consensus"
     repeat_region   45134. .45257
                     /note="L1MB4 repeat: matches 6065. .6184 of consensus"
     repeat_region   45312. .45611
```

```
                     /note="AluJo repeat: matches 1. .293 of consensus"
     repeat_region   45681. .45989
                     /note="AluSg repeat: matches 1. .300 of consensus"
     repeat_region   46204. .46347
                     /note="L1ME2 repeat: matches 6007. .6153 of consensus"
     repeat_region   46876. .47188
                     /note="AluJo repeat: matches 1. .304 of consensus"
     repeat_region   47262. .47560
                     /note="AluSq repeat: matches 1. .301 of consensus"
     repeat_region   47923. .48041
                     /note="L1MC4 repeat: matches 6562. .6679 of consensus"
     repeat_region   48060. .48148
                     /note="L1ME1 repeat: matches 5347. .5428 of consensus"
     repeat_region   48149. .48413
                     /note="AluJo repeat: matches 2. .302 of consensus"
     repeat_region   48414. .48641
                     /note="L1ME1 repeat: matches 5428. .5637 of consensus"
     repeat_region   48642. .48935
                     /note="AluSg repeat: matches 1. .294 of consensus"
     repeat_region   48936. .49372
                     /note="L1ME1 repeat: matches 5637. .6160 of consensus"
     repeat_region   50733. .50947
                     /note="MIR repeat: matches 2. .257 of consensus"
     repeat_region   51470. .51541
                     /note="MIR repeat: matches 44. .124 of consensus"
     repeat_region   51667. .51977
                     /note="AluY repeat: matches 1. .311 of consensus"
     repeat_region   52966. .53214
                     /note="MIR repeat: matches 12. .262 of consensus"
     repeat_region   53551. .54965
                     /note="L2 repeat: matches 505. .2101 of consensus"
     repeat_region   54984. .55272
                     /note="AluSp repeat: matches 1. .291 of consensus"

  Query Match             7.8%;  Score 385.8;  DB 9;  Length 139118;
  Best Local Similarity   77.4%;  Pred. No. 1.4e-67;
  Matches  583;  Conservative    0;  Mismatches  132;  Indels   38;  Gaps    8;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC---AAAA 57
              ||||| ||||||||||  ||||||||||||||||||||||||||||||||    ||||
Db      97498 AGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAAAAA 97557

Qy         58 TTAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAG 117
              ||||   ||| |||||  ||| ||||||| || ||||||| | |||||| ||||||||||
Db      97558 TTAGCTGGGCGTGGTGGCACATGCCTGTAGTCTCAGCTACCTGGGAGGCTGAGGCAGGAG 97617

Qy        118 AATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTC 176
              |||| ||||||||  | || ||||||||||||| | ||| | |||||   ||| |||||||
Db      97618 AATCGCTTGAACC--CAGGAGGTGGAGGTTGCAGTGAGCCGAGATCGCTCCACTGCACTC 97675

Qy        177 CAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCAT---------- 226
              | ||   ||| || |||||    |   | ||||| ||||||||||    ||
Db      97676 CGGCTTGGGCGAC-AGAGCATGACTCCGTCTCAAAAAAAAAAAAATATATATATATATAT 97734

Qy        227 ---------------TTTATATATATATATATATATATACACACACACACATATAT 271
                            | |||||||||| | ||||||| ||| | ||| ||| | ||| ||
Db      97735 ATGTGTGTGTGTGTGTATATATATACACATATATACATATATACATACATATATACAT 97794

Qy        272 ATATACACATATATATACACATATATACATATATACACATATATACACATATATATACAC 331
              ||||||| | |||||||||||||||||||||| ||| |||||||||||||| ||||| | ||
Db      97795 ATATACATACATATATACACATATATACATACATATACATATATACACACATATACATAC 97854

Qy        332 ATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACACAC 391
              ||| | || || ||||||||| ||||| ||  ||| | | ||||| ||||||||||||||
Db      97855 ATATACATATATACATATATACACACACATACATATATACATATATACATATATACACAT 97914

Qy        392 ATATATACACATATATACACACATATATACACATATATACACATATATACACATATACAC 451
              ||||| | | ||| ||| | ||| ||||| ||| | ||| ||||| ||| ||||||| ||
Db      97915 ATATACATATATACATATATACACATATATACACACATATACATACATATACATATATAC 97974

Qy        452 ATATATACACATATATACATATATACACATATATATA-ATATACACACATATATATACAC 510
              ||||||||||||||||||||||||| |||||  || ||||||| | ||| |||||  ||
Db      97975 GTATATACACATATATACATATATACATATATACGTATATATACATATATACATATATAC 98034

Qy        511 ATATATACACACATATATACACATATATACACATATATATACACATATATACACATATAT 570
              |||||| ||| ||||| | | |||  || | ||| ||||| ||||||||| ||||||||
Db      98035 GTATATATACATATATACATATATACGTATATATACATATATACATATATATACATATAT 98094

Qy        571 ACATATATACACATAT----ATATACATATATACACATATATACATATATACACATATAT 626
              ||||||||| ||||||    ||||| | ||||| || ||||| |||||||| ||||||||
Db      98095 ACATATATATACATATATACATATATACATATATACGTATATGCATATATATACATATAT 98154

Qy        627 ACATATATACACACATATATACACATACATATACACACACATAGATATACATATATATAC 686
              |||||||| ||| |||||||| ||| | | | | | ||| | ||||||||| ||||||
Db      98155 ACATATATATACATATATATACATATATACACATATATACACA-ATATACATACATATAC 98213

Qy        687 ACATATATATACGTATATATATGTATATATATA 719
              | ||||| | |  || |||||| || |||||||
Db      98214 ATATATACACATATACATATATATACATATATA 98246


RESULT 14
HUAC004787
LOCUS       HUAC004787             216021 bp    DNA     linear   PRI 30-OCT-2002
DEFINITION  Homo sapiens Chromosome 16 BAC clone CIT987SK-A-952F10, complete
            sequence.
ACCESSION   AC004787
VERSION     AC004787.1  GI:3337381
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 216021)
  AUTHORS   Loftus,B.J., Kim,U.J., Sneddon,V.P., Kalush,F., Brandon,R.,
            Fuhrmann,J., Mason,T., Crosby,M.L., Barnstead,M., Cronin,L.,
            Deslattes Mays,A., Cao,Y., Xu,R.X., Kang,H.L., Mitchell,S.,
            Eichler,E.E., Harris,P.C., Venter,J.C. and Adams,M.D.
  TITLE     Genome duplications and other features in 12 Mb of DNA sequence
            from human chromosome 16p and 16q
  JOURNAL   Genomics 60 (3), 295-308 (1999)
  MEDLINE   99425270
   PUBMED   10493829
REFERENCE   2  (bases 1 to 216021)
  AUTHORS   Adams,M.D., Loftus,B.J., Zhou,L., Crosby,M., Fuhrmann,J.,
            Mason,T.M., Brandon,R., Kim,U.J., Kerlavage,A.R. and Venter,J.C.
  TITLE     Homo sapiens Chromosome 16 BAC clone CIT987SK-A-952F10
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 216021)
  AUTHORS   Adams,M.D. and Loftus,B.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1998) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, Email:
            bjloftus@tigr.org
REFERENCE   4  (bases 1 to 216021)
  AUTHORS   Adams,M.D. and Loftus,B.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JUL-1998) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
COMMENT     On Jul 24, 1998 this sequence version replaced gi:3241936.
            Address all correspondence to: Mark Adams The Institute for Genomic
            Research 9712 Medical Center Dr, Rockville, MD 20850, USA e-mail
            address: humgen@tigr.org.   The orientation of the sequence is from
            SP6 end to T7 end. Genes were identified by a combination of five
            methods including: XGRAIL (available by anonymous ftp from
            arthur.epm.ornl.gov), Genefinder (Phil Green, University of
            Washington), Genscan  (Chris Burge,
            http://gnomic.stanford.edu/~chris/GENSCANW.html)searches of the
            complete sequence against a peptide database,and the Human gene
            Index database at TIGR (http://www.tigr.org/tdb/hgi/hgi.html).
            Genes without pepetide homolgy having spliced EST hits are termed
            'Unknown gene product'. Genes encoding tRNAs are predicted by
            tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
FEATURES             Location/Qualifiers
     source          1. .216021
                     /organism="Homo sapiens"
```

```
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /map="#16q21-22"
                     /clone="A-952F10"
     STS             27765. .27872
                     /note="7766, STS1-cSRL-27g3-uA/cSRL-27g3-uZ, Chr. -, Homo
                     sapiens"
                     /db_xref="dbSTS:G02280"
     STS             73826. .73943
                     /note="7608, STS1-cSRL-24g1-uA/cSRL-24g1-uZ, Chr. -, Homo
                     sapiens"
                     /db_xref="dbSTS:G02122"
     STS             175801. .175945
                     /note="16084, CHLC.GCT10B02, Chr. -, Homo sapiens"
                     /db_xref="dbSTS:G09703"
     STS             175810. .175945
                     /note="16316, CHLC.GCT15C04, Chr. -, Homo sapiens"
                     /db_xref="dbSTS:G09935"
     STS             199463. .199572
                     /note="9824, WI-3555, Chr. 16, Homo sapiens"
                     /db_xref="dbSTS:G04338"
ORIGIN

  Query Match             7.7%;  Score 382.6;  DB 9;  Length 216021;
  Best Local Similarity   76.2%;  Pred. No. 5.7e-67;
  Matches  551;  Conservative    0;  Mismatches  159;  Indels   13;  Gaps    6;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAATTA 60
              ||||| | |||| ||||  |||||| |||||||||||||||||||||||    |||||||
Db      80976 AGTTCGAAACCACCCTGGCCAACATAGTGAAACCCCATCTCTACTAAAATACAAAAATTA 81035

Qy         61 GACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAAT 120
              | ||||| |||||  | |||||||||||||||||||  |||||| |||||||||| ||
Db      81036 GCCAGGCGTGGTGGCAGACGCCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGAGGAT 81095

Qy        121 CACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCCAG 179
              ||||||||  | || ||  |||||||| | ||| | ||||  ||||| || |||||||
Db      81096 TGCTTGAACC--CAGGAGGCAGAGGTTGCAGTGAGCCCAGATTGTGCCACTGGACTCCAG 81153

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  ||  || ||||  | |   | ||||| | | | | |       || | | ||||
Db      81154 CCTGGGAGAC-AGAGTAAGACTCCGTCTCAAATACATATATA-----CATACACACATAT 81207

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              ||||| | ||||||| ||||||| | ||||| | ||| ||||| | | | ||||| | ||
Db      81208 ATATACACATATACATACACACATATATATACACATATACATACACACATACATACACAC 81267

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACAT 359
              |||||||||| ||||||| ||||||| ||| ||| |||  || ||| ||||||||||||
Db      81268 ATATATACACACATATACATATATATACACATATATATACATATACACATATATACACAT 81327

Qy        360 ATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATAT 419
              || || | | | | | | | ||||| | | | | | ||| ||| ||||| ||| |||||
Db      81328 ATATA-ATATACACACATATATATACATATATACACATATACACATATATACATATATAC 81386

Qy        420 ACACATATATACACATATATACACATATACACATATATACACATATATACATATATACAC 479
              |||||||||||||||||||||| ||| |||||||||||| ||| | ||||||| ||||
Db      81387 ACACATATATACACATATATACATATACACACATATATACATATACACACATATACACAC 81446

Qy        480 ATATATATA---ATATACACACATATATATACACATATATACACACATATATACACATAT 536
              ||| | | |    | | |||||||| | | | ||| | | | ||||| | | | |||  |
Db      81447 ATACACACACATACACACACACATACACACATACACACACATACACACACACATACACAC 81506

Qy        537 ATACACATATATATACACATATATACACATATATACATATATACACATATATATACATAT 596
              | | ||| | ||| ||||| ||||||||| ||| |  || | | ||| | ||||||| ||
Db      81507 ACATACACACATACACACACATATACACACATACATGTACACATACACACATATACACAT 81566

Qy        597 ATACACATATATACATATATACACATATATACATATATACACACATATATACACATACAT 656
              ||||||| |||||||||||||||| ||||||||||||||||||||||| ||||| ||
Db      81567 ATACACACATATACATATATACACACATATACATATATACACACATATATATACATATAT 81626

Qy        657 ATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATAT 716
              | ||| | | | | ||||||| |||||||||| |||| | |  |||| |||| || | ||
Db      81627 ACACATATATACATATATACACATATATACATATATGCACATATATACATATATACACAT 81686

Qy        717 ATA 719
              |||
Db      81687 GTA 81689


RESULT 15
AY007685
LOCUS       AY007685               92564 bp    DNA     linear   PRI 22-FEB-2002
DEFINITION  Homo sapiens telomerase catalytic subunit (TERT) and sodium
            channel-like protein genes, complete cds.
ACCESSION   AY007685
VERSION     AY007685.2  GI:15991796
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 92564)
  AUTHORS   Leem,S.H., Londono-Vallejo,J.A., Kim,J.H., Bui,H., Tubacher,E.,
            Solomon,G., Park,J.E., Horikawa,I., Kouprina,N., Barrett,J.C. and
            Larionov,V.
  TITLE     The human telomerase gene: complete genomic sequence and analysis
            of tandem repeat polymorphisms in intronic regions
  JOURNAL   Oncogene 21 (5), 769-777 (2002)
  MEDLINE   21839826
   PUBMED   11850805
REFERENCE   2  (bases 1 to 92564)
  AUTHORS   Londono-Vallejo,J.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-SEP-2000) Centre d'Etudes du Polymorphisme Humain, 27
            rue Juliette Dodu, Paris 75010, France
REFERENCE   3  (bases 1 to 92564)
  AUTHORS   Londono-Vallejo,J.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-OCT-2001) Centre d'Etudes du Polymorphisme Humain, 27
            rue Juliette Dodu, Paris 75010, France
  REMARK    Sequence update by submitter
COMMENT     On Oct 10, 2001 this sequence version replaced gi:12642956.
FEATURES             Location/Qualifiers
     source          1. .92564
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /map="5p13"
     repeat_region   complement(2853. .3034)
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   complement(6174. .6429)
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   complement(7445. .7837)
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   complement(8199. .8341)
                     /rpt_family="MLT1"
                     /rpt_type=dispersed
     repeat_region   complement(10300. .10916)
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   12028. .12301
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   13088. .13597
                     /rpt_family="L1"
                     /rpt_type=dispersed
     repeat_region   13716. .13988
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   14271. .14556
```

```
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   14599. .17711
                /rpt_family="L1"
                /rpt_type=dispersed
repeat_region   18109. .18398
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   complement(20818. .21123)
                /rpt_family="LTR1"
                /rpt_type=dispersed
repeat_region   21650. .21912
                /rpt_family="MLT1"
                /rpt_type=dispersed
repeat_region   complement(22236. .22484)
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   complement(22986. .23283)
                /rpt_family="Alu"
                /rpt_type=dispersed
gene            <24442. .64466
                /gene="TERT"
mRNA            join(<24442. .24660,24765. .26118,35473. .35668,37759. .37939,
                38627. .38806,39301. .39456,45360. .45455,46436. .46521,
                49007. .49120,51105. .51176,53048. .53236,57040. .57166,
                58981. .59042,62229. .62353,63135. .63272,63809. .>63912)
                /gene="TERT"
                /product="telomerase catalytic subunit"
CDS             join(24442. .24660,24765. .26118,35473. .35668,37759. .37939,
                38627. .38806,39301. .39456,45360. .45455,46436. .46521,
                49007. .49120,51105. .51176,53048. .53236,57040. .57166,
                58981. .59042,62229. .62353,63135. .63272,63809. .63912)
                /gene="TERT"
                /codon_start=1
                /product="telomerase catalytic subunit"
                /protein_id="AAG23289.1"
                /db_xref="GI:12642957"
                /translation="MPRAPRCRAVRSLLRSHYREVLPLATFVRRLGPQGWRLVQRGDP
                AAFRALVAQCLVCVPWDARPPPAAPSFRQVSCLKELVARVLQRLCERGAKNVLAFGFA
                LLDGARGGPPEAFTTSVRSYLPNTVTDALRGSGAWGLLLRRVGDDVLVHLLARCALFV
                LVAPSCAYQVCGPPLYQLGAATQARPPPHASGPRRRLGCERAWNHSVREAGVPLGLPA
                PGARRRGGSASRSLPLPKRPRRGAAPEPERTPVGQGSWAHPGRTRGPSDRGFCVVSPA
                RPAEEATSLEGALSGTRHSHPSVGRQHHAGPPSTSRPPRPWDTPCPPVYAETKHFLYS
                SGDKEQLRPSFLLSSLRPSLTGARRLVETIFLGSRPWMPGTPRRLPRLPQRYWQMRPL
                FLELLGNHAQCPYGVLLKTHCPLRAAVTPAAGVCAREKPQGSVAAPEEEDTDPRRLVQ
                LLRQHSSPWQVYGFVRACLRRLVPPGLWGSRHNERRFLRNTKKFISLGKHAKLSLQEL
                TWKMSVRDCAWLRRSPGVGCVPAAEHRLREEILAKFLHWLMSVYVVELLRSFFYVTET
                TFQKNRLFFYRKSVWSKLQSIGIRQHLKRVQLRELSEAEVRQHREARPALLTSRLRFI
                PKPDGLRPIVNMDYVVGARTFRREKRAERLTSRVKALFSVLNYERARRPGLLGASVLG
                LDDIHRAWRTFVLRVRAQDPPPELYFVKVDVTGAYDTIPQDRLTEVIASIIKPQNTYC
                VRRYAVVQKAAHGHVRKAFKSHVSTLTDLQPYMRQFVAHLQETSPLRDAVVIEQSSSL
                NEASSGLFDVFLRFMCHHAVRIRGKSYVQCQGIPQGSILSTLLCSLCYGDMENKLFAG
                IRRDGLLLRLVDDFLLVTPHLTHAKTFLRTLVRGVPEYGCVVNLRKTVVNFPVEDEAL
                GGTAFVQMPAHGLFPWCGLLLDTRTLEVQSDYSSYARTSIRASLTFNRGFKAGRNMRR
                KLFGVLRLKCHSLFLDLQVNSLQTVCTNIYKILLLQAYRFHACVLQLPFHQQVWKNPT
                FFLRVISDTASLCYSILKAKNAGMSLGAKGAAGPLPSEAVQWLCHQAFLLKLTRHRVT
                YVPLLGSLRTAQTQLSRKLPGTTLTALEAAANPALPSDFKTILD"
repeat_region   27341. .27628
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   27650. .29650
                /note="VNTR"
                /rpt_type=tandem
                /rpt_unit="gagtgaggcgccgtccccgggtgtccctgtcacgtgcagggt"
repeat_region   complement(31926. .32203)
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   complement(32555. .32835)
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   33569. .33861
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   complement(33873. .33967)
                /rpt_family="MER33"
                /rpt_type=dispersed
repeat_region   34111. .35453
                /note="VNTR"
                /rpt_type=tandem
repeat_region   complement(35993. .36281)
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   41485. .42324
                /note="VNTR"
                /rpt_type=tandem
                /rpt_unit="gtgggattggttttcatgtgtggggtaggtggggatct"
repeat_region   complement(42656. .42954)
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   43312. .43618
                /rpt_family="MER1"
                /rpt_type=dispersed
repeat_region   44225. .45115
                /note="VNTR"
                /rpt_type=tandem
                /rpt_unit="ggggtctgatgtgtggtgactgtggatggcggtcgt"
repeat_region   complement(48015. .48300)
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   complement(49647. .50276)
                /rpt_family="L1"
                /rpt_type=dispersed
repeat_region   53945. .54234
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   57630. .58735
                /note="VNTR"
                /rpt_type=tandem
                /rpt_unit="tgggcatctgcgtccactccccctctcctg"
repeat_region   61332. .61597
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   complement(66201. .66323)
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   66389. .66661
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   66661. .66736
                /note="microsatellite"
                /rpt_type=tandem
                /rpt_unit="aag"
repeat_region   67140. .67583
                /rpt_family="MLT1"
                /rpt_type=dispersed
mRNA            complement(join(<73264. .73377,73925. .74084,76991. .77080,
                78049. .78146,79551. .79661,81950. .82131,84724. .84861,
                85253. .85393,91909. .>92068))
                /product="sodium channel-like protein"
CDS             complement(join(73264. .73377,73925. .74084,76991. .77080,
                78049. .78146,79551. .79661,81950. .82131,84724. .84861,
                85253. .85393,91909. .92068))
                /codon_start=1
                /evidence=not_experimental
                /product="sodium channel-like protein"
                /protein_id="AAG23290.1"
                /db_xref="GI:12642958"

  Query Match             7.6%;  Score 379.8;  DB 9;  Length 92564;
  Best Local Similarity   75.0%;  Pred. No. 2.4e-66;
  Matches  542;  Conservative    0;  Mismatches  172;  Indels    9;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||| |||||||||||  ||| || | ||||||||||| ||||||||| || ||||||
Db      11380 AGTTCGAGACCAGCCTGGCCAATATAGCGAAACCCCATCTGTACTAAAAACACAAAAATT 11439
```

```
Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   ||  |||||| ||| ||||||||||||||| ||||  ||||||  ||||||||||
Db      11440 AGCTGGGTGTGGTGACACATGCCTGTAATCCCAGGTACTCAGGAGGCTAAGGCAGGAGAA 11499

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              | |||||||  |  ||  | || |||     |   ||||||    ||| ||||||||||
Db      11500 TTGCTTGAACTGGGAGGCAGAGGTGGT-----GAGCCAAGATTGCACCAGTGCACTCCAG 11554

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  ||  || |||| ||||   ||||||| ||| ||| |||   | |  ||||| |||
Db      11555 CCTTGGTGAC-AGAGTGAAACTCCATCTCAAAAACAAACAAACAAAATACATATACATAA 11613

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              ||||||||  | | | | | ||| | ||| |||||||  ||| |||||||  |  |||||||
Db      11614 ATATATATGCACATATATATACATATATAAATATATATACACATATATAAATCTATATAC 11673

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACATATAT-ATACACA 358
              |||||||| ||||| ||| ||||| ||  | ||||||||  || ||||||||  ||| |
Db      11674 ATATATACATATATACACATATATAAATCTATATACATATATATACATATATAATATATT 11733

Qy        359 TATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATA 418
              ||  |||| | |||||||| ||||| | | |||||||| | | ||||| | | ||| |
Db      11734 TACATATATAAATATATACATATATAAATATACATATATAAATACATATATAAATATACA 11793

Qy        419 TACACATATATACACATATATACACATATACACATATATACACATATATACATATATACA 478
              || | | ||||| |||||||| | ||| | | ||| ||||| ||||| | ||| ||| |
Db      11794 TATATAAATATATACATATATAAATATACATATATAAATACATATATAAATATACATATA 11853

Qy        479 CATATATATA-ATATACACACATATATATACACATATATACACACATATATACACATATA 537
               | ||||||| |||||  | | ||| ||||| | |||||||||  | ||| ||| |||||||
Db      11854 TAAATATATACATATATAAATATACATATATAAATATATACATATATAAATATACATATA 11913

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATACATATA 597
              ||||.|||||  ||| |||||||  | | ||| ||| |||||  ||| |||||  ||| | |||
Db      11914 TACATATATAAATATACATATACATATATAAATATATATACACATATATAAATATACATA 11973

Qy        598 TACACATATATACATATATACACATATATACATATATACACACATATATACACATACATA 657
              || | |||||  | |||||  | |  ||| ||| ||| ||| ||| |||||  | |  ||| | |
Db      11974 TATAAATATACATATATAAATATATACATATATAAATATACATATATAAATATATATACA 12033

Qy        658 TACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATA 717
              || | |  ||| | ||| | |  |||| ||| |||||||| |||||| |||||||||
Db      12034 TATATAAATATATAAATATACAAGTATATACAAATATATACATATATAAATGTATATACG 12093

Qy        718 TAT 720
              |||
Db      12094 TAT 12096


RESULT 16
AC114955/c
LOCUS       AC114955              161252 bp    DNA     linear   HTG 14-MAR-2002
DEFINITION  Homo sapiens chromosome 5 clone RP11-117B23, WORKING DRAFT
            SEQUENCE, 9 unordered pieces.
ACCESSION   AC114955
VERSION     AC114955.1  GI:19424416
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 161252)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 161252)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-MAR-2002) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 451340
            Center clone name: RPCI-11_117B23
            -------
            Summary Statistics
            Consensus quality: 148939 bases at least Q40
            Consensus quality: 156608 bases at least Q30
            Consensus quality: 158826 bases at least Q20
            Estimated insert size: 175000; agarose-fp estimation
            Estimated insert size: 160452; sum-of-contigs estimation
            Quality coverage:  7.26 in Q20 bases; agarose-fp estimation
            Quality coverage:  7.92 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *         1     3008: contig of 3008 bp in length
            *      3009     3108: gap of unknown length
            *      3109     9209: contig of 6101 bp in length
            *      9210     9309: gap of unknown length
            *      9310    15613: contig of 6304 bp in length
            *     15614    15713: gap of unknown length
            *     15714    23243: contig of 7530 bp in length
            *     23244    23343: gap of unknown length
            *     23344    39674: contig of 16331 bp in length
            *     39675    39774: gap of unknown length
            *     39775    60001: contig of 20227 bp in length
            *     60002    60101: gap of unknown length
            *     60102    85303: contig of 25202 bp in length
            *     85304    85403: gap of unknown length
            *     85404   117460: contig of 32057 bp in length
            *    117461   117560: gap of unknown length
            *    117561   161252: contig of 43692 bp in length.
FEATURES             Location/Qualifiers
     source          1. .161252
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="RP11-117B23"
                     /clone_lib="RPCI human BAC library 11"
ORIGIN

  Query Match             7.6%;  Score 379.4;  DB 2;  Length 161252;
  Best Local Similarity   68.7%;  Pred. No. 2.6e-66;
  Matches  600;  Conservative    0;  Mismatches  256;  Indels   17;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||| ||||||||||  ||| || | ||||||||||| ||||||||| || ||||||
Db     149753 AGTTCGAGACCAGCCTGGCCAATATAGCGAAACCCCATCTGTACTAAAAACACAAAAATT 149694

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   ||  |||||| ||| |||||||||||||| ||||  ||||||  ||||||||||
Db     149693 AGCTGGGTGTGGTGACACATGCCTGTAATCCCAGGTACTCAGGAGGCTAAGGCAGGAGAA 149634

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              | |||||||  |  ||  | || |||     |   ||||||    ||| ||||||||||
Db     149633 TTGCTTGAACTGGGAGGCAGAGGTGGT-----GAGCCAAGATTGCACCAGTGCACTCCAG 149579

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  ||  || |||| ||||   ||||||| ||| ||| |||   | |  ||||| |||
Db     149578 CCTTGGTGAC-AGAGTGAAACTCCATCTCAAAAACAAACAAACAAAATACATATACATAA 149520

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              ||||||||  | | | | | ||| | ||| |||||||| ||| |||||||  |  |||||||
Db     149519 ATATATATGCACATATATATACATATATAAATATATATACACATATATAAATCTATATAC 149460
```

```
Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACATATATATA----- 354
              ||||||||| ||||| ||| ||||| ||  | |||||||| || | | |||||||
Db     149459 ATATATACATATATACACATATATAAATCTATATACATATATATATACATATATAATATA 149400

Qy        355 --CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACA 412
               |||||| || | | ||| ||| | | ||||||| | ||| ||||| ||||| | | |
Db     149399 TTTACATATATAAATATATACATATATAAATATACATATATAAATACATATATAAATATA 149340

Qy        413 CATATATACACATATATACACATATATACACATATACACATATATACACATATATACATA 472
              |||||||| | ||||| | | ||| ||| ||||||| | ||| ||| | ||| |||||||
Db     149339 CATATATAAATATATACATATATAAATATACATATATAAATACATATATATAAATACATA 149280

Qy        473 TATACACATATATAT---AATATACACACATATATATACACATATATACACACATATATA 529
              |||| | ||| ||||    |||||| ||| ||||| ||| ||||||||| | | ||| |||
Db     149279 TATAAATATACATATATAAATATATACATATATAAATATACATATATAAATATATACATA 149220

Qy        530 CACATATATACACATATATATACACATATATACACATATATACATATATACACACATATATA 589
              | | ||||||| ||||| ||| | | ||||||| ||| ||| ||| ||| | ||| | |
Db     149219 TATAAATATACATATATACATATATAAATATACATATACATATATAAATATATATACACA 149160

Qy        590 TACATATATACACATATATACATATATACACATATATACATATATACACACATATATACA 649
              || ||| ||| ||||||||| ||||| | | ||| ||| ||| ||| | | ||||| | |
Db     149159 TATATAAATATACATATATAAATATACATATATAAATATATACATATATAAATATACATA 149100

Qy        650 CATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATATG 709
              ||| ||||| | ||| ||| | ||| ||| |||||||  |||||| |  |||||| ||
Db     149099 TATAAATATATATACATATATAAATATATAAATATACAAGTATATACAAATATATACATA 149040

Qy        710 TATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGG 769
              |||| | ||||   |     | | ||       | |  |   ||     |   ||| | |
Db     149039 TATAAACATATATACGTATATACGTATATACATGTATATACGTATATACGTATATACATG 148980

Qy        770 AAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGT 829
              | | |    | ||    | ||    | | | | || |     ||     |  ||
Db     148979 TATATACGTATATACGTATATACATGTATATACGTATATATATAAATGTATATACGTATA 148920

Qy        830 AGGGGAAATACATAAATGTGTAATAAAACCTAT 862
                | ||| |||||| | ||| ||||| | |
Db     148919 TACATATATATATAAATATATAAAAAAACTTTT 148887


RESULT 17
AC114291/c
LOCUS       AC114291               202305 bp    DNA     linear   PRI 15-MAR-2003
DEFINITION  Homo sapiens chromosome 5 clone CTD-3080P12, complete sequence.
ACCESSION   AC114291
VERSION     AC114291.2  GI:28973811
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 202305)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 202305)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAR-2002) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 202305)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-MAR-2003) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Mar 15, 2003 this sequence version replaced gi:19224940.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 99.9% of Sequence;
            Estimated Total Number of Errors is 0.2.
            NOTE: Transposon sequencing failed to verify number of repeat
            copies 173000-175780. Unsure number of repeat copies 173000-175780.
            BAC and subclones unstable 76080. Force join at 76080.
FEATURES             Location/Qualifiers
     source          1. .202305
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTD-3080P12"
     misc_feature    76080
                     /note="NOTE: BAC and subclones unstable 76080. Force join
                     at 76080."
     misc_feature    173000. .175780
                     /note="NOTE: Transposon sequencing failed to verify number
                     of repeat copies 173000-175780. Unsure number of repeat
                     copies 173000-175780"
ORIGIN

  Query Match             7.6%;  Score 378;  DB 9;  Length 202305;
  Best Local Similarity   66.3%;  Pred. No. 4.8e-66;
  Matches  640;  Conservative    0;  Mismatches  310;  Indels   16;  Gaps    6;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||| |||||||||||  ||| || | |||||||||||| |||||||| || ||||||
Db     192244 AGTTCGAGACCAGCCTGGCCAATATAGCGAAACCCCATCTGTACTAAAAACACAAAAATT 192185

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   ||  |||||| ||| ||||||||||||| ||||  ||||||  ||||||||||
Db     192184 AGCTGGGTGTGGTGACACATGCCTGTAATCCCAGGTACTCAGGAGGCTAAGGCAGGAGAA 192125

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              | |||||||  |  ||  | ||  |||      |   ||||||    ||| ||||||||||
Db     192124 TTGCTTGAACTGGGAGGCAGAGGTGGT-----GAGCCAAGATTGCACCAGTGCACTCCAG 192070

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  ||  || |||| ||||   ||||||| ||| ||| |||    | |  ||||| |||
Db     192069 CCTTGGTGAC-AGAGTGAAACTCCATCTCAAAAACAAACAAACAAAATACATATACATAA 192011

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              |||||||  | | | | ||| | ||| ||||||| ||| ||||||| |  |||||||
Db     192010 ATATATATGCACATATATATACATATATAAATATATATACACATATATAAATCTATATAC 191951

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACAT-----ATATATA 354
              |||||||| ||||| ||| ||||| ||  | |||||||| || ||||     ||||||
Db     191950 ATATATACATATATACACATATATAAATCTATATACATATATATACATATATAATATATT 191891

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              |||||| || | | ||| ||| | | ||||||| | ||| ||||| ||||| | | |||
Db     191890 TACATATATAAATATATACATATATAAATATACATATATAAATACATATATAAATATACA 191831

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              |||||| | ||||| | | ||| ||| ||||||| | ||| ||| | | ||||||||||
Db     191830 TATATAAATATATACATATATAAATATACATATATAAATACATATATAAATATACATATA 191771

Qy        475 TACACATATATATA---ATATACACACATATATATACACATATATACACACATATATACA 531
              || | ||||| |||    | ||| ||| ||||| ||| | | ||||| | | ||| ||| |
Db     191770 TAAATATATACATATATAAATATACATATATAAATATATACATATATAAATATACATATA 191711

Qy        532 CATATATACACATATATATACACATATATACACATATATACATATATACACATATATATA 591
              | ||||| | ||||| ||| |||||||| | |||||||| ||||| | ||||| |||
Db     191710 TACATATATAAATATACATATACATATATAAATATATATACACATATATAAATATACATA 191651

Qy        592 CATATATACACATATATACATATATACACATATATACATATATACACACATATATACACA 651
              ||| ||| |||||||| |||||||| ||||| | ||| ||| | | ||||||| |||
Db     191650 TATAAATATACATATATAAATATATACATATATAAATATACATATATAAATATATATACA 191591

Qy        652 TACATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTA 711
              || ||| | | | ||| | |   ||||| | | | ||| |||||   | ||||||
Db     191590 TATATAAATATATAAATATACAAGTATATACAAATATATACATATATAAACATATATACG 191531

Qy        712 TATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAA 771
              |||||| |||       ||  ||| |      ||  |   |||    |   ||| | | |
```

```
Db     191530 TATATACATGTATATACGTATATACGTATATACATGTATATACGTACGTATATACATGTA 191471

Qy        772 AAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAG 831

Db     191470 TATACGTATATACGTATATACATGTATATACGTATATACGTATATACATGTATATACGTA 191411

Qy        832 GGGAAATACATAAATGTGTAATAAAACCTATCT-GTAAGGTAGTTAAGAAGGTAACACTA 890

Db     191410 TATACGTATATATATAAATGTATATACGTATATACGTATATATATAAATGTATATACGTA 191351

Qy        891 TATATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTAT 950

Db     191350 TATACATATATATATAAATATATAAAAAAACTTTTGGCTGGGCACCTTTCCAAATCTCAT 191291

Qy        951 AGAAGA 956

Db     191290 GGCACA 191285
```

RESULT 18
AC021955/c

```
LOCUS       AC021955               192139 bp    DNA     linear   PRI 22-MAR-2002
DEFINITION  Homo sapiens chromosome 8, clone RP11-238G17, complete sequence.
ACCESSION   AC021955
VERSION     AC021955.11  GI:19683766
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 192139)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 8, clone RP11-238G17
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 192139)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 192139)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C.,
            Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-FEB-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 192139)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 22, 2002 this sequence version replaced gi:18875268.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L5949
                Center clone name: 238_G_17
            --------------.
FEATURES             Location/Qualifiers
     source          1. .192139
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="RP11-238G17"
                     /clone_lib="RPCI-11 Human Male BAC"
     unsure          441. .445
                     /note="<30 qual SNGL region"
     unsure          455. .461
                     /note="<30 qual SNGL region"
     repeat_region   complement(1301. .1824)
                     /rpt_family="MLT1F"
     repeat_region   complement(1885. .2017)
                     /rpt_family="L2"
     repeat_region   2167. .2264
                     /rpt_family="GA-rich"
     repeat_region   2333. .2853
                     /rpt_family="(TA)n"
     repeat_region   complement(2862. .3156)
                     /rpt_family="AluSq"
     repeat_region   3916. .3974
                     /rpt_family="MLT1F"
     repeat_region   3977. .4182
```

```
                /rpt_family="AluJo"
repeat_region   complement(4553. .4782)
                /rpt_family="MIR"
repeat_region   5326. .5402
                /rpt_family="LTR33"
repeat_region   5570. .5756
                /rpt_family="LTR33"
repeat_region   complement(6278. .7151)
                /rpt_family="MER4D"
repeat_region   7424. .7644
                /rpt_family="MER58A"
repeat_region   7701. .7999
                /rpt_family="AluY"
repeat_region   9337. .9635
                /rpt_family="AluSg"
repeat_region   9723. .9878
                /rpt_family="L1M4c"
repeat_region   9877. .10902
                /rpt_family="L1M4c"
repeat_region   10903. .11141
                /rpt_family="AluSc"
repeat_region   11142. .11261
                /rpt_family="L1M4c"
repeat_region   11261. .11381
                /rpt_family="L1MB4"
repeat_region   11383. .11839
                /rpt_family="MER51A"
repeat_region   11956. .12100
                /rpt_family="L1MB4"
repeat_region   12101. .12132
                /rpt_family="(CA)n"
repeat_region   12133. .12496
                /rpt_family="L1MB4"
repeat_region   12994. .13074
                /rpt_family="L2"
repeat_region   13900. .13930
                /rpt_family="AT_rich"
repeat_region   14087. .14128
                /rpt_family="(TG)n"
repeat_region   complement(15008. .15212)
                /rpt_family="MIR"
repeat_region   15423. .15924
                /rpt_family="MER87"
repeat_region   16247. .16353
                /rpt_family="MLT1I"
repeat_region   17718. .17749
                /rpt_family="(TTA)n"
repeat_region   complement(18035. .18137)
                /rpt_family="MIR3"
repeat_region   18182. .18418
                /rpt_family="MIR"
repeat_region   18433. .18573
                /rpt_family="L3"
repeat_region   complement(18574. .18883)
                /rpt_family="AluSg"
repeat_region   18884. .18963
                /rpt_family="L3"
repeat_region   19497. .19592
                /rpt_family="MIR"
repeat_region   19909. .20082
                /rpt_family="L3"
repeat_region   complement(20319. .20627)
                /rpt_family="AluSx"
repeat_region   20785. .20932
                /rpt_family="MLT1A1"
unsure          20910. .20953
                /note="<30 qual SNGL region"
repeat_region   20933. .21252
                /rpt_family="AluSg"
unsure          20944
                /note="probably G, possibly C"
unsure          20965. .21026
                /note="<30 qual SNGL region"
repeat_region   21253. .21466
                /rpt_family="MLT1A1"
```

```
  Query Match             7.5%;  Score 373.6;  DB 9;  Length 192139;
  Best Local Similarity   74.2%;  Pred. No. 3.7e-65;
  Matches  539;  Conservative    0;  Mismatches  179;  Indels    8;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAAT-ACAAAATT 59
              ||||| |||||  ||||  ||||||| ||||||||||||||||||||||| | ||||||
Db       3081 AGTTCGAGACCCACCTGGCCAACATGATGAAACCCCATCTCTACTAAAAATGAAAAAATT 3022

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGA-GGCCGAGGCAGGAGA 118
              ||   ||  ||||| |   || | ||||||||||||||  ||| ||| ||||||| |||
Db       3021 AGCTGGGTGTGGTGGCAGGTGCATCTAATCCCAGCTACTCTGGAGGGCTGAGGCAGAAGA 2962

Qy        119 ATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCC 177
              ||  |||||||  | || |||||| ||||| | ||| | |||||  ||||||||||||
Db       2961 ATTGCTTGAACC--CAGGAGGTGGACGTTGCAGTGAGCCGAGATCGCGCCATTGCACTCC 2904

Qy        178 AGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATAT 237
              ||||  ||||||||||||| |||     ||||  |||||||||         |||||||||
Db       2903 AGCCTGGGCAACAAGAGCAAAACTCTGTCTCGAAAAAAAAAATTTTATAAATATATATAT 2844

Qy        238 ATATATATATATATACACACACACACATATATATATACACATATATATACACATATAT 297
              | |||||||||| | | | | | | |||| ||||| | | ||||||| | ||||||
Db       2843 AAATATATATATAAATATATATAAATATATATAAATATATATAAATATATATAAATATAT 2784

Qy        298 ACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACAC 357
              | ||| ||| | ||| ||| | ||||| ||| | ||| |||| || | | ||||| | |
Db       2783 ATATAAATATATATAAATATATATATAAATATATATAAATATATATATAAATATAAATAA 2724

Qy        358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417
              |||| |||| | ||| ||| | ||||| | |    ||||| | | ||| ||| | | ||||
Db       2723 ATATATATAAATATAAATAAATATATAAATATGTATATAAATATATAAATATATAAATAT 2664

Qy        418 ATACACATATATACACATATATACACATATACACATATATACACATATATACATATATAC 477
              || | | ||||| | ||| ||| | | |||    ||||| | | ||||| | |||||||
Db       2663 GTATATAAATATATATATAAATATATAAATATGTATATAAATATATATAAATATATATAA 2604

Qy        478 ACATATATATA---ATATACACACATATATATACACATATATACACACATATATACACAT 534
              | || ||||||    ||||| | | || |||||| | ||||||| | | ||||||| | |
Db       2603 ATATGTATATAAATATATATAAATATGTATATAAATATATATATAAAAATATATATAAAA 2544

Qy        535 ATATACACATATATATACACATATATACACATATATACATATATACACATATATATACAT 594
              ||||| | | ||||||| | ||||||| | ||||||| | ||||| | ||||||||| |
Db       2543 ATATATATAAATATATAAATATATATATAAATATATATAAATATATAAATATATATATAA 2484

Qy        595 ATATACACATATATACATATATACACATATATACATATATACACACATATATACACATAC 654
              ||||| | | ||||| ||||| | | ||||| | ||||||| | | ||||||| | |||
Db       2483 ATATATATAAATATATATATAAATATATATAAATATATATAAATATATATATAAATATAT 2424

Qy        655 ATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATAT 714
              ||||| | | | ||| | ||| ||||| ||| | | |||||||  |||||||| || ||
Db       2423 ATATAAATATATATAAAAATATATATAAATATATAAATATATAAATATATATATATAAAT 2364

Qy        715 ATATAT 720
              ||||||
Db       2363 ATATAT 2358
```

```
RESULT 19
AC106873/c
LOCUS       AC106873               99596 bp    DNA     linear   PRI 31-JAN-2004
DEFINITION  Homo sapiens BAC clone RP11-374M7 from 7, complete sequence.
ACCESSION   AC106873
VERSION     AC106873.3  GI:18698877
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 99596)
  AUTHORS   Hillier,L.W., Fulton,R.S., Fulton,L.A., Graves,T.A., Pepin,K.H.,
```

```
            Wagner-McPherson,C., Layman,D., Maas,J., Jaeger,S., Walker,R.,
            Wylie,K., Sekhon,M., Becker,M.C., O'Laughlin,M.D., Schaller,M.E.,
            Fewell,G.A., Delehaunty,K.D., Miner,T.L., Nash,W.E., Cordes,M.,
            Du,H., Sun,H., Edwards,J., Bradshaw-Cordum,H., Ali,J., Andrews,S.,
            Isak,A., Vanbrunt,A., Nguyen,C., Du,F., Lamar,B., Courtney,L.,
            Kalicki,J., Ozersky,P., Bielicki,L., Scott,K., Holmes,A.,
            Harkins,R., Harris,A., Strong,C.M., Hou,S., Tomlinson,C.,
            Dauphin-Kohlberg,S., Kozlowicz-Reilly,A., Leonard,S., Rohlfing,T.,
            Rock,S.M., Tin-Wollam,A.M., Abbott,A., Minx,P., Maupin,R.,
            Strowmatt,C., Latreille,P., Miller,N., Johnson,D., Murray,J.,
            Woessner,J.P., Wendl,M.C., Yang,S.P., Schultz,B.R., Wallis,J.W.,
            Spieth,J., Bieri,T.A., Nelson,J.O., Berkowicz,N., Wohldmann,P.E.,
            Cook,L.L., Hickenbotham,M.T., Eldred,J., Williams,D., Bedell,J.A.,
            Mardis,E.R., Clifton,S.W., Chissoe,S.L., Marra,M.A., Raymond,C.,
            Haugen,E., Gillett,W., Zhou,Y., James,R., Phelps,K., Iadanoto,S.,
            Bubb,K., Simms,E., Levy,R., Clendenning,J., Kaul,R., Kent,W.J.,
            Furey,T.S., Baertsch,R.A., Brent,M.R., Keibler,E., Flicek,P.,
            Bork,P., Suyama,M., Bailey,J.A., Portnoy,M.E., Torrents,D.,
            Chinwalla,A.T., Gish,W.R., Eddy,S.R., McPherson,J.D., Olson,M.V.,
            Eichler,E.E., Green,E.D., Waterston,R.H. and Wilson,R.K.
  TITLE     The DNA sequence of human chromosome 7
  JOURNAL   Nature 424 (6945), 157-164 (2003)
  MEDLINE   22737999
   PUBMED   12853948
REFERENCE   2  (bases 1 to 99596)
  AUTHORS   Cotton,M. and Grewal,N.
  TITLE     The sequence of Homo sapiens BAC clone RP11-374M7
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 99596)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JAN-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 99596)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-FEB-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 99596)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-FEB-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   6  (bases 1 to 99596)
  AUTHORS   Wilson,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-JAN-2004) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Feb 17, 2002 this sequence version replaced gi:18158390.
            ---------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu
                Contact: sapiens@watson.wustl.edu
            ---------- Summary Statistics
                Center project name: H_NH0374M07
            ----------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            The sequence of this clone was established as part of a mapping and
            sequencing collaboration between the NHGRI Chromosome 7 Mapping
            Project (Eric D. Green, Director), John D. McPherson in the
            Department of Genetics (Washington University), and the Washington
            University Genome Sequencing Center. For additional information
            about the map position of this sequence, see
            http://www.nhgri.nih.gov/DIR/GTB/CHR7 , send
            mailto:egreen@nhgri.nih.gov , or see http://genome.wustl.edu

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries. Genomics 51:1-8. The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at http://www.chori.org
            VECTOR:  pBACe3.6.
FEATURES                Location/Qualifiers
     source             1. .99596
                        /organism="Homo sapiens"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9606"
                        /chromosome="7"
                        /map="7"
                        /clone="RP11-374M7"
                        /clone_lib="RPCI-11"
     repeat_region      5. .138
                        /rpt_family="Alu"
     repeat_region      399. .604
                        /rpt_family="L1"
     repeat_region      605. .919
                        /rpt_family="Alu"
     repeat_region      920. .1039
                        /rpt_family="L1"
     repeat_region      1045. .1325
                        /rpt_family="Alu"
     repeat_region      1330. .1444
                        /rpt_family="(GAAA)n"
     repeat_region      1552. .1572
                        /rpt_family="AT_rich"
     repeat_region      1699. .2061
                        /rpt_family="MER2_type"
     repeat_region      2062. .2284
                        /rpt_family="Alu"
     repeat_region      2285. .2351
                        /rpt_family="MER2_type"
     repeat_region      2949. .3225
                        /rpt_family="Alu"
     repeat_region      3237. .3413
                        /rpt_family="Alu"
     repeat_region      3438. .3548
                        /rpt_family="L1"
     repeat_region      3549. .3859
                        /rpt_family="Alu"
     repeat_region      3860. .3891
                        /rpt_family="L1"
     repeat_region      3936. .4009
                        /rpt_family="CR1"
     repeat_region      5320. .5362
                        /rpt_family="CR1"
     repeat_region      5365. .5446
                        /rpt_family="AcHobo"
     repeat_region      5491. .5521
                        /rpt_family="MIR"
     repeat_region      5526. .5571
                        /rpt_family="CR1"
     rRNA               complement(5771. .5860)
                        /product="rRNA-5S_rRNA-related"
     repeat_region      5771. .5860
                        /rpt_family="5S"
     repeat_region      6245. .6555
```

```
                     /rpt_family="Alu"
     repeat_region   6785. .6812
                     /rpt_family="AT_rich"
     repeat_region   7501. .7800
                     /rpt_family="Alu"
     repeat_region   7864. .7884
                     /rpt_family="AT_rich"
     repeat_region   8393. .8707
                     /rpt_family="Alu"
     repeat_region   9733. .10028
                     /rpt_family="Alu"
     repeat_region   10324. .10521
                     /rpt_family="MIR"
     repeat_region   11074. .11279
                     /rpt_family="L2"
     repeat_region   11338. .11642
                     /rpt_family="Alu"
     repeat_region   11690. .12111
                     /rpt_family="CT-rich"
     repeat_region   12118. .12392
                     /rpt_family="Alu"
     repeat_region   12924. .13569
                     /rpt_family="ERV1"
     repeat_region   14427. .14642
                     /rpt_family="MIR"
     misc_feature    14589. .15717
                     /note="CpG_island (%GC=70.4, o/e=0.91, #CpGs=143)"
     gene            14788. .15137
                     /gene="ST7"
     mRNA            14788. .15137
                     /gene="ST7"
     repeat_region   14814. .14913
                     /rpt_family="GC_rich"
     CDS             14987. .>15137
                     /gene="ST7"
                     /note="Homo sapiens suppression of tumorigenicity 7 (ST7),
                     transcript variant a, mRNA.; H_NH0374M07.1
                     This gene was based on gi(11761623 11761625)
                     Continues as H_RG114A06.2.1"
                     /codon_start=1
                     /product="unknown"
                     /protein_id="AAS07529.1"
                     /db_xref="GI:41473937"

  Query Match             7.5%;  Score 372;  DB 9;  Length 99596;
  Best Local Similarity   72.5%;  Pred. No. 8.4e-65;
  Matches  522;  Conservative    0;  Mismatches  195;  Indels    3;  Gaps    3;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              | ||| |||||||||||  ||||||||  ||||| ||||||||||||||||||| ||||||
Db      93552 AATTCGAGACCAGCCTGGTCAACATGGCTAAACCTCATCTCTACTAAAAATACAAAAATT 93493

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              || ||  | |||||  |   ||||||||||||||||||||  ||||||||||||||||
Db      93492 AGCCACACGTGGTGGCATTTGCCTGTAATCCCAGCTACTCAGGAGGCTGAGGCAGGAGAA 93433

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              |  ||||||||||  |  |  | || |   | |   ||| ||||  ||| ||||||||||
Db      93432 TTGCTTGAACCTG-GGAAGCAGAAGTTGCAGTGAGCCAAAATCACACCACTGCACTCCAG 93374

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  ||  ||||      | |    |  || ||||||||||     | |||||||| ||
Db      93373 CCTGGGTGACAAAGCAAGACTCCATTGCCAAAAAAAAAAAAAAAAATATATATATAGAT 93314

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              | ||| ||| ||| | | | | | | | ||||| ||| | ||||| ||| |   |||||
Db      93313 AGATAGATAGATATAGATATAGATATAGATATAGATATAGATATAGATATATGCATATAT 93254

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACAT 359
              ||| | ||| ||||  | | ||||| | | | || |||| || ||| |||||| || ||
Db      93253 ATACACACATATATGCATATATATACACATATATGCATATATATACACATATATGCATAT 93194

Qy        360 ATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATAT 419
              || || ||| ||||  | | ||||  | | | ||  ||| | | | ||| | |||||| |
Db      93193 ATATACACATATATGCATATATATGCATATATATGCATATATACACATATATACATATGT 93134

Qy        420 ACACATATATACACATATATACACATATACACATATATACACATATATACATATATACAC 479
              |||| | | ||||  ||| ||| ||||||||||||| |||||| || ||||||||
Db      93133 GTGTATATACATATATATGCACATATACACACATATATACATATATATGCACATATACAC 93074

Qy        480 ATATATATAATATACACACATATATATACACATATATACACACATATATACACATATATA 539
              | ||||||   ||| | |   | |||||||| |||||  || |||||| ||||||||
Db      93073 ACATATAT-GCATATATATGCACATATACACACATATATGCATATATATGCACATATACA 93015

Qy        540 CACATATATATACACATATATACACATATATACATATATACACATATATATACATATATA 599
              |||||||||  | | ||||  ||| ||| | |||||||| || ||||||  |||||||
Db      93014 CACATATATGCATATATATGCACATATACACACATATATGCATATATATGCACATATACA 92955

Qy        600 CACATATATACATATATACACATATATACATATATACACACATATATACACATACATATA 659
              |||||||| ||||||||  || ||||||| | ||| |  |||||||  || | ||| |
Db      92954 CACATATATGCATATATATGCACATATACACACATATATGCATATATATGCACATATACA 92895

Qy        660 CACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATA 719
              |||| | ||  ||||| ||  | ||| ||| | ||||| |  |||||||| | |||||||
Db      92894 CACATATATGCATATATATGCACATATACACACATATATGCGTATATATGCACATATATA 92835


RESULT 20
BX901949
LOCUS       BX901949               38096 bp    DNA     linear   PRI 24-MAY-2004
DEFINITION  Human DNA sequence from clone WI2-86554F10 on chromosome X,
            complete sequence.
ACCESSION   BX901949
VERSION     BX901949.9  GI:47601709
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 38096)
  AUTHORS   Whitehead,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAY-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On May 24, 2004 this sequence version replaced gi:46559263.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome X, constructed by the Sanger Centre Chromosome X Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/ChrX
            WI2-86554F10 is from a Whitehead human fosmid library VECTOR:
```

```
            pEpiFos-5_Eco72I.
FEATURES             Location/Qualifiers
     source          1. .38096
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /clone="WI2-86554F10"
                     /clone_lib="Whitehead_fosmid"
ORIGIN

  Query Match             7.5%;  Score 371.4;  DB 9;  Length 38096;
  Best Local Similarity   77.9%;  Pred. No. 1.3e-64;
  Matches  540;  Conservative    0;  Mismatches  136;  Indels   17;  Gaps    7;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              |||||||||||||||||  |||  ||||||||||||||||| ||||||||||| |||||
Db      35292 AGTTCAAGACCAGCCTGGCCAATGTGGTGAAACCCCATCTCAACTAAAAATACAAAAATC 35351

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   ||| |||||  |  ||||||||||||||||||| |||||| ||||||||||||
Db      35352 AGCTGGGCGTGGTGGCAGGTGCCTGTAATCCCAGCTACTTGGGAGGCTGAGGCAGGAGAA 35411

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGC-AAGATCATGCCATTGCACTCCA 178
              |  |||||||||  || ||| |||||||| | ||||  ||||     |||||||| || |
Db      35412 TTGCTTGAACCTG--GGAGGTAGAGGTTGCAGTGAGCGGAGATTGCACCATTGCATTCTA 35469

Qy        179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAGGCAT---TTTATATAT 235
              |||  || ||| | || |||||   |||| | ||| ||| | |    ||    | |||||||
Db      35470 GCCTGGGTAACGAAAGTGAAATTCTATCTAAAAAACAAATATATATATACATATATATAT 35529

Qy        236 ATATATATATATATATACACACACACACAT---ATATATATACACATATATATACACA 292
              ||| |  | | | | |       | | ||| ||   ||||||| ||| ||||| ||| | |
Db      35530 ATACACGTGTGTGTGTGTGTATATACATATACCATATATACACATATATACATATATA 35589

Qy        293 TATATACATATATACACATATATACACATATATATACACATACATATGTACACATATATA 352
               |||||||||||||||||||||||||||||| ||| ||||| | || || ||| ||| |
Db      35590 CATATACATATATACACATATATACACATATACATATACATATACATATATACACATACA 35649

Qy        353 TACACATATGTATACACATATATACACATATATACACACATATATACACATATATACACA 412
              |||||||| || | | ||| ||| ||| ||||| |||||| || ||| ||||       |
Db      35650 TACACATATATACATATATACATATACACATATATACACATGTACACATATAT------A 35703

Qy        413 CATATATACACATATATACACATATATACACATATACACATATATACACATATATACATA 472
              |||||||||||||||||||| ||||| ||||||| ||| |||||||||| ||||| |
Db      35704 CATATATACACATATATACACACATATATACATATATACACATATACACATAGATACACA 35763

Qy        473 TATACACATATATATAATATACACACATATATATACACATATATACACACATATATACAC 532
              |||||| |||||||  | ||| | ||||||||| ||| ||| ||| ||| ||||| |||
Db      35764 TATACAGATATATA-CACATATATACATATATACACATATACATATACATATATACACAT 35822

Qy        533 ATATATACACATATATATACACATATATACACATATATACATATATACACATATATATAC 592
              | ||| ||| ||||| ||| | | |||||||||||||||| ||||||||||||||| |
Db      35823 ACATACACATATATACATATATACATATACACATATATACACATATACACATATATACAT 35882

Qy        593 ATATATACACATATATACATATATACACATATATACATATATACACACATATATACACAT 652
              ||||| |||||||||||||||||||||||||| |||||| ||| | ||| | | | ||
Db      35883 ATATACACACATATATACATATATACACATATACACATATACACATAGATACACATATAT 35942

Qy        653 ACATATACACACACATAGATATACATATATATA 685
              ||||||| ||||| ||| |||     | | |||
Db      35943 ACATATATACACATATATATAATTTAAAAAATA 35975


RESULT 21
BX324168_0/c
WPCOMMENT
  Sequence split into 6 fragments  LOCUS BX324168 Accession BX324168
     Fragment Name      Begin     End
     BX324168_0             1    110000
     BX324168_1        100001    210000
     BX324168_2        200001    310000
     BX324168_3        300001    410000
     BX324168_4        400001    510000
     BX324168_5        500001    600124
LOCUS       BX324168               600124 bp    DNA     linear   HTG 13-JUN-2004
DEFINITION  Homo sapiens chromosome X clone XXyac-89H7, WORKING DRAFT SEQUENCE,
            31 unordered pieces.
ACCESSION   BX324168
VERSION     BX324168.8  GI:37777001
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 600124)
  AUTHORS   McLay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 20, 2003 this sequence version replaced gi:37650784.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: yX89H7
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 591433 bases at least Q40
            Consensus quality: 594175 bases at least Q30
            Consensus quality: 595562 bases at least Q20
            Insert size: 597124; sum-of-contigs
            Quality coverage: 6.16x in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 31 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    20165: contig of 20165 bp in length
            *    20166    20265: gap of 100 bp
            *    20266    44493: contig of 24228 bp in length
            *    44494    44593: gap of 100 bp
            *    44594    61605: contig of 17012 bp in length
            *    61606    61705: gap of 100 bp
            *    61706    67907: contig of 6202 bp in length
            *    67908    68007: gap of 100 bp
            *    68008    74311: contig of 6304 bp in length
            *    74312    74411: gap of 100 bp
            *    74412    79880: contig of 5469 bp in length
            *    79881    79980: gap of 100 bp
            *    79981    88557: contig of 8577 bp in length
            *    88558    88657: gap of 100 bp
            *    88658   112320: contig of 23663 bp in length
            *   112321   112420: gap of 100 bp
            *   112421   166393: contig of 53973 bp in length
            *   166394   166493: gap of 100 bp
            *   166494   227100: contig of 60607 bp in length
            *   227101   227200: gap of 100 bp
            *   227201   239356: contig of 12156 bp in length
            *   239357   239456: gap of 100 bp
            *   239457   242526: contig of 3070 bp in length
            *   242527   242626: gap of 100 bp
            *   242627   319013: contig of 76387 bp in length
            *   319014   319113: gap of 100 bp
            *   319114   327413: contig of 8300 bp in length
            *   327414   327513: gap of 100 bp
            *   327514   343545: contig of 16032 bp in length
            *   343546   343645: gap of 100 bp
```

```
          *   343646   391808: contig of 48163 bp in length
          *   391809   391908: gap of 100 bp
          *   391909   426746: contig of 34838 bp in length
          *   426747   426846: gap of 100 bp
          *   426847   440337: contig of 13491 bp in length
          *   440338   440437: gap of 100 bp
          *   440438   460971: contig of 20534 bp in length
          *   460972   461071: gap of 100 bp
          *   461072   465425: contig of 4354 bp in length
          *   465426   465525: gap of 100 bp
          *   465526   488190: contig of 22665 bp in length
          *   488191   488290: gap of 100 bp
          *   488291   497966: contig of 9676 bp in length
          *   497967   498066: gap of 100 bp
          *   498067   517262: contig of 19196 bp in length
          *   517263   517362: gap of 100 bp
          *   517363   567911: contig of 50549 bp in length
          *   567912   568011: gap of 100 bp
          *   568012   582171: contig of 14160 bp in length
          *   582172   582271: gap of 100 bp
          *   582272   584652: contig of 2381 bp in length
          *   584653   584752: gap of 100 bp
          *   584753   586940: contig of 2188 bp in length
          *   586941   587040: gap of 100 bp
          *   587041   589956: contig of 2916 bp in length
          *   589957   590056: gap of 100 bp
          *   590057   592460: contig of 2404 bp in length
          *   592461   592560: gap of 100 bp
          *   592561   597736: contig of 5176 bp in length
          *   597737   597836: gap of 100 bp
          *   597837   600124: contig of 2288 bp in length.
FEATURES             Location/Qualifiers
     source          1. .600124
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /clone="XXyac-89H7"
                     /clone_lib="Human-ICRF-YAC"
     misc_feature    1. .20165
                     /note="assembly_fragment:00356
                     fragment_chain:1"
     misc_feature    20266. .44493
                     /note="assembly_fragment:00589
                     fragment_chain:1"
     misc_feature    44594. .61605
                     /note="assembly_fragment:02258
                     fragment_chain:1"
     misc_feature    61706. .67907
                     /note="assembly_fragment:05479
                     fragment_chain:1"
     misc_feature    68008. .74311
                     /note="assembly_fragment:00274
                     fragment_chain:1"
     misc_feature    74412. .79880
                     /note="assembly_fragment:06012
                     fragment_chain:1"
     misc_feature    79981. .88557
                     /note="assembly_fragment:06208
                     fragment_chain:1"
     misc_feature    88658. .112320
                     /note="assembly_fragment:06075
                     fragment_chain:2"
     misc_feature    112421. .166393
                     /note="assembly_fragment:02111
                     fragment_chain:2"
     misc_feature    166494. .227100
                     /note="assembly_fragment:00544
                     fragment_chain:2"
     misc_feature    227201. .239356
                     /note="assembly_fragment:03674
                     fragment_chain:2"
     misc_feature    239457. .242526
                     /note="assembly_fragment:04283
                     fragment_chain:2"
     misc_feature    242627. .319013
                     /note="assembly_fragment:01306
                     fragment_chain:2"
     misc_feature    319114. .327413
                     /note="assembly_fragment:02688
                     fragment_chain:2"
     misc_feature    327514. .343545
                     /note="assembly_fragment:00658
                     fragment_chain:3"
     misc_feature    343646. .391808
                     /note="assembly_fragment:04025
                     fragment_chain:3"
     misc_feature    391909. .426746
                     /note="assembly_fragment:01126
                     fragment_chain:3"
     misc_feature    426847. .440337
                     /note="assembly_fragment:06618
                     fragment_chain:3"
     misc_feature    440438. .460971
                     /note="assembly_fragment:03859
                     fragment_chain:3"
     misc_feature    461072. .465425
                     /note="assembly_fragment:03896
                     fragment_chain:3"
     misc_feature    465526. .488190
                     /note="assembly_fragment:06799
                     fragment_chain:4"
     misc_feature    488291. .497966
                     /note="assembly_fragment:02743
                     fragment_chain:4"
     misc_feature    498067. .517262
                     /note="assembly_fragment:06602
                     fragment_chain:4"
     misc_feature    517363. .567911
                     /note="assembly_fragment:02958
                     fragment_chain:5"
     misc_feature    568012. .582171
                     /note="assembly_fragment:05773
                     fragment_chain:5"
     misc_feature    582272. .584652
                     /note="assembly_fragment:00466"
     misc_feature    584753. .586940
                     /note="assembly_fragment:02122"
     misc_feature    587041. .589956
                     /note="assembly_fragment:04085"

  Query Match             7.5%;  Score 371.4;  DB 2;  Length 110000;
  Best Local Similarity   77.9%;  Pred. No. 1.1e-64;
  Matches  540;  Conservative    0;  Mismatches  136;  Indels   17;  Gaps    7;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              |||||||||||||||||  |||  |||||||||||||||||| |||||||||| |||||
Db      82809 AGTTCAAGACCAGCCTGGCCAATGTGGTGAAACCCCATCTCAACTAAAAATACAAAAATC 82750

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   ||| |||||   |   ||||||||||||||||||| ||||||  ||||||||||||
Db      82749 AGCTGGGCGTGGTGGCAGGTGCCTGTAATCCCAGCTACTTGGGAGGCTGAGGCAGGAGAA 82690

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGC-AAGATCATGCCATTGCACTCCA 178
              |  ||||||||||  || ||| |||||||| | ||||  ||||    ||||||||  || |
Db      82689 TTGCTTGAACCTG--GGAGGTAGAGGTTGCAGTGAGCGGAGATTGCACCATTGCATTCTA 82632

Qy        179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCAT---TTTATATAT 235
              |||  || ||| | || |||||   |||| | ||| ||| | |   ||   | |||||||
Db      82631 GCCTGGGTAACGAAAGTGAAATTCTATCTAAAAAACAAATATATATATACATATATATAT 82572

Qy        236 ATATATATATATATATACACACACACACACAT---ATATATATACACATATATATACACA 292
              ||| |  | | | | |       | | ||| ||   |||||||  ||| ||||| ||| | |
Db      82571 ATACACGTGTGTGTGTGTGTATATACATATACCATATATACACATATATACATATATA 82512

Qy        293 TATATACATATATACACATATATACACATATATATACACATACATATGTACACATATATA 352
```

```
                 ||||||||||||||||||||||||||||||| ||| ||||| | || || ||| ||| |
Db      82511 CATATACATATATACACATATATACACATATACATATACATATACATATATACACATACA 82452

Qy        353 TACACATATGTATACACATATATACACATATATACACACATATATACACATATATACACA 412
              ||||||||| || | | ||| ||| ||| ||||| |||||| || ||| ||||      |
Db      82451 TACACATATATACATATATACATATACACATATATACACATGTACACATATAT------A 82398

Qy        413 CATATATACACATATATACACATATATACACATATACACATATATACACATATATACATA 472
              |||||||||||||||||||| ||||| ||||||| ||| |||||||||| ||||| |
Db      82397 CATATATACACATATATACACACATATATACATATATACACATATACACATAGATACACA 82338

Qy        473 TATACACATATATATAATATACACACATATATATACACATATATACACACATATATACAC 532
              |||||| |||||||  | ||| | ||||||||| ||| ||| ||| ||| ||||| |||
Db      82337 TATACAGATATATA-CACATATATACATATATACACATATACATATACATATATACACAT 82279

Qy        533 ATATATACACATATATATACACATATATACACATATATACATATATACACATATATATAC 592
              | ||| ||| ||||| ||| | | |||||||||||||||||| ||||||||||||||| |
Db      82278 ACATACACATATATACATATATACATATACACATATATACACATATACACATATATACAT 82219

Qy        593 ATATATACACATATATACATATATACACATATATACATATATACACACATATATACACAT 652
              ||||| |||||||||||||||||||||||||| ||||||| ||| | ||| | | | ||
Db      82218 ATATACACACATATATACATATATACACATATACACATATACACATAGATACACATATAT 82159

Qy        653 ACATATACACACACATAGATATACATATATATA 685
              ||||||| ||||| ||| |||     | | |||
Db      82158 ACATATATACACATATATATAATTTAAAAAATA 82126
```

```
RESULT 22
AC097326
LOCUS       AC097326               252512 bp    DNA     linear   HTG 15-JUN-2002
DEFINITION  Pan troglodytes clone RP43-10G9, WORKING DRAFT SEQUENCE, 14
            unordered pieces.
ACCESSION   AC097326
VERSION     AC097326.3  GI:21397202
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 252512)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 252512)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-OCT-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 252512)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JUN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jun 13, 2002 this sequence version replaced gi:21392442.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: ZUAK
                Center clone name: RP43-10G9
            -------------- Summary Statistics
                Sequencing vector: Plasmid; M77789
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 241130 bases at least Q40
                Consensus quality: 243666 bases at least Q30
                Consensus quality: 245673 bases at least Q20
                Estimated insert size: 259407; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 7.4x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
            * NOTE: This sequence may represent more than one clone.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 14 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    55034: contig of 55034 bp in length
            *    55035    55134: gap of unknown length
            *    55135    92628: contig of 37494 bp in length
            *    92629    92728: gap of unknown length
            *    92729   119376: contig of 26648 bp in length
            *   119377   119476: gap of unknown length
            *   119477   142089: contig of 22613 bp in length
            *   142090   142189: gap of unknown length
            *   142190   165691: contig of 23502 bp in length
            *   165692   165791: gap of unknown length
            *   165792   184091: contig of 18300 bp in length
            *   184092   184191: gap of unknown length
            *   184192   198445: contig of 14254 bp in length
            *   198446   198545: gap of unknown length
            *   198546   211127: contig of 12582 bp in length
            *   211128   211227: gap of unknown length
            *   211228   223003: contig of 11776 bp in length
            *   223004   223103: gap of unknown length
            *   223104   234877: contig of 11774 bp in length
            *   234878   234977: gap of unknown length
```

```
            *   234978   239452: contig of 4475 bp in length
            *   239453   239552: gap of unknown length
            *   239553   246173: contig of 6621 bp in length
            *   246174   246273: gap of unknown length
            *   246274   250219: contig of 3946 bp in length
            *   250220   250319: gap of unknown length
            *   250320   252512: contig of 2193 bp in length.
FEATURES             Location/Qualifiers
     source          1. .252512
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="RP43-10G9"
ORIGIN

  Query Match             7.5%;  Score 370.4;  DB 2;  Length 252512;
  Best Local Similarity   72.5%;  Pred. No. 1.5e-64;
  Matches  535;  Conservative    0;  Mismatches  196;  Indels    7;  Gaps    4;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAATTA 60
              ||||  |||||||||||| |||||||||||||||| |||||||||||| ||||||||||
Db     249133 AGTTTGAGACCAGCCTGACCAACATGGTGAAACCCCGTCTCTACTAAAACTACAAAATTA 249192

Qy         61 GACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAAT 120
              | ||||| ||||| | |||||||||||||  |||||||| ||| || | |||| ||||||
Db     249193 GCCAGGCGTGGTGGTGCACGCCTGTAATATCAGCTACTTGGGAAGCTGGGGCAAGAGAAT 249252

Qy        121 CACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCCAG 179
              |||||||  |  | || |||||||| ||| | ||| |||||||   ||| ||||||||||
Db     249253 AACTTGAAGC--CGGGAGGTGGAGGCTGCAGTGAGCCAAGATCGCACCACTGCACTCCAG 249310

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  | |||||||||| |||   ||||||  | ||  ||        |     ||||  |
Db     249311 CCTGGACAACAAGAGCAAAACTCCATCTCGAATAATTAATTTATTTATAATACTATAATT 249370

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              |       ||  |  |     |    |||   |  |||         |  | | | |||
Db     249371 AACCGGGCATGGAGGCGTGTGCCTGTATAGCCAGCTACTTGGGAGGCTGAAGAAAGATAA 249430

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACAT 359
              |  |  || | |||| | ||| ||||| | ||| |||| ||   ||||||||| | ||
Db     249431 TCACTTGAACCTGTATATATATAGATATATAGATATATATATATCTATATATATATATAT 249490

Qy        360 ATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATAT 419
              || || | | ||||||| | ||||||| | | ||||||| | ||||||| | | ||||||
Db     249491 ATATAGATATATATATATATATATATATATATATATATATATATATATATATATATAT 249550

Qy        420 ACACATATATACACATATATACACATATACACATATATACACATATATACATATATACAC 479
              | | ||||||| |  |||||| | ||||| | ||||||| |  |||||| ||||||| |
Db     249551 ATATATATATATATCTATATATATATATATATATATATATGTATATATATATATATAT 249610

Qy        480 ATATATATA---ATATACACACATATATATACACATATATACACACATATATACACATAT 536
              |||||||||   ||||| | | ||||||||| | ||||||| | | ||||||| | ||||
Db     249611 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 249670

Qy        537 ATACACATATATATACACATATATACACATATATACATATATACACA-TATATATACATA 595
              ||| | ||||||||| | |||||||| | ||||||| ||||||| | | |||||||| |||
Db     249671 ATATATATATATATATATATATATATATATATATATATATATATATATAGTATATATATATA 249730

Qy        596 TATACACATATATACATATATACACATATATACATATATACACACATATATACACATACA 655
              |||| | ||||||| ||||||| | ||||||| ||||||| | | ||||||| | ||| |
Db     249731 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 249790

Qy        656 TATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATA 715
              ||||   | | ||| ||||||||||||||| | |||||||||  | ||||||||||||
Db     249791 TATATTTATATATATATATACATATATATATATATATATATTTTTATATATGTATATC 249850

Qy        716 TATATGCTCCAGAGTTCA 733
              | | |  | |    ||||
Db     249851 TTTCTCTTTCCAGATTCA 249868


RESULT 23
```

```
AC123545/c
LOCUS       AC123545              170646 bp    DNA     linear   HTG 25-JUL-2003
DEFINITION  Pan troglodytes clone RP43-165N9, WORKING DRAFT SEQUENCE, 10
            ordered pieces.
ACCESSION   AC123545
VERSION     AC123545.3  GI:33235839
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 170646)
  AUTHORS   Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
            Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S., Cariaga,K.,
            Chu,G., Coleman,B., Coleman,H., Engle,J., Granite,S., Guan,X.,
            Gupta,J., Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Hu,P.,
            Hurle,B., Idol,J.R., Karlins,E., Kwong,P., Laric,P., Lee-Lin,S.-Q.,
            Legaspi,R., Maduro,Q.L., Maduro,V.B., Margulies,E.H., Masiello,C.,
            Maskeri,B., McDowell,J., Paguirigan,C., Pearson,R., Portnoy,M.E.,
            Prasad,A., Reddix-Dugue,N., Schandler,K., Schueler,M.G., Shah,K.,
            Sison,C., Stantripop,S., Thomas,J.W., Thomas,P.J., Tsipouri,V.,
            Vogt,J.L., Wetherby,K.D., Young,A. and Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 170646)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAY-2002) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE   3  (bases 1 to 170646)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUL-2003) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     On Jul 25, 2003 this sequence version replaced gi:24414279.
            -------------- Genome Center
                 Center: NIH Intramural Sequencing Center
                 Center code: NISC
                 Web site: http://www.nisc.nih.gov
                 Contact: nisc_zoo@nhgri.nih.gov
            -------------- Project Information
                 Center project name: cmz
                 Center clone name: 165N09

            The sequence data in this record represents an 'enhanced'
            version of a Phase 2 submission. Specifically, the indicated
            order and orientation of each sequence contig has been
            established using one or more of the following: read-pair
            data from individual subclones, overlaps with neighboring
            clones, alignment with available reference sequence (e.g.,
            human), and/or confirmation by PCR testing. In addition,
            the sequence assembly is based on at least 8X average
            coverage in Q20 bases and has been reviewed to rule out
            gross misassemblies, the low-quality ends of sequence
            contigs have been trimmed away, and each base is associated
            with a Phrap-derived quality score.
            -------------- Summary Statistics
                 Sequencing vector: plasmid; n/a; 100% of reads
                 Chemistry: Dye-terminator Big Dye; 100% of reads
                 Assembly program: Phrap; version 0.990319
                 Consensus quality: 167110 bases at least Q40
                 Consensus quality: 168559 bases at least Q30
                 Consensus quality: 169363 bases at least Q20
                 Insert size: 152000; agarose-fp
                 Insert size: 169746; sum-of-contigs
                 Quality coverage: 11.31x in Q20 bases; agarose-fp
                 Quality coverage: 10.13x in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
```

```
           * provided by the submittor.
           * This sequence will be replaced
           * by the finished sequence as soon as it is available and
           * the accession number will be preserved.
           *        1     6384: contig of 6384 bp in length
           *     6385     6484: gap of unknown length
           *     6485    37411: contig of 30927 bp in length
           *    37412    37511: gap of unknown length
           *    37512    46351: contig of 8840 bp in length
           *    46352    46451: gap of unknown length
           *    46452    51816: contig of 5365 bp in length
           *    51817    51916: gap of unknown length
           *    51917    67976: contig of 16060 bp in length
           *    67977    68076: gap of unknown length
           *    68077   114680: contig of 46604 bp in length
           *   114681   114780: gap of unknown length
           *   114781   121651: contig of 6871 bp in length
           *   121652   121751: gap of unknown length
           *   121752   125931: contig of 4180 bp in length
           *   125932   126031: gap of unknown length
           *   126032   128356: contig of 2325 bp in length
           *   128357   128456: gap of unknown length
           *   128457   170646: contig of 42190 bp in length.
FEATURES             Location/Qualifiers
     source          1. .170646
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="RP43-165N9"
                     /clone_lib="RP43"
     misc_feature    1. .6384
                     /note="assembly_fragment
                     missing T7 clone end on 5' end of insert"
     misc_feature    6485. .37411
                     /note="assembly_fragment"
     misc_feature    37512. .46351
                     /note="assembly_fragment"
     misc_feature    46452. .51816
                     /note="assembly_fragment"
     misc_feature    51917. .67976
                     /note="assembly_fragment"
     misc_feature    68077. .114680
                     /note="assembly_fragment"
     misc_feature    114781. .121651
                     /note="assembly_fragment"
     misc_feature    121752. .125931
                     /note="assembly_fragment"
     misc_feature    126032. .128356
                     /note="assembly_fragment"
     misc_feature    128457. .170646
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
ORIGIN

  Query Match             7.4%;  Score 368.2;  DB 2;  Length 170646;
  Best Local Similarity   74.9%;  Pred. No. 4.5e-64;
  Matches  543;  Conservative    0;  Mismatches  168;  Indels   14;  Gaps    6;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||| ||||||||||  ||| || | ||||||||||| ||| ||||| || ||||||
Db     103362 AGTTCGAGACCAGCCTGGCCAATATAGCGAAACCCCATCTGTACCAAAAACACAAAAATT 103303

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   ||  |||||| ||| ||||||||||||| ||||  ||||||  ||||||||||
Db     103302 AGCTGGGTGTGGTGACACATGCCTGTAATCCCAGGTACTCAGGAGGCTAAGGCAGGAGAA 103243

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              | |||||||  |  ||  | || |||     |   ||||||    ||| || |||||||
Db     103242 TTGCTTGAACTGGGAGGCAGAGGTGGT-----GAGCCAAGATTGCACCAGTGTACTCCAG 103188

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
               |  ||  || |||| ||||   ||||||| ||| ||| |||   | |  ||||| |||

Db     103187 GCTTGGTGAC-AGAGTGAAACTCCATCTCAAAAACAAACAAACAAAATACATATACATAA 103129

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATA- 298
              ||||||||| | | | | | ||| | ||| ||||||| || ||||||||||| ||||| |
Db     103128 ATATATATACACATATATATACATATATAAATATATATACCTATATATACATATATACAT 103069

Qy        299 ----CATATATACACATATATACACATATATATACACATACATATGTACACATATATATA 354
                  ||||||| |||||||||||| |||||||||   ||| |||  || | | ||||| |
Db     103068 ATATAATATATATACATATATACATATATATATA-TTATATATACATATATAAATATACA 103010

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              | |||  |||||| ||| | | | ||| ||| | | |||||||||  |||| | | |||
Db     103009 TATATAAATATACATATAAAAATATATACATATAAAAATATATACATGTATAAATATACA 102950

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              | |||| | ||| ||| | | ||||| ||||||| ||||||||| | ||| ||| ||| |
Db     102949 TGTATAAATATACATATATAAATATATACATATATACATATATAAATATACATATATAAA 102890

Qy        475 TACACATATATATAATATACACACATATATATACACATATATACACACATATATACACAT 534
              || | | |||||||| ||| |||  |||| ||| ||||||||| | | ||| ||| | |
Db     102889 TATATACATATATAA-ATATACATGTATAAATATACATATATAAATATATACATATATAA 102831

Qy        535 ATATACACATATATATACACATATATACACATATATACATATATACACATATATATACAT 594
              |||||| ||||| ||| | | ||||| | |||| |||| ||||| ||||||||| | ||
Db     102830 ATATACATATATAAATATATAAATATATAAATATGTACAAATATATACATATATAAATAT 102771

Qy        595 ATATACACATATATACATATATACACATATATACATATATACACACATATATACACATAC 654
              |||||||||||||||| ||||||| | |||||||||||||||| | ||| ||| | ||||
Db     102770 ATATACACATATATAGATATATAAATATATATACATATATACATATATAAATATATATAC 102711

Qy        655 ATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATAT 714
              ||||| | | | ||| | ||| ||| ||||||| ||||| |||  |||| |||| || ||
Db     102710 ATATATAAATATATATACATATATAAATATACATATATAAATATATATACATATATAAAT 102651

Qy        715 ATATA 719
              || ||
Db     102650 ATGTA 102646


RESULT 24
HS323B6
LOCUS       HS323B6               112762 bp    DNA     linear   PRI 05-MAR-2003
DEFINITION  Human DNA sequence from clone RP3-323B6 on chromosome Xq11-12,
            complete sequence.
ACCESSION   Z83841
VERSION     Z83841.1  GI:1871528
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 112762)
  AUTHORS   McDougall,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAR-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Mar 10, 1997 this sequence version replaced gi:1772950.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
```

```
          30); an attempt was made to resolve all sequencing problems, such
          as compressions and repeats; all regions were covered by at least
          one plasmid subclone or more than one M13 subclone; and the
          assembly was confirmed by restriction digest, except on the rare
          occasion of the clone being a YAC.
          The following abbreviations are used to associate primary accession
          numbers given in the feature table with their source databases:
          Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
          on the WORMPEP database can be found at
          http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
          was generated from part of bacterial clone contigs of human
          chromosome X, constructed by the Sanger Centre Chromosome X Mapping
          Group.  Further information can be found at
          http://www.sanger.ac.uk/HGP/ChrX
          RP3-323B6 is from the library RPCI-3 constructed by the group of
          Pieter de Jong. For further details see
          http://www.chori.org/bacpac/home.htm
          VECTOR: pCYPAC2.
FEATURES             Location/Qualifiers
     source          1. .112762
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="RZPD:RPCIP704B06323"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="q11-12"
                     /clone="RP3-323B6"
                     /clone_lib="RPCI-3"
ORIGIN

  Query Match             7.4%;  Score 367;  DB 9;  Length 112762;
  Best Local Similarity   67.1%;  Pred. No. 8.3e-64;
  Matches  599;  Conservative    0;  Mismatches  280;  Indels   14;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC---AAAA 57
              ||||||||||||||| ||   |||| ||||||||||||||||||||| ||||||   ||||
Db      39093 AGTTCAAGACCAGCTTGGCTAACACGGTGAAACCCCATCTCTACTATAAATACAAAAAAA 39152

Qy         58 TTAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAG 117
              |||| ||||||||||| | |||||||||| |||| |  ||  ||||| ||||||| ||
Db      39153 TTAGCCAGGCATGGTGGCATGCGCCTGTAATTCCAGTTGTTTGCGAGGCTGAGGCAGAAG 39212

Qy        118 AATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGA-GCAAGATCATGCCATTGCACTC 176
              |||| ||||||||  | || ||||||| |||| | ||  |  |||||   ||| ||||||
Db      39213 AATCGCTTGAACC--CAGGAGGTGGAGCTTGCAGTGAGGTGAGATCGCACCACTGCACTT 39270

Qy        177 CAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATA 236
              |||||  |   |  | |   | |  | |  | | |  |                | ||||| ||
Db      39271 CAGCCTGGATGAACAAA---AGAATTGAATTAATACTATTGCTTTTCACATATATATGTA 39327

Qy        237 TATATATATATATATACACACACACACACATATATATATACACATATATATACACATATA 296
              | | ||||||||||||               |||||| |||| | ||| ||||| ||| |||
Db      39328 TGTGTATATATATATATGTGTGTGTGTGTATATATGTATATATATACATATATACACATA 39387

Qy        297 TACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACA 356
              |  ||||||| |||||||||||||||| |||| | |||||||| |||||||| |||| |
Db      39388 TGTATATATATACATATATACACATATGTATATATATACATATATACACATATGTATATA 39447

Qy        357 CATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATA 416
               |||  |||| ||| ||||  | ||||||||| | ||| | | |  |||| | | |||||
Db      39448 TATACATATATACACATATGTATATATATACATATATACACATATGTATACATATACATA 39507

Qy        417 TATACACATATATACACATATATACACATATACACATATATACACATATATACATATATA 476
              |||||||||| || | ||||||||| |||||||||||| || ||||||| | |||||
Db      39508 TATACACATATGTATATATATATACATATATACACATATGTATACATATACATATATACA 39567

Qy        477 CACATATATATAATATACACACATATATATACACATATATACACACATATATACACACAT 536
              || || |     ||| | | ||||||||| |||  |||| | | ||||||||||||||||
Db      39568 CATATGT-----ATACATATACATATATACACATGTATACATATACATATATACACATAT 39622

Qy        537 ATACACATATATATACACATATATACACATATATACATATATACACATATATATACATAT 596
               || ||||||| ||| | | | |||   ||| ||| | ||||| ||| |||| || | ||
Db      39623 GTATACATATACATATATACACATATGTATACATATACATATATACACATATGTATACAT 39682

Qy        597 ATACACATATATACATATATACACATATATACATATATACACACATATATACACATACAT 656
              |||||  |||||  ||||||  || |||||||  | |||||  |||  |   ||||  | | | ||
Db      39683 ATACATATATACACATATGTATACATATACATATATACACATATGTATACATATACATAT 39742

Qy        657 ATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATAT 716
              |||||||  |  || | |||  | ||||||||||||||| |||||  ||||  |||| || | ||
Db      39743 ATACACATATGTATACATATACATATATACACATATGTATACATATACATATATACACAT 39802

Qy        717 ATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAG 776
              ||||   |     |  |||          ||  | | ||    |  | ||| |    | | |
Db      39803 ATATATACATATATGTATATATATACACATATAAATATATATATGTATATACACACATAT 39862

Qy        777 AGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAA 836
              |  |  | |    |  |     | |   | || |     |      || |||       |
Db      39863 ATATACATATATATACACATATAAATATATATATGTATATACACACATATATATACACAT 39922

Qy        837 ATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACT 889
              ||| ||| || | ||   | |  ||| | ||   |  || ||    || ||
Db      39923 ATATATACATATATATGTATATATATATATTTTTTTTTTGAGATGAAGTCT 39975


RESULT 25
AC108483/c
LOCUS       AC108483              163976 bp    DNA     linear   PRI 29-MAR-2003
DEFINITION  Homo sapiens chromosome 16 clone RP11-170P24, complete sequence.
ACCESSION   AC108483
VERSION     AC108483.3  GI:29366946
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 163976)
  AUTHORS   DOE Joint Genome Institute, Stanford Human Genome Center and Los
            Alamos National Laboratory.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 163976)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JAN-2002) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 163976)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-2002) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   4  (bases 1 to 163976)
  AUTHORS   DOE Joint Genome Institute, Stanford Human Genome Center and Los
            Alamos National Laboratory.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAR-2003) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Mar 29, 2003 this sequence version replaced gi:19073802.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center and Los Alamos
            National Laboratory
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 100% of Sequence;
            Estimated Total Number of Errors is 0.
FEATURES             Location/Qualifiers
     source          1. .163976
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone="RP11-170P24"
ORIGIN

  Query Match             7.4%;  Score 365.4;  DB 9;  Length 163976;
  Best Local Similarity   72.7%;  Pred. No. 1.6e-63;
```

```
  Matches  527;  Conservative    0;  Mismatches  191;  Indels    7;  Gaps    4;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC--AAAAT 58
              || |  |||||| ||||   ||||||||||||  || ||||| ||||||||||  |||||
Db     146516 AGATTGAGACCATCCTGGCTAACATGGTGAAAGTCCGTCTCTCCTAAAAATACAAAAAAT 146457

Qy         59 TAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGA 118
              |||  ||   |||||  ||  |||||||  ||||||||||    |||||  ||||||||||
Db     146456 TAGCTAGTTGTGGTGGCACGTGCCTGTAGTCCCAGCTACTCGAGAGGCTGAGGCAGGAGA 146397

Qy        119 ATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCA 178
              ||  | ||||||    | ||  | ||  |   | |   ||||||  |||||  ||||||||
Db     146396 ATAGCCTGAACC-AAGGAGGCGGAAGTTGCAGTGAGCCAAGATTGTGCCACCGCACTCCA 146338

Qy        179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATA 238
              | || ||||||||| | | | |         | || || || ||     | |||| | | |
Db     146337 G-CCTGGCAACAACAACAACAACAACAACAACAACAACAACAAATATATATATACACACA 146279

Qy        239 TATATATATATATACACACACACACATATATATATACACATATATATACACATATATA 298
              | | | | ||||| | | ||||||| |||||||| | ||||||||||||||||||| | | |
Db     146278 CACACACACATATATATATACACACATATATATACACACACATATATATACACACACACA 146219

Qy        299 CATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACA 358
              |||||||  ||| | | | | ||||| | |||||  | ||||  ||||||  | | | |||  |
Db     146218 TATATATATACACACACATATATATACACACACACATATATATACACACACACACACATA 146159

Qy        359 TATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATA 418
              |||  || ||||| | | | | ||||||||||||||||||| |||  | |  |||  | |||||
Db     146158 TATATACACACACACACATATATATATACACACATATATATACACACACACATATATATA 146099

Qy        419 TACACATATATACACATATATACACATATACACATATATACACATATATACATATATACA 478
              |||||  | | | | ||||||||||| |||  | |||  | |||||  |||||  |||  | |||
Db     146098 CACACACACACATATATATATACACACATATATATACACACACACATATATATACACACA 146039

Qy        479 CATATATATA---ATATACACACATATATATACACATATATACACACATATATACACATA 535
              ||  | | |||   |||||||||||||||||||||||| | | |||  | ||||||||||| |
Db     146038 CACACACATATATATATACACACATATATATACACACACACACATATATATATACACACA 145979

Qy        536 TATACACATATATATACACATATATACACATATATACATATATACACATATATATACATA 595
              ||||  | | | | | |||  |||||||||||  | |  |||||||||  | | | | |||||
Db     145978 TATATATACACACACACATATATATACACACACACACATATATATATACACACACACATA 145919

Qy        596 TATACACATATATACATATATACACATATATACATATATACACACATATATACACATACA 655
              ||||  | | | | | ||||||||||| | |||  |||||  |||||||  |||||  | | |||
Db     145918 TATATATACACACATATATATACACACACATATATATACACACACACATATATATACACA 145859

Qy        656 TATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATA 715
              | |||  | |  |||  | |  ||||||||||||||||  | |||||   ||||  | |    | | |
Db     145858 CACACATATATATATACACACATATATATACACACACATATATATATACACACACACACA 145799

Qy        716 TATAT 720
              |||||
Db     145798 TATAT 145794


RESULT 26
AC022166
LOCUS       AC022166              163998 bp    DNA     linear   PRI 21-JUL-2001
DEFINITION  Homo sapiens chromosome 16 clone RP11-525K10, complete sequence.
ACCESSION   AC022166
VERSION     AC022166.7  GI:14993688
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 163998)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 163998)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JAN-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 163998)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
REFERENCE   4  (bases 1 to 163998)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JUL-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Jul 21, 2001 this sequence version replaced gi:14327778.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 99.8% of Sequence;
            Estimated Total Number of Errors is 0.2.
            STS Content:
            WI-15858 G24147
            SHGC-79420 G52213.
FEATURES             Location/Qualifiers
     source          1. .163998
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone="RP11-525K10"
ORIGIN

  Query Match             7.4%;  Score 365.4;  DB 9;  Length 163998;
  Best Local Similarity   72.7%;  Pred. No. 1.6e-63;
  Matches  527;  Conservative    0;  Mismatches  191;  Indels    7;  Gaps    4;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC--AAAAT 58
              || |  |||||| ||||   ||||||||||||  || ||||| ||||||||||  |||||
Db     130747 AGATTGAGACCATCCTGGCTAACATGGTGAAAGTCCGTCTCTCCTAAAAATACAAAAAAT 130806

Qy         59 TAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGA 118
              |||  ||   |||||  ||  |||||||  ||||||||||    |||||  ||||||||||
Db     130807 TAGCTAGTTGTGGTGGCACGTGCCTGTAGTCCCAGCTACTCGAGAGGCTGAGGCAGGAGA 130866

Qy        119 ATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCA 178
              ||  | ||||||    | ||  | ||  |   | |   ||||||  |||||  ||||||||
Db     130867 ATAGCCTGAACC-AAGGAGGCGGAAGTTGCAGTGAGCCAAGATTGTGCCACCGCACTCCA 130925

Qy        179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATA 238
              | || ||||||||| | | | |         | || || || ||     | |||| | | |
Db     130926 G-CCTGGCAACAACAACAACAACAACAACAACAACAACAACAAATATATATATACACACA 130984

Qy        239 TATATATATATATACACACACACACATATATATATACACATATATATACACATATATA 298
              | | | | ||||| | | ||||||| |||||||| | ||||||||||||||||||| | | |
Db     130985 CACACACACATATATATATACACACATATATATACACACACATATATATACACACACACA 131044

Qy        299 CATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACA 358
              |||||||  ||| | | | | ||||| | |||||  | ||||  ||||||  | | | |||  |
Db     131045 TATATATATACACACACATATATATACACACACACATATATATACACACACACACACATA 131104

Qy        359 TATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATA 418
              |||  || ||||| | | | | ||||||||||||||||||| |||  | |  |||  | |||||
Db     131105 TATATACACACACACACATATATATATACACACATATATATACACACACACATATATATA 131164

Qy        419 TACACATATATACACATATATACACATATACACATATATACACATATATACATATATACA 478
              |||||  | | | | ||||||||||| |||  | |||  | |||||  |||||  |||  | |||
Db     131165 CACACACACACATATATATATACACACATATATATACACACACACATATATATACACACA 131224

Qy        479 CATATATATA---ATATACACACATATATATACACATATATACACACATATATACACATA 535
              ||  | | |||   |||||||||||||||||||||||| | | |||  | ||||||||||| |
Db     131225 CACACACATATATATATACACACATATATATACACACACACACATATATATATACACACA 131284

Qy        536 TATACACATATATATACACATATATACACATATATACATATATACACATATATATACATA 595
              ||||  | | | | | |||  |||||||||||  | |  |||||||||  | | | | |||||
```

```
Db     131285 TATATATACACACACACATATATATACACACACACACATATATATATACACACACACATA 131344

Qy        596 TATACACATATATACATATATACACATATATACATATATACACACATATATACACATACA 655
              |||| | | | | |||||||||||| | ||| ||||| ||||||| ||||| | | |||
Db     131345 TATATATACACACATATATATACACACACATATATATACACACACACATATATATACACA 131404

Qy        656 TATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATA 715
                | ||| | | ||| | | |||||||||||||||| | |||||  |||| | |    | | |
Db     131405 CACACATATATATATACACACATATATATACACACACATATATATATACACACACACACA 131464

Qy        716 TATAT 720
              |||||
Db     131465 TATAT 131469


RESULT 27
AC092863/c
LOCUS       AC092863               153201 bp    DNA     linear   HTG 11-DEC-2002
DEFINITION  Homo sapiens chromosome 12 clone RP11-355N1, WORKING DRAFT
            SEQUENCE, 5 unordered pieces.
ACCESSION   AC092863
VERSION     AC092863.10  GI:26449042
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 153201)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 153201)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-AUG-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 153201)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-DEC-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Dec 11, 2002 this sequence version replaced gi:19172705.
            ------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Drafting Center Code: BCM
                Contact: hgsc-help@bcm.tmc.edu
            ------------- Project Information
                Center project name: HDJU
                Center clone name: RP11-355N1
            ------------- Summary Statistics
                Sequencing vector: Plasmid;
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 150350 bases at least Q40
                Consensus quality: 150698 bases at least Q30
                Consensus quality: 151119 bases at least Q20
                Estimated insert size: 144092; sum-of-contigs estimation
                Quality coverage: 14x in Q20 bases; sum-of-contigs estimation
            -------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     2895: contig of 2895 bp in length
            *    2896     2995: gap of unknown length
            *    2996    16175: contig of 13180 bp in length
            *   16176    16275: gap of unknown length
            *   16276    56442: contig of 40167 bp in length
            *   56443    56542: gap of unknown length
            *   56543    99843: contig of 43301 bp in length
            *   99844    99943: gap of unknown length
            *   99944   153201: contig of 53258 bp in length.
FEATURES             Location/Qualifiers
     source          1. .153201
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="12"
                     /clone="RP11-355N1"
ORIGIN

  Query Match             7.3%;  Score 364.2;  DB 2;  Length 153201;
  Best Local Similarity   74.5%;  Pred. No. 2.9e-63;
  Matches  543;  Conservative    0;  Mismatches  168;  Indels   18;  Gaps    6;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAATTA 60
              |||| |||||| |||| |||||||| |||||||||||| |||||       ||||||
Db       4255 AGTTTGAGACCATCCTGGCCAACATGGCAAAACCCCATCTTTACTA-------AAAATTG 4203

Qy         61 GACAGGCATGGTGATACACGCCTG-----TAATCCCAGCTACTTCGGAGGCCGAGGCAGG 115
              | | ||| |||||| ||| |||||      ||||||||||||||  |||| |  ||||| |
Db       4202 GTCGGGCGTGGTGACACATGCCTGTAATCTAATCCCAGCTACTCAGGAGCCTAAGGCATG 4143

Qy        116 AGAATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCA-AGATCATGCCATTGCAC 174
              |||||| |||||||||| || |||||||||||| | ||||| | |    |||| || ||
Db       4142 AGAATCGCTTGAACCTG--GGAGGTGGAGGTTGCAGTGAGCAGATAATGGGCCACTGTAC 4085
```

```
Qy        175 TCCAGCCCAGGCAAC--AAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATA 232
              ||||||| ||  ||  |  |      |||||    |  ||| |||| | ||| || ||
Db       4084 TCCAGCCTGGGTGACAGAGCAAGACTCTGTCAAAAAAACAAACAAAACATGCAGTTCATC 4025

Qy        233 TATATATATATATATATATACACACACACACATATATATATACACATATATATACACA 292
              | |   |  || ||||  | | |     | | |   | | ||| | || | | || | |
Db       4024 TTTCCCTCAGTAGATATCGAAAGAGTGGGATAAAAGGAAAAATATATATGTGTGTATATA 3965

Qy        293 TATATACATATATACACATATATACACATATATATACACATACATATGTACACATATATA 352
              |||||| ||||||| | ||||||| | |||||||| | ||| |||| || | |||||||
Db       3964 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 3905

Qy        353 TACACATATGTATACACATATATACACATATATACACACATATATACACATATATACACA 412
              || | |||| |||| | ||||||| | ||||||| | | ||||||| | ||||||| | |
Db       3904 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 3845

Qy        413 CATATATACACATATATACACATATATACACATATACACATATATACACATATATACATA 472
              ||||||| | ||||||| | ||||||| | ||||| | ||||||| | ||||||| |||
Db       3844 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 3785

Qy        473 TATACACATATATATA-ATATACACACATATATATACACATATATACACACATATATACA 531
              |||| | |||||||||| ||||| | | ||||||||| | ||||||| | | ||||||| |
Db       3784 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 3725

Qy        532 CATATATACACATATATATACACATATATACACATATATACATATATACACACATATATATA 591
              ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||||||||
Db       3724 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 3665

Qy        592 CATATATACACATATATACATATATACACATATATACATATATACACACATATATACACA 651
              ||||||| | ||||||| ||||||| | ||||||| ||||||| | | ||||||| | |
Db       3664 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 3605

Qy        652 TACATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTA 711
              || ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||||| ||
Db       3604 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 3545

Qy        712 TATATATAT 720
              |||||||||
Db       3544 TATATATAT 3536
```

RESULT 28
AC092694/c

```
LOCUS       AC092694               151549 bp    DNA     linear   PRI 07-MAR-2002
DEFINITION  Homo sapiens 3q BAC RP11-172A10 (Roswell Park Cancer Institute
            Human BAC Library) complete sequence.
ACCESSION   AC092694
VERSION     AC092694.6  GI:18449730
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 151549)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., He,X.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M., Neal,D.,
            Nelson,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N.,
            Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N.,
            Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L.,
            Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M.,
            Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S.,
            Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E.,
            Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A.,
            Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C.,
            Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L.,
            Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S.,
            Warren,R., Washington,C., Watlington,S., Williams,G.,
            Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y.,
            Wu,Y.F., Zhou,J., Zorrilla,S., Naylor,S.L., Weinstock,G. and
            Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 151549)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUL-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 151549)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-JAN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   4  (bases 1 to 151549)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   5  (bases 1 to 151549)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jan 31, 2002 this sequence version replaced gi:17977520.
            INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
            gc-help@bcm.tmc.edu

            CLONE LENGTH: This sequence does not necessarily represent the
            entire insert of this clone. Overlapping regions of clones are only
            sequenced and submitted once, so the sequence for the remainder of
            the insert may be found in the record for the adjacent clones.
            Overlapping clones are noted at the beginning and end of the
            Features listing.

            ANNOTATION OF FEATURES:
               STSs are identified using ePCR (Genome Res. 7:541-550) searches
            of a local database that includes entries from dbSTS, GDB, and
            local mapping efforts.
               Repeats are identified using RepeatMasker (A. Smit and P. Green,
            unpublished.) for Human and Mouse sequences.
               Genes and Region of sequence similarity are identified by BLAST
            (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
            EST and cDNA sequences. Genes demonstrate at least two exons
            flanked by consensus splice sites that maintained sequence
            continuity across the splice junctions. Sequences that are not
            identical matches are annotated as similar.

            SEQUENCING READ COVERAGE:Sequencing is completed to a minimum
```

standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL:
http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

QUALSTAT-REPORT.

```
FEATURES             Location/Qualifiers
     source          1. .151549
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-172A10"
     repeat_region   137. .646
                     /rpt_family="L1M4c"
     repeat_region   696. .1026
                     /rpt_family="L1M4c"
     repeat_region   complement(1065. .1364)
                     /rpt_family="AluSx"
     repeat_region   1365. .1966
                     /rpt_family="L1M1"
     repeat_region   1967. .2309
                     /rpt_family="MLT1A1"
     repeat_region   2310. .3436
                     /rpt_family="L1M1"
     repeat_region   3437. .3473
                     /rpt_family="(TTTA)n"
     repeat_region   complement(3474. .3802)
                     /rpt_family="LTR17"
     repeat_region   complement(3803. .4094)
                     /rpt_family="HERV17"
     repeat_region   complement(8054. .9082)
                     /rpt_family="HERV17"
     repeat_region   complement(8719. .9163)
                     /rpt_family="Harlequin"
     repeat_region   9166. .9611
                     /rpt_family="L1M1"
     repeat_region   complement(9605. .9798)
                     /rpt_family="L1PA16"
     repeat_region   complement(9805. .9857)
                     /rpt_family="L1"
     repeat_region   9880. .10350
                     /rpt_family="L1M1"
     repeat_region   10786. .11108
                     /rpt_family="L1MA2"
     repeat_region   11109. .11492
                     /rpt_family="L1MA2"
     repeat_region   11612. .12041
                     /rpt_family="L1M1"
     repeat_region   12053. .12130
                     /rpt_family="(TA)n"
     misc_feature    12121. .13100
                     /function="unresolved tandem repeat"
     repeat_region   12131. .12462
                     /rpt_family="(CATATA)n"
     repeat_region   12466. .13020
                     /rpt_family="(TATATG)n"
     repeat_region   13040. .13100
                     /rpt_family="(TATATG)n"
     repeat_region   complement(13101. .13387)
                     /rpt_family="AluY"
     repeat_region   13388. .13466
                     /rpt_family="(TATATG)n"
     repeat_region   13469. .13599
                     /rpt_family="L1M1"
     repeat_region   13611. .14584
                     /rpt_family="L1M1"
     repeat_region   14669. .15274
                     /rpt_family="L1M1"
     repeat_region   15275. .17378
                     /rpt_family="L1M1"
     repeat_region   17379. .17684
                     /rpt_family="AluSc"
     repeat_region   17685. .18217
                     /rpt_family="L1M1"
     repeat_region   18200. .18827
                     /rpt_family="L1"
     repeat_region   18829. .19120
                     /rpt_family="AluJb"
     repeat_region   19123. .19483
                     /rpt_family="L1"
     repeat_region   complement(19478. .19591)
                     /rpt_family="L1"
     repeat_region   20015. .20050
                     /rpt_family="(TA)n"
     repeat_region   20531. .20559
                     /rpt_family="AT_rich"
     repeat_region   20570. .20647
                     /rpt_family="AluSx"
```

```
Query Match             7.3%;  Score 363.6;  DB 9;  Length 151549;
Best Local Similarity   68.3%;  Pred. No. 3.8e-63;
Matches  623;  Conservative    0;  Mismatches  274;  Indels   15;  Gaps    8;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              || ||||||||| ||||   |||| |||||||||||||||||||||||||| ||||||
Db      13316 AGATCAAGACCATCCTGGCTAACACGGTGAAACCCCATCTCTACTAAAAATACAAAAATT 13257

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              || ||||||||||      | |||||| |||||||||||| |||||| ||||||||||||
Db      13256 AGCCAGGCATGGTGGCGGGCACCTGTAGTCCCAGCTACTTGGGAGGCTGAGGCAGGAGAA 13197

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGC-AAGATCATGCCATTGCACTCCA 178
              |  | |||||||||  || ||||||| |||| | |||| |||||   |||| |||||||||
Db      13196 TGGCCTGAACCTG--GGAGGTGGAGCTTGCAGTGAGCTAAGATGGCGCCACTGCACTCCA 13139

Qy        179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATA 238
              |||  |||||| ||||| | | | |      ||||||  | |      |  |||| ||||
Db      13138 GCCTGGGCAAC-AGAGCAAGAAGACTCCATCTCAAAAAATATATTATGTACATATGTATA 13080

Qy        239 TATATATATATATACACACACACACATATATATATACACATATATATACACATATATA 298
              | | | |||| |||   | | |||||||| |||| ||| | |  ||| |||| |||| ||
Db      13079 TGTGTCTATACATATGTATATACACACATGTATACATATATACTTATGTACATATATGTA 13020

Qy        299 CATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACA 358
              |||  ||| |||||| || | | |||||  | |  ||  |||  ||  |||||| |||| |||
Db      13019 TATACATATACATATGTATATACATATACATATGTATATACATATACATATGTATATACA 12960

Qy        359 TATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATA 418
              ||||||||     |||||  || |||  |||| | || ||||||||||||  | |    |||||
Db      12959 CATGTATAC----ATATATACTTATGTACATATATGTATACACATATACATATGTATATA 12904

Qy        419 TACACATATATACACATATATACACATATACACATATATACACATATATACATATATACA 478
              ||||||  ||||||||||||    |  || ||||||  ||||||||||||  | || |||| |
Db      12903 CACACATGTATACACATATACTTATGTACACACATGTATACACATATACATATGTATATA 12844

Qy        479 CATATATATA---ATATACACACATATATATACACATATATACACACATATATACACATA 535
              || | || ||    | ||| ||| || ||||||||||  || ||  |||||||||  | |  ||
Db      12843 CACACATGTATACACATATACATATGTATATACACACATGTATACACATATACATATGTA 12784

Qy        536 TATACACATATATATACACATATATACACATATATACATATATACACATATATATACATA 595
              |||||||| || |||||||||||| | |  |||||||  | ||  | | |  |||||||||
Db      12783 TATACACACATGTATACACATATACATATGTATATACACACATGTATACACATATACATA 12724

Qy        596 TATACACATATATACATATATACACATATATACATATATACACACATATATACACATACA 655
              | || | | | |  |||| ||| ||| |||  |||||||||||| ||||||||| |
Db      12723 TGTATATACACACATGTATACACATATACATATGTATATACACACATGTATACACATATA 12664
```

```
Qy        656 TATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATA 715
              |||   | | | | | ||  ||| | ||| ||||||  ||||  ||  | | |||||||  |
Db      12663 CATATGTATATACACACATGTATACACATATACATATGTATA--TACACACATGTATACA 12606

Qy        716 TATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGA 775
              ||||  |    |  |  | |         |              |||   ||   ||| | | | | |
Db      12605 CATATACATATGTATATACACACATGTATACACATATACATATGTACATATATGTATATA 12546

Qy        776 GAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGA 835
              |  |  |||   ||  ||      |   |||| |    ||     || |||       |
Db      12545 CACATATACATATGTACATATATGTATATACATATATACTTATGTACATATATGTATATA 12486

Qy        836 AATACATAAATGTGTAATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTAACACTATATA 894
              ||| | | |||| ||   |  | | | | |  | | || ||| | |  ||    ||||||
Db      12485 CATATACATATGTATATACACCCATGTATACATATATACTTATGTACATATATGTATATA 12426

Qy        895 TATATATAGTGA 906
              |||||| | |
Db      12425 CATATATACTTA 12414
```

RESULT 29
AC021193
LOCUS AC021193 217985 bp DNA linear PRI 21-FEB-2002
DEFINITION Homo sapiens BAC clone RP11-774G5 from 4, complete sequence.
ACCESSION AC021193
VERSION AC021193.7 GI:15321574
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 217985)
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
PUBMED 9847074
REFERENCE 2 (bases 1 to 217985)
AUTHORS Paulson,E., Meyer,R., Elliott,G., Boyer,E. and Dixon,R.
TITLE The sequence of Homo sapiens BAC clone RP11-774G5
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 217985)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (14-JAN-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 217985)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (28-AUG-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 5 (bases 1 to 217985)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (09-JAN-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 6 (bases 1 to 217985)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (21-FEB-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On Aug 28, 2001 this sequence version replaced gi:15022788.
---------- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
---------- Summary Statistics
Center project name: H_NH0774G05
----------.

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at http://www.chori.org
VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is AC079084. Actual start of this clone is at base position 1 of RP11-774G5; actual end is at base position 217985 of RP11-774G5.

FEATURES Location/Qualifiers
source 1. .217985
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone="RP11-774G5"
/clone_lib="RPCI-11"
repeat_region 1077. .1483
/rpt_family="MER2_type"
repeat_region 1511. .1555
/rpt_family="AT_rich"
repeat_region 1669. .1888
/rpt_family="MaLR"
repeat_region 1889. .2198
/rpt_family="Alu"
repeat_region 2172. .2191
/rpt_family="(CAAA)n"
repeat_region 2192. .2219
/rpt_family="(CAAAA)n"
repeat_region 2199. .2388
/rpt_family="MaLR"
repeat_region 2668. .2947
/rpt_family="Alu"
repeat_region 3099. .3125
/rpt_family="AT_rich"
repeat_region 3100. .3462
/rpt_family="L1"
repeat_region 3730. .3751
/rpt_family="AT_rich"
repeat_region 3999. .4659
/rpt_family="L1"
repeat_region 4460. .4551
/rpt_family="(TA)n"
repeat_region 4698. .4897
/rpt_family="L1"
repeat_region 4898. .5162
/rpt_family="L1"

```
repeat_region   5377. .5405
                /rpt_family="AT_rich"
repeat_region   6918. .7257
                /rpt_family="MaLR"
repeat_region   8680. .8986
                /rpt_family="MaLR"
repeat_region   8989. .10092
                /rpt_family="MER2_type"
repeat_region   10216. .10313
                /rpt_family="L1"
repeat_region   11999. .12023
                /rpt_family="AT_rich"
repeat_region   12490. .12719
                /rpt_family="MIR"
repeat_region   13417. .13960
                /rpt_family="L1"
repeat_region   13945. .13969
                /rpt_family="(A)n"
repeat_region   14697. .15010
                /rpt_family="Alu"
repeat_region   14989. .15023
                /rpt_family="AT_rich"
repeat_region   16415. .16850
                /rpt_family="MaLR"
repeat_region   18007. .18299
                /rpt_family="Alu"
repeat_region   19150. .19172
                /rpt_family="AT_rich"
repeat_region   19177. .19375
                /rpt_family="ERV1"
repeat_region   19490. .19640
                /rpt_family="ERV1"
repeat_region   19679. .20025
                /rpt_family="L1"
repeat_region   20302. .20588
                /rpt_family="ERV1"
repeat_region   21525. .21574
                /rpt_family="Mariner"
repeat_region   22586. .22642
                /rpt_family="(TG)n"
repeat_region   22960. .22980
                /rpt_family="AT_rich"
repeat_region   23682. .23709
                /rpt_family="(A)n"
repeat_region   23892. .24010
                /rpt_family="ERV1"
repeat_region   24012. .24039
                /rpt_family="AT_rich"
repeat_region   24187. .24257
                /rpt_family="(TTATA)n"
repeat_region   24472. .24493
                /rpt_family="(T)n"
repeat_region   25473. .25523
                /rpt_family="AT_rich"
repeat_region   25899. .25993
                /rpt_family="MIR"
repeat_region   25994. .27011
                /rpt_family="L1"
repeat_region   26007. .26053
                /rpt_family="AT_rich"
repeat_region   26269. .26421
                /rpt_family="T-rich"
repeat_region   26617. .26647
                /rpt_family="AT_rich"
repeat_region   27102. .27425
                /rpt_family="L1"
repeat_region   27217. .27237
                /rpt_family="AT_rich"
repeat_region   27512. .27828
                /rpt_family="L1"
repeat_region   27595. .27630
                /rpt_family="AT_rich"
repeat_region   27923. .28308
                /rpt_family="L1"
repeat_region   28362. .29158
                /rpt_family="L1"
repeat_region   29254. .29527
```

```
Query Match             7.2%;  Score 357.4;  DB 9;  Length 217985;
Best Local Similarity   78.0%;  Pred. No. 6.2e-62;
Matches  495;  Conservative    0;  Mismatches  131;  Indels    9;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC--AAAAT 58
              || || ||||||  |||   ||||||||||||||| ||||||||||||| |  |||||
Db     146265 AGATCGAGACCATTCTGGCTAACATGGTGAAACCCCGTCTCTACTAAAAATTCCAAAAAT 146324

Qy         59 TAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGA 118
              ||| ||| | |||||     |||||||| ||||||| |||   ||||| ||||||||||
Db     146325 TAGCCAGACGTGGTGGCGGGCGCCTGTAGTCCCAGCCACTCGAGAGGCTGAGGCAGGAGA 146384

Qy        119 ATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGC-AAGATCATGCCATTGCACTCC 177
              ||  | ||||||  | || || ||||  ||| | ||||  |||||  |||| ||||||||
Db     146385 ATGGCGTGAACC--CGGGAGGCGGAGCCTGCAGTGAGCTGAGATCGCGCCACTGCACTCC 146442

Qy        178 AGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATAT 237
              ||||  ||| || ||||||| |   | ||||| ||||||||||     | |||||||||
Db     146443 AGCCTGGGCGAC-AGAGCGAGACTCCGTCTCAAAAAAAAAAAAAAATATATATATATATAT 146501

Qy        238 ATATATATATATATACACACACACACATATATATATACACATATATATACACATATAT 297
              | | |||||||||||||| | | | | ||| ||||||||| ||| ||||||| ||| ||||
Db     146502 ACACATATATATATACATATATATATACACATATATATATACACATATATATACACATAT 146561

Qy        298 ACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACAC 357
              | ||| ||| | ||| | | | ||| ||| | | ||||| || || | | ||||||||||
Db     146562 ATATACATATATATACACATATATACATACATATATACACATATATATACATATATACAC 146621

Qy        358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417
              |||| |||| ||||||||| ||| ||| | | | | ||||| | | ||||| ||| ||||
Db     146622 ATATATATATACATATATATACACATACATATATACATATATATACATATATACATATAT 146681

Qy        418 ATACACATATATACACATATATACACATATACACATATATACACATATATACATATATAC 477
              ||||| ||||||||| ||||| | | ||||||| ||||||||| ||||| ||||||||||
Db     146682 ATACATATATATACATATATACATATATATACATATATATACATATATACACATATATAC 146741

Qy        478 ACATATATATAATATACACACATATATATACACATATATACACACATATATACACATATA 537
              | |||||    ||||| ||| | ||||||||| ||||| ||| ||||||||||| |||||
Db     146742 ATATATA---CATATATACATACATATATACATATATACACATACATATATACATATATA 146798

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATACATATA 597
              |||  | ||||||| ||||||||| ||| ||| | ||||||||| | ||| ||| |||||
Db     146799 TACGTACATATATATACATATATATACACATACATATATATACATACATACATATATATA 146858

Qy        598 TACACATATATACATATATACACATATATACATAT 632
              || | ||||||| ||||||| | ||||||| || |
Db     146859 TATATATATATATATATATATATATATATATGT 146893
```

```
RESULT 30
CR812471/c
LOCUS       CR812471              151424 bp    DNA     linear   HTG 27-SEP-2004
DEFINITION  Danio rerio clone CH211-237A4, *** SEQUENCING IN PROGRESS ***, 9
            unordered pieces.
ACCESSION   CR812471
VERSION     CR812471.2  GI:52748670
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 151424)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
```

```
COMMENT     On Sep 27, 2004 this sequence version replaced gi:52694427.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zC237A4
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 147995 bases at least Q40
            Consensus quality: 148500 bases at least Q30
            Consensus quality: 148901 bases at least Q20
            Insert size: 150624; sum-of-contigs
            Insert size: 152063; 2.5% error; agarose-fp
            Quality coverage: 7.11x in Q20 bases; sum-of-contigs Quality
            coverage: 7.09x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    18105: contig of 18105 bp in length
            *    18106    18205: gap of 100 bp
            *    18206    20634: contig of 2429 bp in length
            *    20635    20734: gap of 100 bp
            *    20735    23916: contig of 3182 bp in length
            *    23917    24016: gap of 100 bp
            *    24017    35672: contig of 11656 bp in length
            *    35673    35772: gap of 100 bp
            *    35773    47596: contig of 11824 bp in length
            *    47597    47696: gap of 100 bp
            *    47697    63363: contig of 15667 bp in length
            *    63364    63463: gap of 100 bp
            *    63464   127690: contig of 64227 bp in length
            *   127691   127790: gap of 100 bp
            *   127791   133366: contig of 5576 bp in length
            *   133367   133466: gap of 100 bp
            *   133467   151424: contig of 17958 bp in length.
FEATURES             Location/Qualifiers
     source          1. .151424
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-237A4"
                     /clone_lib="CHORI-211"
     misc_feature    1. .18105
                     /note="assembly_fragment:00546
                     fragment_chain:1
                     clone_end:SP6
                     vector_side:left"
     misc_feature    18206. .20634
                     /note="assembly_fragment:00036
                     fragment_chain:1"
     misc_feature    20735. .23916
                     /note="assembly_fragment:00008
                     fragment_chain:1"
     misc_feature    24017. .35672
                     /note="assembly_fragment:00104
                     fragment_chain:2"
     misc_feature    35773. .47596
                     /note="assembly_fragment:00228
                     fragment_chain:2"
     misc_feature    47697. .63363
                     /note="assembly_fragment:00380
                     fragment_chain:2"
     misc_feature    63464. .127690
                     /note="assembly_fragment:01023
                     fragment_chain:2"
     misc_feature    127791. .133366
                     /note="assembly_fragment:00065"
     misc_feature    133467. .151424
                     /note="assembly_fragment:00782
                     clone_end:T7
                     vector_side:right"
ORIGIN

  Query Match             7.2%;  Score 357.2;  DB 2;  Length 151424;
  Best Local Similarity   65.2%;  Pred. No. 7.2e-62;
  Matches  468;  Conservative    0;  Mismatches  248;  Indels    2;  Gaps    2;

Qy        209 AGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATATACACACACACACACATA 268
              | ||||||||||||     || |||||||||||||||||||||||| | | | | | |||
Db      21248 AAAAAAAAAAAAAGCATATTCATATATATATATATATATATATATATATATATATATATA 21189

Qy        269 TATATATACACATATATATACACATATATACATATATACACATATATACACATATATATA 328
              |||||||| | ||||||||| | ||||||| ||||||| | ||||||| | |||||||||
Db      21188 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 21129

Qy        329 CACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACA 388
              | ||| |||| || | |||||||||| | |||| |||| | ||||||| | ||||||| |
Db      21128 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 21069

Qy        389 CACATATATACACATATATACACACATATATACACATATATACACATATATACACATATA 448
              | ||||||| | ||||||| | | ||||||| | ||||||| | ||||||| | |||||
Db      21068 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 21009

Qy        449 CACATATATACACATATATACATATATACACATATATATA-ATATACACACATATATATA 507
              | ||||||| | ||||||| ||||||| | ||||||||| ||||| | | |||||||||
Db      21008 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 20949

Qy        508 CACATATATACACACATATATACACATATATACACATATATATACACATATATACACATA 567
              | ||||||| | | ||||||| | ||||||| | ||||||||| | ||||||| | |||
Db      20948 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 20889

Qy        568 TATACATATATACACATATATATACATATATACACATATATACATATATACACATATATA 627
              |||| ||||||| | |||||||| ||||||| | ||||||| ||||||| | |||||||
Db      20888 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 20829

Qy        628 CATATATACACACATATATACACATACATATACACACACATAGATATACATATATATACA 687
              ||||||| | | ||||||| | ||| ||||| | | | ||| ||||| ||||||||| |
Db      20828 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 20769

Qy        688 CATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGT 747
              |||||||||  |||||||| ||||||||||
Db      20768 TATATATATATATATATATATATATATATATATANNNNNNNNNNNNNNNNNNNNNNNN 20709

Qy        748 TGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACA 807

Db      20708 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 20649

Qy        808 TTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA 867
                          | |||       | ||| ||| || | ||   | |  ||| | ||
Db      20648 NNNNNNNNNNNNNNTATATATATATATATATATATATATATATATATATATATATATA 20589

Qy        868 -AGGTAGTTAAGAAGGTAACACTATATATATATATAGTGAAAGCAGTGTAAACCTAAA 924
               |  ||  ||   |  ||    |||||||||||||   | |    | || |  || |
Db      20588 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 20531


RESULT 31
HS141H5/c
LOCUS       HS141H5               121600 bp    DNA     linear   PRI 05-JUN-2003
DEFINITION  Human DNA sequence from clone RP6-141H5 on chromosome Xq22.1-23
            Contains the 3' end of the gene for a novel protein with von
            Willebrand factor type C domains, complete sequence.
ACCESSION   AL049176
VERSION     AL049176.3  GI:4808226
KEYWORDS    HTG; von Willebrand factor type C.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
```

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 121600)
  AUTHORS   Pearce,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JUN-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On May 12, 1999 this sequence version replaced gi:4495007.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome X, constructed by the Sanger Centre Chromosome X Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/ChrX
            RP6-141H5 is from the library RPCI-6 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pPAC4
            This sequence is the entire insert of clone RP6-141H5.
FEATURES              Location/Qualifiers
     source           1. .121600
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                      /db_xref="RZPD:RPCIP709H05141"
                      /db_xref="taxon:9606"
                      /chromosome="X"
                      /map="q22.1-23"
                      /clone="RP6-141H5"
                      /clone_lib="RPCI-6"
     repeat_region    1. .665
                      /note="L1P repeat: matches 2. .667 of consensus"
     repeat_region    825. .1034
                      /note="MER58A repeat: matches 2. .216 of consensus"
     repeat_region    1103. .1634
                      /note="L1ME repeat: matches 572. .1075 of consensus"
     repeat_region    1635. .2006
                      /note="MLT1A1 repeat: matches 1. .365 of consensus"
     repeat_region    2007. .2521
                      /note="L1ME repeat: matches 1075. .1707 of consensus"
     repeat_region    2526. .2715
                      /note="L1MEc repeat: matches 1961. .2159 of consensus"
     repeat_region    2761. .4193
                      /note="Tigger1 repeat: matches 902. .2357 of consensus"
     repeat_region    4195. .5277
                      /note="L1PA16 repeat: matches 5057. .6133 of consensus"
     repeat_region    5291. .6213
                      /note="Tigger1 repeat: matches 1. .931 of consensus"
     repeat_region    6576. .8350
                      /note="L1MA8 repeat: matches 4428. .6279 of consensus"
     repeat_region    8501. .11123
                      /note="L1PA10 repeat: matches 3561. .6159 of consensus"
     repeat_region    11119. .11399
                      /note="L1M4 repeat: matches 2315. .2614 of consensus"
     repeat_region    11408. .17541
                      /note="L1PA6 repeat: matches 11. .6143 of consensus"
     repeat_region    17545. .17592
                      /note="24 copies 2 mer aa 77% conserved"
     repeat_region    17716. .17752
                      /note="L1PA16 repeat: matches 6117. .6153 of consensus"
     repeat_region    18281. .18366
                      /note="L1MB8 repeat: matches 6083. .6169 of consensus"
     repeat_region    20394. .20549
                      /note="MER97c repeat: matches 19. .177 of consensus"
     repeat_region    20895. .21189
                      /note="L2 repeat: matches 2428. .2709 of consensus"
     repeat_region    21669. .21960
                      /note="AluJo repeat: matches 1. .310 of consensus"
     repeat_region    21977. .22111
                      /note="MIR repeat: matches 107. .250 of consensus"
     repeat_region    23514. .23619
                      /note="53 copies 2 mer at 83% conserved"
     repeat_region    23520. .23619
                      /note="25 copies 4 mer atat 85% conserved"
     repeat_region    23621. .23821
                      /note="AluJo repeat: matches 92. .283 of consensus"
     repeat_region    24261. .24306
                      /note="L2 repeat: matches 2659. .2705 of consensus"
     repeat_region    24403. .24543
                      /note="MIR repeat: matches 116. .262 of consensus"
     repeat_region    24579. .24823
                      /note="L2 repeat: matches 2255. .2481 of consensus"
     repeat_region    25586. .25873
                      /note="MER34 repeat: matches 9. .297 of consensus"
     repeat_region    25874. .26288
                      /note="MSTB repeat: matches 1. .426 of consensus"
     repeat_region    26289. .26372
                      /note="MER34 repeat: matches 297. .380 of consensus"
     repeat_region    26373. .26768
                      /note="MSTA repeat: matches 1. .426 of consensus"
     repeat_region    26769. .26932
                      /note="MER34 repeat: matches 380. .543 of consensus"
     repeat_region    27373. .27775
                      /note="L2 repeat: matches 1058. .1475 of consensus"
     repeat_region    28030. .28249
                      /note="MIR repeat: matches 1. .243 of consensus"
     gene             complement(28358. .117309)
                      /gene="dA141H5.1"
     mRNA             complement(join(28358. .30863,33838. .33927,35984. .36151,
                      43096. .43305,48665. .48833,55156. .55223,74337. .74496,
                      75887. .76032,114163. .114256,117197. .>117309))
                      /gene="dA141H5.1"
                      /product="dA141H5.1 (novel protein with von Willebrand
                      factor type C domains)"
                      /note="match: cDNAs: Em:AF039656
                      match: ESTs: Em:AA302430 Em:AI246227 Em:H77289 Em:AA156742
                      Em:AA724151 Em:AA843324 Em:AA057564 Em:AI337909 Em:N39848
                      Em:AA555017 Em:AA135450 Em:N26819 Em:AI123481 Em:AI139649
                      Em:AA018704 Em:R52421 Em:AA724185 Em:AI138345 Em:AA057563
                      Em:AI004265 Em:AA150072 Em:AA244080 Em:AA854359
                      Em:AI420184 Em:R59551 Em:W52860 Em:AA724504 Em:AA635439
                      Em:AA018487 Em:Z42301 Em:AA327755 Em:C00874 Em:N43181
                      Em:AA177053 Em:AA015694 Em:W96276 Em:AA029156 Em:AA463650
                      Em:AA013040 Em:W28849 Em:AA918837"
                      /evidence=not_experimental
     polyA_signal     complement(28361. .28366)
                      /gene="dA141H5.1"
     polyA_site       complement(28384)
                      /gene="dA141H5.1"
     repeat_region    28960. .29007
                      /note="24 copies 2 mer ca 100% conserved"
     misc_feature     complement(28995. .29451)
                      /gene="dA141H5.1"
```

```
                     /note="match: GSS: Em:AQ442614"
     misc_feature    29462. .29764
                     /note="match: GSS: Em:AQ618076"
     repeat_region   30162. .30217
                     /note="28 copies 2 mer tg 89% conserved"
     CDS             complement(join(30733. .30863,33838. .33927,35984. .36151,
                     43096. .43305,48665. .48833,55156. .55223,74337. .74496,
                     75887. .76032,114163. .114256,117197. .>117309))
                     /gene="dA141H5.1"
                     /note="continued from bA814C6.1.1 in Em:AL591489
                     match: proteins: Tr:O57472 Sw:Q24025 Tr:O57465 Sw:Q91713"
                     /codon_start=3
                     /evidence=not_experimental
                     /product="dA141H5.1 (novel protein with von Willebrand
                     factor type C domains)"
                     /protein_id="CAD92488.1"
                     /db_xref="GI:31455411"
                     /translation="SETYCMFQDKKYRVGERWHPYLEPYGLVYCVNCICSENGNVLCS
                     RVRCPNVHCLSPVHIPHLCCPRCPEDSLPPVNNKVTSKSCEYNGTTYQHGELFVAEGL
                     FQNRQPNQCTQCSCSVLSEAFTLILDASNLNSVSPWQEGNVYCGLKTCPKLTCAFPVS
                     VPDSCCRVCRGDGELSWEHSDGDIFRQPANREARHSYHRSHYDPPPSRQAGGLSRFPG
                     ARSHRGALMDSQQASGTIVQIVINNKHKHGQVCVSNGKTYSHGESWHPNLRAFGIVEC
                     VLCTCNVTKQECKKIHCPNRYPCKYPQKIDGKCCKVCPGKKAKEELPGQSFDNKGYFC
                     GEETMPVYESVFMEDGETTRKIALETERPPQVEVHVWTIRKGILQHFHIEKISKRMFE
                     ELPHFKLVTRTTLSQWKIFTEGEAQISQMCSSRVCRTELEDLVKVLYLERSEKGHC"
     misc_feature    complement(31798. .32170)
                     /gene="dA141H5.1"
                     /note="match: GSS: Em:AQ694976"
     repeat_region   32245. .32729
                     /note="L2 repeat: matches 1728. .2263 of consensus"
     repeat_region   34978. .35165
                     /note="MER5A repeat: matches 1. .189 of consensus"
     misc_feature    35093. .35438
                     /note="match: GSS: Em:AQ205639"
     repeat_region   35552. .35734
                     /note="MER5B repeat: matches 6. .177 of consensus"
     misc_feature    37351. .37723
                     /note="match: GSS: Em:B99326"
     repeat_region   38099. .38585
                     /note="LTR32 repeat: matches 15. .471 of consensus"
     repeat_region   38934. .39248
                     /note="L2 repeat: matches 2418. .2748 of consensus"
     repeat_region   39760. .40064
                     /note="AluJb repeat: matches 1. .301 of consensus"
     repeat_region   40389. .40698

  Query Match             7.2%;  Score 356.6;  DB 9;  Length 121600;
  Best Local Similarity   74.7%;  Pred. No. 9.8e-62;
  Matches  543;  Conservative    0;  Mismatches  169;  Indels   15;  Gaps    7;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAA--- 57
              || || |||||| ||||   ||||  |||||||||| ||||||||||||||| |||||
Db      44776 AGATCGAGACCATCCTGGCTAACACAGTGAAACCCCGTCTCTACTAAAAATCCAAAAAAA 44717

Qy         58 --TTAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGG 115
                |||| | ||| |||||      | |||||| |||||||||||  |||||| ||||||||
Db      44716 AGTTAGCCGGGCGTGGTGGCGGGCACCTGTAGTCCCAGCTACTCTGGAGGCTGAGGCAGG 44657

Qy        116 AGAATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCAC 174
              |||||  | |||||  ||  || || |||| |||| | ||| | ||||   |||| |||||
Db      44656 AGAATGGCGTGAACTTG--GGAGGCGGAGCTTGCAGTGAGCCGAGATAGCGCCACTGCAC 44599

Qy        175 TCCAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATA 234
              |||||  |  ||| || ||||||| |   ||||||| |||||||| |      | ||||||
Db      44598 TCCAGTCTGGGCGAC-AGAGCGAGACTCCATCTCAAAAAAAAAATA----TATATATATA 44544

Qy        235 TATATATATATATATATACACACACACACATATATATATACACATATATATACACATA 294
              || |||||||| |||||| | | | | | |||||||| ||| | | |||||||| | |||
Db      44543 TAGATATATATAGATATACAGATATATAGATATATATAGATATATAGATATATATAGATA 44484

Qy        295 TATACATATATACACATATATACACATATATATACACATACATATGTACACATATATATA 354
              |||| ||||||| | ||||||| | ||||||| | | ||| |||| || | ||||||| |
Db      44483 TATAGATATATATAGATATATAGATATATATAGATATATAGATATATATAGATATATAGA 44424

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              | ||||  ||| | | ||||| | | ||||| | | ||||| | | ||||| | | | |
Db      44423 TATATATAGATATATAGATATATATAGATATATAGATATATAGATATATATAGATATATA 44364

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              ||||| | | ||||| | ||||||| | ||||| | ||||||| | ||||||| |||||
Db      44363 GATATATATAGATATATAGATATATATAGATATATAGATATATAGATATATATAGATATA 44304

Qy        475 TACACATATATATAATATACACACATATATATACACATATATACACACATATATACACAT 534
              || |  |||| || ||||| | | ||||| ||| | ||| ||| | | |||||||| | ||
Db      44303 TAGATGTATAGAT-ATATATAGATATATAGATATATATAGATATATAGATATATAGATAT 44245

Qy        535 ATATACACATATATATACACATATATACACATATATACATATATACACATATATATACAT 594
              ||||| | ||||| ||| | ||| ||| | ||||| | ||||| | | ||| |||||  |
Db      44244 ATATAGATATATAGATATATATAGATATATATATAGATATATAGATATATAGATATATAG 44185

Qy        595 ATATACACATATATACATATATACA-CATATATACATATATACACACATATATACACATA 653
              ||||| | ||||||| ||||||| |  ||||||| ||||||| | | ||||| | | |||
Db      44184 ATATATAGATATATAGATATATAGATGATATATAGATATATATAGATATATAGATATATA 44125

Qy        654 CATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATA 713
              ||||| | | | ||||||||| ||| ||||| | ||||||| |  |||| |||| ||||
Db      44124 GATATATAGATATATAGATATATATAGATATATAGATATATAGATATATAGATATATATA 44065

Qy        714 TATATAT 720
              |||||||
Db      44064 GATATAT 44058


RESULT 32
CR788239/c
LOCUS       CR788239              127172 bp    DNA     linear   HTG 18-SEP-2004
DEFINITION  Danio rerio clone CH211-105M19, *** SEQUENCING IN PROGRESS ***, 26
            unordered pieces.
ACCESSION   CR788239
VERSION     CR788239.1  GI:52345234
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 127172)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zC105M19
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 116372 bases at least Q40
            Consensus quality: 119969 bases at least Q30
            Consensus quality: 122718 bases at least Q20
            Insert size: 124672; sum-of-contigs
            Insert size: 168914; 5.4% error; agarose-fp
            Quality coverage: 2.88x in Q20 bases; sum-of-contigs Quality
            coverage: 2.51x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 26 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
```

```
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     3262: contig of 3262 bp in length
            *     3263     3362: gap of 100 bp
            *     3363     5810: contig of 2448 bp in length
            *     5811     5910: gap of 100 bp
            *     5911     8259: contig of 2349 bp in length
            *     8260     8359: gap of 100 bp
            *     8360    11266: contig of 2907 bp in length
            *    11267    11366: gap of 100 bp
            *    11367    14728: contig of 3362 bp in length
            *    14729    14828: gap of 100 bp
            *    14829    18192: contig of 3364 bp in length
            *    18193    18292: gap of 100 bp
            *    18293    25841: contig of 7549 bp in length
            *    25842    25941: gap of 100 bp
            *    25942    28449: contig of 2508 bp in length
            *    28450    28549: gap of 100 bp
            *    28550    32988: contig of 4439 bp in length
            *    32989    33088: gap of 100 bp
            *    33089    37208: contig of 4120 bp in length
            *    37209    37308: gap of 100 bp
            *    37309    41754: contig of 4446 bp in length
            *    41755    41854: gap of 100 bp
            *    41855    48505: contig of 6651 bp in length
            *    48506    48605: gap of 100 bp
            *    48606    69764: contig of 21159 bp in length
            *    69765    69864: gap of 100 bp
            *    69865    77035: contig of 7171 bp in length
            *    77036    77135: gap of 100 bp
            *    77136    80534: contig of 3399 bp in length
            *    80535    80634: gap of 100 bp
            *    80635    85265: contig of 4631 bp in length
            *    85266    85365: gap of 100 bp
            *    85366    87432: contig of 2067 bp in length
            *    87433    87532: gap of 100 bp
            *    87533    92877: contig of 5345 bp in length
            *    92878    92977: gap of 100 bp
            *    92978   100746: contig of 7769 bp in length
            *   100747   100846: gap of 100 bp
            *   100847   103064: contig of 2218 bp in length
            *   103065   103164: gap of 100 bp
            *   103165   106145: contig of 2981 bp in length
            *   106146   106245: gap of 100 bp
            *   106246   108396: contig of 2151 bp in length
            *   108397   108496: gap of 100 bp
            *   108497   114275: contig of 5779 bp in length
            *   114276   114375: gap of 100 bp
            *   114376   121458: contig of 7083 bp in length
            *   121459   121558: gap of 100 bp
            *   121559   124719: contig of 3161 bp in length
            *   124720   124819: gap of 100 bp
            *   124820   127172: contig of 2353 bp in length.
FEATURES             Location/Qualifiers
     source          1. .127172
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-105M19"
                     /clone_lib="CHORI-211"
     misc_feature    1. .3262
                     /note="assembly_fragment:00259
                     fragment_chain:1"
     misc_feature    3363. .5810
                     /note="assembly_fragment:00160
                     fragment_chain:1
                     clone_end:T7
                     vector_side:left"
     misc_feature    5911. .8259
                     /note="assembly_fragment:00081
                     fragment_chain:2"
     misc_feature    8360. .11266
                     /note="assembly_fragment:00224
                     fragment_chain:2"
     misc_feature    11367. .14728
                     /note="assembly_fragment:00278
                     fragment_chain:2"
     misc_feature    14829. .18192
                     /note="assembly_fragment:00208
                     fragment_chain:2"
     misc_feature    18293. .25841
                     /note="assembly_fragment:00560
                     fragment_chain:2"
     misc_feature    25942. .28449
                     /note="assembly_fragment:00192
                     fragment_chain:2"
     misc_feature    28550. .32988
                     /note="assembly_fragment:00144
                     fragment_chain:2"
     misc_feature    33089. .37208
                     /note="assembly_fragment:00176
                     fragment_chain:2"
     misc_feature    37309. .41754
                     /note="assembly_fragment:00343
                     fragment_chain:2"
     misc_feature    41855. .48505
                     /note="assembly_fragment:00443
                     fragment_chain:2"
     misc_feature    48606. .69764
                     /note="assembly_fragment:00603
                     fragment_chain:2"
     misc_feature    69865. .77035
                     /note="assembly_fragment:00417
                     fragment_chain:3"
     misc_feature    77136. .80534
                     /note="assembly_fragment:00367
                     fragment_chain:3"
     misc_feature    80635. .85265
                     /note="assembly_fragment:00391
                     fragment_chain:3"
     misc_feature    85366. .87432
                     /note="assembly_fragment:00132
                     fragment_chain:3"
     misc_feature    87533. .92877
                     /note="assembly_fragment:00321
                     fragment_chain:3"
     misc_feature    92978. .100746
                     /note="assembly_fragment:00479
                     fragment_chain:3"
     misc_feature    100847. .103064
                     /note="assembly_fragment:00049
                     fragment_chain:3"
     misc_feature    103165. .106145
                     /note="assembly_fragment:00241
                     fragment_chain:3"
     misc_feature    106246. .108396
                     /note="assembly_fragment:00090
                     fragment_chain:3"
     misc_feature    108497. .114275
                     /note="assembly_fragment:00299
                     fragment_chain:4"
     misc_feature    114376. .121458
                     /note="assembly_fragment:00518
                     fragment_chain:4"
     misc_feature    121559. .124719
                     /note="assembly_fragment:00110"
     misc_feature    124820. .127172
                     /note="assembly_fragment:00100
                     clone_end:SP6
                     vector_side:right"
ORIGIN

  Query Match             7.2%;  Score 356.6;  DB 2;  Length 127172;
  Best Local Similarity   64.9%;  Pred. No. 9.7e-62;
  Matches  454;  Conservative    0;  Mismatches  244;  Indels    1;  Gaps    1;
```

```
Qy        227 TTTATATATATATATATATATATATACACACACACACATATATATATACACATATATA 286
              |||||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db      42349 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 42290

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | |||||||| | ||| |||| || | |
Db      42289 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 42230

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db      42229 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 42170

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db      42169 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 42110

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db      42109 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 42050

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db      42049 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 41990

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db      41989 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 41930

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db      41929 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 41870

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||||
Db      41869 TATATATATATATATNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 41810

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
                                                                      | |
Db      41809 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTTATA 41750

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885
              ||       | ||| ||| || | ||   | |  ||| | ||   || ||   |  ||
Db      41749 TATATATATATATATATATATATATATATATATATATATATATATTATATATATATATAT 41690

Qy        886 CACTATATATATATAGTGAAAGCAGTGTAAACCTAAA 924
                |||||||||||||   | |   | || |  || |
Db      41689 ATATATATATATATATATATATATATATATATATATATA 41651


RESULT 33
CR753865
LOCUS       CR753865              282240 bp    DNA     linear   HTG 04-SEP-2004
DEFINITION  Danio rerio clone DKEY-172B2, WORKING DRAFT SEQUENCE, 17 unordered
            pieces.
ACCESSION   CR753865
VERSION     CR753865.4  GI:51890091
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 282240)
  AUTHORS   Burton,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 4, 2004 this sequence version replaced gi:51870573.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK172B2
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 274515 bases at least Q40
            Consensus quality: 275937 bases at least Q30
            Consensus quality: 276861 bases at least Q20
            Insert size: 280640; sum-of-contigs
            Insert size: 261465; 4.0% error; agarose-fp
            Quality coverage: 9.23x in Q20 bases; sum-of-contigs Quality
            coverage: 10.00x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 17 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     6899: contig of 6899 bp in length
            *     6900     6999: gap of 100 bp
            *     7000    20575: contig of 13576 bp in length
            *    20576    20675: gap of 100 bp
            *    20676    38280: contig of 17605 bp in length
            *    38281    38380: gap of 100 bp
            *    38381    76364: contig of 37984 bp in length
            *    76365    76464: gap of 100 bp
            *    76465    95438: contig of 18974 bp in length
            *    95439    95538: gap of 100 bp
            *    95539   114600: contig of 19062 bp in length
            *   114601   114700: gap of 100 bp
            *   114701   134153: contig of 19453 bp in length
            *   134154   134253: gap of 100 bp
            *   134254   167898: contig of 33645 bp in length
            *   167899   167998: gap of 100 bp
            *   167999   171913: contig of 3915 bp in length
            *   171914   172013: gap of 100 bp
            *   172014   176193: contig of 4180 bp in length
            *   176194   176293: gap of 100 bp
            *   176294   206625: contig of 30332 bp in length
            *   206626   206725: gap of 100 bp
            *   206726   209648: contig of 2923 bp in length
            *   209649   209748: gap of 100 bp
            *   209749   213682: contig of 3934 bp in length
            *   213683   213782: gap of 100 bp
            *   213783   230655: contig of 16873 bp in length
            *   230656   230755: gap of 100 bp
            *   230756   239434: contig of 8679 bp in length
            *   239435   239534: gap of 100 bp
            *   239535   241595: contig of 2061 bp in length
            *   241596   241695: gap of 100 bp
            *   241696   282240: contig of 40545 bp in length.
FEATURES             Location/Qualifiers
     source          1. .282240
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-172B2"
                     /clone_lib="DanioKey"
     misc_feature    1. .6899
                     /note="assembly_fragment:00130
                     fragment_chain:1"
     misc_feature    7000. .20575
                     /note="assembly_fragment:00501
                     fragment_chain:1"
     misc_feature    20676. .38280
                     /note="assembly_fragment:00282
                     fragment_chain:1"
     misc_feature    38381. .76364
                     /note="assembly_fragment:02629
```

```
                     fragment_chain:1"
     misc_feature    76465. .95438
                     /note="assembly_fragment:00723
                     fragment_chain:1"
     misc_feature    95539. .114600
                     /note="assembly_fragment:01220
                     fragment_chain:1"
     misc_feature    114701. .134153
                     /note="assembly_fragment:01489
                     fragment_chain:1"
     misc_feature    134254. .167898
                     /note="assembly_fragment:01785
                     fragment_chain:1"
     misc_feature    167999. .171913
                     /note="assembly_fragment:00062
                     fragment_chain:1"
     misc_feature    172014. .176193
                     /note="assembly_fragment:00105
                     fragment_chain:1"
     misc_feature    176294. .206625
                     /note="assembly_fragment:02143
                     fragment_chain:1"
     misc_feature    206726. .209648
                     /note="assembly_fragment:00050
                     fragment_chain:2"
     misc_feature    209749. .213682
                     /note="assembly_fragment:00082
                     fragment_chain:2"
     misc_feature    213783. .230655
                     /note="assembly_fragment:00958
                     fragment_chain:2"
     misc_feature    230756. .239434
                     /note="assembly_fragment:00171
                     fragment_chain:2"
     misc_feature    239535. .241595
                     /note="assembly_fragment:00042"
     misc_feature    241696. .282240
                     /note="assembly_fragment:03238.0"
ORIGIN

  Query Match             7.2%;  Score 355.4;  DB 2;  Length 282240;
  Best Local Similarity   43.6%;  Pred. No. 1.5e-61;
  Matches  564;  Conservative    0;  Mismatches  728;  Indels    2;  Gaps    2;

Qy        182 CAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATAT 241
              || |||| || |   ||||         | || || || ||     | |||||||||||||
Db       5750 CATGCAAAAAAAAATAAATAAATAAAAATAATAATAATAATAATATATATATATATATAT 5809

Qy        242 ATATATATATACACACACACACATATATATATACACATATATATACACATATATACAT 301
              |||||||||| | | | | | | |||||||||| | ||||||||| | ||||||| ||
Db       5810 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 5869

Qy        302 ATATACACATATATACACATATATATACACATACATATGTACACATATATATACACATAT 361
              ||||| | ||||||| | ||||||||| | ||| |||| || | ||||||||| | ||||
Db       5870 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 5929

Qy        362 GTATACACATATATACACATATATACACACATATATACACATATATACACACATATATAC 421
              |||| | ||||||| | ||||||| | | ||||||| | ||||||| | | |||||||
Db       5930 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 5989

Qy        422 ACATATATACACATATATACACATATACACATATATACACATATATACATATATACACAT 481
              | ||||||| | ||||||| | ||||| | ||||||| | ||||||| ||||||| | ||
Db       5990 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 6049

Qy        482 ATATATAATATACACACATATATATACACATATATACACACATATATACACATATATACA 541
              |||||| ||||| | | ||||||||| | ||||||| | | ||||||| | ||||||| |
Db       6050 ATATAT-ATATATATATATATATATATATATATATATATATATATATATATATATATA 6108

Qy        542 CATATATATACACATATATACACATATATACATATATACACATATATATACATATATACA 601
              ||||||||| | ||||||| | ||||||| ||||||| | ||||||||| ||||||| |
Db       6109 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 6168

Qy        602 CATATATACATATATACACATATATACATATATACACACATATATACACATACATATACA 661
```

```
              ||||||| ||||||| | ||||||| ||||||| | | ||||||| | ||| ||||| |
Db       6169 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 6228

Qy        662 CACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATATG 721
              | | ||| ||||| ||||||||| | ||||||||| |||||||||| ||||||||||
Db       6229 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 6288

Qy        722 CTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTT 781
                  | |||       | |   ||    |    ||| |   | | | |  |
Db       6289 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 6348

Qy        782 GACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACA 841
              | |    | |||    | |   | || |      |       | |||       | ||| |
Db       6349 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 6408

Qy        842 TAAATGTGTAATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTAACACTATATATATATA 900
              || || | ||  || |  ||| | || |  ||  ||    |  ||     ||||||||||||
Db       6409 TATATATATATAAATATATATATATATATATATATATATANATATATATATATATA 6468

Qy        901 TAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTATAGAAGAATGG 960
              |   |     | || |   |  |          || || || |   | || |
Db       6469 AATATATNTNTATNTANATAANTAAAAATATNNNNNAATATANATATANNAAAANANANN 6528

Qy        961 CTAAGATGCCAAAGCTCAGTGTATGTGGCAGAGGCATGGTGTAGGGTGTGTCCAGGTTCA 1020
                |     |                                   | | |
Db       6529 NNNATTNNNAANNNNNTNTNNNNNTANNNNNNNTNNATNNNNNNTNNTTTNNANANNNN 6588

Qy       1021 TATATTGCATTAAGTGTGAGAACACCCTGGAGTATGAACCAAGAAAATGCAAAAGCCAGA 1080
                                  ||        |   |  ||  |           |
Db       6589 NNNNNNNNNNNNNNNNAANNAAANNNNANNNNNAANNAAAAANNANNNNNNNNNNANN 6648

Qy       1081 AGTGATGGAGGAAATGAGACACAATAATGAAGATATTGAGAGGAGGGTGTGGGCCTAGAG 1140
                    |
Db       6649 NTNNNNNNNANNNANNNTNNNNNNNNNNNNTNNNNNNNNNNNNNNNNNNTNNNNNNN 6708

Qy       1141 TGAAGCTTTTCGTGCCAGTACTTCTTTTGAAGGCCCAGTTCTCTTCTCTCTCGGGGGCTC 1200
                 ||                                | |         |
Db       6709 NNTNAATTNNNNNNNNNNNNNNNNNNNTNNNNNNAANTNNNNNNNNNTANTNNNNNNN 6768

Qy       1201 CTTCATCTCTCATAGAGTCCACAGCTTTTAAGGGCCAACACTTGAGGTCAGCCTGGCTCT 1260
                       |              ||
Db       6769 NNNNNNNNNNANNNNNNNNNANNNNNNAANNNNNNNNNNNNNNNAANTNNNNNNNNN 6828

Qy       1261 CTCATTTGAGCTGGATAGAACATTTTAGAGCACCATCTATTCTTCAAGAGGAAGTTTAAA 1320
                       |    ||                  |
Db       6829 NNNNNNTNNNANTAATNNNNNTNNNNNNNANNNNNNNNNNNTNNNNNNNNTNNNNNNT 6888

Qy       1321 AATAAAAGAACCTTGAAGAGGAAAAAATGTAGACATTCAATCTAACCTTTTCATTTTACT 1380
                 ||
Db       6889 NNNNNNNNAAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 6948

Qy       1381 AGCCAAAGCTAAATAGAATGCAGGTTACCTGTTTTTCAGCCAGGCACCATCATTTCCTAA 1440
                                                                |   ||
Db       6949 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTATAAATAT 7008

Qy       1441 TTGTTATAAAATTTATTATTATTGTTGTTATTAT 1474
                 | ||   | || |||  | ||  |   |  |||
Db       7009 ATATTTATATATATATTAATATTATATTATAT 7042


RESULT 34
AC024180/c
LOCUS       AC024180              149969 bp    DNA     linear   HTG 01-SEP-2000
DEFINITION  Homo sapiens chromosome 3 clone RP11-172A10, WORKING DRAFT
            SEQUENCE, 17 unordered pieces.
ACCESSION   AC024180
VERSION     AC024180.3  GI:8569838
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 149969)
```

```
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 149969)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-FEB-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Jun 16, 2000 this sequence version replaced gi:7630874.

            ---------- Genome Center ----------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            ---------- Project Information ----------
            Center project name: H_NH0172A10
            ---------- Summary Statistics  ----------
            Sequencing vector: M13; 100%
            Sequencing vector: plasmid; 0%
            Chemistry: Dye-primer ET; 97% of reads
            Chemistry: Dye-terminator Big Dye; 3% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 141584 bases at least Q40
            Consensus quality: 144367 bases at least Q30
            Consensus quality: 145803 bases at least Q20
            Insert size: 150000; agarose-fp
            Insert size: 148369; sum-of-contigs
            Quality coverage: 3.76 in Q20 bases; agarose-fp
            Quality coverage: 4.04 in Q20 bases; sum-of-contigs
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 17 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     1731: contig of 1731 bp in length
            *    1732     1831: gap of unknown length
            *    1832     4275: contig of 2444 bp in length
            *    4276     4375: gap of unknown length
            *    4376     6562: contig of 2187 bp in length
            *    6563     6662: gap of unknown length
            *    6663     9713: contig of 3051 bp in length
            *    9714     9813: gap of unknown length
            *    9814    13035: contig of 3222 bp in length
            *   13036    13135: gap of unknown length
            *   13136    17265: contig of 4130 bp in length
            *   17266    17365: gap of unknown length
            *   17366    23146: contig of 5781 bp in length
            *   23147    23246: gap of unknown length
            *   23247    28187: contig of 4941 bp in length
            *   28188    28287: gap of unknown length
            *   28288    35880: contig of 7593 bp in length
            *   35881    35980: gap of unknown length
            *   35981    46076: contig of 10096 bp in length
            *   46077    46176: gap of unknown length
            *   46177    56492: contig of 10316 bp in length
            *   56493    56592: gap of unknown length
            *   56593    67536: contig of 10944 bp in length
            *   67537    67636: gap of unknown length
            *   67637    79604: contig of 11968 bp in length
            *   79605    79704: gap of unknown length
            *   79705    94913: contig of 15209 bp in length
            *   94914    95013: gap of unknown length
            *   95014   112048: contig of 17035 bp in length
            *  112049   112148: gap of unknown length
            *  112149   130002: contig of 17854 bp in length
            *  130003   130102: gap of unknown length
            *  130103   149969: contig of 19867 bp in length.
FEATURES             Location/Qualifiers
     source          1. .149969
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-172A10"
     misc_feature    1. .1731
                     /note="assembly_name:Contig10"
     misc_feature    1832. .4275
                     /note="assembly_name:Contig11"
     misc_feature    4376. .6562
                     /note="assembly_name:Contig12
                     clone_end:T7
                     vector_side:left"
     misc_feature    6663. .9713
                     /note="assembly_name:Contig13"
     misc_feature    9814. .13035
                     /note="assembly_name:Contig14"
     misc_feature    13136. .17265
                     /note="assembly_name:Contig15"
     misc_feature    17366. .23146
                     /note="assembly_name:Contig16"
     misc_feature    23247. .28187
                     /note="assembly_name:Contig17
                     clone_end:SP6
                     vector_side:left"
     misc_feature    28288. .35880
                     /note="assembly_name:Contig18"
     misc_feature    35981. .46076
                     /note="assembly_name:Contig19"
     misc_feature    46177. .56492
                     /note="assembly_name:Contig20"
     misc_feature    56593. .67536
                     /note="assembly_name:Contig21"
     misc_feature    67637. .79604
                     /note="assembly_name:Contig22"
     misc_feature    79705. .94913
                     /note="assembly_name:Contig23"
     misc_feature    95014. .112048
                     /note="assembly_name:Contig24"
     misc_feature    112149. .130002
                     /note="assembly_name:Contig25"
     misc_feature    130103. .149969
                     /note="assembly_name:Contig26"
ORIGIN

  Query Match             7.1%;  Score 355.2;  DB 2;  Length 149969;
  Best Local Similarity   68.3%;  Pred. No. 1.8e-61;
  Matches  596;  Conservative    0;  Mismatches  263;  Indels   13;  Gaps    7;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              || ||||||||| ||||   |||| ||||||||||||||||||||||||||| ||||||
Db      74523 AGATCAAGACCATCCTGGCTAACACGGTGAAACCCCATCTCTACTAAAAATACAAAAATT 74464

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              || ||||||||||      | |||||| |||||||||||| |||||| ||||||||||||
Db      74463 AGCCAGGCATGGTGGCGGGCACCTGTAGTCCCAGCTACTTGGGAGGCTGAGGCAGGAGAA 74404

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGC-AAGATCATGCCATTGCACTCCA 178
              |  | ||||||||  || ||||||| |||| | |||| |||||   |||| |||||||||
Db      74403 TGGCCTGAACCTG--GGAGGTGGAGCTTGCAGTGAGCTAAGATGGCGCCACTGCACTCCA 74346

Qy        179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATA 238
              |||  |||||| ||||| | | | |  |      |||||| | |      |  |||| ||||
Db      74345 GCCTGGGCAAC-AGAGCAAGAAGACTCCATCTCAAAAAATATATTATGTACATATGTATA 74287

Qy        239 TATATATATATATACACACACACACACATATATATATACACATATATATACACATATATA 298
              | | | |||| |||   | | |||||||| |||| ||| | |  ||| |||| |||| ||
Db      74286 TGTGTCTATACATATGTATATACACACATGTATACATATATACTTATGTACATATATGTA 74227

Qy        299 CATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACA 358
              ||| ||| |||||| || | | ||||| | |  || |||  || |||||| |||| |||
Db      74226 TATACATATACATATGTATATACATATACATATGTATATACATATACATATGTATATACA 74167
```

```
Qy        359 TATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATA 418
              ||||||||    ||||| || ||| |||| | || ||||||||||||| | |   |||||
Db      74166 CATGTATAC----ATATATACTTATGTACATATATGTATACACATATACATATGTATATA 74111

Qy        419 TACACATATATACACATATATACACATATACACATATATACACATATATACATATATACA 478
              ||||||  |||||||||||||    | || |||||| |||||||||||| | || |||| |
Db      74110 CACACATGTATACACATATACTTATGTACACACATGTATACACATATACATATGTATATA 74051

Qy        479 CATATATATA---ATATACACACATATATATACACATATATACACACATATATACACATA 535
              ||  || ||    | ||| ||| || |||||||||| || || ||||||||| | |  ||
Db      74050 CACACATGTATACACATATACATATGTATATACACACATGTATACACATATACATATGTA 73991

Qy        536 TATACACATATATATACACATATATACACATATATACATATATACACATATATATACATA 595
              |||||||| || |||||||||||| | |  |||||||| | ||  | | | |||||||||
Db      73990 TATACACACATGTATACACATATACATATGTATATACACACATGTATACACATATACATA 73931

Qy        596 TATACACATATATACATATATACACATATATACATATATACACACATATATACACATACA 655
              | || | | | | |  |||| ||| ||| ||| || |||||   ||||| ||| |||||
Db      73930 TGTATATACACACATGTATACACATATACATATGTACATACATGTATATACACATATACA 73871

Qy        656 TATACACACACATAGATATACATATATATACAC-ATATATATACGTATATATATGTATAT 714
              |||  ||| | ||  |||||||| ||| |   ||  || ||||| ||||||| || || ||
Db      73870 TATGTACATATATGTATATACATACATACTTACTGTACATATATGTATATACATATACAT 73811

Qy        715 ATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAG 774
              || |||   |     |  |||        |   ||   |    ||  ||| |   | |
Db      73810 ATGTATATACACCCATGTATACATATATACTTATGTACATATATGTATATACATATATAC 73751

Qy        775 AGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGG 834
                |  | | |   | ||     |    || |  |  |  |    | | |||
Db      73750 TTATGTACATATATGAATACATATATGTAGATACATGTGTATATGTGTAGATACATATAT 73691

Qy        835 AAATACATAAATGTGTAATAAAACCTATCTGT 866
              ||||||||  ||| ||   |    ||| |||
Db      73690 GAATACATATGTGTATATAGATGTGTATGTGT 73659
```

```
RESULT 35
CR759776/c
LOCUS       CR759776              151047 bp    DNA     linear   HTG 08-SEP-2004
DEFINITION  Danio rerio clone DKEY-205B10, *** SEQUENCING IN PROGRESS ***, 5
            unordered pieces.
ACCESSION   CR759776
VERSION     CR759776.3  GI:51965222
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 151047)
  AUTHORS   McLay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 9, 2004 this sequence version replaced gi:51890366.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK205B10
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 149837 bases at least Q40
            Consensus quality: 150168 bases at least Q30
            Consensus quality: 150366 bases at least Q20
            Insert size: 150647; sum-of-contigs
            Insert size: 157117; 1.3% error; agarose-fp
            Quality coverage: 8.56x in Q20 bases; sum-of-contigs Quality
            coverage: 8.40x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    21342: contig of 21342 bp in length
            *   21343    21442: gap of 100 bp
            *   21443    23542: contig of 2100 bp in length
            *   23543    23642: gap of 100 bp
            *   23643    56824: contig of 33182 bp in length
            *   56825    56924: gap of 100 bp
            *   56925   140273: contig of 83349 bp in length
            *  140274   140373: gap of 100 bp
            *  140374   151047: contig of 10674 bp in length.
FEATURES             Location/Qualifiers
     source          1. .151047
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-205B10"
                     /clone_lib="DanioKey"
     misc_feature    1. .21342
                     /note="assembly_fragment:00167
                     fragment_chain:1"
     misc_feature    21443. .23542
                     /note="assembly_fragment:00022
                     fragment_chain:1"
     misc_feature    23643. .56824
                     /note="assembly_fragment:00416
                     fragment_chain:1"
     misc_feature    56925. .140273
                     /note="assembly_fragment:00819
                     fragment_chain:1"
     misc_feature    140374. .151047
                     /note="assembly_fragment:00054"
ORIGIN
```

```
  Query Match             7.1%;  Score 355.2;  DB 2;  Length 151047;
  Best Local Similarity   64.9%;  Pred. No. 1.8e-61;
  Matches  470;  Conservative    0;  Mismatches  248;  Indels    6;  Gaps    2;

Qy        227 TTTATATATATATATATATATATATACACACACACACATATATATATACACATATATA 286
              |||||||||||||||||||||||||| | | | | | |||||||||| | |||||||
Db     140869 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 140810

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | |||||||||| | ||| |||| || | |
Db     140809 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 140750

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db     140749 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 140690

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db     140689 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 140630

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db     140629 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 140570

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db     140569 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 140510

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
```

```
Db     140509 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 140450

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | ||||| ||||| |||||||||| | |||||||||  ||||||
Db     140449 TATATATATATATATATATACATATATATATATATATATATATATATATATATATATATA 140390

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||
Db     140389 TATATATATATATATANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 140330

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
                                                                      |||
Db     140329 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNAAAA 140270

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA-----AGGTAGTTAAGAA 880
              |  |    | ||| ||| || | || | | |  ||| | ||     |  ||  ||   |
Db     140269 AAAAAAATTATATATATATATTTATATTTATATATATATATATATATATATATATATATA 140210

Qy        881 GGTAACACTATATATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATT 940
                ||   |||||||||||||   | || | |  |||  | |||    |   | | ||
Db     140209 TATATATATATATATATATATATATATAGTACTTAAAAAATGAAGACATGAAGACAGCTT 140150

Qy        941 TAAA 944
              ||||
Db     140149 TAAA 140146
```

```
RESULT 36
AC006441/c
LOCUS       AC006441               162996 bp    DNA     linear   PRI 14-MAY-1999
DEFINITION  Homo sapiens chromosome 17, clone hRPC.56_K_13, complete sequence.
ACCESSION   AC006441
VERSION     AC006441.13  GI:4760435
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 162996)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone hRPC.56_K_13
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 162996)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 162996)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-MAY-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 8, 1999 this sequence version replaced gi:4755216.
            All repeats were identified using RepeatMasker: Smit, A.F.A. &
            Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html.
FEATURES             Location/Qualifiers
     source          1. .162996
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="hRPC.56_K_13"
                     /clone_lib="Peter de Jong/ human PAC library"
     repeat_region   90. .376
                     /rpt_family="AluSx"
     repeat_region   1084. .1390
                     /rpt_family="AluSq"
     unsure          1092. .1093
                     /note="Unresolved compression. Probably GC."
     repeat_region   1424. .1557
                     /rpt_family="AluSg"
     repeat_region   complement(2260. .2414)
                     /rpt_family="AluSc"
     repeat_region   2562. .2857
                     /rpt_family="AluY"
     repeat_region   3102. .3129
                     /rpt_family="(CAAAAA)n"
     repeat_region   complement(3499. .3790)
                     /rpt_family="AluSg"
     repeat_region   complement(3931. .4252)
                     /rpt_family="LTR28"
     repeat_region   complement(4253. .4693)
                     /rpt_family="Tigger1"
     repeat_region   4694. .4996
                     /rpt_family="AluSq"
     repeat_region   complement(4997. .5181)
                     /rpt_family="Tigger1"
     repeat_region   complement(5186. .5466)
                     /rpt_family="AluSg"
     repeat_region   complement(5472. .6215)
                     /rpt_family="Tigger1"
     repeat_region   7119. .7436
                     /rpt_family="AluSq"
     repeat_region   7451. .7521
                     /rpt_family="MADE1"
     repeat_region   7945. .8251
                     /rpt_family="AluSq"
     repeat_region   complement(8354. .8748)
                     /rpt_family="MER21A"
     repeat_region   8743. .9125
                     /rpt_family="LTR1B"
     repeat_region   complement(9126. .9419)
                     /rpt_family="AluSg"
     repeat_region   9420. .9581
                     /rpt_family="LTR1B"
     repeat_region   complement(9784. .9915)
                     /rpt_family="MER21B"
     repeat_region   complement(9916. .10213)
                     /rpt_family="AluSg"
     repeat_region   complement(10214. .10249)
                     /rpt_family="MER21B"
     repeat_region   complement(10305. .10409)
                     /rpt_family="MER21A"
     repeat_region   complement(10410. .11116)
                     /rpt_family="LTR8"
     repeat_region   complement(11117. .11155)
```

```
                     /rpt_family="MER21A"
repeat_region        11215. .11433
                     /rpt_family="MER33"
repeat_region        11434. .11718
                     /rpt_family="AluSx"
repeat_region        11719. .11827
                     /rpt_family="MER33"
repeat_region        complement(11851. .12153)
                     /rpt_family="AluSx"
repeat_region        complement(12155. .12281)
                     /rpt_family="MIR"
repeat_region        12282. .12619
                     /rpt_family="MER1B"
repeat_region        complement(12620. .12649)
                     /rpt_family="MIR"
repeat_region        12656. .12809
                     /rpt_family="L1MA2"
repeat_region        complement(12941. .13659)
                     /rpt_family="LTR1B"
repeat_region        13662. .13779
                     /rpt_family="L1MA3"
repeat_region        13780. .14075
                     /rpt_family="AluJb"
repeat_region        14076. .14240
                     /rpt_family="L1MA3"
repeat_region        complement(14333. .14627)
                     /rpt_family="AluY"
repeat_region        14628. .14651
                     /rpt_family="AT_rich"
repeat_region        complement(14691. .14761)
                     /rpt_family="MIR"
repeat_region        15905. .16202
                     /rpt_family="AluJo"
repeat_region        complement(16333. .16724)
                     /rpt_family="MSTC"
repeat_region        complement(16777. .16993)
                     /rpt_family="MIR"
repeat_region        complement(17020. .17125)
                     /rpt_family="L1MC5"
repeat_region        complement(17141. .17428)
                     /rpt_family="AluSg1"
repeat_region        complement(17432. .17570)
                     /rpt_family="AluSq/x"
repeat_region        complement(17582. .17653)
                     /rpt_family="L1MC5"
repeat_region        17954. .18218
                     /rpt_family="AluSp"
repeat_region        18219. .18289
                     /rpt_family="(GGGAA)n"
repeat_region        18359. .18405
                     /rpt_family="(TG)n"
repeat_region        18639. .18676
                     /rpt_family="(TG)n"
repeat_region        18878. .19179
                     /rpt_family="AluJb"
repeat_region        19202. .19503
                     /rpt_family="AluSx"
repeat_region        complement(20063. .20356)
                     /rpt_family="AluSx"
repeat_region        20835. .20877
                     /rpt_family="L2"
repeat_region        20878. .20952
                     /rpt_family="GA-rich"
repeat_region        complement(21223. .21533)
                     /rpt_family="AluSx"
repeat_region        complement(21991. .22245)
                     /rpt_family="L2"
repeat_region        complement(22402. .22474)
                     /rpt_family="MIR"
repeat_region        complement(22472. .22790)
                     /rpt_family="L2"
repeat_region        complement(22924. .23524)
                     /rpt_family="L2"
repeat_region        23525. .23829
                     /rpt_family="AluJb"
repeat_region        complement(23830. .23861)
                     /rpt_family="L2"
repeat_region        23936. .24244
                     /rpt_family="AluSx"
repeat_region        24247. .24443
                     /rpt_family="AluSg/x"
repeat_region        25664. .25971
                     /rpt_family="AluY"

Query Match           7.1%;  Score 355.2;  DB 9;  Length 162996;
Best Local Similarity  66.4%;  Pred. No. 1.8e-61;
Matches  616;  Conservative    0;  Mismatches  303;  Indels    9;  Gaps    7;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC--AAAAT 58
              || || |||||| ||||   |||| |||||||||  ||||||||||| ||||  |||||
Db      79561 AGATCGAGACCAACCTGGCTAACACGGTGAAACCCTGTCTCTACTAAATATACAAAAAAT 79502

Qy         59 TAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGA 118
               || |  ||  || ||      |||||||| |||||||||||| |||| | ||||||||||
Db      79501 AAGCCGGGAGTGATGGCGGGCGCCTGTAGTCCCAGCTACTTGGGAGACTGAGGCAGGAGA 79442

Qy        119 ATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCC 177
              ||  | | ||||  | || |||||||| ||| | ||| ||||||| ||||| ||||||||
Db      79441 ATGGCGTTAACC--CGGGAGGTGGAGCTTGTAGTGAGCCAAGATCGTGCCACTGCACTCC 79384

Qy        178 AGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATAT 237
              ||||  ||  || ||||||| |      |||| ||||||||||      | |||||||||
Db      79383 AGCCTGGGTGAC-AGAGCGAGACTCTGGCTCAAAAAAAAAAAAAAATATATATATATATAT 79325

Qy        238 ATATATATATATATACACACACACACATATATATATACACATATATATACACATATAT 297
              | ||||||| ||||| | ||| | | | | ||||||||| | | ||| ||| | | ||||
Db      79324 AAATATATAAATATATATACAAATATATAAATATATATAAATACATAAATATAAAAATAT 79265

Qy        298 ACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACAC 357
              | | ||||| | | ||||| | | | ||||| | |||||||   || | || |||||| |
Db      79264 AAAAATATATATAAATATATATA-AAATATATAAATACATAAATATATATTTATATATAA 79206

Qy        358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417
              |||  || | | |||| || | ||| ||||| | ||| ||| | ||| ||| | | ||||
Db      79205 ATACATAAATATATATTTATATATAAATACATAAATAAATATATATACATATATAAATAT 79146

Qy        418 ATACACATATATACACATATATACACATATACACATATATACACATATATACATATATAC 477
              ||| | | ||| | | |||||| | | ||| | | ||||||| | ||||| | |||||||
Db      79145 ATAAATACATAAAAATATATAAATATATAAAAATATATATAAATATATAAAAATATATAT 79086

Qy        478 ACATATATATAATATACACACATATATATACACATATATACACACATATATACACATATA 537
              | ||||||| || | | | | ||| ||||| ||| ||||| | | ||||| | | || ||
Db      79085 AAATATATACAA-AAATATATATAAATATATACAAATATATATAAATATAAAAATATGTA 79027

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATACATATA 597
              || | ||||||| | | ||||||| | ||||| | ||||||| | ||||||| | |||||
Db      79026 TAAATATATATAAAAATATATATAAATATATAAAAATATATATAAATATATAAATATATA 78967

Qy        598 TACACATATATACATATATACACATATATACATATATACACACATATATACACATACATA 657
              || | ||||| | ||||||| | ||||| | ||||||| | | ||| ||| | ||| | |
Db      78966 TAAATATATAAATATATATAAATATATAAATATATATAAATATATAGATATATATATAAA 78907

Qy        658 TACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATA 717
              || | | |  || ||||| ||||||| | | |||||||    |||| |||| |||  |||
Db      78906 TATATATATCTATATATAAATATATAGATATATATATAACTATATAAATATATATGAATA 78847

Qy        718 TATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGA 777
              |||    |   | |||      | |  |         |   ||| | ||| | | |
Db      78846 TATATAACTATATAAATATATATGAATATATATAACTATATAAATATATATGAATATATA 78787

Qy        778 GTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAA 837
                | |  | | |||    | |   | |  |             |   |        ||
Db      78786 TAACTATATAAATATATATGAATATATATAACTATATAAATATATATGAATATATATAAC 78727

Qy        838 TACATAAATGTGTAATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTAACACTATATATA 896
              || |||||| | ||  || |  || || || |  ||  | |  |  ||  |||||||| |
```

```
Db      78726 TATATAAATATATATGAATATATAACTATATAAATATATGAATATATATAACTATATAAA 78667

Qy        897 TATATAGTGAAAGCAGTGTAAACCTAAA 924
              ||||||   | |      || |  ||||
Db      78666 TATATATGAATATATAAATATATATAAA 78639


RESULT 37
AL590381/c
LOCUS       AL590381              182430 bp    DNA     linear   PRI 16-MAY-2001
DEFINITION  Human DNA sequence from clone RP11-373O10 on chromosome 9, complete
            sequence.
ACCESSION   AL590381
VERSION     AL590381.4  GI:13872470
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 182430)
  AUTHORS   Griffiths,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Apr 27, 2001 this sequence version replaced gi:13752046.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr9
            RP11-373O10 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            This sequence is the entire insert of clone RP11-373O10 The true
            left end of clone RP11-82L2 is at 128987 in this sequence. The true
            right end of clone RP11-346G21 is at 9227 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .182430
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /clone="RP11-373O10"
                     /clone_lib="RPCI-11.2"
     repeat_region   8. .317
                     /note="L1M4 repeat: matches 2333. .2656 of consensus"
     repeat_region   348. .683
                     /note="L1M4 repeat: matches 2747. .3082 of consensus"
     repeat_region   763. .1062
                     /note="AluSc repeat: matches 1. .296 of consensus"
     repeat_region   1568. .1826
                     /note="L1M4c repeat: matches 1694. .1953 of consensus"
     repeat_region   2178. .2434
                     /note="L1MB8 repeat: matches 5132. .5403 of consensus"
     repeat_region   2454. .2584
                     /note="MLT1E repeat: matches 35. .157 of consensus"
     repeat_region   2638. .2888
                     /note="MLT1E repeat: matches 312. .561 of consensus"
     repeat_region   2895. .3648
                     /note="L1MB8 repeat: matches 5403. .6171 of consensus"
     repeat_region   4203. .4252
                     /note="L1MC2 repeat: matches 6263. .6312 of consensus"
     repeat_region   4262. .4578
                     /note="L1M4 repeat: matches 2683. .3032 of consensus"
     repeat_region   4866. .5228
                     /note="L1PB3 repeat: matches 5785. .6149 of consensus"
     repeat_region   5844. .5891
                     /note="24 copies 2 mer ta 75% conserved"
     repeat_region   5893. .6306
                     /note="23 copies 18 mer 82% conserved"
     repeat_region   5909. .6298
                     /note="13 copies 30 mer 83% conserved"
     repeat_region   6301. .6597
                     /note="AluY repeat: matches 1. .296 of consensus"
     repeat_region   6743. .6839
                     /note="L1M4 repeat: matches 4041. .4137 of consensus"
     repeat_region   7150. .7519
                     /note="THE1C repeat: matches 1. .371 of consensus"
     repeat_region   8154. .8217
                     /note="THE1B repeat: matches 1. .62 of consensus"
     repeat_region   8237. .8467
                     /note="MIR repeat: matches 2. .253 of consensus"
     repeat_region   8679. .8814
                     /note="AluJo/FRAM repeat: matches 159. .295 of consensus"
     repeat_region   9551. .9910
                     /note="THE1B repeat: matches 1. .363 of consensus"
     repeat_region   11221. .11700
                     /note="L1MC4 repeat: matches 7390. .7847 of consensus"
     repeat_region   13929. .14162
                     /note="L1PA4 repeat: matches 1430. .1662 of consensus"
     repeat_region   14156. .18641
                     /note="L1PA4 repeat: matches 1666. .6146 of consensus"
     repeat_region   18690. .19243
                     /note="L1ME3A repeat: matches 5523. .6091 of consensus"
     repeat_region   19861. .20102
                     /note="L1MEc repeat: matches 1817. .2070 of consensus"
     repeat_region   20203. .20330
                     /note="L1MEc repeat: matches 2167. .2303 of consensus"
     repeat_region   20430. .21121
                     /note="L1 repeat: matches 2894. .3643 of consensus"
     repeat_region   21332. .22100
                     /note="L1M4 repeat: matches 3912. .4721 of consensus"
     repeat_region   22457. .22551
                     /note="L1MC/D repeat: matches 5374. .5466 of consensus"
     repeat_region   23344. .23992
                     /note="L1PA4 repeat: matches 5496. .6144 of consensus"
     repeat_region   26070. .26233
                     /note="MLT1G repeat: matches 12. .179 of consensus"
     repeat_region   26946. .27101
                     /note="LTR1 repeat: matches 631. .785 of consensus"
     repeat_region   27115. .27210
                     /note="LTR1 repeat: matches 1. .95 of consensus"
     repeat_region   27634. .27945
                     /note="AluSx repeat: matches 1. .312 of consensus"
     repeat_region   29014. .29302
                     /note="AluJb repeat: matches 1. .290 of consensus"
     repeat_region   30332. .30492
                     /note="L1MEc repeat: matches 1749. .1897 of consensus"
     repeat_region   31684. .32009
                     /note="MLT1A2 repeat: matches 34. .374 of consensus"
     repeat_region   32010. .32313
                     /note="AluY repeat: matches 1. .305 of consensus"
     repeat_region   33320. .33629
                     /note="AluSx repeat: matches 1. .308 of consensus"
     repeat_region   33683. .33975
                     /note="AluSc repeat: matches 1. .295 of consensus"
     repeat_region   33986. .34132
```

```
                /note="MIR repeat: matches 100. .243 of consensus"
     repeat_region   34156. .34309
                /note="L2 repeat: matches 2575. .2734 of consensus"
     repeat_region   34739. .34862
                /note="MIR repeat: matches 2. .139 of consensus"
     repeat_region   35497. .35578
                /note="41 copies 2 mer tt 67% conserved"
     repeat_region   36076. .36464
                /note="L1M4c repeat: matches 1599. .1989 of consensus"
     repeat_region   36471. .36974
                /note="MLT1D repeat: matches 3. .499 of consensus"
     repeat_region   36977. .37504
                /note="L1M4c repeat: matches 1001. .1604 of consensus"
     repeat_region   37507. .38001
                /note="L1ME2 repeat: matches 5652. .6156 of consensus"
     repeat_region   38998. .39313
                /note="AluSx repeat: matches 1. .289 of consensus"
     repeat_region   39775. .40319
                /note="MLT2D repeat: matches 1. .553 of consensus"
     repeat_region   41011. .41583
                /note="MLT2D repeat: matches 1. .545 of consensus"
     repeat_region   42023. .42458
                /note="MLT2G repeat: matches 1. .386 of consensus"
     repeat_region   42910. .42964
                /note="11 copies 5 mer atata 72% conserved"
     repeat_region   43137. .43251
                /note="L1MB6 repeat: matches 5937. .6053 of consensus"
     repeat_region   43240. .43428
                /note="L1MB6 repeat: matches 5584. .5776 of consensus"
     repeat_region   43431. .44672
                /note="L1MA4 repeat: matches 5053. .6289 of consensus"
     repeat_region   44661. .44816
                /note="L1MA4 repeat: matches 4931. .5090 of consensus"
     repeat_region   44810. .45261
                /note="L1 repeat: matches 3313. .3738 of consensus"
     repeat_region   45270. .45407
                /note="L1M4 repeat: matches 5474. .5595 of consensus"
     repeat_region   45415. .45634
                /note="AluSg repeat: matches 1. .301 of consensus"
     repeat_region   45635. .45677
                /note="AluSg repeat: matches 257. .299 of consensus"
     repeat_region   45680. .45891
                /note="L1M4 repeat: matches 5254. .5470 of consensus"
     repeat_region   46033. .47277
                /note="L1PA14 repeat: matches 4885. .6149 of consensus"
     repeat_region   47274. .47631
                /note="L1MB8 repeat: matches 5816. .6169 of consensus"
     repeat_region   47654. .47961
                /note="AluSx repeat: matches 1. .308 of consensus"
     repeat_region   47964. .49294
                /note="L1MA4 repeat: matches 4951. .6299 of consensus"
     repeat_region   49298. .49946
                /note="L1MD3 repeat: matches 6922. .7588 of consensus"
     repeat_region   51174. .51342
                /note="MER90 repeat: matches 4. .169 of consensus"
     repeat_region   52038. .52075
                /note="19 copies 2 mer tt 81% conserved"
     repeat_region   52894. .53118
                /note="MIR repeat: matches 14. .245 of consensus"
     repeat_region   53389. .53773
                /note="L1ME1 repeat: matches 5778. .6159 of consensus"
     repeat_region   54054. .54383
                /note="L2 repeat: matches 2370. .2709 of consensus"

  Query Match             7.1%;  Score 354.8;  DB 9;  Length 182430;
  Best Local Similarity   73.8%;  Pred. No. 2.1e-61;
  Matches  549;  Conservative    0;  Mismatches  177;  Indels   18;  Gaps    7;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAA---A 57
              || |  |||||| |||    ||||  |||||||||| ||||||| ||||||| ||   |
Db       6524 AGATTGAGACCAACCTAGCGAACACTGTGAAACCCCGTCTCTACCAAAAATACCAAAATA 6465

Qy         58 TTAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAG 117
                |||| ||| ||||     |||||||| ||| ||||||| |||||| ||||||||||
Db       6464 TTAGCTGGGCGCGGTGGCCGGCGCCTGTAGGCCCGGCTACTTGGGAGGCTGAGGCAGGAG 6405

Qy        118 AATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTC 176
              |||    ||||||||  || ||  ||| |||| | ||| | |||    ||.|| |||||||
Db       6404 AATGGTGTGAACCTG--GGAGGCAGAGCTTGCAGTGAGCCGAGACTGCGCCACTGCACTC 6347

Qy        177 CAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATA 236
              |||||  |||||||        | | |      | ||||||||||||      | |||| |||
Db       6346 CAGCCTGGGCAACAGAGCAAGACTCTGTCTCAAAAAAAAAAAAAAAATATATATATACATA 6287

Qy        237 TATATATATATATATACACACACACACATATATATATACACATATATATACACATATA 296
              |||||||| | ||||| | ||||| | | ||||| | ||| | ||| | ||| | ||| |
Db       6286 TATATATACACATATATATACACATATATATATACACATATATATACACATATATATACA 6227

Qy        297 TACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACA 356
              ||||||||| ||| ||||| | | ||||||||||  |||||| || ||||||||||||
Db       6226 CACATATATATACACATATATATACATATATACAC--ACATATATATACATATATATACA 6169

Qy        357 CATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATA 416
              ||||| |||||||| |||||||||||||||| ||| ||||||||| ||||||||| | |||
Db       6168 CATATATATACACACATATACACATATATATACATATATATACATATATATACATATATA 6109

Qy        417 TATACACATATATACACATATATACACATATACAC------ATATATACACATATATACA 470
              | ||| ||||| ||| |||||||||||||||| ||      ||||| ||| |||||||||
Db       6108 CACACATATATACACATATATATACACATATATACACATATATATACACATATATATACA 6049

Qy        471 TATATACACATATATATA-ATATACACACATATATATACACATATATACACACATATATA 529
              |||||||||||||||||| ||||| | ||| ||||||||||||||| ||| | ||||| |
Db       6048 TATATACACATATATATACATATATATACACATATATACACATATACACATATATATACA 5989

Qy        530 CACATATATACACATATATATACACATATATACACATATATACATATATACACATATATA 589
              | |||| |||| |||||| |||| |||||||||||||||  | |||||   | | |||
Db       5988 TATATATGTACATATATATGTACATATATATACACATATATGTACATATATGTACACATA 5929

Qy        590 TACATATATACACATATATACATATATACACATATATACATATATACACACATATATACA 649
              || | ||||| |||||||||||||||||||| |||||   |||||||||| ||||| | |
Db       5928 TATACATATATACATATATACATATATACATATA---ATATATACACATATATACATA 5872

Qy        650 CATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATATG 709
              ||| ||||||| | | | ||| | | ||    |   ||||  || | |   |||
Db       5871 TATATATATACAGAGAGAGAGAGAGAGATGGTGAATTGCATAAGTAAATTTTCTGATAAT 5812

Qy        710 TATATATATATGCTCCAGAGTTCA 733
                ||   | | | |||| | |
Db       5811 GGATAATTATTTCACTAGAGATAA 5788


RESULT 38
CR753861/c
LOCUS       CR753861              252450 bp    DNA     linear   HTG 11-OCT-2004
DEFINITION  Danio rerio clone DKEY-192K11, WORKING DRAFT SEQUENCE, 9 unordered
            pieces.
ACCESSION   CR753861
VERSION     CR753861.4  GI:54035031
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 252450)
  AUTHORS   McLaren,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 11, 2004 this sequence version replaced gi:51861477.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
```

```
            ---------- Project Information
            Center project name: zK192K11
            ---------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 247462 bases at least Q40
            Consensus quality: 248434 bases at least Q30
            Consensus quality: 249070 bases at least Q20
            Insert size: 251650; sum-of-contigs
            Insert size: 239884; 3.5% error; agarose-fp
            Quality coverage: 10.19x in Q20 bases; sum-of-contigs Quality
            coverage: 10.69x in Q20 bases; agarose-fp
            ----------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     5351: contig of 5351 bp in length
            *     5352     5451: gap of 100 bp
            *     5452    19343: contig of 13892 bp in length
            *    19344    19443: gap of 100 bp
            *    19444    58067: contig of 38624 bp in length
            *    58068    58167: gap of 100 bp
            *    58168    75897: contig of 17730 bp in length
            *    75898    75997: gap of 100 bp
            *    75998   166203: contig of 90206 bp in length
            *   166204   166303: gap of 100 bp
            *   166304   216140: contig of 49837 bp in length
            *   216141   216240: gap of 100 bp
            *   216241   224264: contig of 8024 bp in length
            *   224265   224364: gap of 100 bp
            *   224365   228110: contig of 3746 bp in length
            *   228111   228210: gap of 100 bp
            *   228211   252450: contig of 24240 bp in length.
FEATURES             Location/Qualifiers
     source          1. .252450
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-192K11"
                     /clone_lib="DanioKey"
     misc_feature    1. .5351
                     /note="assembly_fragment:00110
                     fragment_chain:1"
     misc_feature    5452. .19343
                     /note="assembly_fragment:00179
                     fragment_chain:1"
     misc_feature    19444. .58067
                     /note="assembly_fragment:00986
                     fragment_chain:1"
     misc_feature    58168. .75897
                     /note="assembly_fragment:00384
                     fragment_chain:1"
     misc_feature    75998. .166203
                     /note="assembly_fragment:02209
                     fragment_chain:2"
     misc_feature    166304. .216140
                     /note="assembly_fragment:01459
                     fragment_chain:2"
     misc_feature    216241. .224264
                     /note="assembly_fragment:00050
                     fragment_chain:2"
     misc_feature    224365. .228110
                     /note="assembly_fragment:00015"
     misc_feature    228211. .252450
                     /note="assembly_fragment:00628.0"
ORIGIN

  Query Match             7.1%;  Score 354;  DB 2;  Length 252450;
  Best Local Similarity   64.7%;  Pred. No. 2.9e-61;
  Matches  452;  Conservative    0;  Mismatches  246;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db     166799 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 166740

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | |||||||||  | ||| |||| || | |
Db     166739 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 166680

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db     166679 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 166620

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db     166619 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 166560

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db     166559 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 166500

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db     166499 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 166440

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||| | ||||||| | ||||||| ||||||| | |  ||||
Db     166439 TATATATATATATATATATATACATATATATATATATATATATATATATATTTATA 166380

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || ||||| ||||| | | | ||| ||||| |||||||||  | |||||||||  ||||||
Db     166379 TATACATATATATAGATATATATATATATATATATATATATATATATATATATATATA 166320

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||  ||||||||||||
Db     166319 TAAATATATATATATGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 166260

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
                                                                      |||
Db     166259 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNAAAA 166200

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885
                |       | | | ||||||   ||   | |  ||| | ||   |  |||   |  ||
Db     166199 AATATTTATATAAAAATAAATANATATATATATATTTATATTTTTTTTATTTATATAT 166140

Qy        886 CACTATATATATATATAGTGAAAGCAGTGTAAACCTAAA 924
                |||||||||||||   | |   | || |  || |
Db     166139 ATATATATATATATATATATATATATATATATATATA 166101


RESULT 39
CR790381
LOCUS       CR790381              163651 bp    DNA     linear   HTG 24-SEP-2004
DEFINITION  Danio rerio clone DKEY-202K24, *** SEQUENCING IN PROGRESS ***, 10
            unordered pieces.
ACCESSION   CR790381
VERSION     CR790381.3  GI:52673626
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 163651)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 24, 2004 this sequence version replaced gi:52626686.
            ---------- Genome Center
```

```
          Center: Wellcome Trust Sanger Institute
          Center code: SC
          Web site: http://www.sanger.ac.uk
          Contact: zfish-help@sanger.ac.uk
          -------------- Project Information
          Center project name: zK202K24
          -------------- Summary Statistics
          Assembly program: XGAP4; version 4.5
          Chemistry: Dye-terminator; 100% of reads
          Consensus quality: 159309 bases at least Q40
          Consensus quality: 160146 bases at least Q30
          Consensus quality: 160896 bases at least Q20
          Insert size: 162751; sum-of-contigs
          Insert size: 165246; 2.0% error; agarose-fp
          Quality coverage: 8.55x in Q20 bases; sum-of-contigs Quality
          coverage: 8.63x in Q20 bases; agarose-fp
          --------------.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 10 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *      1     4775: contig of 4775 bp in length
          *   4776     4875: gap of 100 bp
          *   4876    26055: contig of 21180 bp in length
          *  26056    26155: gap of 100 bp
          *  26156    28584: contig of 2429 bp in length
          *  28585    28684: gap of 100 bp
          *  28685    34192: contig of 5508 bp in length
          *  34193    34292: gap of 100 bp
          *  34293    48345: contig of 14053 bp in length
          *  48346    48445: gap of 100 bp
          *  48446    55079: contig of 6634 bp in length
          *  55080    55179: gap of 100 bp
          *  55180   119903: contig of 64724 bp in length
          * 119904   120003: gap of 100 bp
          * 120004   140199: contig of 20196 bp in length
          * 140200   140299: gap of 100 bp
          * 140300   150280: contig of 9981 bp in length
          * 150281   150380: gap of 100 bp
          * 150381   163651: contig of 13271 bp in length.
FEATURES             Location/Qualifiers
     source          1. .163651
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-202K24"
                     /clone_lib="DanioKey"
     misc_feature    1. .4775
                     /note="assembly_fragment:00038
                     fragment_chain:1"
     misc_feature    4876. .26055
                     /note="assembly_fragment:00583
                     fragment_chain:1"
     misc_feature    26156. .28584
                     /note="assembly_fragment:00027
                     fragment_chain:1"
     misc_feature    28685. .34192
                     /note="assembly_fragment:00076
                     fragment_chain:1"
     misc_feature    34293. .48345
                     /note="assembly_fragment:00435
                     fragment_chain:1"
     misc_feature    48446. .55079
                     /note="assembly_fragment:00118
                     fragment_chain:1"
     misc_feature    55180. .119903
                     /note="assembly_fragment:01052
                     fragment_chain:1"
     misc_feature    120004. .140199
                     /note="assembly_fragment:00806
                     fragment_chain:1"
     misc_feature    140300. .150280
                     /note="assembly_fragment:00174
                     fragment_chain:1"
     misc_feature    150381. .163651
                     /note="assembly_fragment:00293.0"
ORIGIN

  Query Match             7.1%;  Score 353.6;  DB 2;  Length 163651;
  Best Local Similarity   68.5%;  Pred. No. 3.7e-61;
  Matches  502;  Conservative    0;  Mismatches  230;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACATATATATATACACATATATA 286
              |||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db     139185 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 139244

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | |||||||||| | ||| |||| || | |
Db     139245 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 139304

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db     139305 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 139364

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db     139365 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 139424

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | |||||||||| ||||| | | |||||||||| | ||||||| | | |||
Db     139425 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 139484

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db     139485 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 139544

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db     139545 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 139604

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db     139605 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 139664

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||||          | |||     | |   ||    |    |||
Db     139665 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 139724

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
              |   | | | |  | | |    | |||    | |   | ||  |     |      | |
Db     139725 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 139784

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885
              ||        | ||| ||| || | ||   | |  ||| | ||   |  ||  ||  ||
Db     139785 TATATATATATATATATATATATATATATATATATATATATATATAATATATCAATATAT 139844

Qy        886 CACTATATATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAAT 945
              | ||||||||||||||  | |   | ||||  ||||      |   || ||||
Db     139845 AAATATATATATATATATATATATATATAAAAATAAATATATATATATAAATNTAAAA 139904

Qy        946 GTTATAGAAGAAT 958
                || | |  |||
Db     139905 AATAAATATTAAT 139917


RESULT 40
AC141288/c
LOCUS       AC141288               164994 bp    DNA     linear   HTG 11-MAR-2003
DEFINITION  Homo sapiens chromosome 16 clone RP11-357A20, WORKING DRAFT
            SEQUENCE, 2 unordered pieces.
ACCESSION   AC141288
VERSION     AC141288.1  GI:28913068
```

```
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 164994)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 164994)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-MAR-2003) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 543473
            Center clone name: RPCI-11_357A20
            -------
            Summary Statistics
            Consensus quality: 164710 bases at least Q40
            Consensus quality: 164791 bases at least Q30
            Consensus quality: 164821 bases at least Q20
            Estimated insert size: 175000; agarose-fp estimation
            Estimated insert size: 164894; sum-of-contigs estimation
            Quality coverage:  8.51 in Q20 bases; agarose-fp estimation
            Quality coverage:  9.03 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    66545: contig of 66545 bp in length
            *   66546    66645: gap of unknown length
            *   66646   164994: contig of 98349 bp in length.
FEATURES             Location/Qualifiers
     source          1. .164994
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone="RP11-357A20"
                     /clone_lib="RPCI human BAC library 11"
ORIGIN

  Query Match             7.1%;  Score 353.6;  DB 2;  Length 164994;
  Best Local Similarity   72.6%;  Pred. No. 3.7e-61;
  Matches  527;  Conservative    0;  Mismatches  189;  Indels   10;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAA--- 57
              || || |||||| ||||   ||||  |||||||||| ||||||||||||| |||||
Db     149586 AGATCGAGACCATCCTGGCTAACACAGTGAAACCCCGTCTCTACTAAAAATCCAAAAAAA 149527

Qy         58 -TTAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGG 115
               |||| | ||| |||||     | |||||| |||||||||||  |||||| ||||||||
Db     149526 AGTTAGCCGGGCGTGGTGGCGGGCACCTGTAGTCCCAGCTACTCTGGAGGCTGAGGCAGG 149467

Qy        116 AGAATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCAC 174
              |||||  | ||||| ||  || ||  ||| |||| | ||| | ||||  ||||| |||||
Db     149466 AGAATGGCGTGAACTTG--GGAGGCAGAGCTTGCAGTGAGCCGAGATGGTGCCACTGCAC 149409

Qy        175 TCCAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATA 234
              |||||  |  ||| || ||||||| |   ||||||| |||||| | |       | |||| |
Db     149408 TCCAGTCTGGGCGAC-AGAGCGAGACTCCATCTCAAAAAAAATATATATATATATATAGA 149350

Qy        235 TATATATATATATATATACACACACACACATATATATATACACATATATATACACATA 294
              ||||||| ||||||| | | | | | | ||| ||||||| | ||||||| | | |||
Db     149349 TATATATAGATATATAGATATATATAGATATATAGATATATATAGATATATAGATATATA 149290

Qy        295 TATACATATATACACATATATACACATATATATACACATACATATGTACACATATATATA 354
              || | ||||| | | ||||||| | ||||||| | | ||| | || || | |||||||
Db     149289 TAGATATATAGATATATATATATAGATATATAGATATATATAGATATATAGATATATAGA 149230

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              | ||||  ||| | | ||||| | ||||||| | | ||| ||| | |||||||  | | |
Db     149229 TATATATAGATATATAGATATATAGATATATAGATATATAGATATAGATATATAGATATA 149170

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              |||| | | ||| ||| | | ||||| | | ||| | | ||||| | |||||||  |  ||
Db     149169 TATAGATATATAGATATATAGATATATATAGATATATAGATATATAGATATATATAGTTA 149110

Qy        475 TACACATATATATAATATACACACATATATATACACATATATACACACATATATACACAT 534
              || | |||||||  ||||| | | ||||| | ||| ||| ||| | | | ||||| | ||
Db     149109 TATAGATATATA-GATATATATAGATATAGAGACATATAGATATATATAGATATAGAGAT 149051

Qy        535 ATATACACATATATATACACATATATACACATATATACATATATACACATATATATACAT 594
              ||||| | ||||| | | | ||| ||| | | ||||| | | ||| | ||| |||||  |
Db     149050 ATATAGATATATAGAGATATATAGATATATAGATATATAGAGATATATATAGATATATAG 148991

Qy        595 ATATACACATATATACATATATACACATATATACATATATACACACATATATACACATAC 654
              ||||| | ||||||| ||||||| | ||||||| |||||||  | | |||||   | |||
Db     148990 ATATATAGATATATAGATATATATAGATATATAGATATATAGATATATATAGTTATATAG 148931

Qy        655 ATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATAT 714
              ||||| | | | ||| | ||| | | | ||| | |||||||||  |||| | || ||||
Db     148930 ATATATAGATATATATAGATATAGAGACATATAGATATATATAGATATAGAGATATATAG 148871

Qy        715 ATATAT 720
              ||||||
Db     148870 ATATAT 148865


RESULT 41
CR376741/c
LOCUS       CR376741              165979 bp    DNA     linear   HTG 24-SEP-2004
DEFINITION  Danio rerio clone DKEY-114B16, *** SEQUENCING IN PROGRESS ***, 12
            unordered pieces.
ACCESSION   CR376741
VERSION     CR376741.5  GI:52673643
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 165979)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 24, 2004 this sequence version replaced gi:52546082.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK114B16
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 159153 bases at least Q40
            Consensus quality: 160217 bases at least Q30
            Consensus quality: 161345 bases at least Q20
            Insert size: 164879; sum-of-contigs
            Insert size: 169971; 1.8% error; agarose-fp
            Quality coverage: 8.37x in Q20 bases; sum-of-contigs Quality
            coverage: 8.39x in Q20 bases; agarose-fp
```

```
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 12 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     9842: contig of 9842 bp in length
            *    9843     9942: gap of 100 bp
            *    9943    22622: contig of 12680 bp in length
            *   22623    22722: gap of 100 bp
            *   22723    45138: contig of 22416 bp in length
            *   45139    45238: gap of 100 bp
            *   45239    70145: contig of 24907 bp in length
            *   70146    70245: gap of 100 bp
            *   70246    73805: contig of 3560 bp in length
            *   73806    73905: gap of 100 bp
            *   73906    77192: contig of 3287 bp in length
            *   77193    77292: gap of 100 bp
            *   77293   101185: contig of 23893 bp in length
            *  101186   101285: gap of 100 bp
            *  101286   127529: contig of 26244 bp in length
            *  127530   127629: gap of 100 bp
            *  127630   143030: contig of 15401 bp in length
            *  143031   143130: gap of 100 bp
            *  143131   145591: contig of 2461 bp in length
            *  145592   145691: gap of 100 bp
            *  145692   159780: contig of 14089 bp in length
            *  159781   159880: gap of 100 bp
            *  159881   165979: contig of 6099 bp in length.
FEATURES             Location/Qualifiers
     source          1. .165979
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-114B16"
                     /clone_lib="DanioKey"
     misc_feature    1. .9842
                     /note="assembly_fragment:00238
                     fragment_chain:1
                     clone_end:T7
                     vector_side:left"
     misc_feature    9943. .22622
                     /note="assembly_fragment:00377
                     fragment_chain:1"
     misc_feature    22723. .45138
                     /note="assembly_fragment:00739
                     fragment_chain:2"
     misc_feature    45239. .70145
                     /note="assembly_fragment:01657
                     fragment_chain:2"
     misc_feature    70246. .73805
                     /note="assembly_fragment:00125
                     fragment_chain:2"
     misc_feature    73906. .77192
                     /note="assembly_fragment:00093
                     fragment_chain:2"
     misc_feature    77293. .101185
                     /note="assembly_fragment:01273
                     fragment_chain:2"
     misc_feature    101286. .127529
                     /note="assembly_fragment:02046
                     fragment_chain:3"
     misc_feature    127630. .143030
                     /note="assembly_fragment:00558
                     fragment_chain:3"
     misc_feature    143131. .145591
                     /note="assembly_fragment:00062
                     fragment_chain:3"
     misc_feature    145692. .159780
                     /note="assembly_fragment:00999
                     fragment_chain:3"
     misc_feature    159881. .165979
                     /note="assembly_fragment:00165.0"
ORIGIN
```

```
  Query Match             7.1%;  Score 353;  DB 2;  Length 165979;
  Best Local Similarity   68.8%;  Pred. No. 4.9e-61;
  Matches  499;  Conservative    0;  Mismatches  225;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db      74825 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 74766

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db      74765 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 74706

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db      74705 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 74646

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db      74645 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 74586

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db      74585 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 74526

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db      74525 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 74466

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db      74465 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 74406

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db      74405 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 74346

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||          | |||          | |  | |  ||    |  |||
Db      74345 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 74286

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
              |  | | | |  | | |  | |||  | |  | || |  |  | |
Db      74285 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 74226

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885
              ||       | ||| |||||| | ||   |   ||| | ||      ||||  |  |||
Db      74225 TATATATATATATATATAAATATATATATATTTATATATATATATATATTAATTATATAA 74166

Qy        886 CACTATATATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAAT 945
              ||||||||||||||| | | |  | | ||.|  ||||  |      |  |   ||||
Db      74165 ATATATATATATATATAATTATAAAATTATATAAATAAATAAAATTAATAAAAAAAAAAT 74106

Qy        946 GTTAT 950
              | ||
Db      74105 ATAAT 74101
```

```
RESULT 42
CR627491/c
LOCUS       CR627491              248167 bp    DNA     linear   HTG 07-AUG-2004
DEFINITION  Danio rerio clone DKEY-163B6, *** SEQUENCING IN PROGRESS ***, 11
            unordered pieces.
ACCESSION   CR627491
VERSION     CR627491.2  GI:51036473
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
          Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
          Cypriniformes; Cyprinidae; Danio.
REFERENCE 1  (bases 1 to 248167)
  AUTHORS McLay,K.
  TITLE   Direct Submission
  JOURNAL Submitted (06-AUG-2004) Wellcome Trust Sanger Institute, Hinxton,
          Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
          zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT   On Aug 7, 2004 this sequence version replaced gi:50978439.
          -------------- Genome Center
          Center: Wellcome Trust Sanger Institute
          Center code: SC
          Web site: http://www.sanger.ac.uk
          Contact: zfish-help@sanger.ac.uk
          -------------- Project Information
          Center project name: zK163B6
          -------------- Summary Statistics
          Assembly program: XGAP4; version 4.5
          Chemistry: Dye-terminator; 100% of reads
          Consensus quality: 244397 bases at least Q40
          Consensus quality: 245102 bases at least Q30
          Consensus quality: 245733 bases at least Q20
          Insert size: 247167; sum-of-contigs
          Insert size: 234721; 4.3% error; agarose-fp
          Quality coverage: 5.83x in Q20 bases; sum-of-contigs Quality
          coverage: 6.34x in Q20 bases; agarose-fp
          --------------.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 11 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1     5680: contig of 5680 bp in length
          *     5681     5780: gap of 100 bp
          *     5781    34320: contig of 28540 bp in length
          *    34321    34420: gap of 100 bp
          *    34421    48965: contig of 14545 bp in length
          *    48966    49065: gap of 100 bp
          *    49066    71965: contig of 22900 bp in length
          *    71966    72065: gap of 100 bp
          *    72066    76303: contig of 4238 bp in length
          *    76304    76403: gap of 100 bp
          *    76404   114940: contig of 38537 bp in length
          *   114941   115040: gap of 100 bp
          *   115041   163084: contig of 48044 bp in length
          *   163085   163184: gap of 100 bp
          *   163185   176987: contig of 13803 bp in length
          *   176988   177087: gap of 100 bp
          *   177088   189934: contig of 12847 bp in length
          *   189935   190034: gap of 100 bp
          *   190035   220399: contig of 30365 bp in length
          *   220400   220499: gap of 100 bp
          *   220500   248167: contig of 27668 bp in length.
FEATURES             Location/Qualifiers
     source          1. .248167
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-163B6"
                     /clone_lib="DanioKey"
     misc_feature    1. .5680
                     /note="assembly_fragment:00062
                     fragment_chain:1"
     misc_feature    5781. .34320
                     /note="assembly_fragment:01271
                     fragment_chain:1"
     misc_feature    34421. .48965
                     /note="assembly_fragment:00097
                     fragment_chain:1"
     misc_feature    49066. .71965
                     /note="assembly_fragment:00433
                     fragment_chain:1"
     misc_feature    72066. .76303
                     /note="assembly_fragment:00035
                     fragment_chain:2"
     misc_feature    76404. .114940
                     /note="assembly_fragment:01575
                     fragment_chain:2"
     misc_feature    115041. .163084
                     /note="assembly_fragment:01979
                     fragment_chain:2"
     misc_feature    163185. .176987
                     /note="assembly_fragment:00318
                     fragment_chain:3"
     misc_feature    177088. .189934
                     /note="assembly_fragment:00207
                     fragment_chain:3"
     misc_feature    190035. .220399
                     /note="assembly_fragment:00968
                     fragment_chain:3"
     misc_feature    220500. .248167
                     /note="assembly_fragment:00686.0"
ORIGIN

  Query Match             7.1%;  Score 353;  DB 2;  Length 248167;
  Best Local Similarity   65.2%;  Pred. No. 4.6e-61;
  Matches  462;  Conservative    0;  Mismatches  245;  Indels    2;  Gaps    2;

Qy        218 AAAAGGCATTTTATATATATATATATATATATACACACACACACATATATATATAC 277
              ||||  ||   ||||||||||||||||||||||||| | | | | | ||||||||||
Db     115544 AAAAAACAGCATATATATATATATATATATATATATATATATATATATATATATAT 115485

Qy        278 ACATATATATACACATATATACATATATACACATATATACACATATATATACACATACAT 337
              | ||||||||| | ||||||| |||||| | ||||||| | ||||||||| | ||| ||
Db     115484 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 115425

Qy        338 ATGTACACATATATATACACATATGTATACACATATATACACATATATACACACATATAT 397
              || || | ||||||||| | |||| |||| | ||||||| | ||||||| | | ||||||
Db     115424 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 115365

Qy        398 ACACATATATACACACATATATACACATATATACACATATATACACATATACACATATAT 457
              | | ||||||| | | ||||||| | ||||||| | ||||||| | ||||| | ||||||
Db     115364 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 115305

Qy        458 ACACATATATACATATATACACATATATATA-ATATACACACATATATATACACATATAT 516
              | | ||||||| ||||||| | ||||||||| ||||| | | ||||||||| | ||||||
Db     115304 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 115245

Qy        517 ACACACATATATACACATATATACACATATATATACACATATATACACATATATACATAT 576
              | | | ||||||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||
Db     115244 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 115185

Qy        577 ATACACATATATATACATATATACACATATATACATATATACACATATATACATATATAC 636
              ||| | ||||||||| ||||||| | ||||||| ||||||| | ||||||| |||||||
Db     115184 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 115125

Qy        637 ACACATATATACACATACATATACACACACATAGATATACATATATATACACATATATAT 696
              | | ||||||| | ||| ||||| | | | ||| ||||| ||||||||| | ||||||||
Db     115124 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 115065

Qy        697 ACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCA 756
              |  |||||||||  ||||||||||
Db     115064 ATATATATATATATATATATATATNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 115005

Qy        757 CTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTT 816

Db     115004 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 114945

Qy        817 GATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA-AGGTAGTT 875
                  |||||       | ||| |||||| | ||   | |  ||| | || |   ||  |
Db     114944 NNNNAAAAATATATATATATATATATAAATATATATATATATATATATATATATAT 114885

Qy        876 AAGAAGGTAACACTATATATATATATAGTGAAAGCAGTGTAAACCTAAA 924
```

```
Db     114884 ATATATATATATATATATATATATATATATATATATATATATATATATA 114836


RESULT 43
AL731541
LOCUS       AL731541              178254 bp    DNA     linear   PRI 20-JUN-2002
DEFINITION  Human DNA sequence from clone RP11-472K8 on chromosome 10, complete
            sequence.
ACCESSION   AL731541 AC016823
VERSION     AL731541.6  GI:21537521
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 178254)
  AUTHORS   Heath,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAY-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jun 21, 2002 this sequence version replaced gi:21065395.
            Draft Sequence Produced by Genome Therapeutics Corp, 100 Beaver
            Street, Waltham, MA 02453, USA
            http://www.genomecorp.com
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 10, constructed by the Sanger Centre Chromosome 10
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr10
            RP11-472K8 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6.
FEATURES             Location/Qualifiers
     source          1. .178254
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP11-472K8"
                     /clone_lib="RPCI-11.2"
ORIGIN

  Query Match             7.1%;  Score 352.4;  DB 9;  Length 178254;
  Best Local Similarity   70.2%;  Pred. No. 6.4e-61;
  Matches  559;  Conservative    0;  Mismatches  226;  Indels   11;  Gaps    6;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||  ||||||||||  |||||||||||||||||| ||| ||||||||||| ||||||
Db     104371 AGTTTGAGACCAGCCTGGCCAACATGGTGAAACCCCAGCTCCACTAAAAATACAAAAATT 104430

Qy         60 AGACAGGCATGGTGATACACGCCTGTA-----ATCCCAGCTACTTCGGAGGCCGAGGCAG 114
              || ||||||| ||| | ||||||||||     |||| |||||||  |||||| |||||||
Db     104431 AGCCAGGCATTGTGGTGCACGCCTGTAATCTCATCCTAGCTACTCGGGAGGCTGAGGCAG 104490

Qy        115 GAGAATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCA 173
              |||||| ||||||  | | || || ||||||||| | ||| | |||||  |||| ||||
Db     104491 AAGAATCGCTTGAA--TCCGGGAGGCGGAGGTTGCAGTGAGCCCAGATCGCGCCACTGCA 104548

Qy        174 CTCCAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATAT 233
              |||||||  || ||| |||| ||||      | |||||||| | | |          | |
Db     104549 CTCCAGCCTGGGTAAC-AGAGTGAAACCGTGTTTCAGAAAATATACACACACACACACAC 104607

Qy        234 ATATATATATATATATATACACACACACACATATATATATACACATATATATACACAT 293
              | | | | ||||| | |||||||||| | ||||| | | ||||| ||||| |||||
Db     104608 ACACACACACATATACACACACACACATATATATACACACACACACATATACACACAC 104667

Qy        294 ATATACATATATACACATATATACACATATATATACACATACATATGTACACATATATAT 353
              | ||| ||| | ||||| ||||| ||| | | | | | ||||| |  ||| |||||||
Db     104668 ACATATATACACACACACATATATACACACACACATATATACACACACACATATATATAC 104727

Qy        354 ACACATATGTATACACATATATACACATATATACACACATATATACACATATATACACAC 413
              ||||| || |||| ||| | | ||| |||||||||||||| | ||||| ||||||||||||
Db     104728 ACACACATATATATACACACACACATATATATACACACACACATACATATATATACACAC 104787

Qy        414 ATATATACACATATATACACATATATACACATATACACATATATACACATATATACATAT 473
              | | ||| | | | | ||||| ||||| ||| | |||||||||||| ||| | | ||||||
Db     104788 ACACATATATACACACACACACATATATACACACACACATATATATACACACACACATAT 104847

Qy        474 ATACACATATATATAATATACACACATATATATACACATATATACACACATATATACACA 533
              ||| ||| | | || ||| ||||||| ||| ||||| |||| |    | ||||||  | |
Db     104848 ATATACACACACAT-ATACACACACACATACATACATATATGTGTGTATATATATGTATA 104906

Qy        534 TATATACACATATATATACACATATATACACATATATACATATATACACACATATATATACA 593
              || ||| | || | | || | |||| || | | ||||| || | |  | |||||| || |
Db     104907 TACATATATATGTGTGTATATATATGTATATACATATATATGTGTGTATATATATGTATA 104966

Qy        594 TATATACACATATATACATATATACACATATATACATATATACACACATATATACACATA 653
              || ||| | || | |  ||||||    ||||||| |||||||| | |||| ||   |||
Db     104967 TACATATATATGTGTGTATATATGTGTATATATATGTATATACATATATATGTATGTATA 105026

Qy        654 CATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATA 713
              ||||  | | ||||| ||||  || ||||||    |||||||| ||||||| |   | |
Db     105027 TATATGTATATACATATATATGTATGTATATATGTGTATATATATGTATATACACACACA 105086

Qy        714 TATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAA 773
              |||||||          | | | |||          | | |  ||| |   || | |
Db     105087 TATATATATATATATATCCCTGAGATAATAATACTTAAAAAGACTTAGTCCAGTGCCCAG 105146

Qy        774 GAGAGTTTGACAGCAG 789
              | |||  ||   |||
Db     105147 CACAGTGGGAACTCAG 105162


RESULT 44
CR752655
LOCUS       CR752655              204486 bp    DNA     linear   HTG 22-SEP-2004
DEFINITION  Danio rerio clone DKEY-107O6, WORKING DRAFT SEQUENCE, 11 unordered
            pieces.
ACCESSION   CR752655
VERSION     CR752655.5  GI:52546093
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 204486)
  AUTHORS   Burton,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 23, 2004 this sequence version replaced gi:52353233.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
```

```
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK107O6
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 201681 bases at least Q40
            Consensus quality: 202171 bases at least Q30
            Consensus quality: 202651 bases at least Q20
            Insert size: 203486; sum-of-contigs
            Insert size: 204129; 3.5% error; agarose-fp
            Quality coverage: 11.12x in Q20 bases; sum-of-contigs Quality
            coverage: 11.09x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    10267: contig of 10267 bp in length
            *    10268    10367: gap of 100 bp
            *    10368    14914: contig of 4547 bp in length
            *    14915    15014: gap of 100 bp
            *    15015    25928: contig of 10914 bp in length
            *    25929    26028: gap of 100 bp
            *    26029    68196: contig of 42168 bp in length
            *    68197    68296: gap of 100 bp
            *    68297   108045: contig of 39749 bp in length
            *   108046   108145: gap of 100 bp
            *   108146   114266: contig of 6121 bp in length
            *   114267   114366: gap of 100 bp
            *   114367   118079: contig of 3713 bp in length
            *   118080   118179: gap of 100 bp
            *   118180   139935: contig of 21756 bp in length
            *   139936   140035: gap of 100 bp
            *   140036   151139: contig of 11104 bp in length
            *   151140   151239: gap of 100 bp
            *   151240   189955: contig of 38716 bp in length
            *   189956   190055: gap of 100 bp
            *   190056   204486: contig of 14431 bp in length.
FEATURES             Location/Qualifiers
     source          1. .204486
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-107O6"
                     /clone_lib="DanioKey"
     misc_feature    1. .10267
                     /note="assembly_fragment:00170
                     fragment_chain:1"
     misc_feature    10368. .14914
                     /note="assembly_fragment:00029
                     fragment_chain:1"
     misc_feature    15015. .25928
                     /note="assembly_fragment:00569
                     fragment_chain:1"
     misc_feature    26029. .68196
                     /note="assembly_fragment:03030
                     fragment_chain:1"
     misc_feature    68297. .108045
                     /note="assembly_fragment:02194
                     fragment_chain:1"
     misc_feature    108146. .114266
                     /note="assembly_fragment:00286
                     fragment_chain:2"
     misc_feature    114367. .118079
                     /note="assembly_fragment:00088
                     fragment_chain:2"
     misc_feature    118180. .139935
                     /note="assembly_fragment:00956
                     fragment_chain:2"
     misc_feature    140036. .151139
                     /note="assembly_fragment:00410
                     fragment_chain:2"
     misc_feature    151240. .189955
                     /note="assembly_fragment:01419
                     fragment_chain:2"
     misc_feature    190056. .204486
                     /note="assembly_fragment:00735.0"
ORIGIN

  Query Match             7.1%;  Score 352.4;  DB 2;  Length 204486;
  Best Local Similarity   69.8%;  Pred. No. 6.3e-61;
  Matches  504;  Conservative    0;  Mismatches  216;  Indels    2;  Gaps    2;

Qy        209 AGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATATACACACACACACATA 268
              ||||||||||||||     | |||||||||||||||||||||||| | | | | | ||| |
Db     107299 AGAAAAAAAAAAAAAAATATATATATATATATATATATATATATATATATATATATATA 107358

Qy        269 TATATATACACATATATATACACATATATACATATATACACATATATACACATATATATA 328
              |||||||| | ||||||||| | ||||||| ||||||| | ||||||| | |||||||||
Db     107359 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 107418

Qy        329 CACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACA 388
              | ||| |||| || | ||||||||| | |||| |||| | ||||||| | ||||||| |
Db     107419 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 107478

Qy        389 CACATATATACACATATATACACACATATATACACATATATACACATATATACACATATA 448
              | ||||||| | ||||||| | | ||||||| | ||||||| | ||||||| | |||||
Db     107479 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 107538

Qy        449 CACATATATACACATATATACATATATACACATATATATA-ATATACACACATATATATA 507
              | ||||||| | ||||||| ||||||| | ||||||||| ||||| | | |||||||||
Db     107539 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 107598

Qy        508 CACATATATACACACATATATACACATATATACACATATATATACACATATATACACATA 567
              | ||||||| | | ||||||| | ||||||| | ||||||||| | ||||||| | |||
Db     107599 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 107658

Qy        568 TATACATATATACACATATATATACATATATACACATATATACATATATACACATATATA 627
              |||| ||||||| | ||||||||| ||||||| | ||||||| ||||||| | |||||||
Db     107659 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 107718

Qy        628 CATATATACACACATATATACACATACATATACACACACATAGATATACATATATATACA 687
              ||||||| | | ||||||| | ||| ||||| | | | ||| ||||| ||||||||| |
Db     107719 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 107778

Qy        688 CATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGT 747
              |||||||||  |||||||||  ||||||||||          |  |||          |
Db     107779 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 107838

Qy        748 TGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACA 807
              |   ||     |    ||| |   | | | |  | | |    | |||    | |   | |
Db     107839 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 107898

Qy        808 TTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA 867
              | |     |      | |||      | ||| ||| || | ||   | |  ||| | ||
Db     107899 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 107958

Qy        868 -AGGTAGTTAAGAAGGTAACACTATATATATATATAGTGAAAGCAGTGTAAACCTAAAGG 926
               | ||  ||  ||  ||    |||||| |||||||   ||| || | || |  || |
Db     107959 TATATATATATAAATATATATATATATACATATATATATAAATCAATATATATATATATA 108018

Qy        927 AT 928
              ||
Db     108019 AT 108020


RESULT 45
CR392006/c
LOCUS       CR392006              178427 bp    DNA     linear   HTG 23-AUG-2004
DEFINITION  Danio rerio clone DKEYP-57F11, *** SEQUENCING IN PROGRESS ***, 9
```

```
            unordered pieces.
ACCESSION   CR392006
VERSION     CR392006.3  GI:51510714
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 178427)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-AUG-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 23, 2004 this sequence version replaced gi:46879107.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zKp57F11
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 173955 bases at least Q40
            Consensus quality: 174654 bases at least Q30
            Consensus quality: 175292 bases at least Q20
            Insert size: 177627; sum-of-contigs
            Insert size: 177796; 2.7% error; agarose-fp
            Quality coverage: 7.97x in Q20 bases; sum-of-contigs Quality
            coverage: 8.05x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    11505: contig of 11505 bp in length
            *    11506    11605: gap of 100 bp
            *    11606    24931: contig of 13326 bp in length
            *    24932    25031: gap of 100 bp
            *    25032    31819: contig of 6788 bp in length
            *    31820    31919: gap of 100 bp
            *    31920    50297: contig of 18378 bp in length
            *    50298    50397: gap of 100 bp
            *    50398    83841: contig of 33444 bp in length
            *    83842    83941: gap of 100 bp
            *    83942   122559: contig of 38618 bp in length
            *   122560   122659: gap of 100 bp
            *   122660   166383: contig of 43724 bp in length
            *   166384   166483: gap of 100 bp
            *   166484   175663: contig of 9180 bp in length
            *   175664   175763: gap of 100 bp
            *   175764   178427: contig of 2664 bp in length.
FEATURES             Location/Qualifiers
     source          1. .178427
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEYP-57F11"
                     /clone_lib="DanioKeypilot"
     misc_feature    1. .11505
                     /note="assembly_fragment:00415
                     fragment_chain:1"
     misc_feature    11606. .24931
                     /note="assembly_fragment:00266
                     fragment_chain:1"
     misc_feature    25032. .31819
                     /note="assembly_fragment:00041
                     fragment_chain:1"
     misc_feature    31920. .50297
                     /note="assembly_fragment:00579
                     fragment_chain:1"
     misc_feature    50398. .83841
                     /note="assembly_fragment:00848
                     fragment_chain:1"
     misc_feature    83942. .122559
                     /note="assembly_fragment:01340
                     fragment_chain:1"
     misc_feature    122660. .166383
                     /note="assembly_fragment:01877
                     fragment_chain:1"
     misc_feature    166484. .175663
                     /note="assembly_fragment:00124
                     fragment_chain:1"
     misc_feature    175764. .178427
                     /note="assembly_fragment:00006"
ORIGIN

  Query Match             7.1%;  Score 352.2;  DB 2;  Length 178427;
  Best Local Similarity   60.3%;  Pred. No. 7e-61;
  Matches  502;  Conservative    0;  Mismatches  328;  Indels    3;  Gaps    1;

Qy        194 GCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATATAC 253
              || ||||  |   | |||  ||||||||   |  ||||||||||||||||||||||||
Db      12310 GCAAAATAACTGGTAAGATGAAAAAAAAAAAAAAATATATATATATATATATATATATAT 12251

Qy        254 ACACACACACACATATATATATACACATATATATACACATATATACATATATACACATAT 313
              | | | | | | |||||||||| | |||||||| | ||||||| |||||| | ||||
Db      12250 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 12191

Qy        314 ATACACATATATATACACATACATATGTACACATATATATACACATATGTATACACATAT 373
              ||| | ||||||||| | ||| |||| || | |||||||| | |||| |||| | ||||
Db      12190 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 12131

Qy        374 ATACACATATATACACACATATATACACATATATACACACATATATACACATATATACAC 433
              ||| | ||||||| | | ||||||| | ||||||| | | ||||||| | ||||||| |
Db      12130 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 12071

Qy        434 ATATATACACATATACACATATATACACATATATACATATATACACATATATATAATATA 493
              ||||||| | ||||| | ||||||| | ||||||| ||||||| | ||||||   |||||
Db      12070 ATATATATATATATATATATATATATATATATATATATATATATATATATATAT---ATATA 12014

Qy        494 CACACATATATATACACATATATACACACATATATACACATATATACACATATATATACA 553
              | | ||||||||| | ||||||| | | ||||||| | ||||||| | |||||||| |
Db      12013 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 11954

Qy        554 CATATATACACATATATACATATATACACATATATATACATATATACACATATATACATA 613
              ||||||| | ||||||| ||||||| | |||||||| |||||||| | ||||||| |||
Db      11953 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 11894

Qy        614 TATACACATATATACATATATACACACATATATACACATACATATACACACACATAGATA 673
              |||| | ||||||| ||||||| | | ||||||| | ||| ||||| | | | ||| |||
Db      11893 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 11834

Qy        674 TACATATATATACACATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCA 733
              || ||||||||| | |||||||||  |||||||| ||||||||||          | |
Db      11833 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 11774

Qy        734 TAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTA 793
              ||          | |  ||    |  ||| |  | | | |  | | |   | ||
Db      11773 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 11714

Qy        794 TTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAAT 853
              |   | |  | || |      |  |||       | ||| ||| || | ||
Db      11713 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 11654

Qy        854 AAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATATATATATAGTGAAAGCAGT 913
              | |  ||||    |  ||  ||  |  ||   ||||||||||||
Db      11653 TATATATATCATATATATATATATATATATATATATATATATATNNNNNNNNNNNN 11594
```

```
Qy        914 GTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTATAGAAGAATGGCTAAGATGCCAAA 973

Db      11593 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 11534

Qy        974 GCTCAGTGTATGTGGCAGAGGCATGGTGTAGGGTGTGTCCAGGTTCATATATT 1026

Db      11533 NNNNNNNNNNNNNNNNNNNNNNNNNNNNTTTATAATATATAATTATATATATT 11481


RESULT 46
CR548632/c
LOCUS       CR548632              184630 bp    DNA     linear   HTG 08-JUL-2004
DEFINITION  Danio rerio clone DKEY-175H17, WORKING DRAFT SEQUENCE, 8 unordered
            pieces.
ACCESSION   CR548632
VERSION     CR548632.3  GI:50080050
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 184630)
  AUTHORS   Burton,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUL-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 8, 2004 this sequence version replaced gi:50057714.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK175H17
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 181205 bases at least Q40
            Consensus quality: 181924 bases at least Q30
            Consensus quality: 182467 bases at least Q20
            Insert size: 183930; sum-of-contigs
            Insert size: 181082; 1.5% error; agarose-fp
            Quality coverage: 10.04x in Q20 bases; sum-of-contigs Quality
            coverage: 10.36x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    14714: contig of 14714 bp in length
            *    14715    14814: gap of 100 bp
            *    14815    95419: contig of 80605 bp in length
            *    95420    95519: gap of 100 bp
            *    95520   132623: contig of 37104 bp in length
            *   132624   132723: gap of 100 bp
            *   132724   165602: contig of 32879 bp in length
            *   165603   165702: gap of 100 bp
            *   165703   169699: contig of 3997 bp in length
            *   169700   169799: gap of 100 bp
            *   169800   174998: contig of 5199 bp in length
            *   174999   175098: gap of 100 bp
            *   175099   177397: contig of 2299 bp in length
            *   177398   177497: gap of 100 bp
            *   177498   184630: contig of 7133 bp in length.
FEATURES             Location/Qualifiers
     source          1. .184630
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-175H17"
                     /clone_lib="DanioKey"
     misc_feature    1. .14714
                     /note="assembly_fragment:00296
                     fragment_chain:1"
     misc_feature    14815. .95419
                     /note="assembly_fragment:01430
                     fragment_chain:1"
     misc_feature    95520. .132623
                     /note="assembly_fragment:00810
                     fragment_chain:1"
     misc_feature    132724. .165602
                     /note="assembly_fragment:00444
                     fragment_chain:1"
     misc_feature    165703. .169699
                     /note="assembly_fragment:00028
                     fragment_chain:1"
     misc_feature    169800. .174998
                     /note="assembly_fragment:00068
                     fragment_chain:2"
     misc_feature    175099. .177397
                     /note="assembly_fragment:00017
                     fragment_chain:2"
     misc_feature    177498. .184630
                     /note="assembly_fragment:00176.0"
ORIGIN

  Query Match             7.1%;  Score 352.2;  DB 2;  Length 184630;
  Best Local Similarity   69.2%;  Pred. No. 7e-61;
  Matches  491;  Conservative    0;  Mismatches  218;  Indels    1;  Gaps    1;

Qy        216 AAAAAAGGCATTTTATATATATATATATATATATACACACACACACATATATATAT 275
              ||| ||   ||||||||||||||||||||||||||| | | | | | ||||||||||
Db        744 AAATAAAAGATTTTATATATATATATATATATATATATATATATATATATATATAT 685

Qy        276 ACACATATATATACACATATATACATATATACACATATATACACATATATATACACATAC 335
              | | ||||||||| | ||||||| ||||||| | ||||||| | ||||||||| | |||
Db        684 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 625

Qy        336 ATATGTACACATATATATACACATATGTATACACATATATACACATATATACACACATAT 395
              |||| || | ||||||||| | |||| |||| | ||||||| | ||||||| | | ||||
Db        624 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 565

Qy        396 ATACACATATATACACACATATATACACATATATACACATATATACACATATACACATAT 455
              ||| | |||||||| | | ||||||| | ||||||| | ||||||| | ||||| | ||||
Db        564 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 505

Qy        456 ATACACATATATACATATATACACATATATATA-ATATACACACATATATATACACATAT 514
              ||| | ||||||| ||||||| | ||||||||| ||||| | | ||||||||| | ||||
Db        504 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 445

Qy        515 ATACACACATATATACACATATATACACATATATATACACATATATACACATATATACAT 574
              ||| | | ||||||| | ||||||| | ||||||||| | ||||||| | ||||||| ||
Db        444 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 385

Qy        575 ATATACACATATATATACATATATACACATATATACATATATACACATATATACATATAT 634
              ||||| | ||||||||| ||||||| | ||||||| ||||||| | ||||||| ||||||
Db        384 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 325

Qy        635 ACACACATATATACACATACATATACACACACATAGATATACATATATATACACATATAT 694
              | | | |||||||| | ||| ||||| | | | ||| ||||| |||||||||| | ||||||
Db        324 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 265

Qy        695 ATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTAC 754
              |||  |||||||| ||||||||||          | |||          |  |   ||
Db        264 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 205

Qy        755 CACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGG 814
                  |    ||| |   | | | |  | | |    | |||   | |   | || |
Db        204 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 145
```

```
Qy        815 TTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGT 874
              |      | |||       | ||| ||| || | ||   | |  ||| | | |  ||
Db        144 ATATATATATATATATATATATATATATATATATATATATATATATATATNTATTATATATA 85

Qy        875 TAAGAAGGTAACACTATATATATATATAGTGAAAGCAGTGTAAACCTAAA 924
              |||  |  ||    |||| ||||||  |   | |    | || |  || |
Db         84 TAAATATATATATATATANATATATNNATATATATATATATATATNTATA 35


RESULT 47
CR762385
LOCUS       CR762385              195644 bp    DNA     linear   HTG 19-SEP-2004
DEFINITION  Danio rerio clone DKEY-245L10, *** SEQUENCING IN PROGRESS ***, 7
            unordered pieces.
ACCESSION   CR762385
VERSION     CR762385.3  GI:52346294
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 195644)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 19, 2004 this sequence version replaced gi:51988239.
            ---------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            ---------- Project Information
            Center project name: zK245L10
            ---------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 193251 bases at least Q40
            Consensus quality: 193792 bases at least Q30
            Consensus quality: 194212 bases at least Q20
            Insert size: 195044; sum-of-contigs
            Insert size: 196613; 1.8% error; agarose-fp
            Quality coverage: 8.31x in Q20 bases; sum-of-contigs Quality
            coverage: 8.42x in Q20 bases; agarose-fp
            ----------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *      1    38651: contig of 38651 bp in length
            *  38652    38751: gap of 100 bp
            *  38752    58158: contig of 19407 bp in length
            *  58159    58258: gap of 100 bp
            *  58259   117292: contig of 59034 bp in length
            * 117293   117392: gap of 100 bp
            * 117393   123058: contig of 5666 bp in length
            * 123059   123158: gap of 100 bp
            * 123159   150654: contig of 27496 bp in length
            * 150655   150754: gap of 100 bp
            * 150755   160809: contig of 10055 bp in length
            * 160810   160909: gap of 100 bp
            * 160910   195644: contig of 34735 bp in length.
FEATURES             Location/Qualifiers
     source          1. .195644
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-245L10"
                     /clone_lib="DanioKey"
     misc_feature    1. .38651
                     /note="assembly_fragment:01153
                     fragment_chain:1"
     misc_feature    38752. .58158
                     /note="assembly_fragment:00183
                     fragment_chain:1"
     misc_feature    58259. .117292
                     /note="assembly_fragment:01608
                     fragment_chain:1"
     misc_feature    117393. .123058
                     /note="assembly_fragment:00020
                     fragment_chain:2"
     misc_feature    123159. .150654
                     /note="assembly_fragment:00406
                     fragment_chain:2"
     misc_feature    150755. .160809
                     /note="assembly_fragment:00070.0"
     misc_feature    160910. .195644
                     /note="assembly_fragment:00721"
ORIGIN

  Query Match             7.1%;  Score 352.2;  DB 2;  Length 195644;
  Best Local Similarity   59.1%;  Pred. No. 6.9e-61;
  Matches  518;  Conservative    0;  Mismatches  358;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||||| | | | | | |||||||||| | |||||||
Db     122359 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 122418

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db     122419 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 122478

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              ||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db     122479 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 122538

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db     122539 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 122598

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db     122599 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 122658

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db     122659 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 122718

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db     122719 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 122778

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| |||||||||| | |||||||||  ||||||
Db     122779 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 122838

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||          | |||          | |  ||    |   |||
Db     122839 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 122898

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
              |  | | | |  | | |  | |||   | |  | || |  |     |    | |
Db     122899 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 122958

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885
              ||       | ||| ||| || | ||   | |  ||| |   |  ||  ||  |     |
Db     122959 TATATATAAATATAAATATATATATATATATATATATAATATATATATATAATAAATA 123018

Qy        886 CACTATATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAAT 945
              | ||||||||||||| | | |    | || |  ||
```

```
Db     123019 AAATATATATATATATAATTATATATATATATAAATATTTNNNNNNNNNNNNNNNNNNNN 123078

Qy        946 GTTATAGAAGAATGGCTAAGATGCCAAAGCTCAGTGTATGTGGCAGAGGCATGGTGTAGG 1005

Db     123079 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 123138

Qy       1006 GTGTGTCCAGGTTCATATATTGCATTAAGTGTGAGAACACCCTGGAGTATGAACCAAGAA 1065

Db     123139 NNNNNNNNNNNNNNNNNNNCGTTTTATGGATAATATTTTGCGGAACAAGGCCCTTAAAA 123198

Qy       1066 AATGCAAAAGCCAGAAGTGATGGAGGAAATGAGACAC 1102

Db     123199 AGTGGGGAGGGAACACCTTATTGGGTCTAGGTAACCC 123235


RESULT 48
CR749751/c
LOCUS       CR749751              224393 bp    DNA     linear   HTG 27-AUG-2004
DEFINITION  Danio rerio clone DKEY-24H18, WORKING DRAFT SEQUENCE, 10 unordered
            pieces.
ACCESSION   CR749751
VERSION     CR749751.4  GI:51591749
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 224393)
  AUTHORS   Burton,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-AUG-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 27, 2004 this sequence version replaced gi:51510864.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK24H18
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 221304 bases at least Q40
            Consensus quality: 221921 bases at least Q30
            Consensus quality: 222377 bases at least Q20
            Insert size: 223493; sum-of-contigs
            Insert size: 225057; 1.0% error; agarose-fp
            Quality coverage: 10.31x in Q20 bases; sum-of-contigs Quality
            coverage: 10.28x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2873: contig of 2873 bp in length
            *     2874     2973: gap of 100 bp
            *     2974    66389: contig of 63416 bp in length
            *    66390    66489: gap of 100 bp
            *    66490   106977: contig of 40488 bp in length
            *   106978   107077: gap of 100 bp
            *   107078   145207: contig of 38130 bp in length
            *   145208   145307: gap of 100 bp
            *   145308   153395: contig of 8088 bp in length
            *   153396   153495: gap of 100 bp
            *   153496   167269: contig of 13774 bp in length
            *   167270   167369: gap of 100 bp
            *   167370   175339: contig of 7970 bp in length
            *   175340   175439: gap of 100 bp
            *   175440   179204: contig of 3765 bp in length
            *   179205   179304: gap of 100 bp
            *   179305   200083: contig of 20779 bp in length
            *   200084   200183: gap of 100 bp
            *   200184   224393: contig of 24210 bp in length.
FEATURES             Location/Qualifiers
     source          1. .224393
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-24H18"
                     /clone_lib="DanioKey"
     misc_feature    1. .2873
                     /note="assembly_fragment:00089
                     fragment_chain:1"
     misc_feature    2974. .66389
                     /note="assembly_fragment:02891
                     fragment_chain:1"
     misc_feature    66490. .106977
                     /note="assembly_fragment:01374
                     fragment_chain:1"
     misc_feature    107078. .145207
                     /note="assembly_fragment:02083
                     fragment_chain:1"
     misc_feature    145308. .153395
                     /note="assembly_fragment:00265
                     fragment_chain:1"
     misc_feature    153496. .167269
                     /note="assembly_fragment:00404
                     fragment_chain:2"
     misc_feature    167370. .175339
                     /note="assembly_fragment:00155
                     fragment_chain:2"
     misc_feature    175440. .179204
                     /note="assembly_fragment:00108
                     fragment_chain:2"
     misc_feature    179305. .200083
                     /note="assembly_fragment:00671"
     misc_feature    200184. .224393
                     /note="assembly_fragment:00942.0"
ORIGIN

  Query Match             7.1%;  Score 352.2;  DB 2;  Length 224393;
  Best Local Similarity   79.4%;  Pred. No. 6.8e-61;
  Matches  417;  Conservative    0;  Mismatches  108;  Indels    0;  Gaps    0;

Qy        196 GAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACAC 255

Db      19100 GTAATTTTATTTCTTTAGTTAAAAAAAAAAATATATATATATATATATATATATATAT 19041

Qy        256 ACACACACACATATATATATACACATATATATACACATATATACATATATACACATATAT 315

Db      19040 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 18981

Qy        316 ACACATATATATACACATACATATGTACACATATATATACACATATGTATACACATATAT 375

Db      18980 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 18921

Qy        376 ACACATATATACACACATATATACACATATATACACACATATATACACATATATACACAT 435

Db      18920 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 18861

Qy        436 ATATACACATATACACATATATACACATATATACATATATACACATATATATAATATACA 495

Db      18860 ATATATATATATATATATATATATATATATATATATATATATACATATATATCATATATA 18801

Qy        496 CACATATATATACACATATATACACACATATATACACATATATACACATATATATACACA 555

Db      18800 TATATATATATATATATATATATATATACATATATACATATATATATATATATATACATATA 18741

Qy        556 TATATACACATATATACATATATACACATATATATACATATATACACATATATACATATA 615
```

```
Db      18740 TATATATATATATATATATATATATATATATATATATATATATATATGTATATATATATA 18681

Qy        616 TACACATATATACATATATACACACATATATACACATACATATACACACACATAGATATA 675

Db      18680 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 18621

Qy        676 CATATATATACACATATATATACGTATATATATGTATATATATAT 720

Db      18620 TATATATATATATATATATATATATATATATATATATATATATAT 18576


RESULT 49
CR762494
LOCUS       CR762494               238804 bp    DNA     linear   HTG 10-OCT-2004
DEFINITION  Danio rerio clone DKEYP-78B11, WORKING DRAFT SEQUENCE, 7 unordered
            pieces.
ACCESSION   CR762494
VERSION     CR762494.3  GI:52345306
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 238804)
  AUTHORS   McLaren,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 18, 2004 this sequence version replaced gi:52214011.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zKp78B11
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 236234 bases at least Q40
            Consensus quality: 236747 bases at least Q30
            Consensus quality: 237190 bases at least Q20
            Insert size: 238204; sum-of-contigs
            Insert size: 231499; 2.2% error; agarose-fp
            Quality coverage: 11.55x in Q20 bases; sum-of-contigs Quality
            coverage: 11.97x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    21508: contig of 21508 bp in length
            *    21509    21608: gap of 100 bp
            *    21609    47490: contig of 25882 bp in length
            *    47491    47590: gap of 100 bp
            *    47591    54540: contig of 6950 bp in length
            *    54541    54640: gap of 100 bp
            *    54641   140186: contig of 85546 bp in length
            *   140187   140286: gap of 100 bp
            *   140287   163133: contig of 22847 bp in length
            *   163134   163233: gap of 100 bp
            *   163234   166432: contig of 3199 bp in length
            *   166433   166532: gap of 100 bp
            *   166533   238804: contig of 72272 bp in length.
FEATURES             Location/Qualifiers
     source          1. .238804
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEYP-78B11"
                     /clone_lib="DanioKeypilot"
     misc_feature    1. .21508
                     /note="assembly_fragment:00516
                     fragment_chain:1
                     clone_end:SP6
                     vector_side:left"
     misc_feature    21609. .47490
                     /note="assembly_fragment:00888
                     fragment_chain:1"
     misc_feature    47591. .54540
                     /note="assembly_fragment:00086
                     fragment_chain:2"
     misc_feature    54641. .140186
                     /note="assembly_fragment:01354
                     fragment_chain:2"
     misc_feature    140287. .163133
                     /note="assembly_fragment:00178
                     fragment_chain:2"
     misc_feature    163234. .166432
                     /note="assembly_fragment:00035
                     fragment_chain:3"
     misc_feature    166533. .238804
                     /note="assembly_fragment:02624
                     fragment_chain:3"
ORIGIN

  Query Match             7.1%;  Score 352;  DB 2;  Length 238804;
  Best Local Similarity   68.8%;  Pred. No. 7.4e-61;
  Matches  512;  Conservative    0;  Mismatches  230;  Indels    2;  Gaps    2;

Qy        182 CAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATAT 241

Db     165682 CAAAAAAAAAAAAAAAAAAAACATGACATAAAAAAAAAAAAAAAAAAAATATATATATATAT 165741

Qy        242 ATATATATATACACACACACACATATATATATACACATATATATACACATATATACAT 301

Db     165742 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 165801

Qy        302 ATATACACATATATACACATATATATACACATACATATGTACACATATATATACACATAT 361

Db     165802 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 165861

Qy        362 GTATACACATATATACACATATATACACACATATATACACATATATACACACATATATAC 421

Db     165862 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 165921

Qy        422 ACATATATACACATATATACACATATACACATATATACACATATATACATATATACACAT 481

Db     165922 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 165981

Qy        482 ATATATAATATACACACATATATATACACATATATACACACATATATACACATATATACA 541

Db     165982 ATATAT-ATATATATATATATATATATATATATATATATATATATATATATATATATATA 166040

Qy        542 CATATATATACACATATATACACATATATACATATATACACATATATATACATATATACA 601

Db     166041 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 166100

Qy        602 CATATATACATATATACACATATATACATATATACACACATATATACACATACATATACA 661

Db     166101 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 166160

Qy        662 CACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATATG 721

Db     166161 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 166220

Qy        722 CTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTT 781

Db     166221 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 166280
```

```
Qy        782 GACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACA 841
              | |    | |||    | |   | || |     |       | |||       | ||| |
Db     166281 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 166340

Qy        842 TAAATGTGTAATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTAACACTATATATATATA 900
              || || | ||   | |  ||| | || |  ||  ||   |  ||    ||||||||||||
Db     166341 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 166400

Qy        901 TAGTGAAAGCAGTGTAAACCTAAA 924
              ||   |||    | || |  || |
Db     166401 TATATAAATATATATATATATA 166424


RESULT 50
CR790372/c
LOCUS       CR790372              299106 bp    DNA     linear   HTG 05-OCT-2004
DEFINITION  Danio rerio clone DKEY-70K13, *** SEQUENCING IN PROGRESS ***, 25
            unordered pieces.
ACCESSION   CR790372
VERSION     CR790372.4  GI:53828302
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 299106)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 5, 2004 this sequence version replaced gi:53755940.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK70K13
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 289331 bases at least Q40
            Consensus quality: 291856 bases at least Q30
            Consensus quality: 293478 bases at least Q20
            Insert size: 296706; sum-of-contigs
            Insert size: 225769; 2.9% error; agarose-fp
            Quality coverage: 6.22x in Q20 bases; sum-of-contigs Quality
            coverage: 9.16x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 25 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     3928: contig of 3928 bp in length
            *     3929     4028: gap of 100 bp
            *     4029     8417: contig of 4389 bp in length
            *     8418     8517: gap of 100 bp
            *     8518    12283: contig of 3766 bp in length
            *    12284    12383: gap of 100 bp
            *    12384    16946: contig of 4563 bp in length
            *    16947    17046: gap of 100 bp
            *    17047    19834: contig of 2788 bp in length
            *    19835    19934: gap of 100 bp
            *    19935    26863: contig of 6929 bp in length
            *    26864    26963: gap of 100 bp
            *    26964    29135: contig of 2172 bp in length
            *    29136    29235: gap of 100 bp
            *    29236    43722: contig of 14487 bp in length
            *    43723    43822: gap of 100 bp
            *    43823   109813: contig of 65991 bp in length
            *   109814   109913: gap of 100 bp
            *   109914   140749: contig of 30836 bp in length
            *   140750   140849: gap of 100 bp
            *   140850   156362: contig of 15513 bp in length
            *   156363   156462: gap of 100 bp
            *   156463   165292: contig of 8830 bp in length
            *   165293   165392: gap of 100 bp
            *   165393   243285: contig of 77893 bp in length
            *   243286   243385: gap of 100 bp
            *   243386   245562: contig of 2177 bp in length
            *   245563   245662: gap of 100 bp
            *   245663   250094: contig of 4432 bp in length
            *   250095   250194: gap of 100 bp
            *   250195   256405: contig of 6211 bp in length
            *   256406   256505: gap of 100 bp
            *   256506   258788: contig of 2283 bp in length
            *   258789   258888: gap of 100 bp
            *   258889   262780: contig of 3892 bp in length
            *   262781   262880: gap of 100 bp
            *   262881   265140: contig of 2260 bp in length.
            *   265141   265240: gap of 100 bp
            *   265241   269635: contig of 4395 bp in length
            *   269636   269735: gap of 100 bp
            *   269736   272991: contig of 3256 bp in length
            *   272992   273091: gap of 100 bp
            *   273092   291385: contig of 18294 bp in length
            *   291386   291485: gap of 100 bp
            *   291486   293629: contig of 2144 bp in length
            *   293630   293729: gap of 100 bp
            *   293730   296461: contig of 2732 bp in length
            *   296462   296561: gap of 100 bp
            *   296562   299106: contig of 2545 bp in length.
FEATURES             Location/Qualifiers
     source          1. .299106
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-70K13"
                     /clone_lib="DanioKey"
     misc_feature    1. .3928
                     /note="assembly_fragment:00025
                     fragment_chain:1"
     misc_feature    4029. .8417
                     /note="assembly_fragment:00178
                     fragment_chain:1"
     misc_feature    8518. .12283
                     /note="assembly_fragment:00051
                     fragment_chain:1"
     misc_feature    12384. .16946
                     /note="assembly_fragment:00068
                     fragment_chain:1"
     misc_feature    17047. .19834
                     /note="assembly_fragment:03721
                     fragment_chain:1"
     misc_feature    19935. .26863
                     /note="assembly_fragment:00206
                     fragment_chain:1"
     misc_feature    26964. .29135
                     /note="assembly_fragment:03693
                     fragment_chain:1"
     misc_feature    29236. .43722
                     /note="assembly_fragment:00405
                     fragment_chain:2"
     misc_feature    43823. .109813
                     /note="assembly_fragment:01445
                     fragment_chain:2"
     misc_feature    109914. .140749
                     /note="assembly_fragment:01030
                     fragment_chain:2"
```

```
     misc_feature    140850. .156362
                     /note="assembly_fragment:00596
                     fragment_chain:3"
     misc_feature    156463. .165292
                     /note="assembly_fragment:00282
                     fragment_chain:3"
     misc_feature    165393. .243285
                     /note="assembly_fragment:02358
                     fragment_chain:3"
     misc_feature    243386. .245562
                     /note="assembly_fragment:03730
                     fragment_chain:4"
     misc_feature    245663. .250094
                     /note="assembly_fragment:00104
                     fragment_chain:4"
     misc_feature    250195. .256405
                     /note="assembly_fragment:00151
                     fragment_chain:4"
     misc_feature    256506. .258788
                     /note="assembly_fragment:00124
                     fragment_chain:5"
     misc_feature    258889. .262780
                     /note="assembly_fragment:00234
                     fragment_chain:5"
     misc_feature    262881. .265140
                     /note="assembly_fragment:03713
                     fragment_chain:6"
     misc_feature    265241. .269635
                     /note="assembly_fragment:00038
                     fragment_chain:6"
     misc_feature    269736. .272991
                     /note="assembly_fragment:00085"
     misc_feature    273092. .291385
                     /note="assembly_fragment:00801.0"
     misc_feature    291486. .293629
                     /note="assembly_fragment:03699"
     misc_feature    293730. .296461
                     /note="assembly_fragment:03706"
     misc_feature    296562. .299106
                     /note="assembly_fragment:03750"
ORIGIN

  Query Match             7.1%;  Score 351.8;  DB 2;  Length 299106;
  Best Local Similarity   66.8%;  Pred. No. 7.8e-61;
  Matches  531;  Conservative    0;  Mismatches  262;  Indels    2;  Gaps    2;

Qy        171 GCACTCCAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTA 230
              | | ||||    | || |||   | |   ||   |   | ||||||||||||     | ||
Db     166223 GTATTCCAATTTAAGCTACAGATGGGGCCTGCGCTAAAAAAAAAAAAAAAAAATATATATA 166164

Qy        231 TATATATATATATATATATATACACACACACACATATATATATACACATATATATACA 290
              ||||||||||||||||||||| | | | | | |||||||||||| | |||||||||| |
Db     166163 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 166104

Qy        291 CATATATACATATATACACATATATACACATATATATACACATACATATGTACACATATA 350
              ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |||||
Db     166103 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 166044

Qy        351 TATACACATATGTATACACATATATACACATATATACACACATATATACACATATATACA 410
              |||| | |||| |||| | ||||||| | ||||||| | | ||||||| | ||||||| |
Db     166043 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 165984

Qy        411 CACATATATACACATATATACACATATATACACATATACACATATATACACATATATACA 470
              | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | ||||||| |
Db     165983 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 165924

Qy        471 TATATACACATATATATA-ATATACACACATATATATACACATATATACACACATATATA 529
              |||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||||||
Db     165923 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 165864

Qy        530 CACATATATACACATATATATACACATATATACACATATATACATATATACACATATATA 589
              | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||||||
Db     165863 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 165804

Qy        590 TACATATATACACATATATACATATATACACATATATACATATATACACACATATATACA 649
              || ||||||| | ||||||| ||||||| | ||||||| ||||||| | | ||||||| |
Db     165803 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 165744

Qy        650 CATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATATG 709
              ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  |||||||||
Db     165743 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 165684

Qy        710 TATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGG 769
              ||||||||||    |  |||      |  |   ||    |    ||| |
Db     165683 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 165624

Qy        770 AAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGT 829
              | | | |  | | |    | |||    | |   | || |     |       | |||
Db     165623 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 165564

Qy        830 AGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTAACAC 888
              | ||| ||| || | ||   | |  ||| | || |  ||  ||   | ||
Db     165563 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 165504

Qy        889 TATATATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTT 948
              |||||||||||||   | |    | || |  || |       |   || ||||| |
Db     165503 TATATATATATATATATATATATATATATATATATATATATATATATATATAAATATA 165444

Qy        949 ATAGAAGAATGGCTA 963
              || |   ||   ||
Db     165443 TTATATTTATAAATA 165429


RESULT 51
CR762407/c
LOCUS       CR762407              149106 bp    DNA     linear   HTG 13-SEP-2004
DEFINITION  Danio rerio clone DKEY-262N17, *** SEQUENCING IN PROGRESS ***, 5
            unordered pieces.
ACCESSION   CR762407
VERSION     CR762407.2  GI:52000335
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 149106)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 13, 2004 this sequence version replaced gi:51988183.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK262N17
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 147385 bases at least Q40
            Consensus quality: 147795 bases at least Q30
            Consensus quality: 148053 bases at least Q20
            Insert size: 148706; sum-of-contigs
            Insert size: 147741; 3.4% error; agarose-fp
            Quality coverage: 9.30x in Q20 bases; sum-of-contigs Quality
            coverage: 9.36x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
```

```
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *         1    36781: contig of 36781 bp in length
            *     36782    36881: gap of 100 bp
            *     36882    62357: contig of 25476 bp in length
            *     62358    62457: gap of 100 bp
            *     62458    68981: contig of 6524 bp in length
            *     68982    69081: gap of 100 bp
            *     69082    89839: contig of 20758 bp in length
            *     89840    89939: gap of 100 bp
            *     89940   149106: contig of 59167 bp in length.
FEATURES             Location/Qualifiers
     source          1. .149106
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-262N17"
                     /clone_lib="DanioKey"
     misc_feature    1. .36781
                     /note="assembly_fragment:00719
                     fragment_chain:1"
     misc_feature    36882. .62357
                     /note="assembly_fragment:00059
                     fragment_chain:1"
     misc_feature    62458. .68981
                     /note="assembly_fragment:00007"
     misc_feature    69082. .89839
                     /note="assembly_fragment:00379.0"
     misc_feature    89940. .149106
                     /note="assembly_fragment:01240"
ORIGIN

  Query Match             7.1%;  Score 351.6;  DB 2;  Length 149106;
  Best Local Similarity   68.8%;  Pred. No. 9.5e-61;
  Matches  497;  Conservative    0;  Mismatches  224;  Indels    1;  Gaps    1;

Qy        197 AAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATATACACA 256
              || |||  | ||| ||||||||||||   |    | ||||||||||||||||||||| | |
Db      90716 AACTGTATTTTCAAAAAAAAAAAAAAAAAAAAAAAAATATATATATATATATATATATATA 90657

Qy        257 CACACACACATATATATATACACATATATATACACATATATACATATATACACATATATA 316
              | | | | |||||||||||| | ||||||||| | ||||||| ||||||| | |||||||
Db      90656 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 90597

Qy        317 CACATATATATACACATACATATGTACACATATATATACACATATGTATACACATATATA 376
              | ||||||||| | ||| |||| || | ||||||||| | |||| |||| | |||||||
Db      90596 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 90537

Qy        377 CACATATATACACACATATATACACATATATACACACATATATACACATATATACACATA 436
              | ||||||| | | ||||||| | ||||||| | | ||||||| | ||||||| | |||
Db      90536 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 90477

Qy        437 TATACACATATACACATATATACACATATATACATATATACACATATATATAATATACAC 496
              ||||  | |||||  | ||||||| | ||||||| ||||||| | |||||||| ||||| |
Db      90476 TATATATATATATATATATATATATATATATATATATATATATATATATATATAT-ATATATAT 90418

Qy        497 ACATATATATACACATATATACACACATATATACACATATATACACATATATATACACAT 556
              | ||||||||| | ||||||| | | ||||||| | ||||||| | ||||||||| | ||
Db      90417 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 90358

Qy        557 ATATACACATATATACATATATACACATATATATACATATATACACATATATACATATAT 616
              ||||| | ||||||| ||||||| | |||||||| ||||||| | ||||||| ||||||
Db      90357 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 90298

Qy        617 ACACATATATACATATATACACACATATATACACATACATATACACACACATAGATATAC 676
              | | ||||||| ||||||| | | ||||||| | ||| ||||| | | | ||| |||||
Db      90297 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 90238

Qy        677 ATATATATACACATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAA 736
              ||||||||| | |||||||||  |||||||| ||||||||||          |  |||
Db      90237 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 90178

Qy        737 GAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTG 796
              |  |   ||    |   ||| |  | | | |  |  | |    | |||
Db      90177 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 90118

Qy        797 TGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAA 856
              | |   | || |      |      | |||       | ||| ||| || | ||   |
Db      90117 ATATATATATATATATATATATATAAATATATATATATATATATATATATATATATAT 90058

Qy        857 ACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATATATATATAGTGAAAGCAGTGTA 916
              |  ||| |   |  ||  ||   |  ||    |||||| ||| |||   | |  | | ||
Db      90057 ATATATATAATATATATATATATATATACACATATACATAAATACACATAAAATTATA 89998

Qy        917 AA 918
              ||
Db      89997 TA 89996
```

```
RESULT 52
CR762414
LOCUS       CR762414            171302 bp    DNA    linear   HTG 13-SEP-2004
DEFINITION  Danio rerio clone DKEY-197K18, *** SEQUENCING IN PROGRESS ***, 6
            unordered pieces.
ACCESSION   CR762414
VERSION     CR762414.2  GI:52000341
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 171302)
  AUTHORS   McLay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 13, 2004 this sequence version replaced gi:51988190.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK197K18
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 168888 bases at least Q40
            Consensus quality: 169406 bases at least Q30
            Consensus quality: 169716 bases at least Q20
            Insert size: 170802; sum-of-contigs
            Insert size: 170040; 4.9% error; agarose-fp
            Quality coverage: 7.32x in Q20 bases; sum-of-contigs Quality
            coverage: 7.59x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *         1    15887: contig of 15887 bp in length
            *     15888    15987: gap of 100 bp
            *     15988    42962: contig of 26975 bp in length
            *     42963    43062: gap of 100 bp
            *     43063    58072: contig of 15010 bp in length
            *     58073    58172: gap of 100 bp
            *     58173    60197: contig of 2025 bp in length
            *     60198    60297: gap of 100 bp
            *     60298   112076: contig of 51779 bp in length
            *    112077   112176: gap of 100 bp
```

```
          *   112177   171302: contig of 59126 bp in length.
FEATURES             Location/Qualifiers
     source          1. .171302
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-197K18"
                     /clone_lib="DanioKey"
     misc_feature    1. .15887
                     /note="assembly_fragment:00045
                     fragment_chain:1"
     misc_feature    15988. .42962
                     /note="assembly_fragment:00332
                     fragment_chain:1"
     misc_feature    43063. .58072
                     /note="assembly_fragment:00177
                     fragment_chain:2"
     misc_feature    58173. .60197
                     /note="assembly_fragment:00025
                     fragment_chain:2"
     misc_feature    60298. .112076
                     /note="assembly_fragment:00601.0"
     misc_feature    112177. .171302
                     /note="assembly_fragment:01241"
ORIGIN

  Query Match             7.1%;  Score 351.4;  DB 2;  Length 171302;
  Best Local Similarity   69.0%;  Pred. No. 1e-60;
  Matches  486;  Conservative    0;  Mismatches  217;  Indels    1;  Gaps    1;

Qy        226 TTTTATATATATATATATATATATATACACACACACACACATATATATATACACATATAT 285
              || ||||||||||||||||||||||| | | | | | | |||||||||||| | ||||||
Db      42255 TTATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 42314

Qy        286 ATACACATATATACATATATACACATATATACACATATATATACACATACATATGTACAC 345
              ||| | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || |
Db      42315 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 42374

Qy        346 ATATATATACACATATGTATACACATATATACACATATATACACACATATATACACATAT 405
              ||||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | ||||
Db      42375 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 42434

Qy        406 ATACACACATATATACACATATATACACATATATACACATATACACATATATACACATAT 465
              ||| | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | ||||
Db      42435 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 42494

Qy        466 ATACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACAT 524
              ||| ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | ||
Db      42495 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 42554

Qy        525 ATATACACATATATACACATATATATACACATATATACACATATATACATATATACACAT 584
              ||||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | ||
Db      42555 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 42614

Qy        585 ATATATACATATATACACATATATACATATATACACATATATACATATATACACACATAT 644
              ||||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | ||||
Db      42615 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 42674

Qy        645 ATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATAT 704
              ||| | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  |||||
Db      42675 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 42734

Qy        705 ATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGAT 764
              |||| ||||||||||          | |||          | |   ||    |    ||
Db      42735 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 42794

Qy        765 AGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAA 824
              | |  | | | |  | |   | |||   | |  | || |    |     |      |
Db      42795 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 42854

Qy        825 ATAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTA 884
              |||  |    | ||| |||||| | ||  || |  ||| | ||    |  ||  ||   |
Db      42855 ATATAAAAAAATATATATAAATATATATAAATATATATATNTATATAAAATANAAANNAA 42914

Qy        885 ACACTATATATATATATAGTGAAAGCAGTGTAAACCTAAAGGAT 928
              | |||| ||  | ||| | ||   | |       || |  ||
Db      42915 TAAATATAAATTAANATAATAAANNAAATAATNNATTATAATAT 42958


RESULT 53
CR749761
LOCUS       CR749761              216259 bp    DNA     linear   HTG 10-OCT-2004
DEFINITION  Danio rerio clone CH211-215F19, WORKING DRAFT SEQUENCE, 10
            unordered pieces.
ACCESSION   CR749761
VERSION     CR749761.4  GI:54021792
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 216259)
  AUTHORS   McLaren,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 10, 2004 this sequence version replaced gi:51988273.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zC215F19
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 212820 bases at least Q40
            Consensus quality: 213560 bases at least Q30
            Consensus quality: 214140 bases at least Q20
            Insert size: 215359; sum-of-contigs
            Insert size: 205140; 3.4% error; agarose-fp
            Quality coverage: 7.94x in Q20 bases; sum-of-contigs Quality
            coverage: 8.53x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    48219: contig of 48219 bp in length
            *    48220    48319: gap of 100 bp
            *    48320    65856: contig of 17537 bp in length
            *    65857    65956: gap of 100 bp
            *    65957    86815: contig of 20859 bp in length
            *    86816    86915: gap of 100 bp
            *    86916   108880: contig of 21965 bp in length
            *   108881   108980: gap of 100 bp
            *   108981   118104: contig of 9124 bp in length
            *   118105   118204: gap of 100 bp
            *   118205   185027: contig of 66823 bp in length
            *   185028   185127: gap of 100 bp
            *   185128   193579: contig of 8452 bp in length
            *   193580   193679: gap of 100 bp
            *   193680   207482: contig of 13803 bp in length
            *   207483   207582: gap of 100 bp
            *   207583   211897: contig of 4315 bp in length
            *   211898   211997: gap of 100 bp
            *   211998   216259: contig of 4262 bp in length.
FEATURES             Location/Qualifiers
     source          1. .216259
```

```
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-215F19"
                     /clone_lib="CHORI-211"
     misc_feature    1. .48219
                     /note="assembly_fragment:01191
                     fragment_chain:1
                     clone_end:T7
                     vector_side:left"
     misc_feature    48320. .65856
                     /note="assembly_fragment:00391
                     fragment_chain:1
                     clone_end:SP6
                     vector_side:right"
     misc_feature    65957. .86815
                     /note="assembly_fragment:00649
                     fragment_chain:2"
     misc_feature    86916. .108880
                     /note="assembly_fragment:00908
                     fragment_chain:2"
     misc_feature    108981. .118104
                     /note="assembly_fragment:00069
                     fragment_chain:2"
     misc_feature    118205. .185027
                     /note="assembly_fragment:01823
                     fragment_chain:2"
     misc_feature    185128. .193579
                     /note="assembly_fragment:00271
                     fragment_chain:2"
     misc_feature    193680. .207482
                     /note="assembly_fragment:00167
                     fragment_chain:3"
     misc_feature    207583. .211897
                     /note="assembly_fragment:00018
                     fragment_chain:3"
     misc_feature    211998. .216259
                     /note="assembly_fragment:00040
                     fragment_chain:3"
ORIGIN

  Query Match             7.1%;  Score 351.4;  DB 2;  Length 216259;
  Best Local Similarity   69.0%;  Pred. No. 9.8e-61;
  Matches  492;  Conservative    0;  Mismatches  220;  Indels    1;  Gaps    1;

Qy        191 AGAGCGAAATGTCATCTCAGAAAAAAAAAAAGGCATTTTATATATATATATATATATATA 250
              ||||   |||         | ||||||||||      | ||||||||||||||||||||
Db     215544 AGAGATTAATAAGTCTGTAAAAAAAAAAAAAAATATATATATATATATATATATATATA 215603

Qy        251 TACACACACACACATATATATATACACATATATATACACATATATACATATATACACA 310
              || | | | | | | ||||||||||| | ||||||||| | ||||||| ||||||| | |
Db     215604 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 215663

Qy        311 TATATACACATATATATACACATACATATGTACACATATATATACACATATGTATACACA 370
              |||||| | ||||||||| | ||| |||| || | ||||||||| | |||| |||| | |
Db     215664 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 215723

Qy        371 TATATACACATATATACACACATATATACACATATATACACACATATATACACATATATA 430
              |||||| | ||||||| | | ||||||| | ||||||| | | ||||||| | |||||||
Db     215724 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 215783

Qy        431 CACATATATACACATATACACATATATACACATATATACATATATACACATATATATA-A 489
               | ||||||| | ||||| | ||||||| | ||||||| ||||||| | ||||||||| |
Db     215784 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 215843

Qy        490 TATACACACATATATATACACATATATACACACATATATACACATATATACACATATATA 549
              |||| | | ||||||||| | ||||||| | | ||||||| | ||||||| | |||||||
Db     215844 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 215903

Qy        550 TACACATATATACACATATATACATATATACACATATATATACATATATACACATATATA 609
              || | ||||||| | ||||||| ||||||| | ||||||||| ||||||| | |||||||
Db     215904 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 215963

Qy        610 CATATATACACATATATACATATATACACACATATATACACATACATATACACACACATA 669
              ||||||| | ||||||| ||||||| | | ||||||| | ||| ||||| | | | |||
Db     215964 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 216023

Qy        670 GATATACATATATATACACATATATATACGTATATATATGTATATATATATGCTCCAGAG 729
              ||||| ||||||||| | ||||||||| |||||||||| ||||||||||
Db     216024 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 216083

Qy        730 TTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAG 789
              |  |||          |  |   ||    |   ||| |   | | | |  |  | |
Db     216084 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 216143

Qy        790 TGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTG 849
              | |||    | |   | || |      |       | |||       | ||| ||| | |
Db     216144 TATATATATATATATATATATATATATATATATATATATATATATATATATATAAATA 216203

Qy        850 TAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATATATATATA 902
              || | | |  ||| | ||    | |  |     ||  | ||||| || |||||
Db     216204 TATTTATATATATTTATATANAAATATAANNANTATAAATATATTTAAATATA 216256


RESULT 54
CR548622
LOCUS       CR548622              200752 bp    DNA     linear   HTG 06-JUL-2004
DEFINITION  Danio rerio clone DKEY-189A9, WORKING DRAFT SEQUENCE, 8 unordered
            pieces.
ACCESSION   CR548622
VERSION     CR548622.2  GI:50057705
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 200752)
  AUTHORS   Burton,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUL-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 8, 2004 this sequence version replaced gi:49658985.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK189A9
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 198295 bases at least Q40
            Consensus quality: 198730 bases at least Q30
            Consensus quality: 199141 bases at least Q20
            Insert size: 200052; sum-of-contigs
            Insert size: 193380; 3.4% error; agarose-fp
            Quality coverage: 9.52x in Q20 bases; sum-of-contigs Quality
            coverage: 9.92x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     5799: contig of 5799 bp in length
            *     5800     5899: gap of 100 bp
            *     5900    61323: contig of 55424 bp in length
            *    61324    61423: gap of 100 bp
            *    61424    94914: contig of 33491 bp in length
            *    94915    95014: gap of 100 bp
```

```
            *    95015   169245: contig of 74231 bp in length
            *   169246   169345: gap of 100 bp
            *   169346   171607: contig of 2262 bp in length
            *   171608   171707: gap of 100 bp
            *   171708   174537: contig of 2830 bp in length
            *   174538   174637: gap of 100 bp
            *   174638   183583: contig of 8946 bp in length
            *   183584   183683: gap of 100 bp
            *   183684   200752: contig of 17069 bp in length.
FEATURES             Location/Qualifiers
     source          1. .200752
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-189A9"
                     /clone_lib="DanioKey"
     misc_feature    1. .5799
                     /note="assembly_fragment:00060
                     fragment_chain:1"
     misc_feature    5900. .61323
                     /note="assembly_fragment:00964
                     fragment_chain:1"
     misc_feature    61424. .94914
                     /note="assembly_fragment:00439
                     fragment_chain:1"
     misc_feature    95015. .169245
                     /note="assembly_fragment:01748
                     fragment_chain:2"
     misc_feature    169346. .171607
                     /note="assembly_fragment:00009
                     fragment_chain:2"
     misc_feature    171708. .174537
                     /note="assembly_fragment:00025"
     misc_feature    174638. .183583
                     /note="assembly_fragment:00114"
     misc_feature    183684. .200752
                     /note="assembly_fragment:00228.0"
ORIGIN

  Query Match             7.1%;  Score 351.2;  DB 2;  Length 200752;
  Best Local Similarity   63.0%;  Pred. No. 1.1e-60;
  Matches  466;  Conservative    0;  Mismatches  273;  Indels    1;  Gaps    1;

Qy        186 CAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAGGCATTTTATATATATATATATAT 245
              ||| || |   | |  | |    | ||||||||||||   | | ||||||||||||||||||
Db      60737 CAAAAATAATAATAAATAAATAAAAAAAAAAAAAAAAAAAAAAATATATATATATATATAT 60796

Qy        246 ATATATACACACACACACATATATATATACACATATATATACACATATATACATATAT 305
              ||||||| | | | | | | |||||||||||| | |||||||| | ||||||| ||||||
Db      60797 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 60856

Qy        306 ACACATATATACACATATATATACACATACATATGTACACATATATATACACATATGTAT 365
              | | ||||||| | |||||||| | ||| |||| || | |||||||| | |||| |||
Db      60857 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 60916

Qy        366 ACACATATATACACATATATACACACATATATACACATATATACACACATATATACACAT 425
              | | ||||||| | ||||||| | | ||||||| | ||||||| | | ||||||| | ||
Db      60917 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 60976

Qy        426 ATATACACATATATACACATATACACATATATACACATATATACATATATACACATATAT 485
              ||||| | ||||||| | ||||| | ||||||| | ||||||| ||||||| | ||||||
Db      60977 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 61036

Qy        486 ATA-ATATACACACATATATATACACATATATACACACATATATACACATATATACACAT 544
              ||| ||||| | | ||||||||| | ||||||| | | ||||||| | ||||||| | ||
Db      61037 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 61096

Qy        545 ATATATACACATATATACACATATATACATATATACACATATATATACATATATACACAT 604
              ||||||| | ||||||| | ||||||| ||||||| | ||||||||| ||||||| | ||
Db      61097 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 61156

Qy        605 ATATACATATATACACATATATACATATATACACACATATATACACATACATATACACAC 664
              ||||| ||||||| | ||||||| ||||||| | | ||||||| | ||| ||||| | |
```

```
Db      61157 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 61216

Qy        665 ACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATATGCTC 724
              | ||| ||||| |||||||||| | |||||||||  |||||||||  ||||||||||
Db      61217 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 61276

Qy        725 CAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGAC 784
                |  |||          |  |   ||     |     ||| |   |
Db      61277 ATATATATATATATATATATATATATATATATATATATATATATATANNNNNNNNNNNNN 61336

Qy        785 AGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAA 844

Db      61337 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 61396

Qy        845 ATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATATATATATAGT 904
                                        |  ||   |  ||    ||||||||||||||
Db      61397 NNNNNNNNNNNNNNNNNNNNNNNNNNNATATATATATATATATATATATATATATA 61456

Qy        905 GAAAGCAGTGTAAACCTAAA 924
              | |    | || |  || |
Db      61457 TATATATATATATATATATA 61476


RESULT 55
CR786577/c
LOCUS       CR786577              156578 bp    DNA     linear   HTG 17-SEP-2004
DEFINITION  Danio rerio clone CH211-285C10, *** SEQUENCING IN PROGRESS ***, 11
            unordered pieces.
ACCESSION   CR786577
VERSION     CR786577.1  GI:52313179
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 156578)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zC285C10
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 152392 bases at least Q40
            Consensus quality: 153620 bases at least Q30
            Consensus quality: 154402 bases at least Q20
            Insert size: 155578; sum-of-contigs
            Insert size: 156602; 1.5% error; agarose-fp
            Quality coverage: 6.30x in Q20 bases; sum-of-contigs Quality
            coverage: 6.26x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    28552: contig of 28552 bp in length
            *    28553    28652: gap of 100 bp
            *    28653    32448: contig of 3796 bp in length
            *    32449    32548: gap of 100 bp
            *    32549    46601: contig of 14053 bp in length
```

```
            *    46602    46701: gap of 100 bp
            *    46702    54418: contig of 7717 bp in length
            *    54419    54518: gap of 100 bp
            *    54519    79156: contig of 24638 bp in length
            *    79157    79256: gap of 100 bp
            *    79257    90677: contig of 11421 bp in length
            *    90678    90777: gap of 100 bp
            *    90778    95922: contig of 5145 bp in length
            *    95923    96022: gap of 100 bp
            *    96023   100602: contig of 4580 bp in length
            *   100603   100702: gap of 100 bp
            *   100703   125015: contig of 24313 bp in length
            *   125016   125115: gap of 100 bp
            *   125116   150124: contig of 25009 bp in length
            *   150125   150224: gap of 100 bp
            *   150225   156578: contig of 6354 bp in length.
FEATURES             Location/Qualifiers
     source          1. .156578
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-285C10"
                     /clone_lib="CHORI-211"
     misc_feature    1. .28552
                     /note="assembly_fragment:01088
                     fragment_chain:1
                     clone_end:SP6
                     vector_side:left"
     misc_feature    28653. .32448
                     /note="assembly_fragment:00002
                     fragment_chain:1"
     misc_feature    32549. .46601
                     /note="assembly_fragment:00430
                     fragment_chain:1"
     misc_feature    46702. .54418
                     /note="assembly_fragment:00210
                     fragment_chain:1"
     misc_feature    54519. .79156
                     /note="assembly_fragment:00576
                     fragment_chain:1"
     misc_feature    79257. .90677
                     /note="assembly_fragment:00303
                     fragment_chain:1"
     misc_feature    90778. .95922
                     /note="assembly_fragment:00082
                     fragment_chain:1"
     misc_feature    96023. .100602
                     /note="assembly_fragment:00039
                     fragment_chain:1"
     misc_feature    100703. .125015
                     /note="assembly_fragment:00813
                     fragment_chain:1"
     misc_feature    125116. .150124
                     /note="assembly_fragment:01376"
     misc_feature    150225. .156578
                     /note="assembly_fragment:00130
                     clone_end:T7
                     vector_side:right"
ORIGIN

  Query Match             7.1%;  Score 351;  DB 2;  Length 156578;
  Best Local Similarity   69.8%;  Pred. No. 1.2e-60;
  Matches  488;  Conservative    0;  Mismatches  210;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | |||||||||||||||||||||||| | | | | | | |||||||||||| | |||||||
Db     100582 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 100523

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db     100522 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 100463

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
```

```
              |||||||| | |||| |||| | ||||||| | ||||||| | | |||||||| | |||||
Db     100462 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 100403

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db     100402 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 100343

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db     100342 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 100283

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db     100282 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 100223

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db     100222 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 100163

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db     100162 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 100103

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||||         | |||          |  |   ||    |    |||
Db     100102 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 100043

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
              |   | | | |   | | |    | |||     | |    | || |     |       | |
Db     100042 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 99983

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885
              ||       | ||| ||| || | ||   | |  ||| | ||      |||| ||||||
Db      99982 TATATATATATATATATATATATATATATATATATATATATATATATATTAATAAGGTAT 99923

Qy        886 CACTATATATATATATAGTGAAAGCAGTGTAAACCTAAA 924
              |||||||||||||||   | |    | || |  || |
Db      99922 ATATATATATATATATATATATATATATATATATATA 99884


RESULT 56
CR762422/c
LOCUS       CR762422              159603 bp    DNA     linear   HTG 13-SEP-2004
DEFINITION  Danio rerio clone CH211-184I23, *** SEQUENCING IN PROGRESS ***, 13
            unordered pieces.
ACCESSION   CR762422
VERSION     CR762422.2  GI:52000347
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 159603)
  AUTHORS   McLay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 13, 2004 this sequence version replaced gi:51988198.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zC184I23
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 153793 bases at least Q40
            Consensus quality: 154824 bases at least Q30
            Consensus quality: 155824 bases at least Q20
```

```
            Insert size: 158403; sum-of-contigs
            Insert size: 158676; 6.9% error; agarose-fp
            Quality coverage: 8.08x in Q20 bases; sum-of-contigs Quality
            coverage: 8.13x in Q20 bases; agarose-fp
            ----------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 13 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    33117: contig of 33117 bp in length
            *    33118    33217: gap of 100 bp
            *    33218    38465: contig of 5248 bp in length
            *    38466    38565: gap of 100 bp
            *    38566    47730: contig of 9165 bp in length
            *    47731    47830: gap of 100 bp
            *    47831    83298: contig of 35468 bp in length
            *    83299    83398: gap of 100 bp
            *    83399    86467: contig of 3069 bp in length
            *    86468    86567: gap of 100 bp
            *    86568    88634: contig of 2067 bp in length
            *    88635    88734: gap of 100 bp
            *    88735   103656: contig of 14922 bp in length
            *   103657   103756: gap of 100 bp
            *   103757   120853: contig of 17097 bp in length
            *   120854   120953: gap of 100 bp
            *   120954   132618: contig of 11665 bp in length
            *   132619   132718: gap of 100 bp
            *   132719   146433: contig of 13715 bp in length
            *   146434   146533: gap of 100 bp
            *   146534   152937: contig of 6404 bp in length
            *   152938   153037: gap of 100 bp
            *   153038   156530: contig of 3493 bp in length
            *   156531   156630: gap of 100 bp
            *   156631   159603: contig of 2973 bp in length.
FEATURES             Location/Qualifiers
     source          1. .159603
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-184I23"
                     /clone_lib="CHORI-211"
     misc_feature    1. .33117
                     /note="assembly_fragment:01129
                     fragment_chain:1
                     clone_end:SP6
                     vector_side:left"
     misc_feature    33218. .38465
                     /note="assembly_fragment:00128
                     fragment_chain:1"
     misc_feature    38566. .47730
                     /note="assembly_fragment:00252
                     fragment_chain:1"
     misc_feature    47831. .83298
                     /note="assembly_fragment:01535
                     fragment_chain:1"
     misc_feature    83399. .86467
                     /note="assembly_fragment:00044
                     fragment_chain:1"
     misc_feature    86568. .88634
                     /note="assembly_fragment:00029
                     fragment_chain:2"
     misc_feature    88735. .103656
                     /note="assembly_fragment:00682
                     fragment_chain:2"
     misc_feature    103757. .120853
                     /note="assembly_fragment:00894"
     misc_feature    120954. .132618
                     /note="assembly_fragment:00349
                     fragment_chain:3"
     misc_feature    132719. .146433
                     /note="assembly_fragment:00481
                     fragment_chain:3"
     misc_feature    146534. .152937
                     /note="assembly_fragment:00172
                     fragment_chain:3"
     misc_feature    153038. .156530
                     /note="assembly_fragment:00068
                     fragment_chain:3"
     misc_feature    156631. .159603
                     /note="assembly_fragment:00094
                     fragment_chain:3
                     clone_end:T7
                     vector_side:right"
ORIGIN

  Query Match             7.1%;  Score 350.8;  DB 2;  Length 159603;
  Best Local Similarity   70.0%;  Pred. No. 1.4e-60;
  Matches  486;  Conservative    0;  Mismatches  207;  Indels    1;  Gaps    1;

Qy        209 AGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACACATA 268
              |||||||||||    | | ||||||||||||||||||||||||| | | | | | | |||
Db     121963 AAAAAAAAAAAAAAAAAATATATATATATATATATATATATATATATATATATATATA 121904

Qy        269 TATATATACACATATATATACACATATATACATATATACACATATATACACATATATATA 328
              |||||||| | ||||||||| | ||||||| ||||||| | ||||||| | |||||||||
Db     121903 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 121844

Qy        329 CACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACA 388
              | ||| |||| || | ||||||||| | |||| |||| | ||||||| | ||||||| |
Db     121843 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 121784

Qy        389 CACATATATACACATATATACACACATATATACACATATATACACATATATACACATATA 448
              | ||||||| | ||||||| | | ||||||| | ||||||| | ||||||| | |||||
Db     121783 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 121724

Qy        449 CACATATATACACATATATACATATATACACATATATATA-ATATACACACATATATATA 507
              | ||||||| | ||||||| ||||||| | ||||||||| ||||| | | |||||||||
Db     121723 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 121664

Qy        508 CACATATATACACACATATATACACATATATACACATATATATACACATATATACACATA 567
              | ||||||| | | ||||||| | ||||||| | ||||||||| | ||||||| | |||
Db     121663 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 121604

Qy        568 TATACATATATACACATATATATACATATATACACATATATACATATATACACATATATA 627
              |||| ||||||| | ||||||||| ||||||| | ||||||| ||||||| | |||||||
Db     121603 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 121544

Qy        628 CATATATACACACATATATACACATACATATACACACACATAGATATACATATATATACA 687
              ||||||| | | ||||||| | ||| ||||| | | | ||| ||||| |||||||||
Db     121543 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 121484

Qy        688 CATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGT 747
              |||||||||  |||||||||| ||||||||||                ||| |
Db     121483 TATATATATATATATATATATATATATATATATATATATATAAATATATATATATATA 121424

Qy        748 TGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACA 807
              |   ||     |    ||| |   | | | |  | | |    | |||    | |  | |
Db     121423 TATATATATATATATATATATATACATATATATATATAAATATATATATATATATAAA 121364

Qy        808 TTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA 867
              | |     |   | |||       | ||| |||||| | |||||| |  ||| | ||
Db     121363 TATATATATATATATATATATATATATATATATATAAATATATAATAATAAATATTTATA 121304

Qy        868 AGGTAGTTAAGAAGGTAACACTATATATATATAT 901
              |  |  | | ||  ||  |    |||   | ||
Db     121303 ATATTATATATAATTTATAATATAATAATAAAAT 121270


RESULT 57
CR762498/c
LOCUS       CR762498              200607 bp    DNA     linear   HTG 10-OCT-2004
DEFINITION  Danio rerio clone DKEY-4J1, WORKING DRAFT SEQUENCE, 9 unordered
```

```
            pieces.
ACCESSION   CR762498
VERSION     CR762498.4  GI:52346306
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 200607)
  AUTHORS   McLaren,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 19, 2004 this sequence version replaced gi:52345308.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK4J1
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 196926 bases at least Q40
            Consensus quality: 197654 bases at least Q30
            Consensus quality: 198332 bases at least Q20
            Insert size: 199807; sum-of-contigs
            Insert size: 210583; 10.0% error; agarose-fp
            Quality coverage: 10.81x in Q20 bases; sum-of-contigs Quality
            coverage: 10.33x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    18795: contig of 18795 bp in length
            *   18796    18895: gap of 100 bp
            *   18896    73470: contig of 54575 bp in length
            *   73471    73570: gap of 100 bp
            *   73571    76405: contig of 2835 bp in length
            *   76406    76505: gap of 100 bp
            *   76506    89061: contig of 12556 bp in length
            *   89062    89161: gap of 100 bp
            *   89162   100153: contig of 10992 bp in length
            *  100154   100253: gap of 100 bp
            *  100254   157131: contig of 56878 bp in length
            *  157132   157231: gap of 100 bp
            *  157232   161046: contig of 3815 bp in length
            *  161047   161146: gap of 100 bp
            *  161147   167068: contig of 5922 bp in length
            *  167069   167168: gap of 100 bp
            *  167169   200607: contig of 33439 bp in length.
FEATURES             Location/Qualifiers
     source          1. .200607
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-4J1"
                     /clone_lib="DanioKey"
     misc_feature    1. .18795
                     /note="assembly_fragment:00545
                     fragment_chain:1"
     misc_feature    18896. .73470
                     /note="assembly_fragment:02336
                     fragment_chain:1"
     misc_feature    73571. .76405
                     /note="assembly_fragment:00024
                     fragment_chain:1"
     misc_feature    76506. .89061
                     /note="assembly_fragment:00349
                     fragment_chain:1"
     misc_feature    89162. .100153
                     /note="assembly_fragment:00174
                     fragment_chain:2"
     misc_feature    100254. .157131
                     /note="assembly_fragment:01357
                     fragment_chain:2"
     misc_feature    157232. .161046
                     /note="assembly_fragment:00051.0"
     misc_feature    161147. .167068
                     /note="assembly_fragment:00099"
     misc_feature    167169. .200607
                     /note="assembly_fragment:00747"
ORIGIN

  Query Match             7.1%;  Score 350.6;  DB 2;  Length 200607;
  Best Local Similarity   62.3%;  Pred. No. 1.4e-60;
  Matches  472;  Conservative    0;  Mismatches  285;  Indels    1;  Gaps    1;

Qy        197 AAATGTCATCTCAGAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACA 256
              |||||| |    | |||||||| |      | ||||||||||||||||||||||| | |
Db      77043 AAATGTAAAAAAAAAAAAAAAAATATATATATATATATATATATATATATATATATA 76984

Qy        257 CACACACACATATATATATACACATATATATACACATATATACATATATACACATATATA 316
              | | | | |||||||||||| | ||||||||| | ||||||| ||||||| | |||||||
Db      76983 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 76924

Qy        317 CACATATATATACACATACATATGTACACATATATATACACATATGTATACACATATATA 376
              | ||||||||| | ||| |||| || | ||||||||| | |||| |||| | |||||||
Db      76923 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 76864

Qy        377 CACATATATACACACATATATACACATATATACACACATATATACACATATATACACATA 436
              | ||||||| | | ||||||| | ||||||| | | ||||||| | ||||||| | |||
Db      76863 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 76804

Qy        437 TATACACATATACACATATATACACATATATACATATATACACATATATATAATATACAC 496
              |||| | ||||| | ||||||| | ||||||| ||||||| | ||||||||| ||||| |
Db      76803 TATATATATATATATATATATATATATATATATATATATATATATATATATATAT-ATATATAT 76745

Qy        497 ACATATATATACACATATATACACACATATATACACATATATACACATATATATACACAT 556
              | ||||||||| | ||||||| | | ||||||| | ||||||| | ||||||||| | ||
Db      76744 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 76685

Qy        557 ATATACACATATATACATATATACACATATATATACATATATACACATATATACATATAT 616
              ||||| | ||||||| ||||||| | ||||||||| ||||||| | ||||||| ||||||
Db      76684 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 76625

Qy        617 ACACATATATACATATATACACACATATATACACATACATATACACACACATAGATATAC 676
              | | ||||||| ||||||| | | |||||||| | ||| ||||| | | | ||| |||||
Db      76624 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 76565

Qy        677 ATATATATACACATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAA 736
              |||||||| | |||||||||  |||||||||||||||||          |  |||
Db      76564 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAN 76505

Qy        737 GAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTG 796

Db      76504 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 76445

Qy        797 TGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAA 856
                                                  ||| ||| || | ||   |
Db      76444 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTATATATATATATATATATAT 76385

Qy        857 ACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATATATATATAGTGAAAGCAGTGTA 916
              |  ||| | ||      | | |   |     | | | | | || | |  |  ||   |
Db      76384 ATATATATATATATATATATATTTTTTTTTTTTTTNTCTTTACAATATATACAAATTT 76325

Qy        917 AACCTAAAGGATGGGCCAAGGATTTAAATGTTATAGAA 954
              |   | |         |  || |||  | |||| ||
```

```
Db      76324 ACAAAATATATATATATATATATATAATGGATATATAA 76287


RESULT 58
CR589875/c
LOCUS       CR589875              227441 bp    DNA     linear   HTG 19-JUL-2004
DEFINITION  Danio rerio clone DKEY-15K16, WORKING DRAFT SEQUENCE, 17 unordered
            pieces.
ACCESSION   CR589875
VERSION     CR589875.2  GI:50399773
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 227441)
  AUTHORS   Burton,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JUL-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK15K16
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 221759 bases at least Q40
            Consensus quality: 222952 bases at least Q30
            Consensus quality: 223650 bases at least Q20
            Insert size: 225841; sum-of-contigs
            Insert size: 225442; 2.6% error; agarose-fp
            Quality coverage: 7.98x in Q20 bases; sum-of-contigs Quality
            coverage: 8.44x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 17 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     7750: contig of 7750 bp in length
            *    7751     7850: gap of 100 bp
            *    7851    17808: contig of 9958 bp in length
            *   17809    17908: gap of 100 bp
            *   17909    26300: contig of 8392 bp in length
            *   26301    26400: gap of 100 bp
            *   26401    37193: contig of 10793 bp in length
            *   37194    37293: gap of 100 bp
            *   37294    42790: contig of 5497 bp in length
            *   42791    42890: gap of 100 bp
            *   42891    75037: contig of 32147 bp in length
            *   75038    75137: gap of 100 bp
            *   75138    77164: contig of 2027 bp in length
            *   77165    77264: gap of 100 bp
            *   77265    93266: contig of 16002 bp in length
            *   93267    93366: gap of 100 bp
            *   93367   109748: contig of 16382 bp in length
            *  109749   109848: gap of 100 bp
            *  109849   156408: contig of 46560 bp in length
            *  156409   156508: gap of 100 bp
            *  156509   168441: contig of 11933 bp in length
            *  168442   168541: gap of 100 bp
            *  168542   193830: contig of 25289 bp in length
            *  193831   193930: gap of 100 bp
            *  193931   212755: contig of 18825 bp in length
            *  212756   212855: gap of 100 bp
            *  212856   215964: contig of 3109 bp in length
            *  215965   216064: gap of 100 bp
            *  216065   218399: contig of 2335 bp in length
            *  218400   218499: gap of 100 bp
            *  218500   221549: contig of 3050 bp in length
            *  221550   221649: gap of 100 bp
            *  221650   227441: contig of 5792 bp in length.
FEATURES             Location/Qualifiers
     source          1. .227441
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-15K16"
                     /clone_lib="DanioKey"
     misc_feature    1. .7750
                     /note="assembly_fragment:00294
                     fragment_chain:1"
     misc_feature    7851. .17808
                     /note="assembly_fragment:00581
                     fragment_chain:1"
     misc_feature    17909. .26300
                     /note="assembly_fragment:00382
                     fragment_chain:1"
     misc_feature    26401. .37193
                     /note="assembly_fragment:00471
                     fragment_chain:1"
     misc_feature    37294. .42790
                     /note="assembly_fragment:00222
                     fragment_chain:1"
     misc_feature    42891. .75037
                     /note="assembly_fragment:02032
                     fragment_chain:1"
     misc_feature    75138. .77164
                     /note="assembly_fragment:00092
                     fragment_chain:1"
     misc_feature    77265. .93266
                     /note="assembly_fragment:01346
                     fragment_chain:1"
     misc_feature    93367. .109748
                     /note="assembly_fragment:00863
                     fragment_chain:1"
     misc_feature    109849. .156408
                     /note="assembly_fragment:02587
                     fragment_chain:1"
     misc_feature    156509. .168441
                     /note="assembly_fragment:00719
                     fragment_chain:1"
     misc_feature    168542. .193830
                     /note="assembly_fragment:01629
                     fragment_chain:1"
     misc_feature    193931. .212755
                     /note="assembly_fragment:01104
                     fragment_chain:2"
     misc_feature    212856. .215964
                     /note="assembly_fragment:00130
                     fragment_chain:2"
     misc_feature    216065. .218399
                     /note="assembly_fragment:00077"
     misc_feature    218500. .221549
                     /note="assembly_fragment:00107.0"
     misc_feature    221650. .227441
                     /note="assembly_fragment:00155.0"
ORIGIN

  Query Match             7.1%;  Score 350.4;  DB 2;  Length 227441;
  Best Local Similarity   64.4%;  Pred. No. 1.5e-60;
  Matches  466;  Conservative    0;  Mismatches  256;  Indels    2;  Gaps    2;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db      18405 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 18346
```

```
Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db      18345 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 18286

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db      18285 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 18226

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db      18225 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 18166

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db      18165 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 18106

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db      18105 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 18046

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db      18045 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 17986

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db      17985 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 17926

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||||
Db      17925 TATATATATATATATAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 17866

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
                                                                      |||
Db      17865 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNAAA 17806

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTA 884
              ||     | ||| ||| || | ||   | |  ||| | || |  ||  ||   |  ||
Db      17805 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 17746

Qy        885 ACACTATATATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAA 944
                  ||||| |||||||| |  | |  || | |  |        || |||   | ||   |
Db      17745 TATATATATGTATATATATTCTATGATGTTTTAGTCATTTTTGTGTGCCTCTGTTTCTCA 17686

Qy        945 TGTT 948
              | ||
Db      17685 TATT 17682


RESULT 59
CR788315
LOCUS       CR788315             153947 bp    DNA     linear   HTG 24-SEP-2004
DEFINITION  Danio rerio clone DKEY-64E1, *** SEQUENCING IN PROGRESS ***, 14
            unordered pieces.
ACCESSION   CR788315
VERSION     CR788315.3  GI:52673623
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 153947)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 24, 2004 this sequence version replaced gi:52626672.
            ------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            ------------- Project Information
            Center project name: zK64E1
            ------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 150089 bases at least Q40
            Consensus quality: 150989 bases at least Q30
            Consensus quality: 151632 bases at least Q20
            Insert size: 152647; sum-of-contigs
            Insert size: 169881; 7.1% error; agarose-fp
            Quality coverage: 10.12x in Q20 bases; sum-of-contigs Quality
            coverage: 9.68x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 14 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     9129: contig of 9129 bp in length
            *     9130     9229: gap of 100 bp
            *     9230    14190: contig of 4961 bp in length
            *    14191    14290: gap of 100 bp
            *    14291    46311: contig of 32021 bp in length
            *    46312    46411: gap of 100 bp
            *    46412    54860: contig of 8449 bp in length
            *    54861    54960: gap of 100 bp
            *    54961    62361: contig of 7401 bp in length
            *    62362    62461: gap of 100 bp
            *    62462    70465: contig of 8004 bp in length
            *    70466    70565: gap of 100 bp
            *    70566    85672: contig of 15107 bp in length
            *    85673    85772: gap of 100 bp
            *    85773    90034: contig of 4262 bp in length
            *    90035    90134: gap of 100 bp
            *    90135   104444: contig of 14310 bp in length
            *   104445   104544: gap of 100 bp
            *   104545   111290: contig of 6746 bp in length
            *   111291   111390: gap of 100 bp
            *   111391   123791: contig of 12401 bp in length
            *   123792   123891: gap of 100 bp
            *   123892   125954: contig of 2063 bp in length
            *   125955   126054: gap of 100 bp
            *   126055   129586: contig of 3532 bp in length
            *   129587   129686: gap of 100 bp
            *   129687   153947: contig of 24261 bp in length.
FEATURES             Location/Qualifiers
     source          1. .153947
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-64E1"
                     /clone_lib="DanioKey"
     misc_feature    1. .9129
                     /note="assembly_fragment:00638
                     fragment_chain:1"
     misc_feature    9230. .14190
                     /note="assembly_fragment:00181
                     fragment_chain:1"
     misc_feature    14291. .46311
                     /note="assembly_fragment:01886
                     fragment_chain:1"
     misc_feature    46412. .54860
                     /note="assembly_fragment:00323
                     fragment_chain:1"
     misc_feature    54961. .62361
                     /note="assembly_fragment:00252
                     fragment_chain:1"
     misc_feature    62462. .70465
                     /note="assembly_fragment:00518
```

```
                     fragment_chain:1"
     misc_feature    70566. .85672
                     /note="assembly_fragment:00985
                     fragment_chain:1"
     misc_feature    85773. .90034
                     /note="assembly_fragment:00086
                     fragment_chain:1"
     misc_feature    90135. .104444
                     /note="assembly_fragment:00797
                     fragment_chain:1"
     misc_feature    104545. .111290
                     /note="assembly_fragment:00417
                     fragment_chain:2"
     misc_feature    111391. .123791
                     /note="assembly_fragment:01185
                     fragment_chain:2"
     misc_feature    123892. .125954
                     /note="assembly_fragment:00065"
     misc_feature    126055. .129586
                     /note="assembly_fragment:00133.0"
     misc_feature    129687. .153947
                     /note="assembly_fragment:01427"
ORIGIN

  Query Match             7.0%;  Score 350.2;  DB 2;  Length 153947;
  Best Local Similarity   64.4%;  Pred. No. 1.8e-60;
  Matches  450;  Conservative    0;  Mismatches  248;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | |||||||||||||||||||||||| | | | | | | |||||||||||| | |||||||
Db     123283 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 123342

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| ||| |||| || | |
Db     123343 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 123402

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db     123403 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 123462

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db     123463 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 123522

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db     123523 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 123582

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db     123583 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 123642

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db     123643 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 123702

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db     123703 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 123762

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              |||  ||||||||||         |  ||
Db     123763 TATAAATATATATATATAAATATATATATNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 123822

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825

Db     123823 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 123882

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885
                      |   || || | || |||||  ||| | ||   || ||    | |||
Db     123883 NNNNNNNNNTTTTTTTTATATATATATTAAAAAATATATATATATTATTTTTATATATAA 123942

Qy        886 CACTATATATATATATAGTGAAAGCAGTGTAAACCTAAA 924
```

```
                  | ||||||||||||   | |    | || |  || |
Db     123943 ATAAAAATATATATATATATATATATATATATATATA 123981


RESULT 60
CR788302/c
LOCUS       CR788302              175303 bp    DNA     linear   HTG 24-SEP-2004
DEFINITION  Danio rerio clone DKEYP-38A10, *** SEQUENCING IN PROGRESS ***, 9
            unordered pieces.
ACCESSION   CR788302
VERSION     CR788302.3  GI:52673620
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 175303)
  AUTHORS   McLay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 24, 2004 this sequence version replaced gi:52420996.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zKp38A10
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 173044 bases at least Q40
            Consensus quality: 173386 bases at least Q30
            Consensus quality: 173663 bases at least Q20
            Insert size: 174503; sum-of-contigs
            Insert size: 174235; 1.3% error; agarose-fp
            Quality coverage: 8.55x in Q20 bases; sum-of-contigs Quality
            coverage: 8.62x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     6846: contig of 6846 bp in length
            *     6847     6946: gap of 100 bp
            *     6947    39633: contig of 32687 bp in length
            *    39634    39733: gap of 100 bp
            *    39734    44270: contig of 4537 bp in length
            *    44271    44370: gap of 100 bp
            *    44371    81545: contig of 37175 bp in length
            *    81546    81645: gap of 100 bp
            *    81646    84574: contig of 2929 bp in length
            *    84575    84674: gap of 100 bp
            *    84675   103084: contig of 18410 bp in length
            *   103085   103184: gap of 100 bp
            *   103185   122912: contig of 19728 bp in length
            *   122913   123012: gap of 100 bp
            *   123013   130331: contig of 7319 bp in length
            *   130332   130431: gap of 100 bp
            *   130432   175303: contig of 44872 bp in length.
FEATURES             Location/Qualifiers
     source          1. .175303
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEYP-38A10"
```

```
                     /clone_lib="DanioKeypilot"
     misc_feature    1. .6846
                     /note="assembly_fragment:00101
                     fragment_chain:1"
     misc_feature    6947. .39633
                     /note="assembly_fragment:01279
                     fragment_chain:1"
     misc_feature    39734. .44270
                     /note="assembly_fragment:00052
                     fragment_chain:1"
     misc_feature    44371. .81545
                     /note="assembly_fragment:00801
                     fragment_chain:1"
     misc_feature    81646. .84574
                     /note="assembly_fragment:00024
                     fragment_chain:1"
     misc_feature    84675. .103084
                     /note="assembly_fragment:00248
                     fragment_chain:2"
     misc_feature    103185. .122912
                     /note="assembly_fragment:00522
                     fragment_chain:2"
     misc_feature    123013. .130331
                     /note="assembly_fragment:00171.0"
     misc_feature    130432. .175303
                     /note="assembly_fragment:01761"
ORIGIN

  Query Match             7.0%;  Score 350.2;  DB 2;  Length 175303;
  Best Local Similarity   63.5%;  Pred. No. 1.8e-60;
  Matches  459;  Conservative    0;  Mismatches  263;  Indels    1;  Gaps    1;

Qy        202 TCATCTCAGAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACAC 261
              | | ||   ||||||||||||     | ||||||||||||||||||||||||| | | | |
Db     123573 TTACCTTTAAAAAAAAAAAAATATATATATATATATATATATATATATATATATATAT 123514

Qy        262 ACACATATATATATACACATATATATACACATATATACATATATACACATATATACACAT 321
              | | ||||||||||| | ||||||||| | ||||||| ||||||| | ||||||| | ||
Db     123513 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 123454

Qy        322 ATATATACACATACATATGTACACATATATATACACATATGTATACACATATATACACAT 381
              ||||||| | ||| |||| || | ||||||||| | |||| |||| | ||||||| | ||
Db     123453 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 123394

Qy        382 ATATACACACATATATACACATATATACACACATATATACACATATATACACATATATAC 441
              ||||| | | ||||||| | ||||||| | | ||||||| | ||||||| | |||||||
Db     123393 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 123334

Qy        442 ACATATACACATATATACACATATATACATATATACACATATATATAATATACACACATA 501
              | ||||| | ||||||| | ||||||| ||||||| | |||||||| ||||| | | |||
Db     123333 ATATATATATATATATATATATATATATATATATATATATATATAT-ATATATATATATA 123275

Qy        502 TATATACACATATATACACACATATATACACATATATACACATATATATACACATATATA 561
              |||||| | ||||||| | | ||||||| | ||||||| | |||||||| | |||||||
Db     123274 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 123215

Qy        562 CACATATATACATATATACACATATATATACATATATACACATATATACATATATACACA 621
               | ||||||| ||||||| | |||||||| ||||||| | ||||||| ||||||| | |
Db     123214 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 123155

Qy        622 TATATACATATATACACACATATATACACATACATATACACACACATAGATATACATATA 681
              |||||| ||||||| | | ||||||| | ||| ||||| | | | ||| ||||| |||||
Db     123154 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 123095

Qy        682 TATACACATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGT 741
              |||| | |||||||||  |||||||| |||||||||||          |  |||  |
Db     123094 TATATATATATATATATATATATATATATATATATATATATATATATATATATAAATA 123035

Qy        742 AGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGA 801
                |  |  ||   |  |
Db     123034 TATATATATATATATATAAAAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 122975

Qy        802 AGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTA 861

Db     122974 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 122915

Qy        862 TCTGTAAGGTAGTTAAGAAGGTAACACTATATATATATAGTGAAAGCAGTGTAAACCT 921
              | |||  |  ||    |  ||   ||||||||||||||  | |   | || | |
Db     122914 NNTTTAATATTATTTTTTATATATATATATATATATATATATATATATATATATAT 122855

Qy        922 AAA 924
              |||
Db     122854 AAA 122852


RESULT 61
CR753838
LOCUS       CR753838              204662 bp    DNA     linear   HTG 10-OCT-2004
DEFINITION  Danio rerio clone DKEY-189H5, WORKING DRAFT SEQUENCE, 10 unordered
            pieces.
ACCESSION   CR753838
VERSION     CR753838.3  GI:54021757
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 204662)
  AUTHORS   McLaren,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 10, 2004 this sequence version replaced gi:51592199.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK189H5
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 200213 bases at least Q40
            Consensus quality: 201277 bases at least Q30
            Consensus quality: 202286 bases at least Q20
            Insert size: 203762; sum-of-contigs
            Insert size: 192359; 2.6% error; agarose-fp
            Quality coverage: 8.93x in Q20 bases; sum-of-contigs Quality
            coverage: 10.24x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    54213: contig of 54213 bp in length
            *    54214    54313: gap of 100 bp
            *    54314    67044: contig of 12731 bp in length
            *    67045    67144: gap of 100 bp
            *    67145   118416: contig of 51272 bp in length
            *   118417   118516: gap of 100 bp
            *   118517   132097: contig of 13581 bp in length
            *   132098   132197: gap of 100 bp
            *   132198   135006: contig of 2809 bp in length
            *   135007   135106: gap of 100 bp
            *   135107   137155: contig of 2049 bp in length
            *   137156   137255: gap of 100 bp
            *   137256   139284: contig of 2029 bp in length
            *   139285   139384: gap of 100 bp
            *   139385   141755: contig of 2371 bp in length
```

```
            *   141756   141855: gap of 100 bp
            *   141856   146655: contig of 4800 bp in length
            *   146656   146755: gap of 100 bp
            *   146756   204662: contig of 57907 bp in length.
FEATURES             Location/Qualifiers
     source          1. .204662
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-189H5"
                     /clone_lib="DanioKey"
     misc_feature    1. .54213
                     /note="assembly_fragment:01901
                     fragment_chain:1"
     misc_feature    54314. .67044
                     /note="assembly_fragment:00280
                     fragment_chain:1"
     misc_feature    67145. .118416
                     /note="assembly_fragment:00441
                     fragment_chain:1"
     misc_feature    118517. .132097
                     /note="assembly_fragment:00148
                     fragment_chain:1"
     misc_feature    132198. .135006
                     /note="assembly_fragment:00046
                     fragment_chain:2"
     misc_feature    135107. .137155
                     /note="assembly_fragment:00042
                     fragment_chain:2"
     misc_feature    137256. .139284
                     /note="assembly_fragment:00024"
     misc_feature    139385. .141755
                     /note="assembly_fragment:00059"
     misc_feature    141856. .146655
                     /note="assembly_fragment:00081.0"
     misc_feature    146756. .204662
                     /note="assembly_fragment:01145"
ORIGIN

  Query Match             7.0%;  Score 350.2;  DB 2;  Length 204662;
  Best Local Similarity   68.7%;  Pred. No. 1.7e-60;
  Matches  497;  Conservative    0;  Mismatches  223;  Indels    3;  Gaps    1;

Qy        183 AGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATA 242
              | | |||||  |   |||    ||   | || ||||  |      | ||||||||||||||
Db     203771 ATGAAACAATAAAAAAAAACAATAAAAAAATAAAATTATATATATATATATATATATATA 203830

Qy        243 TATATATATACACACACACACATATATATATACACATATATATACACATATATACATA 302
              |||||||||| | | | | | ||||||||||| | |||||||| | ||||||| |||
Db     203831 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 203890

Qy        303 TATACACATATATACACATATATATACACATACATATGTACACATATATATACACATATG 362
              |||| | ||||||| | |||||||| | ||| |||| || | |||||||| | ||||
Db     203891 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 203950

Qy        363 TATACACATATATACACATATATACACACATATATACACATATATACACACATATATACA 422
              |||| | ||||||| | ||||||| | | ||||||| | ||||||| | | ||||||| |
Db     203951 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 204010

Qy        423 CATATATACACATATATACACATATACACATATATACACATATATACATATATACACATA 482
              ||||||| | ||||||| | ||||| | ||||||| | ||||||| ||||||| | |||
Db     204011 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 204070

Qy        483 TATATA---ATATACACACATATATATACACATATATACACACATATATACACATATATA 539
              ||||||   ||||| | | ||||||||| | ||||||| | | ||||||| | |||||||
Db     204071 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 204130

Qy        540 CACATATATATACACATATATACACATATATACATATATACACATATATATACATATATA 599
              | ||||||||| | ||||||| | ||||||| ||||||| | ||||||||| |||||||
Db     204131 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 204190

Qy        600 CACATATATACATATATACACATATATACATATATACACACATATATACACATACATATA 659
               | ||||||| ||||||| | ||||||| ||||||| | | ||||||| | ||| |||||

Db     204191 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 204250

Qy        660 CACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATA 719
              | | | ||| ||||| |||||||||| | |||||||||  |||||||| ||||||||||
Db     204251 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 204310

Qy        720 TGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGT 779
              |      | |||          | |   ||    |    ||| |   | | | |
Db     204311 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 204370

Qy        780 TTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATA 839
              | | |    | |||    | |   | || |     |      | |||      | |||
Db     204371 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 204430

Qy        840 CATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATATATAT 899
              ||| || | ||  |||| ||  | |||  ||  ||      ||    |||| ||||||
Db     204431 TATATATATATATAAAAATATAAATATAATATATATATTTTAATATATATATAAATATAT 204490

Qy        900 ATA 902
              |||
Db     204491 ATA 204493


RESULT 62
CR626907
LOCUS       CR626907              218701 bp    DNA     linear   HTG 30-JUL-2004
DEFINITION  Danio rerio clone DKEY-52D15, WORKING DRAFT SEQUENCE, 7 unordered
            pieces.
ACCESSION   CR626907
VERSION     CR626907.2  GI:50871854
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 218701)
  AUTHORS   Burton,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUL-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 30, 2004 this sequence version replaced gi:50787983.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK52D15
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 216083 bases at least Q40
            Consensus quality: 216675 bases at least Q30
            Consensus quality: 216971 bases at least Q20
            Insert size: 218101; sum-of-contigs
            Insert size: 195570; 5.6% error; agarose-fp
            Quality coverage: 8.49x in Q20 bases; sum-of-contigs Quality
            coverage: 9.59x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     6174: contig of 6174 bp in length
            *     6175     6274: gap of 100 bp
            *     6275    10497: contig of 4223 bp in length
            *    10498    10597: gap of 100 bp
```

```
            *    10598    49931: contig of 39334 bp in length
            *    49932    50031: gap of 100 bp
            *    50032   118150: contig of 68119 bp in length
            *   118151   118250: gap of 100 bp
            *   118251   146821: contig of 28571 bp in length
            *   146822   146921: gap of 100 bp
            *   146922   186335: contig of 39414 bp in length
            *   186336   186435: gap of 100 bp
            *   186436   218701: contig of 32266 bp in length.
FEATURES             Location/Qualifiers
     source          1. .218701
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-52D15"
                     /clone_lib="DanioKey"
     misc_feature    1. .6174
                     /note="assembly_fragment:00016
                     fragment_chain:1"
     misc_feature    6275. .10497
                     /note="assembly_fragment:00055
                     fragment_chain:1"
     misc_feature    10598. .49931
                     /note="assembly_fragment:00497
                     fragment_chain:1"
     misc_feature    50032. .118150
                     /note="assembly_fragment:01902
                     fragment_chain:1"
     misc_feature    118251. .146821
                     /note="assembly_fragment:00934"
     misc_feature    146922. .186335
                     /note="assembly_fragment:01398"
     misc_feature    186436. .218701
                     /note="assembly_fragment:00096
                     clone_end:SP6
                     vector_side:right"
ORIGIN

  Query Match             7.0%;  Score 350.2;  DB 2;  Length 218701;
  Best Local Similarity   44.5%;  Pred. No. 1.7e-60;
  Matches  554;  Conservative    0;  Mismatches  689;  Indels    2;  Gaps    2;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              |||||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db       5167 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 5226

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db       5227 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 5286

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db       5287 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 5346

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db       5347 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 5406

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db       5407 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 5466

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db       5467 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 5526

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db       5527 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 5586

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db       5587 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 5646
```

```
Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| |||||||||||          | |||          |  |   ||    |   |||
Db       5647 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 5706

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
              |  | | | |  | | |    | |||   | |   | || |     |      | |
Db       5707 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 5766

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTA 884
              ||       | ||| ||| || | ||   | |  ||| |  | |  ||  |||  |   |
Db       5767 TATATATATATATATATATATATATATATATANATATATAAATATATATATAAANANANA 5826

Qy        885 ACACTATATATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAA 944
              |||||||||  ||      ||  | |    |   |          |          ||
Db       5827 TATATATATATATNTAANNATNAAAAAATNNNTATNNNANAANNNNNNNANNNNNNNAA 5886

Qy        945 TGTTATAGAAGAATGGCTAAGATGCCAAAGCTCAGTGTATGTGGCAGAGGCATGGTGTAG 1004
                | |          |   | |              |     |   |
Db       5887 ANNNNTTTANNNNNNANNNNTNTNTTNNANNTTNNNNNNNNNNNTAANNNNNTNNNTTTN 5946

Qy       1005 GGTGTGTCCAGGTTCATATATTGCATTAAGTGTGAGAACACCCTGGAGTATGAACCAAGA 1064
                           | ||     ||  |   |         | | |     |
Db       5947 TNNANNNAATAAAAANNNTTTTANNNNAATNNTNNTTATNNANNNNAATNTTNTNTANNN 6006

Qy       1065 AAATGCAAAAGCCAGAAGTGATGGAGGAAATGAGACACAATAATGAAGATATTGAGAGGA 1124
                | |                  |  |   |              | |
Db       6007 TNAAAAANNNNNNNNNNNANNTNNNNNNNTNNNANNNNANNNNNNNNNNNNNNANAA 6066

Qy       1125 GGGTGTGGGCCTAGAGTGAAGCTTTTCGTGCCAGTACTTCTTTTGAAGGCCCAGTTCTCT 1184
                |              ||||      | || || |          |
Db       6067 AATATTNNTTTNTTNNNNNNNTTTTTTNNNTAATTATTTTTNNNTTNNNNTAATNNNNNN 6126

Qy       1185 TCTCTCTCGGGGGCTCCTTCATCTCTCATAGAGTCCACAGCTTTTAAGGGCCAACACTTG 1244
                |                                        || ||
Db       6127 NATNANNNTNNTAANAAANAAANNNNNNNNNNNNNNNNNTNTTNAAANNNNNNNNNNNN 6186

Qy       1245 AGGTCAGCCTGGCTCTCTCATTTGAGCTGGATAGAACATTTTAGAGCACCATCTATTCTT 1304

Db       6187 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 6246

Qy       1305 CAAGAGGAAGTTTAAAAATAAAAGAACCTTGAAGAGGAAAAAATGTAGACATTCAATCTA 1364
                                          ||          |   | | |     || |
Db       6247 NNNNNNNNNNNNNNNNNNNNNNNNNNNTTANNNNNNNTATTNTANATANNNNTNATTTT 6306

Qy       1365 ACCTTTTCATTTTACTAGCCAAAGCTAAATAGAATGCAGGTTACCTGTTTTTCAGCCAGG 1424
                |    ||          |
Db       6307 TNNTNNNNATNNNTTNNNNANANNNNNNNNNNNNNNNNNANNNNNNNNNNNNNNNNNN 6366

Qy       1425 CACCATCATTTCCTAATTGTTATAAAATTTATTATTATTGTTGTT 1469
                       |   | || |    ||| ||  |||     ||
Db       6367 NNNNNNNNNNNNNTTNNTTTTTTTTTTTTTTTTNNTATNNNNNTT 6411
```

```
RESULT 63
CR769779
LOCUS       CR769779              219074 bp    DNA     linear   HTG 18-SEP-2004
DEFINITION  Danio rerio clone DKEY-18C12, *** SEQUENCING IN PROGRESS ***, 5
            unordered pieces.
ACCESSION   CR769779
VERSION     CR769779.3  GI:52345314
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 219074)
  AUTHORS   McLay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
```

```
COMMENT     On Sep 18, 2004 this sequence version replaced gi:52214022.
            ------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            ------------- Project Information
            Center project name: zK18C12
            ------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 217273 bases at least Q40
            Consensus quality: 217586 bases at least Q30
            Consensus quality: 217870 bases at least Q20
            Insert size: 218674; sum-of-contigs
            Insert size: 217168; 4.7% error; agarose-fp
            Quality coverage: 8.00x in Q20 bases; sum-of-contigs Quality
            coverage: 8.05x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    31758: contig of 31758 bp in length
            *   31759    31858: gap of 100 bp
            *   31859    46986: contig of 15128 bp in length
            *   46987    47086: gap of 100 bp
            *   47087    62942: contig of 15856 bp in length
            *   62943    63042: gap of 100 bp
            *   63043   117588: contig of 54546 bp in length
            *  117589   117688: gap of 100 bp
            *  117689   219074: contig of 101386 bp in length.
FEATURES            Location/Qualifiers
     source         1. .219074
                    /organism="Danio rerio"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:7955"
                    /clone="DKEY-18C12"
                    /clone_lib="DanioKey"
     misc_feature   1. .31758
                    /note="assembly_fragment:00384
                    fragment_chain:1"
     misc_feature   31859. .46986
                    /note="assembly_fragment:00008
                    fragment_chain:1"
     misc_feature   47087. .62942
                    /note="assembly_fragment:00176.0"
     misc_feature   63043. .117588
                    /note="assembly_fragment:00812"
     misc_feature   117689. .219074
                    /note="assembly_fragment:01508"
ORIGIN

  Query Match             7.0%;  Score 350.2;  DB 2;  Length 219074;
  Best Local Similarity   67.3%;  Pred. No. 1.7e-60;
  Matches  524;  Conservative    0;  Mismatches  253;  Indels    2;  Gaps    2;

Qy        187 AACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATA 246
              | ||| |   |||     |     | ||||||||||||   | | ||||||||||||||||||
Db     218052 ACCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATATATATATATATATATA 218111

Qy        247 TATATACACACACACACACATATATATATACACATATATATACACATATATACACATATATA 306
              |||||| | | | | | |||||||||||| | ||||||||| | ||||||| |||||||
Db     218112 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 218171

Qy        307 CACATATATACACATATATATACACATACATATGTACACATATATATACACATATGTATA 366
              | |||||||| | ||||||||| | ||| |||| || | ||||||||| | |||| ||||
Db     218172 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 218231

Qy        367 CACATATATACACATATATACACACATATATACACATATATACACACATATATACACATA 426
              | |||||||| | |||||||| | | ||||||| | ||||||| | | ||||||| | |||
Db     218232 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 218291

Qy        427 TATACACATATATACACATATACACATATATACACATATATACATATATACACATATATA 486
              |||| | |||||||| | ||||| | ||||||| | ||||||| ||||||| | |||||||
Db     218292 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 218351

Qy        487 TAATATACACACATATATATACACATATATACACACATATATACACATATATACACATAT 546
              | ||||| | | |||||||||| | ||||||| | | ||||||| | ||||||| | ||||
Db     218352 T-ATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 218410

Qy        547 ATATACACATATATACACATATATACATATATACACATATATATACATATATACACATAT 606
              ||||| | |||||||| | ||||||| ||||||| | ||||||||| ||||||| | ||||
Db     218411 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 218470

Qy        607 ATACATATATACACATATATACATATATACACACATATATACACATACATATACACACAC 666
              ||| ||||||| | ||||||| ||||||| | | ||||||| | ||| ||||| | | |
Db     218471 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 218530

Qy        667 ATAGATATACATATATATACACATATATATACGTATATATATGTATATATATATGCTCCA 726
              ||| ||||| ||||||||| | |||||||||  |||||||||  ||||||||||
Db     218531 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 218590

Qy        727 GAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAG 786
              |  |||  |    |  |   ||   |   |||  |  | | | |  |  | |
Db     218591 ATATATATATAAATATATATATATATATATATATATATATATATATATATATATATAT 218650

Qy        787 CAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAAT 846
              | |||    | |   | || |     |      | |||       | ||| ||| ||
Db     218651 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 218710

Qy        847 GTGTAATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTAACACTATATATATATATAGTG 905
              | |||  | |  ||| | || |  ||    |  |  ||    || ||||| |||||
Db     218711 ATATAAATATATATATATATATATAAAATAATATATATATTTAAATATAAATATAAAT 218770

Qy        906 AAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTATAGAAGAATGGCTAA 964
              | |     |||   |  ||    ||  || ||||| ||  |  | |||   |||
Db     218771 ATAAATTATAAAAATATATATATATATTAATAATATAAATATTTAAATAAAATATATAA 218829


RESULT 64
CR749742
LOCUS       CR749742              194235 bp    DNA     linear   HTG 23-AUG-2004
DEFINITION  Danio rerio clone DKEYP-90H9, *** SEQUENCING IN PROGRESS ***, 9
            unordered pieces.
ACCESSION   CR749742
VERSION     CR749742.3  GI:51510860
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 194235)
  AUTHORS   McLay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-AUG-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 23, 2004 this sequence version replaced gi:51491327.
            ------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            ------------- Project Information
            Center project name: zKp90H9
            ------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 190869 bases at least Q40
            Consensus quality: 191868 bases at least Q30
```

```
            Consensus quality: 192427 bases at least Q20
            Insert size: 193435; sum-of-contigs
            Insert size: 191728; 8.4% error; agarose-fp
            Quality coverage: 8.95x in Q20 bases; sum-of-contigs Quality
            coverage: 9.11x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     2462: contig of 2462 bp in length
            *    2463     2562: gap of 100 bp
            *    2563    19346: contig of 16784 bp in length
            *   19347    19446: gap of 100 bp
            *   19447    64582: contig of 45136 bp in length
            *   64583    64682: gap of 100 bp
            *   64683    99973: contig of 35291 bp in length
            *   99974   100073: gap of 100 bp
            *  100074   123327: contig of 23254 bp in length
            *  123328   123427: gap of 100 bp
            *  123428   130685: contig of 7258 bp in length
            *  130686   130785: gap of 100 bp
            *  130786   142959: contig of 12174 bp in length
            *  142960   143059: gap of 100 bp
            *  143060   157214: contig of 14155 bp in length
            *  157215   157314: gap of 100 bp
            *  157315   194235: contig of 36921 bp in length.
FEATURES             Location/Qualifiers
     source          1. .194235
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEYP-90H9"
                     /clone_lib="DanioKeypilot"
     misc_feature    1. .2462
                     /note="assembly_fragment:00014
                     fragment_chain:1"
     misc_feature    2563. .19346
                     /note="assembly_fragment:00268
                     fragment_chain:1"
     misc_feature    19447. .64582
                     /note="assembly_fragment:02137
                     fragment_chain:1"
     misc_feature    64683. .99973
                     /note="assembly_fragment:01580
                     fragment_chain:1"
     misc_feature    100074. .123327
                     /note="assembly_fragment:00761
                     fragment_chain:1"
     misc_feature    123428. .130685
                     /note="assembly_fragment:00032
                     fragment_chain:1"
     misc_feature    130786. .142959
                     /note="assembly_fragment:00136.0"
     misc_feature    143060. .157214
                     /note="assembly_fragment:00500"
     misc_feature    157315. .194235
                     /note="assembly_fragment:01065"
ORIGIN

  Query Match             7.0%;  Score 350;  DB 2;  Length 194235;
  Best Local Similarity   69.2%;  Pred. No. 1.9e-60;
  Matches  504;  Conservative    0;  Mismatches  222;  Indels    2;  Gaps    2;

Qy        175 TCCAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATA 234
              | ||  |||   |   |||   | |   || ||   ||||||||||||   |   ||||||
Db     122601 TTCATTCCATATATTCAGACTTAGAATCCAACTTGAAAAAAAAAAAAAAAAAAAATATATA 122660

Qy        235 TATATATATATATATATACACACACACACACATATATATATACACATATATATACACATA 294
              |||||||||||||||||| | | | | | ||||||||||| | ||||||||| | |||
Db     122661 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 122720

Qy        295 TATACATATATACACATATATACACATATATATACACATACATATGTACACATATATATA 354
              |||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |||||||||
Db     122721 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 122780

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              | |||| |||| | ||||||| | ||||||| | | ||||||| | ||||||| | | |
Db     122781 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 122840

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              |||||| | ||||||| | ||||||| | ||||| | ||||||| | ||||||| |||||
Db     122841 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 122900

Qy        475 TACACATATATATAATATACACACATATATATACACATATATACACACATATATACACAT 534
              || | |||||||| ||||| | | ||||||||| | ||||||| | | ||||||| | ||
Db     122901 TATATATATATAT-ATATATATATATATATATATATATATATATATATATATATATATAT 122959

Qy        535 ATATACACATATATATACACATATATACACATATATACATATATACACATATATATACAT 594
              ||||| | |||||||||| | ||||||| | ||||||| ||||||| | |||||||||| ||
Db     122960 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 123019

Qy        595 ATATACACATATATACATATATACACATATATACATATATACACACATATATACACATAC 654
              ||||| | ||||||| ||||||| | ||||||| ||||||| | | ||||||| | |||
Db     123020 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 123079

Qy        655 ATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATAT 714
              ||||| | | | ||| ||||| ||||||||| | ||||||||| |||||||||| |||||
Db     123080 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 123139

Qy        715 ATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAG 774
              ||||||          | |||          | |   ||    |    ||| |   | |
Db     123140 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 123199

Qy        775 AGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGG 834
              | |  |  | |   | |||   | |   | || |    |    | |||
Db     123200 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 123259

Qy        835 AAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATA 894
              | ||| ||| || | ||   | |  ||| | | |  ||  ||  ||  |||  ||||||
Db     123260 ATATAAATATATATATATATATATATATAT-ANATATNTATAAATATAAATTTATATA 123318

Qy        895 TATATATA 902
              ||||||||
Db     123319 TATATATA 123326


RESULT 65
AL513530/c
LOCUS       AL513530               50612 bp    DNA     linear   PRI 15-MAY-2001
DEFINITION  Human DNA sequence from clone RP11-447M22 on chromosome 6, complete
            sequence.
ACCESSION   AL513530
VERSION     AL513530.8  GI:14133035
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 50612)
  AUTHORS   Sycamore,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On May 16, 2001 this sequence version replaced gi:13751553.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
```

```
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            RP11-447M22 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-447M22 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true left end of clone RP4-708F5 is at 50513 in this sequence.
            The true right end of clone RP4-682J15 is at 100 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .50612
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP11-447M22"
                     /clone_lib="RPCI-11.2"
     repeat_region   1485. .1570
                     /note="MIR repeat: matches 88. .174 of consensus"
     repeat_region   1745. .1774
                     /note="15 copies 2 mer tt 86% conserved"
     repeat_region   3540. .3811
                     /note="MLT1E repeat: matches 265. .561 of consensus"
     repeat_region   3800. .4122
                     /note="MLT1D repeat: matches 1. .322 of consensus"
     repeat_region   4169. .4648
                     /note="L1M4 repeat: matches 3589. .4085 of consensus"
     repeat_region   5046. .5161
                     /note="58 copies 2 mer aa 59% conserved"
     repeat_region   6172. .6350
                     /note="MER5A repeat: matches 1. .189 of consensus"
     repeat_region   6352. .6435
                     /note="42 copies 2 mer ta 63% conserved"
     repeat_region   6783. .8591
                     /note="TIGGER2 repeat: matches 895. .2718 of consensus"
     repeat_region   8592. .8882
                     /note="AluSx repeat: matches 1. .291 of consensus"
     repeat_region   8883. .9301
                     /note="TIGGER2 repeat: matches 474. .895 of consensus"
     repeat_region   9302. .9598
                     /note="AluJb repeat: matches 1. .291 of consensus"
     repeat_region   9599. .10107
                     /note="TIGGER2 repeat: matches 4. .474 of consensus"
     repeat_region   10470. .10756
                     /note="AluSc repeat: matches 1. .288 of consensus"
     repeat_region   13276. .13732
                     /note="MLT1D repeat: matches 1. .505 of consensus"
     repeat_region   13898. .14119
                     /note="MLT1-INTERNAL repeat: matches 951. .1182 of
                     consensus"
     repeat_region   14554. .15103
                     /note="MLT1-INTERNAL repeat: matches 36. .592 of consensus"
     repeat_region   15130. .15428
                     /note="MLT2E repeat: matches 136. .397 of consensus"
     repeat_region   15401. .15553
                     /note="MLT1D repeat: matches 2. .154 of consensus"
     repeat_region   16010. .16185
                     /note="L2 repeat: matches 2191. .2372 of consensus"
     repeat_region   16302. .16343
                     /note="L1MB8 repeat: matches 6117. .6163 of consensus"
     repeat_region   16353. .16656
                     /note="AluSx repeat: matches 1. .304 of consensus"
     repeat_region   16841. .16908
                     /note="L2 repeat: matches 2641. .2709 of consensus"
     repeat_region   17903. .17997
                     /note="L2 repeat: matches 2575. .2680 of consensus"
     repeat_region   18264. .19239
                     /note="488 copies 2 mer ta 87% conserved"
     misc_feature    18959
                     /note="Tandem repeat. Forced join. Gap size estimated to
                     be approximately 250bp by restriction digest data."
     repeat_region   19240. .19524
                     /note="AluY repeat: matches 1. .285 of consensus"
     repeat_region   19735. .20045
                     /note="AluY repeat: matches 1. .308 of consensus"
     repeat_region   20344. .20626
                     /note="L2 repeat: matches 2002. .2294 of consensus"
     repeat_region   20849. .21066
                     /note="MIR repeat: matches 34. .244 of consensus"
     repeat_region   21538. .21606
                     /note="MER5A repeat: matches 38. .103 of consensus"
     repeat_region   23617. .23800
                     /note="MER5A repeat: matches 2. .186 of consensus"
     repeat_region   23953. .24005
                     /note="MADE1 repeat: matches 28. .80 of consensus"
     repeat_region   24271. .24453
                     /note="MER5A repeat: matches 2. .189 of consensus"
     repeat_region   26159. .26323
                     /note="MIR repeat: matches 78. .249 of consensus"
     repeat_region   26313. .26567
                     /note="L2 repeat: matches 397. .654 of consensus"
     repeat_region   26731. .26876
                     /note="L2 repeat: matches 879. .1028 of consensus"
     repeat_region   27304. .27596
                     /note="AluY repeat: matches 1. .292 of consensus"
     repeat_region   28752. .29042
                     /note="L1 repeat: matches 2111. .2413 of consensus"
     repeat_region   29124. .30099
                     /note="L1 repeat: matches 2504. .3534 of consensus"
     repeat_region   30174. .31422
                     /note="L1M4 repeat: matches 3279. .4581 of consensus"
     repeat_region   31424. .32086
                     /note="L1MC3 repeat: matches 7067. .7739 of consensus"
     repeat_region   32092. .32407
                     /note="AluSq repeat: matches 1. .306 of consensus"
     repeat_region   32417. .32806
                     /note="L1M4 repeat: matches 4603. .5002 of consensus"
     repeat_region   32818. .33181
                     /note="L1MA1 repeat: matches -1388. .-1025 of consensus"
     repeat_region   33182. .33493
                     /note="AluSp repeat: matches 1. .313 of consensus"
     repeat_region   33494. .34874
                     /note="L1MA1 repeat: matches -1025. .356 of consensus"
     repeat_region   34876. .34981
                     /note="L1P repeat: matches 3001. .3103 of consensus"
     repeat_region   34966. .35806
                     /note="L1PA2 repeat: matches 5275. .6144 of consensus"
     repeat_region   35825. .41028
                     /note="L1MA1 repeat: matches 347. .5615 of consensus"
     repeat_region   41029. .41334
                     /note="AluSq repeat: matches 6. .310 of consensus"
     repeat_region   41335. .41728
                     /note="L1MA1 repeat: matches 5615. .6005 of consensus"
     repeat_region   41727. .41963
                     /note="L1MC2 repeat: matches 5825. .6063 of consensus"
     repeat_region   41976. .42095
                     /note="60 copies 2 mer at 85% conserved"
     repeat_region   42097. .42330
                     /note="L1MC1 repeat: matches 6081. .6322 of consensus"
     repeat_region   42547. .42818
```

```
                     /note="L1MB7 repeat: matches 5778. .6033 of consensus"
     repeat_region   42819. .43022
                     /note="AluY repeat: matches 102. .303 of consensus"
     repeat_region   43023. .43157
                     /note="L1MB7 repeat: matches 6033. .6171 of consensus"
     repeat_region   43283. .43474
                     /note="MIR repeat: matches 21. .247 of consensus"
     repeat_region   45140. .45454
                     /note="AluY repeat: matches 1. .311 of consensus"
     repeat_region   45522. .45828
                     /note="AluSg repeat: matches 1. .307 of consensus"
     repeat_region   47590. .47773
                     /note="HAL1 repeat: matches 1309. .1491 of consensus"
     repeat_region   47984. .48257
                     /note="HAL1 repeat: matches 630. .931 of consensus"
     repeat_region   49161. .49464
                     /note="AluSq repeat: matches 1. .305 of consensus"
     repeat_region   50177. .50264
                     /note="22 copies 4 mer agag 70% conserved"
     repeat_region   50186. .50263
                     /note="39 copies 2 mer ga 74% conserved"
     repeat_region   50465. .50490
                     /note="13 copies 2 mer gt 96% conserved"
ORIGIN

  Query Match             7.0%;  Score 349.6;  DB 9;  Length 50612;
  Best Local Similarity   72.2%;  Pred. No. 2.8e-60;
  Matches  523;  Conservative    0;  Mismatches  194;  Indels    7;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC--AAAAT 58
              || |  |||||| ||||   |||| |||||||||| ||||||||||||||||  |||||
Db      19451 AGATTGAGACCATCCTGGCTAACACGGTGAAACCCCGTCTCTACTAAAAATACAAAAAAT 19392

Qy         59 TAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGA 118
              |||   ||| |||||  ||  ||||||| |||||||||| ||||||||||||||||||||
Db      19391 TAGCAGGGCGTGGTGGCACGTGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGA 19332

Qy        119 ATCACTTGAACCTGCTGGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCC 177
              ||| ||||||||||  || ||||| |||||||| ||| |||||||  || | ||||||||
Db      19331 ATCGCTTGAACCTG--GGAGGTGGGGGTTGCGGTGAGCCAAGATCGCGCTACTGCACTCC 19274

Qy        178 AGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATAT 237
              ||||  ||| || ||||| | |   | ||||| || | | |       | ||||| |||
Db      19273 AGCCTGGGCGAC-AGAGCAAGACCCCGTCTCAAAATATATATATATTTATATATATTTAT 19215

Qy        238 ATATATATATATATACACACACACACATATATATATACACATATATATACACATATAT 297
              ||||||||  ||||||    | | |   ||||||||||  | |||||| || | || |||
Db      19214 ATATATATTTATATATTTATATATATTTATATATATATTTATATATATTTATATATTTAT 19155

Qy        298 ACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACAC 357
              | |||| || | ||||||  | ||||||| || | |||  ||| || | || ||||||
Db      19154 ATATATTTATATATATATTTATATATATTTATATATATTTATATATATATTTATATATTT 19095

Qy        358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417
              |||| |||  | ||||||  | |||| || | | || |||| |  ||||||  | |  |||
Db      19094 ATATATATTTATATATATTTATATATTTATATATATTTATATATTTATATATATATTTAT 19035

Qy        418 ATACACATATATACACATATATACACATATACACATATATACACATATATACATATATAC 477
              |||   |||||||  |  |||||| |  |||| |  ||||||   |||| || ||||||
Db      19034 ATATTTATATATATATTTATATATATCTATATATTTATATATGTATATTTATATATATCT 18975

Qy        478 ACATATATATAATATACACACATATATATACACATATATACACACATATATACACATATA 537
              | |||| ||| ||||| | | |||| |||| | || |||| |   ||||||  | |||||
Db      18974 ATATATTTAT-ATATATATATATATTTATATATATTTATATATTTATATATTTATATATA 18916

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATACATATA 597
              |  | ||||||||  | |||| || | |||| || |||||| || ||||||||  |||||
Db      18915 TTTATATATATATTTATATATTTATATATATTTATATATATTCATATATATATTTATATA 18856

Qy        598 TACACATATATACATATATACACATATATACATATATACACACATATATACACATACATA 657
              |  |||||| |   ||||||   |||||||  ||||||   |||||| ||   ||| ||
Db      18855 TTTACATATGTGTTTATATATTTATATATATTTATATATTTACATATGTATTTATATATT 18796
```

```
Qy        658 TACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATA 717
              || | | |  || ||||  |||||| ||   |||||| ||  ||||| ||| ||| ||||
Db      18795 TATATATATTTATATATTTATATATGTATTTATATATTTATATATATTTATATATTTATA 18736

Qy        718 TATG 721
              ||||
Db      18735 TATG 18732


RESULT 66
AF438327_3
WPCOMMENT
  Sequence split into 4 fragments  LOCUS AF438327 Accession AF438327
     Fragment Name      Begin     End
     AF438327_0             1     110000
     AF438327_1        100001     210000
     AF438327_2        200001     310000
     AF438327_3        300001     381696
Continuation (4 of 4) of AF438327 from base 300001 (AF438327 Homo sapiens alpha 1 type X)

  Query Match             7.0%;  Score 349.6;  DB 9;  Length 81696;
  Best Local Similarity   72.2%;  Pred. No. 2.6e-60;
  Matches  523;  Conservative    0;  Mismatches  194;  Indels    7;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC--AAAAT 58
              || |  |||||| ||||   |||| |||||||||| ||||||||||||||||  |||||
Db      8385 AGATTGAGACCATCCTGGCTAACACGGTGAAACCCCGTCTCTACTAAAAATACAAAAAAT 8444

Qy         59 TAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGA 118
              |||   ||| |||||  ||  ||||||| |||||||||| ||||||||||||||||||||
Db      8445 TAGCAGGGCGTGGTGGCACGTGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGA 8504

Qy        119 ATCACTTGAACCTGCTGGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCC 177
              ||| ||||||||||  || ||||| |||||||| ||| |||||||  || | ||||||||
Db      8505 ATCGCTTGAACCTG--GGAGGTGGGGGTTGCGGTGAGCCAAGATCGCGCTACTGCACTCC 8562

Qy        178 AGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATAT 237
              ||||  ||| || ||||| | |   | ||||| || | | |       | ||||| |||
Db      8563 AGCCTGGGCGAC-AGAGCAAGACCCCGTCTCAAAATATATATATATTTATATATATTTAT 8621

Qy        238 ATATATATATATATACACACACACACATATATATATACACATATATATACACATATAT 297
              ||||||||  ||||||    | | |   ||||||||||  | |||||| || | || |||
Db      8622 ATATATATTTATATATTTATATATATTTATATATATATTTATATATATTTATATATTTAT 8681

Qy        298 ACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACAC 357
              | |||| || | ||||||  | ||||||| || | |||  ||| || | || ||||||
Db      8682 ATATATTTATATATATATTTATATATATTTATATATATTTATATATATATTTATATATTT 8741

Qy        358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417
              |||| |||  | ||||||  | |||| || | | || |||| |  ||||||  | |  |||
Db      8742 ATATATATTTATATATATTTATATATTTATATATATTTATATATTTATATATATATTTAT 8801

Qy        418 ATACACATATATACACATATATACACATATACACATATATACACATATATACATATATAC 477
              |||   |||||||  |  |||||| |  |||| |  ||||||   |||| || ||||||
Db      8802 ATATTTATATATATATTTATATATATCTATATATTTATATATGTATATTTATATATATCT 8861

Qy        478 ACATATATATAATATACACACATATATATACACATATATACACACATATATACACATATA 537
              | |||| ||| ||||| | | |||| |||| | || |||| |   ||||||  | |||||
Db      8862 ATATATTTAT-ATATATATATATATTTATATATATTTATATATTTATATATTTATATATA 8920

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATACATATA 597
              |  | ||||||||  | |||| || | |||| || |||||| || ||||||||  |||||
Db      8921 TTTATATATATATTTATATATTTATATATATTTATATATATTCATATATATATTTATATA 8980

Qy        598 TACACATATATACATATATACACATATATACATATATACACACATATATACACATACATA 657
              |  |||||| |   ||||||   |||||||  ||||||   |||||| ||   ||| ||
Db      8981 TTTACATATGTGTTTATATATTTATATATATTTATATATTTACATATGTATTTATATATT 9040

Qy        658 TACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATA 717
              || | | |  || ||||  |||||| ||   |||||| ||  ||||| ||| ||| ||||
Db      9041 TATATATATTTATATATTTATATATGTATTTATATATTTATATATATTTATATATTTATA 9100

Qy        718 TATG 721
```

```
                ||||
Db       9101 TATG 9104


RESULT 67
AF438327_2
WPCOMMENT
  Sequence split into 4 fragments  LOCUS AF438327 Accession AF438327
     Fragment Name      Begin     End
     AF438327_0             1    110000
     AF438327_1        100001    210000
     AF438327_2        200001    310000
     AF438327_3        300001    381696
Continuation (3 of 4) of AF438327 from base 200001 (AF438327 Homo sapiens alpha 1 type X

  Query Match             7.0%;  Score 349.6;  DB 9;  Length 110000;
  Best Local Similarity   72.2%;  Pred. No. 2.5e-60;
  Matches  523;  Conservative    0;  Mismatches  194;  Indels    7;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC--AAAAT 58
              || |  |||||| ||||   |||| |||||||||||| ||||||||||||||  |||||
Db     108385 AGATTGAGACCATCCTGGCTAACACGGTGAAACCCCGTCTCTACTAAAAATACAAAAAAT 108444

Qy         59 TAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGA 118
              |||   ||| |||||  ||  ||||||| |||||||||||  ||||||  ||||||||||
Db     108445 TAGCAGGGCGTGGTGGCACGTGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGA 108504

Qy        119 ATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCC 177
              ||| ||||||||||  || ||||| |||||||| ||| |||||||  || | ||||||||
Db     108505 ATCGCTTGAACCTG--GGAGGTGGGGGTTGCGGTGAGCCAAGATCGCGCTACTGCACTCC 108562

Qy        178 AGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATAT 237
              ||||  ||| || ||||| | |   | ||||| || | | | |      | ||||| |||
Db     108563 AGCCTGGGCGAC-AGAGCAAGACCCCGTCTCAAAATATATATATATTTATATATATTTAT 108621

Qy        238 ATATATATATATATACACACACACACATATATATATACACATATATATACACATATAT 297
              |||||||| ||||||   | | | |   |||||||||  | |||||| || | || |||
Db     108622 ATATATATTTATATATTTATATATATTTATATATATATTTATATATATTTATATATTTAT 108681

Qy        298 ACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACAC 357
              | |||| || | ||||||  | |||||| || | |||  ||| || | || ||||||
Db     108682 ATATATTTATATATATATTTATATATATTTATATATATTTATATATATATTTATATATTT 108741

Qy        358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417
              |||| |||  | ||||||  | |||| || | | || |||| |  |||||| | |  |||
Db     108742 ATATATATTTATATATATTTATATATTTATATATATTTATATATTTATATATATATTTAT 108801

Qy        418 ATACACATATATACACATATATACACATATACACATATATACACATATATACATATATAC 477
              |||  ||||||| |  |||||| |  |||| |  ||||||   |||| || ||||||
Db     108802 ATATTTATATATATATTTATATATATCTATATATTTATATATGTATATTTATATATATCT 108861

Qy        478 ACATATATATAATATACACACATATATATACACATATATACACACATATATACACATATA 537
              | |||| ||| ||||| | | |||| |||| | || |||| |   ||||||  | |||||
Db     108862 ATATATTTAT-ATATATATATATATTTATATATATTTATATATTTATATATTTATATATA 108920

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATACATATA 597
              |  | ||||||||  | |||| || | |||| || |||||| || ||||||||  |||||
Db     108921 TTTATATATATATTTATATATTTATATATATTTATATATATTCATATATATATTTATATA 108980

Qy        598 TACACATATATACATATATACACATATATACATATATACACACATATATACACATACATA 657
              |  |||||| |   ||||||   |||||||  ||||||   ||||||  ||   ||| ||
Db     108981 TTTACATATGTGTTTATATATTTATATATATTTATATATTTACATATGTATTTATATATT 109040

Qy        658 TACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATA 717
              || | |  || ||||  |||||| ||   |||||| ||  ||||| ||| ||| ||||
Db     109041 TATATATATTTATATATTTATATATGTATTTATATATTTATATATATTTATATATTTATA 109100

Qy        718 TATG 721
                ||||
Db     109101 TATG 109104


RESULT 68
```

```
CR759826/c
LOCUS       CR759826              157042 bp    DNA     linear   HTG 05-SEP-2004
DEFINITION  Danio rerio clone CH211-266K13, *** SEQUENCING IN PROGRESS ***, 10
            unordered pieces.
ACCESSION   CR759826
VERSION     CR759826.1  GI:51890341
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 157042)
  AUTHORS   McLay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zC266K13
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 153933 bases at least Q40
            Consensus quality: 154487 bases at least Q30
            Consensus quality: 154866 bases at least Q20
            Insert size: 156142; sum-of-contigs
            Insert size: 159217; 0.8% error; agarose-fp
            Quality coverage: 6.01x in Q20 bases; sum-of-contigs Quality
            coverage: 5.90x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    44418: contig of 44418 bp in length
            *   44419    44518: gap of 100 bp
            *   44519    52984: contig of 8466 bp in length
            *   52985    53084: gap of 100 bp
            *   53085    65205: contig of 12121 bp in length
            *   65206    65305: gap of 100 bp
            *   65306    67571: contig of 2266 bp in length
            *   67572    67671: gap of 100 bp
            *   67672   110374: contig of 42703 bp in length
            *  110375   110474: gap of 100 bp
            *  110475   142025: contig of 31551 bp in length
            *  142026   142125: gap of 100 bp
            *  142126   146693: contig of 4568 bp in length
            *  146694   146793: gap of 100 bp
            *  146794   151219: contig of 4426 bp in length
            *  151220   151319: gap of 100 bp
            *  151320   154853: contig of 3534 bp in length
            *  154854   154953: gap of 100 bp
            *  154954   157042: contig of 2089 bp in length.
FEATURES             Location/Qualifiers
     source          1. .157042
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-266K13"
                     /clone_lib="CHORI-211"
     misc_feature    1. .44418
                     /note="assembly_fragment:00615
                     fragment_chain:1"
```

```
     misc_feature    44519. .52984
                     /note="assembly_fragment:00139
                     fragment_chain:1"
     misc_feature    53085. .65205
                     /note="assembly_fragment:00214
                     fragment_chain:1"
     misc_feature    65306. .67571
                     /note="assembly_fragment:00005
                     fragment_chain:1"
     misc_feature    67672. .110374
                     /note="assembly_fragment:01059
                     fragment_chain:1"
     misc_feature    110475. .142025
                     /note="assembly_fragment:00318
                     fragment_chain:2"
     misc_feature    142126. .146693
                     /note="assembly_fragment:00071
                     fragment_chain:2"
     misc_feature    146794. .151219
                     /note="assembly_fragment:00104"
     misc_feature    151320. .154853
                     /note="assembly_fragment:00042
                     fragment_chain:3"
     misc_feature    154954. .157042
                     /note="assembly_fragment:00020
                     fragment_chain:3
                     clone_end:SP6
                     vector_side:right"
ORIGIN

  Query Match             7.0%;  Score 349.6;  DB 2;  Length 157042;
  Best Local Similarity   63.1%;  Pred. No. 2.4e-60;
  Matches  462;  Conservative    0;  Mismatches  269;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db      68181 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 68122

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db      68121 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 68062

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db      68061 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 68002

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db      68001 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 67942

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db      67941 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 67882

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db      67881 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 67822

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db      67821 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 67762

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||  |||||  | | | ||| ||||| ||||||||| | |||||||||  || |||
Db      67761 TATATATTTATATATATATATATATATATATATATATATATATATATATATATATATACATA 67702

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||| ||||    |  |
Db      67701 TATATATATATATATGATCCATCTCTACTCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 67642

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825

Db      67641 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 67582

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885
                        | | | || | ||   | |  ||| | ||   |  ||       ||
Db      67581 NNNNNNNNNNTTTTTAAATATATATATATATATATATATTTTTTTTTTTTTTATAT 67522

Qy        886 CACTATATATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAAT 945
                |||||||||||||   | |    | || |  || |         |   || || ||
Db      67521 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 67462

Qy        946 GTTATAGAAGAA 957
                | | || ||
Db      67461 AAAAAAAAAAAA 67450


RESULT 69
CR589883/c
LOCUS       CR589883              180560 bp    DNA     linear   HTG 27-JUL-2004
DEFINITION  Danio rerio clone CH211-51N18, WORKING DRAFT SEQUENCE, 6 unordered
            pieces.
ACCESSION   CR589883
VERSION     CR589883.2  GI:50724901
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 180560)
  AUTHORS   Burton,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JUL-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 27, 2004 this sequence version replaced gi:50399732.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zC51N18
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 177647 bases at least Q40
            Consensus quality: 178065 bases at least Q30
            Consensus quality: 178456 bases at least Q20
            Insert size: 180060; sum-of-contigs
            Insert size: 184777; 2.8% error; agarose-fp
            Quality coverage: 8.72x in Q20 bases; sum-of-contigs Quality
            coverage: 8.60x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     7987: contig of 7987 bp in length
            *     7988     8087: gap of 100 bp
            *     8088    33232: contig of 25145 bp in length
            *    33233    33332: gap of 100 bp
            *    33333    70114: contig of 36782 bp in length
            *    70115    70214: gap of 100 bp
            *    70215   126158: contig of 55944 bp in length
            *   126159   126258: gap of 100 bp
            *   126259   162306: contig of 36048 bp in length
            *   162307   162406: gap of 100 bp
            *   162407   180560: contig of 18154 bp in length.
FEATURES             Location/Qualifiers
     source          1. .180560
```

```
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-51N18"
                     /clone_lib="CHORI-211"
     misc_feature    1. .7987
                     /note="assembly_fragment:00018
                     clone_end:SP6
                     vector_side:left"
     misc_feature    8088. .33232
                     /note="assembly_fragment:00362.0"
     misc_feature    33333. .70114
                     /note="assembly_fragment:01176
                     fragment_chain:1"
     misc_feature    70215. .126158
                     /note="assembly_fragment:01644
                     fragment_chain:1"
     misc_feature    126259. .162306
                     /note="assembly_fragment:00726
                     fragment_chain:1"
     misc_feature    162407. .180560
                     /note="assembly_fragment:00143
                     fragment_chain:1
                     clone_end:T7
                     vector_side:right"
ORIGIN

  Query Match             7.0%;  Score 349.6;  DB 2;  Length 180560;
  Best Local Similarity   43.8%;  Pred. No. 2.3e-60;
  Matches  560;  Conservative    0;  Mismatches  717;  Indels    2;  Gaps    2;

Qy        205 TCTCAGAAAAAAAAAAAGGCATTTTATATATATATATATATATACACACACACACA 264
              | | || | | | | |      | |||||||||||||||||||||||| | | | | | |
Db     127006 TTTAAGTATATATATATATATATATATATATATATATATATATATATATATATATA 126947

Qy        265 CATATATATATACACATATATATACACATATATACATATATACACATATATACACATATA 324
              ||||||||||| | |||||||||| | ||||||| ||||||| | ||||||| | |||||
Db     126946 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 126887

Qy        325 TATACACATACATATGTACACATATATATACACATATGTATACACATATATACACATATA 384
              |||| | ||| |||| || | ||||||||| | |||| |||| | ||||||| | |||||
Db     126886 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 126827

Qy        385 TACACACATATATACACATATATACACACATATATACACATATATACACATATATACACA 444
              || | | ||||||| | ||||||| | | ||||||| | ||||||| | ||||||| | |
Db     126826 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 126767

Qy        445 TATACACATATATACACATATATACATATATACACATATATATAATATACACACATATAT 504
              |||| | ||||||| | ||||||| ||||||| | |||||||| ||||| | | ||||||
Db     126766 TATATATATATATATATATATATATATATATATATATATATAT-ATATATATATATAT 126708

Qy        505 ATACACATATATACACACATATATACACATATATACACATATATATACACATATATACAC 564
              ||| | ||||||| | | ||||||| | ||||||| | ||||||||| | ||||||| |
Db     126707 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 126648

Qy        565 ATATATACATATATACACATATATATACATATATACACATATATACATATATACACATAT 624
              ||||||| ||||||| | ||||||||| ||||||| | ||||||| ||||||| | ||||
Db     126647 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 126588

Qy        625 ATACATATATACACACATATATACACATACATATACACACACATAGATATACATATATAT 684
              ||| ||||||| | | ||||||| | ||| ||||| | | | ||| ||||| |||||||||
Db     126587 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 126528

Qy        685 ACACATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGC 744
              | | |||||||||  ||||||||| ||||||||||        |  |||
Db     126527 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 126468

Qy        745 AGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGG 804
              |  |   ||    |   ||| |   | | |  |  | |    | |||    | |
Db     126467 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 126408

Qy        805 ACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCT 864
              | || |    |      | |||       | ||| ||| || | ||   | |  ||| |
```

```
Db     126407 ATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 126348

Qy        865 GTA-AGGTAGTTAAGAAGGTAACACTATATATATATAGTGAAAGCAGTGTAAACCTAA 923
              ||| |  ||  ||   |  ||    |||||||||||||   | |    | ||    ||
Db     126347 ATATATATATATATATATATATATATATATATATATATATATATATCTATATGAATAT 126288

Qy        924 AGGATGGGCCAAGGATTTAAATGTTATAGAAGAATGGCTAAGATGCCAAAGCTCAGTGTA 983
              |       |   || || || | |
Db     126287 ATATATATATATATATATATATATATANNNNNNNNNNNNNNNNNNNNNNNNNNNNN 126228

Qy        984 TGTGGCAGAGGCATGGTGTAGGGTGTGTCCAGGTTCATATATTGCATTAAGTGTGAGAAC 1043

Db     126227 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 126168

Qy       1044 ACCCTGGAGTATGAACCAAGAAAATGCAAAAGCCAGAAGTGATGGAGGAAATGAGACACA 1103
                  ||  ||       |                |
Db     126167 NNNNNNNNNTAAAAANNNNNNANNATNNNNNNNNNNANNNNNNNNNNTNNNNNNNTN 126108

Qy       1104 ATAATGAAGATATTGAGAGGAGGGTGTGGGCCTAGAGTGAAGCTTTTCGTGCCAGTACTT 1163
              |                                    ||
Db     126107 TTNNNTNNNNNNNNNNNNANNTNNNNNTNNNNNNNNNNAAANNNNNNNNNNNNNNNNN 126048

Qy       1164 CTTTTGAAGGCCCAGTTCTCTTCTCTCTCGGGGGCTCCTTCATCTCTCATAGAGTCCACA 1223
                           |
Db     126047 NNANNNNNNNNTANNNNNTNNANNTNNNTNNNNNNNNNNNNNNNNNANNNNNNNNNN 125988

Qy       1224 GCTTTTAAGGGCCAACACTTGAGGTCAGCCTGGCTCTCTCATTTGAGCTGGATAGAACAT 1283

Db     125987 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 125928

Qy       1284 TTTAGAGCACCATCTATTCTTCAAGAGGAAGTTTAAAAATAAAAGAACCTTGAAGAGGAA 1343
                    | | || ||  | |  |  |  ||||     |     |
Db     125927 NNNNNNNTTNNNTTTTTTTTTNTATAANANATANTAAAAANNNNNANNNNNANNNNANT 125868

Qy       1344 AAAATGTAGACATTCAATCTAACCTTTTCATTTTACTAGCCAAAGCTAAATAGAATGCAG 1403
                  |                |          |     |   || ||      |
Db     125867 TNNNNNTTNNNNNNNNNNNNNNNTAAANNNNNNNNAANNANNATTAAATTANTTAAAAT 125808

Qy       1404 GTTACCTGTTTTTCAGCCAGGCACCATCATTTCCTAATTGTTATAAAATTTATTATTATT 1463
                | |     ||||                |        || || |     ||    |   |
Db     125807 ATAANTNNNTTTTTTTNTNTNNNNNTNTNTNNNNNTTTTTTTNTNTNTNTTTNNNTNNNT 125748

Qy       1464 GTTGTTATTATTATTATTT 1482
                ||    ||    || | |
Db     125747 NTTNNNTTTNNNTTTNTNT 125729
```

```
RESULT 70
CR556714
LOCUS       CR556714               210384 bp    DNA     linear   HTG 12-JUL-2004
DEFINITION  Danio rerio clone DKEY-29P13, WORKING DRAFT SEQUENCE, 5 unordered
            pieces.
ACCESSION   CR556714
VERSION     CR556714.2  GI:50251122
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 210384)
  AUTHORS   Burton,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUL-2004) Wellcome Trust Sanger.Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 12, 2004 this sequence version replaced gi:50234128.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
```

```
            Center project name: zK29P13
            ------------ Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 209301 bases at least Q40
            Consensus quality: 209711 bases at least Q30
            Consensus quality: 209847 bases at least Q20
            Insert size: 209984; sum-of-contigs
            Insert size: 209994; 2.1% error; agarose-fp
            Quality coverage: 10.01x in Q20 bases; sum-of-contigs Quality
            coverage: 10.13x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    19750: contig of 19750 bp in length
            *    19751    19850: gap of 100 bp
            *    19851    37399: contig of 17549 bp in length
            *    37400    37499: gap of 100 bp
            *    37500   112088: contig of 74589 bp in length
            *   112089   112188: gap of 100 bp
            *   112189   196376: contig of 84188 bp in length
            *   196377   196476: gap of 100 bp
            *   196477   210384: contig of 13908 bp in length.
FEATURES             Location/Qualifiers
     source          1. .210384
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-29P13"
                     /clone_lib="DanioKey"
     misc_feature    1. .19750
                     /note="assembly_fragment:00422
                     fragment_chain:1"
     misc_feature    19851. .37399
                     /note="assembly_fragment:00191
                     fragment_chain:1"
     misc_feature    37500. .112088
                     /note="assembly_fragment:00690
                     fragment_chain:2"
     misc_feature    112189. .196376
                     /note="assembly_fragment:01732
                     fragment_chain:2"
     misc_feature    196477. .210384
                     /note="assembly_fragment:00009.0"
ORIGIN

  Query Match             7.0%;  Score 349.6;  DB 2;  Length 210384;
  Best Local Similarity   68.8%;  Pred. No. 2.3e-60;
  Matches  509;  Conservative    0;  Mismatches  229;  Indels    2;  Gaps    2;

Qy        187 AACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATA 246
              ||  | |   |||    | |  |  |||| ||||||  | | ||||||||||||||||||
Db     111340 AAGTAAAAAAAAAGTAAAACAAAATAAAATAAAAAGTAAATATATATATATATATATATA 111399

Qy        247 TATATACACACACACACATATATATATACACATATATATACACATATATACATATATA 306
              |||||| | | | | | |||||||||||| | ||||||||| | ||||||| |||||||
Db     111400 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 111459

Qy        307 CACATATATACACATATATATACACATACATATGTACACATATATATACACATATGTATA 366
              | ||||||| | ||||||||| | ||| |||| || | ||||||||| | |||| ||||
Db     111460 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 111519

Qy        367 CACATATATACACATATATACACACATATATACACATATATACACACATATATACACATA 426
              | ||||||| | ||||||| | | ||||||| | ||||||| | | ||||||| | |||
Db     111520 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 111579

Qy        427 TATACACATATATACACATATACACATATATACACATATATACATATATACACATATATA 486
              |||| | ||||||| | ||||| | ||||||| | ||||||| ||||||| | |||||||
Db     111580 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 111639

Qy        487 TA-ATATACACACATATATATACACATATATACACACATATATACACATATATACACATA 545
              || ||||| | | |||||||||| | ||||||| | | ||||||| | ||||||| | |||
Db     111640 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 111699

Qy        546 TATATACACATATATACACATATATACATATATACACATATATATACATATATACACATA 605
              |||||| | ||||||| | ||||||| ||||||| | ||||||||| ||||||| | |||
Db     111700 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 111759

Qy        606 TATACATATATACACATATATACATATATACACACATATATACACATACATATACACACA 665
              |||| ||||||| | ||||||| ||||||| | | ||||||| | ||| ||||| | | |
Db     111760 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 111819

Qy        666 CATAGATATACATATATATACACATATATATACGTATATATATGTATATATATATGCTCC 725
              ||| ||||| ||||||||| | ||||||||| |||||||||| ||||||||||
Db     111820 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 111879

Qy        726 AGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACA 785
              |  |||          |  |   ||     |   ||| |   | | | |  | | |
Db     111880 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 111939

Qy        786 GCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAA 845
              | |||   | |   | || |     |     | |||       | ||| ||| |
Db     111940 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 111999

Qy        846 TGTGTAATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTAACACTATATATATATATAGT 904
              | | ||   | |  ||| | || |  ||  ||   |  |||   |||||||||||||
Db     112000 TATATATATATATATATATATATATATATATATATATAAATATATATATATATATATA 112059

Qy        905 GAAAGCAGTGTAAACCTAAA 924
              | |  | | || |  || |
Db     112060 TATATAAATATATATATA 112079


RESULT 71
CR790372
LOCUS       CR790372              299106 bp    DNA     linear   HTG 05-OCT-2004
DEFINITION  Danio rerio clone DKEY-70K13, *** SEQUENCING IN PROGRESS ***, 25
            unordered pieces.
ACCESSION   CR790372
VERSION     CR790372.4  GI:53828302
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 299106)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 5, 2004 this sequence version replaced gi:53755940.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK70K13
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 289331 bases at least Q40
            Consensus quality: 291856 bases at least Q30
            Consensus quality: 293478 bases at least Q20
            Insert size: 296706; sum-of-contigs
            Insert size: 225769; 2.9% error; agarose-fp
            Quality coverage: 6.22x in Q20 bases; sum-of-contigs Quality
            coverage: 9.16x in Q20 bases; agarose-fp
```

```
            ----------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 25 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     3928: contig of 3928 bp in length
            *     3929     4028: gap of 100 bp
            *     4029     8417: contig of 4389 bp in length
            *     8418     8517: gap of 100 bp
            *     8518    12283: contig of 3766 bp in length
            *    12284    12383: gap of 100 bp
            *    12384    16946: contig of 4563 bp in length
            *    16947    17046: gap of 100 bp
            *    17047    19834: contig of 2788 bp in length
            *    19835    19934: gap of 100 bp
            *    19935    26863: contig of 6929 bp in length
            *    26864    26963: gap of 100 bp
            *    26964    29135: contig of 2172 bp in length
            *    29136    29235: gap of 100 bp
            *    29236    43722: contig of 14487 bp in length
            *    43723    43822: gap of 100 bp
            *    43823   109813: contig of 65991 bp in length
            *   109814   109913: gap of 100 bp
            *   109914   140749: contig of 30836 bp in length
            *   140750   140849: gap of 100 bp
            *   140850   156362: contig of 15513 bp in length
            *   156363   156462: gap of 100 bp
            *   156463   165292: contig of 8830 bp in length
            *   165293   165392: gap of 100 bp
            *   165393   243285: contig of 77893 bp in length
            *   243286   243385: gap of 100 bp
            *   243386   245562: contig of 2177 bp in length
            *   245563   245662: gap of 100 bp
            *   245663   250094: contig of 4432 bp in length
            *   250095   250194: gap of 100 bp
            *   250195   256405: contig of 6211 bp in length
            *   256406   256505: gap of 100 bp
            *   256506   258788: contig of 2283 bp in length
            *   258789   258888: gap of 100 bp
            *   258889   262780: contig of 3892 bp in length
            *   262781   262880: gap of 100 bp
            *   262881   265140: contig of 2260 bp in length
            *   265141   265240: gap of 100 bp
            *   265241   269635: contig of 4395 bp in length
            *   269636   269735: gap of 100 bp
            *   269736   272991: contig of 3256 bp in length
            *   272992   273091: gap of 100 bp
            *   273092   291385: contig of 18294 bp in length
            *   291386   291485: gap of 100 bp
            *   291486   293629: contig of 2144 bp in length
            *   293630   293729: gap of 100 bp
            *   293730   296461: contig of 2732 bp in length
            *   296462   296561: gap of 100 bp
            *   296562   299106: contig of 2545 bp in length.
FEATURES             Location/Qualifiers
     source          1. .299106
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-70K13"
                     /clone_lib="DanioKey"
     misc_feature    1. .3928
                     /note="assembly_fragment:00025
                     fragment_chain:1"
     misc_feature    4029. .8417
                     /note="assembly_fragment:00178
                     fragment_chain:1"
     misc_feature    8518. .12283
                     /note="assembly_fragment:00051
                     fragment_chain:1"
     misc_feature    12384. .16946
                     /note="assembly_fragment:00068
                     fragment_chain:1"
     misc_feature    17047. .19834
                     /note="assembly_fragment:03721
                     fragment_chain:1"
     misc_feature    19935. .26863
                     /note="assembly_fragment:00206
                     fragment_chain:1"
     misc_feature    26964. .29135
                     /note="assembly_fragment:03693
                     fragment_chain:1"
     misc_feature    29236. .43722
                     /note="assembly_fragment:00405
                     fragment_chain:2"
     misc_feature    43823. .109813
                     /note="assembly_fragment:01445
                     fragment_chain:2"
     misc_feature    109914. .140749
                     /note="assembly_fragment:01030
                     fragment_chain:2"
     misc_feature    140850. .156362
                     /note="assembly_fragment:00596
                     fragment_chain:3"
     misc_feature    156463. .165292
                     /note="assembly_fragment:00282
                     fragment_chain:3"
     misc_feature    165393. .243285
                     /note="assembly_fragment:02358
                     fragment_chain:3"
     misc_feature    243386. .245562
                     /note="assembly_fragment:03730
                     fragment_chain:4"
     misc_feature    245663. .250094
                     /note="assembly_fragment:00104
                     fragment_chain:4"
     misc_feature    250195. .256405
                     /note="assembly_fragment:00151
                     fragment_chain:4"
     misc_feature    256506. .258788
                     /note="assembly_fragment:00124
                     fragment_chain:5"
     misc_feature    258889. .262780
                     /note="assembly_fragment:00234
                     fragment_chain:5"
     misc_feature    262881. .265140
                     /note="assembly_fragment:03713
                     fragment_chain:6"
     misc_feature    265241. .269635
                     /note="assembly_fragment:00038
                     fragment_chain:6"
     misc_feature    269736. .272991
                     /note="assembly_fragment:00085"
     misc_feature    273092. .291385
                     /note="assembly_fragment:00801.0"
     misc_feature    291486. .293629
                     /note="assembly_fragment:03699"
     misc_feature    293730. .296461
                     /note="assembly_fragment:03706"
     misc_feature    296562. .299106
                     /note="assembly_fragment:03750"
ORIGIN

  Query Match             7.0%;  Score 349.6;  DB 2;  Length 299106;
  Best Local Similarity   60.4%;  Pred. No. 2.1e-60;
  Matches  492;  Conservative    0;  Mismatches  321;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||||| | | | | | | |||||||||||| | |||||||
Db     242607 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 242666

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
```

```
Db     242667 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 242726

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406

Db     242727 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 242786

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466

Db     242787 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 242846

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525

Db     242847 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 242906

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585

Db     242907 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 242966

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645

Db     242967 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 243026

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705

Db     243027 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 243086

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765

Db     243087 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 243146

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825

Db     243147 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 243206

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885

Db     243207 TATATATATATATATATATATATATAAATATATATATATTTATTGTTGTTATATANANAA 243266

Qy        886 CACTATATATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAAT 945

Db     243267 AAAAAAAAATATATATATANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 243326

Qy        946 GTTATAGAAGAATGGCTAAGATGCCAAAGCTCAGTGTATGTGGCAGAGGCATGGTGTAGG 1005

Db     243327 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNC 243386

Qy       1006 GTGTGTCCAGGTTCATATATTGCATTAAGTGTGA 1039

Db     243387 TGAGCAAAGGCTGTGAATACTTCTGTACGTGTGA 243420


RESULT 72
CR790384/c
LOCUS       CR790384              182223 bp    DNA     linear   HTG 08-OCT-2004
DEFINITION  Danio rerio clone DKEY-208I6, *** SEQUENCING IN PROGRESS ***, 7
            unordered pieces.
ACCESSION   CR790384
VERSION     CR790384.3  GI:54019829
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 182223)
  AUTHORS   McLay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 9, 2004 this sequence version replaced gi:52694441.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK208I6
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 180226 bases at least Q40
            Consensus quality: 180685 bases at least Q30
            Consensus quality: 180956 bases at least Q20
            Insert size: 181623; sum-of-contigs
            Insert size: 177450; 3.6% error; agarose-fp
            Quality coverage: 17.50x in Q20 bases; sum-of-contigs Quality
            coverage: 18.50x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    27274: contig of 27274 bp in length
            *    27275    27374: gap of 100 bp
            *    27375    66205: contig of 38831 bp in length
            *    66206    66305: gap of 100 bp
            *    66306   114861: contig of 48556 bp in length
            *   114862   114961: gap of 100 bp
            *   114962   118489: contig of 3528 bp in length
            *   118490   118589: gap of 100 bp
            *   118590   147487: contig of 28898 bp in length
            *   147488   147587: gap of 100 bp
            *   147588   158792: contig of 11205 bp in length
            *   158793   158892: gap of 100 bp
            *   158893   182223: contig of 23331 bp in length.
FEATURES             Location/Qualifiers
     source          1. .182223
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-208I6"
                     /clone_lib="DanioKey"
     misc_feature    1. .27274
                     /note="assembly_fragment:01073
                     fragment_chain:1"
     misc_feature    27375. .66205
                     /note="assembly_fragment:01954
                     fragment_chain:1"
     misc_feature    66306. .114861
                     /note="assembly_fragment:03886
                     fragment_chain:1"
     misc_feature    114962. .118489
                     /note="assembly_fragment:00044
                     fragment_chain:2"
     misc_feature    118590. .147487
                     /note="assembly_fragment:02902
                     fragment_chain:2"
     misc_feature    147588. .158792
                     /note="assembly_fragment:00096.0"
     misc_feature    158893. .182223
                     /note="assembly_fragment:00385"
ORIGIN

  Query Match             7.0%;  Score 349.4;  DB 2;  Length 182223;
  Best Local Similarity   68.3%;  Pred. No. 2.5e-60;
  Matches  499;  Conservative    0;  Mismatches  231;  Indels    1;  Gaps    1;

Qy        199 ATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACA 258

Db      67159 ATTTAAGCAAGTAAAAAAAAAATATATATATATATATATATATATATATATATATATA 67100
```

```
Qy        259 CACACACATATATATATACACATATATATACACATATATACATATATACACATATATACA 318
              | | | |||||||||||| | ||||||||| | ||||||| ||||||| | ||||||| |
Db      67099 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 67040

Qy        319 CATATATATACACATACATATGTACACATATATATACACATATGTATACACATATATACA 378
              ||||||||| | ||| |||| || | ||||||||| | |||| |||| | ||||||| |
Db      67039 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 66980

Qy        379 CATATATACACACATATATACACATATATACACACATATATACACATATATACACATATA 438
              ||||||| | | ||||||| | ||||||| | | ||||||| | ||||||| | |||||
Db      66979 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 66920

Qy        439 TACACATATACACATATATACACATATATACATATATACACATATATATA-ATATACACA 497
              || | ||||| | ||||||| | ||||||| ||||||| | ||||||||| ||||| |.|
Db      66919 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 66860

Qy        498 CATATATATACACATATATACACACATATATACACATATATACACATATATATACACATA 557
              ||||||||| | ||||||| | | ||||||| | ||||||| | ||||||||| | |||
Db      66859 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 66800

Qy        558 TATACACATATATACATATATACACATATATATACATATATACACATATATACATATATA 617
              |||| | ||||||| ||||||| | ||||||||| ||||||| | ||||||| |||||||
Db      66799 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 66740

Qy        618 CACATATATACATATATACACACATATATACACATACATATACACACACATAGATATACA 677
              | ||||||| ||||||| | | ||||||| | ||| ||||| | | | ||| ||||| |
Db      66739 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 66680

Qy        678 TATATATACACATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAG 737
              |||||||| | ||||||||| |||||||||| |||||||||||           |  |||
Db      66679 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 66620

Qy        738 AGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGT 797
              |   ||  | |   ||| |  ||| | |  |  | ||   | |||
Db      66619 TATATATATAATATATATAATATATATATATAAATAAATATATAAATATATATATATA 66560

Qy        798 GAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAA 857
              | |   | |||| |  |  ||    |  |        | ||| | |||      ||  ||
Db      66559 TATATATATATTTTAAATATATATATATAATATATAAATATATAAAAAATAAATATTTAA 66500

Qy        858 CCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATATATATATAGTGAAAGCAGTGTAA 917
              ||   | |   | ||| |  ||| |  || ||  ||||   | ||   |   |||
Db      66499 ATAATAATTTATAAATATAAATATATAATAAAATTTAATTATAATATAAATTAATAATAA 66440

Qy        918 ACCTAAAGGAT 928
              |  | ||  ||
Db      66439 ATATTAAATAT 66429
```

```
RESULT 73
CR790362
LOCUS       CR790362              223166 bp    DNA     linear   HTG 05-OCT-2004
DEFINITION  Danio rerio clone DKEY-219H14, *** SEQUENCING IN PROGRESS ***, 13
            unordered pieces.
ACCESSION   CR790362
VERSION     CR790362.3  GI:53828288
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 223166)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 5, 2004 this sequence version replaced gi:52626676.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK219H14
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 218162 bases at least Q40
            Consensus quality: 219035 bases at least Q30
            Consensus quality: 219848 bases at least Q20
            Insert size: 221966; sum-of-contigs
            Insert size: 231484; 3.1% error; agarose-fp
            Quality coverage: 9.07x in Q20 bases; sum-of-contigs Quality
            coverage: 9.03x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 13 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    45629: contig of 45629 bp in length
            *    45630    45729: gap of 100 bp
            *    45730    47743: contig of 2014 bp in length
            *    47744    47843: gap of 100 bp
            *    47844    71492: contig of 23649 bp in length
            *    71493    71592: gap of 100 bp
            *    71593    80196: contig of 8604 bp in length
            *    80197    80296: gap of 100 bp
            *    80297    94270: contig of 13974 bp in length
            *    94271    94370: gap of 100 bp
            *    94371   105308: contig of 10938 bp in length
            *   105309   105408: gap of 100 bp
            *   105409   113062: contig of 7654 bp in length
            *   113063   113162: gap of 100 bp
            *   113163   125638: contig of 12476 bp in length
            *   125639   125738: gap of 100 bp
            *   125739   129703: contig of 3965 bp in length
            *   129704   129803: gap of 100 bp
            *   129804   143077: contig of 13274 bp in length
            *   143078   143177: gap of 100 bp
            *   143178   148174: contig of 4997 bp in length
            *   148175   148274: gap of 100 bp
            *   148275   180179: contig of 31905 bp in length
            *   180180   180279: gap of 100 bp
            *   180280   223166: contig of 42887 bp in length.
FEATURES             Location/Qualifiers
     source          1. .223166
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-219H14"
                     /clone_lib="DanioKey"
     misc_feature    1. .45629
                     /note="assembly_fragment:02641
                     fragment_chain:1"
     misc_feature    45730. .47743
                     /note="assembly_fragment:00088
                     fragment_chain:1"
     misc_feature    47844. .71492
                     /note="assembly_fragment:01125
                     fragment_chain:1"
     misc_feature    71593. .80196
                     /note="assembly_fragment:00289
                     fragment_chain:1"
     misc_feature    80297. .94270
                     /note="assembly_fragment:00715
                     fragment_chain:1"
     misc_feature    94371. .105308
                     /note="assembly_fragment:00568
                     fragment_chain:2"
```

```
     misc_feature    105409. .113062
                     /note="assembly_fragment:00191
                     fragment_chain:2"
     misc_feature    113163. .125638
                     /note="assembly_fragment:00919
                     fragment_chain:2"
     misc_feature    125739. .129703
                     /note="assembly_fragment:00071
                     fragment_chain:3"
     misc_feature    129804. .143077
                     /note="assembly_fragment:00426
                     fragment_chain:3"
     misc_feature    143178. .148174
                     /note="assembly_fragment:00118.0"
     misc_feature    148275. .180179
                     /note="assembly_fragment:01481"
     misc_feature    180280. .223166
                     /note="assembly_fragment:01968"
ORIGIN

  Query Match             7.0%;  Score 349.4;  DB 2;  Length 223166;
  Best Local Similarity   79.2%;  Pred. No. 2.5e-60;
  Matches  427;  Conservative    0;  Mismatches  111;  Indels    1;  Gaps    1;

Qy        188 ACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATAT 247
              | || |  ||||| |  | | |||||||||| |||   ||| ||||||||||||||||||
Db     201702 AAAAAAAAGAAATTTTTTTTTAGAAAAAAATAAATAAATTATATATATATATATATATAT 201761

Qy        248 ATATACACACACACACATATATATATACACATATATATACACATATATACATATATAC 307
              ||||| | | | | | |||||||||||| | |||||||| | |||||||| |||||||
Db     201762 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 201821

Qy        308 ACATATATACACATATATATACACATACATATGTACACATATATATACACATATGTATAC 367
              | ||||||| | |||||||| | ||| |||| || | |||||||| | |||| ||||
Db     201822 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 201881

Qy        368 ACATATATACACATATATACACACATATATACACATATATACACACATATATACACATAT 427
              | ||||||| | ||||||| | | ||||||| | ||||||| | | ||||||| | ||||
Db     201882 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 201941

Qy        428 ATACACATATATACACATATACACATATATACACATATATACATATATACACATATATAT 487
              ||| | ||||||| | ||||| | ||||||| | ||||||| ||||||| | ||||||||
Db     201942 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 202001

Qy        488 A-ATATACACACATATATATACACATATATACACACATATATACACATATATACACATAT 546
              | ||||| | | |||||||| | ||||||| | | ||||||| | ||||||| | ||||
Db     202002 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 202061

Qy        547 ATATACACATATATACACATATATACATATATACACATATATATACATATATACACATAT 606
              ||||| | ||||||| | ||||||| ||||||| | ||||||||| ||||||| | ||||
Db     202062 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 202121

Qy        607 ATACATATATACACATATATACATATATACACACATATATACACATACATATACACACAC 666
              ||| || |||| | ||||||| ||||||| | | ||||||| | ||| ||||| | | |
Db     202122 ATATATCTATATATATATATATATATATATATATATATATATATATATATATATATAT 202181

Qy        667 ATAGATATACATATATATACACATATATATACGTATATATATGTATATATATATGCTCC 725
              ||| ||||| |||||||| | |||||||||  |||||||| |||||||||  || |
Db     202182 ATATATATATATATATATATATATATATATATATATATATATATATATATATAACCTTC 202240


RESULT 74
CR759946
LOCUS.       CR759946               232824 bp    DNA     linear   HTG 11-SEP-2004
DEFINITION   Danio rerio clone DKEY-79C1, *** SEQUENCING IN PROGRESS ***, 8
             unordered pieces.
ACCESSION    CR759946
VERSION      CR759946.2  GI:51988227
KEYWORDS     HTG; HTGS_PHASE1.
SOURCE       Danio rerio (zebrafish)
  ORGANISM   Danio rerio
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
             Cypriniformes; Cyprinidae; Danio.
REFERENCE    1  (bases 1 to 232824)
  AUTHORS    Sims,S.
  TITLE      Direct Submission
  JOURNAL    Submitted (10-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
             Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
             zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT      On Sep 11, 2004 this sequence version replaced gi:51965771.
             -------------- Genome Center
             Center: Wellcome Trust Sanger Institute
             Center code: SC
             Web site: http://www.sanger.ac.uk
             Contact: zfish-help@sanger.ac.uk
             -------------- Project Information
             Center project name: zK79C1
             -------------- Summary Statistics
             Assembly program: XGAP4; version 4.5
             Chemistry: Dye-terminator; 100% of reads
             Consensus quality: 229932 bases at least Q40
             Consensus quality: 230704 bases at least Q30
             Consensus quality: 231171 bases at least Q20
             Insert size: 232124; sum-of-contigs
             Insert size: 219964; 3.3% error; agarose-fp
             Quality coverage: 8.73x in Q20 bases; sum-of-contigs Quality
             coverage: 9.26x in Q20 bases; agarose-fp
             --------------.
             * NOTE: This is a 'working draft' sequence. It currently
             * consists of 8 contigs. The true order of the pieces
             * is not known and their order in this sequence record is
             * arbitrary. Gaps between the contigs are represented as
             * runs of N, but the exact sizes of the gaps are unknown.
             * This record will be updated with the finished sequence
             * as soon as it is available and the accession number will
             * be preserved.
             *        1    33972: contig of 33972 bp in length
             *    33973    34072: gap of 100 bp
             *    34073    37589: contig of 3517 bp in length
             *    37590    37689: gap of 100 bp
             *    37690    45659: contig of 7970 bp in length
             *    45660    45759: gap of 100 bp
             *    45760    54370: contig of 8611 bp in length
             *    54371    54470: gap of 100 bp
             *    54471    87298: contig of 32828 bp in length
             *    87299    87398: gap of 100 bp
             *    87399   111650: contig of 24252 bp in length
             *   111651   111750: gap of 100 bp
             *   111751   170131: contig of 58381 bp in length
             *   170132   170231: gap of 100 bp
             *   170232   232824: contig of 62593 bp in length.
FEATURES             Location/Qualifiers
     source          1. .232824
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-79C1"
                     /clone_lib="DanioKey"
     misc_feature    1. .33972
                     /note="assembly_fragment:00912
                     fragment_chain:1"
     misc_feature    34073. .37589
                     /note="assembly_fragment:00025
                     fragment_chain:1"
     misc_feature    37690. .45659
                     /note="assembly_fragment:00053
                     fragment_chain:1"
     misc_feature    45760. .54370
                     /note="assembly_fragment:00127
                     fragment_chain:1"
     misc_feature    54471. .87298
                     /note="assembly_fragment:00512
                     fragment_chain:2"
     misc_feature    87399. .111650
                     /note="assembly_fragment:00209
```

```
                     fragment_chain:2"
     misc_feature    111751. .170131
                     /note="assembly_fragment:01393"
     misc_feature    170232. .232824
                     /note="assembly_fragment:02188.0"
ORIGIN

  Query Match             7.0%;  Score 349.4;  DB 2;  Length 232824;
  Best Local Similarity   59.8%;  Pred. No. 2.4e-60;
  Matches  502;  Conservative    0;  Mismatches  336;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db     169438 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 169497

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db     169498 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 169557

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db     169558 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 169617

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db     169618 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 169677

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db     169678 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 169737

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db     169738 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 169797

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| |||||||| | ||||||| |||||||| | ||||||| ||||||| | | |||||
Db     169798 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 169857

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db     169858 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 169917

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||||          | |||          | |   ||     |    |||
Db     169918 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 169977

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
              |   | | | |  | | |     | |||     | |   | || |      |      | |
Db     169978 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 170037

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885
              ||       | ||| ||| || | ||   | |  ||| | ||   ||  ||       ||
Db     170038 TATATATATATATATATATATATATATATATATATATATATATAATATATATTTATAT 170097

Qy        886 CACTATATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAAT 945
                |||||||||||||   | |      || |
Db     170098 ATATATATATATATATATATATATACATATATNNNNNNNNNNNNNNNNNNNNNNNNN 170157

Qy        946 GTTATAGAAGAATGGCTAAGATGCCAAAGCTCAGTGTATGTGGCAGAGGCATGGTGTAGG 1005

Db     170158 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 170217

Qy       1006 GTGTGTCCAGGTTCATATATTGCATTAAGTGTGAGAACACCCTGGAGTATGAACCAAGA 1064
                  ||| |||   | |      | ||   |||   || |||  | | |
Db     170218 NNNNNNNNNNNNNATACATTTTGTCACTAAAAACAATGTCCTCTGGTGTGACACTATA 170276


RESULT 75
CR318636
LOCUS       CR318636              233296 bp    DNA     linear   HTG 31-AUG-2004
DEFINITION  Danio rerio clone DKEY-249N13, *** SEQUENCING IN PROGRESS ***, 10
            unordered pieces.
ACCESSION   CR318636
VERSION     CR318636.3  GI:51861411
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 233296)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-AUG-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 2, 2004 this sequence version replaced gi:44909414.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK249N13
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 229955 bases at least Q40
            Consensus quality: 230652 bases at least Q30
            Consensus quality: 231088 bases at least Q20
            Insert size: 232396; sum-of-contigs
            Insert size: 229559; 5.6% error; agarose-fp
            Quality coverage: 7.04x in Q20 bases; sum-of-contigs Quality
            coverage: 7.26x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2974: contig of 2974 bp in length
            *     2975     3074: gap of 100 bp
            *     3075    13370: contig of 10296 bp in length
            *    13371    13470: gap of 100 bp
            *    13471    23811: contig of 10341 bp in length
            *    23812    23911: gap of 100 bp
            *    23912    37932: contig of 14021 bp in length
            *    37933    38032: gap of 100 bp
            *    38033    54166: contig of 16134 bp in length
            *    54167    54266: gap of 100 bp
            *    54267   102290: contig of 48024 bp in length
            *   102291   102390: gap of 100 bp
            *   102391   125681: contig of 23291 bp in length
            *   125682   125781: gap of 100 bp
            *   125782   142646: contig of 16865 bp in length
            *   142647   142746: gap of 100 bp
            *   142747   165218: contig of 22472 bp in length
            *   165219   165318: gap of 100 bp
            *   165319   233296: contig of 67978 bp in length.
FEATURES             Location/Qualifiers
     source          1. .233296
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-249N13"
                     /clone_lib="DanioKey"
     misc_feature    1. .2974
                     /note="assembly_fragment:00024
                     fragment_chain:1"
     misc_feature    3075. .13370
                     /note="assembly_fragment:00163
                     fragment_chain:1"
```

```
     misc_feature    13471. .23811
                     /note="assembly_fragment:00045
                     fragment_chain:1"
     misc_feature    23912. .37932
                     /note="assembly_fragment:00283
                     fragment_chain:2"
     misc_feature    38033. .54166
                     /note="assembly_fragment:00602
                     fragment_chain:2"
     misc_feature    54267. .102290
                     /note="assembly_fragment:01245
                     fragment_chain:2"
     misc_feature    102391. .125681
                     /note="assembly_fragment:01003
                     fragment_chain:3"
     misc_feature    125782. .142646
                     /note="assembly_fragment:00432
                     fragment_chain:3"
     misc_feature    142747. .165218
                     /note="assembly_fragment:00787
                     fragment_chain:3"
     misc_feature    165319. .233296
                     /note="assembly_fragment:01715.0"
ORIGIN

  Query Match             7.0%;  Score 349.4;  DB 2;  Length 233296;
  Best Local Similarity   69.7%;  Pred. No. 2.4e-60;
  Matches  487;  Conservative    0;  Mismatches  211;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||||| | | | | | | |||||||||||| | |||||||
Db     164502 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 164561

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db     164562 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 164621

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db     164622 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 164681

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db     164682 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 164741

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db     164742 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 164801

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db     164802 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 164861

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db     164862 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 164921

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db     164922 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 164981

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||||          | |||          | |    ||      |   |||
Db     164982 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 165041

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
              |  | | | |  | | |  | |||    | |  | || |    |      | | |
Db     165042 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 165101

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885
              |       | ||| ||| || | ||    | |  ||| | ||   ||  |||  |  |||
Db     165102 AATATATATATATATATATATATATATATATATATATATATATAATATATAAATATATAA 165161

Qy        886 CACTATATATATATATAGTGAAAGCAGTGTAAACCTAAA 924
              | ||||  ||||||||   | |  | | || |  ||||
Db     165162 AAATATAATTATATATATATATATAAATATATATATAAA 165200


RESULT 76
CR631122
LOCUS       CR631122              162977 bp    DNA     linear   HTG 13-AUG-2004
DEFINITION  Danio rerio clone DKEYP-36H11, *** SEQUENCING IN PROGRESS ***, 7
            unordered pieces.
ACCESSION   CR631122
VERSION     CR631122.2  GI:51241630
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 162977)
  AUTHORS   McLay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-AUG-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 13, 2004 this sequence version replaced gi:51127401.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zKp36H11
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 159934 bases at least Q40
            Consensus quality: 160525 bases at least Q30
            Consensus quality: 161089 bases at least Q20
            Insert size: 162377; sum-of-contigs
            Insert size: 159156; 13.8% error; agarose-fp
            Quality coverage: 7.39x in Q20 bases; sum-of-contigs Quality
            coverage: 7.54x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    11615: contig of 11615 bp in length
            *    11616    11715: gap of 100 bp
            *    11716    30875: contig of 19160 bp in length
            *    30876    30975: gap of 100 bp
            *    30976    47841: contig of 16866 bp in length
            *    47842    47941: gap of 100 bp
            *    47942   115881: contig of 67940 bp in length
            *   115882   115981: gap of 100 bp
            *   115982   134454: contig of 18473 bp in length
            *   134455   134554: gap of 100 bp
            *   134555   137147: contig of 2593 bp in length
            *   137148   137247: gap of 100 bp
            *   137248   162977: contig of 25730 bp in length.
FEATURES             Location/Qualifiers
     source          1. .162977
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEYP-36H11"
                     /clone_lib="DanioKeypilot"
     misc_feature    1. .11615
                     /note="assembly_fragment:00030
```

```
                     fragment_chain:1"
     misc_feature    11716. .30875
                     /note="assembly_fragment:00862
                     fragment_chain:1"
     misc_feature    30976. .47841
                     /note="assembly_fragment:00397
                     fragment_chain:1"
     misc_feature    47942. .115881
                     /note="assembly_fragment:01121
                     fragment_chain:1"
     misc_feature    115982. .134454
                     /note="assembly_fragment:00177
                     fragment_chain:1"
     misc_feature    134555. .137147
                     /note="assembly_fragment:00008"
     misc_feature    137248. .162977
                     /note="assembly_fragment:00618.0"
ORIGIN

  Query Match             7.0%;  Score 349.2;  DB 2;  Length 162977;
  Best Local Similarity   69.9%;  Pred. No. 2.8e-60;
  Matches  485;  Conservative    0;  Mismatches  208;  Indels    1;  Gaps    1;

Qy        209 AGAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACATA 268
              ||| |||  | |       | ||||||||||||||||||||||| | | | | | |||
Db     115076 AGAGAAACTATATATATATATATATATATATATATATATATATATATATATATATA 115135

Qy        269 TATATATACACATATATATACACATATATACATATATACACATATATACACATATATATA 328
              |||||||| | ||||||||| | ||||||| ||||||| | ||||||| | |||||||||
Db     115136 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 115195

Qy        329 CACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACA 388
              | ||| |||| || | ||||||||| | |||| |||| | ||||||| | ||||||| |
Db     115196 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 115255

Qy        389 CACATATATACACATATATACACACATATATACACATATATACACATATATACACATATA 448
              | ||||||| | ||||||| | | ||||||| | ||||||| | ||||||| | |||||
Db     115256 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 115315

Qy        449 CACATATATACACATATATACATATATACACATATATATAATATACACACATATATATAC 508
              | |||||||| | ||||||| ||||||| | |||||||| ||||| | | |||||||||
Db     115316 TATATATATATATATATATATATATATATATATATATAT-ATATATATATATATATAT 115374

Qy        509 ACATATATACACACATATATACACATATATACACATATATATACACATATATACACATAT 568
              | ||||||| | | ||||||| | ||||||| | ||||||||| | ||||||| | ||||
Db     115375 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 115434

Qy        569 ATACATATATACACATATATATACATATATACACATATATACATATATACACATATATAC 628
              ||| ||||||| | ||||||||| ||||||| | ||||||| ||||||| | |||||||
Db     115435 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 115494

Qy        629 ATATATACACACATATATACACATACATATACACACACATAGATATACATATATATACAC 688
              ||||||| | | ||||||| | ||| ||||| | | | ||| ||||| |||||||||
Db     115495 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 115554

Qy        689 ATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTT 748
              |||||||||  |||||||||| ||||||||||          |  |||        |  |
Db     115555 ATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 115614

Qy        749 GATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACAT 808
                ||    |    ||| |  | | |  |  | |    | |||    | |   | ||
Db     115615 ATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 115674

Qy        809 TTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAA 868
              |    |    | |||    | ||| ||| || | ||   | |  ||| | ||
Db     115675 ATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 115734

Qy        869 GGTAGTTAAGAAGGTAACACTATATATATATATA 902
                ||  ||   |  ||| | ||||||||||||||
Db     115735 TATATATATATATATAATATTATATATATATATA 115768


RESULT 77
```

```
AC021590/c
LOCUS       AC021590              171427 bp    DNA     linear   PRI 01-JAN-2002
DEFINITION  Homo sapiens chromosome 8, clone RP11-143H8, complete sequence.
ACCESSION   AC021590
VERSION     AC021590.10  GI:17432882
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 171427)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 8, clone RP11-143H8
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 171427)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 171427)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JAN-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Dec 10, 2001 this sequence version replaced gi:17402804.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L5777
                Center clone name: 143_H_8
```

```
                     ----------.
FEATURES             Location/Qualifiers
     source          1. .171427
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="RP11-143H8"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   complement(2020. .2137)
                     /rpt_family="L2"
     repeat_region   complement(2160. .2295)
                     /rpt_family="L2"
     repeat_region   complement(2349. .2479)
                     /rpt_family="FLAM_C"
     repeat_region   complement(2483. .2835)
                     /rpt_family="L2"
     repeat_region   complement(3975. .4107)
                     /rpt_family="FLAM_C"
     repeat_region   complement(4639. .4747)
                     /rpt_family="L2"
     repeat_region   complement(5060. .5878)
                     /rpt_family="L1PA16"
     repeat_region   complement(6034. .6096)
                     /rpt_family="MIR"
     repeat_region   6248. .6555
                     /rpt_family="AluJb"
     repeat_region   6716. .7017
                     /rpt_family="L2"
     repeat_region   7035. .7070
                     /rpt_family="(CA)n"
     repeat_region   7333. .7572
                     /rpt_family="MIR"
     repeat_region   7627. .7903
                     /rpt_family="L2"
     repeat_region   8277. .8484
                     /rpt_family="AluJo"
     repeat_region   8718. .8877
                     /rpt_family="L3"
     repeat_region   complement(9247. .9483)
                     /rpt_family="MER46A"
     repeat_region   10206. .10263
                     /rpt_family="AT_rich"
     repeat_region   10421. .10493
                     /rpt_family="AT_rich"
     repeat_region   complement(10891. .11060)
                     /rpt_family="L2"
     repeat_region   11256. .11305
                     /rpt_family="AT_rich"
     repeat_region   11809. .12090
                     /rpt_family="AluSx"
     repeat_region   12091. .12189
                     /rpt_family="(TAGA)n"
     repeat_region   12406. .12512
                     /rpt_family="FRAM"
     repeat_region   12513. .12539
                     /rpt_family="AT_rich"
     repeat_region   12566. .12755
                     /rpt_family="MER61A"
     repeat_region   12883. .18519
                     /rpt_family="L1PA6"
     repeat_region   18929. .18994
                     /rpt_family="MIR"
     repeat_region   20160. .20182
                     /rpt_family="(TGGA)n"
     repeat_region   complement(20466. .20667)
                     /rpt_family="MIR"
     repeat_region   20671. .20850
                     /rpt_family="FAM"
     repeat_region   20880. .20909
                     /rpt_family="AT_rich"
     repeat_region   complement(20910. .21037)
                     /rpt_family="FRAM"
     repeat_region   21121. .21419
                     /rpt_family="AluJo"
     repeat_region   complement(21424. .21508)
                     /rpt_family="L1ME"
     repeat_region   21934. .22046
                     /rpt_family="MIR3"
     repeat_region   23344. .23388
                     /rpt_family="AT_rich"
     repeat_region   23449. .23483
                     /rpt_family="(TA)n"
     repeat_region   complement(23484. .23779)
                     /rpt_family="AluSx"
     repeat_region   24093. .24190
                     /rpt_family="L2"
     repeat_region   24985. .25019
                     /rpt_family="L1MA1"
     repeat_region   25020. .25319
                     /rpt_family="AluSx"
     repeat_region   25320. .25557
                     /rpt_family="L1MA1"
     repeat_region   25558. .25581
                     /rpt_family="AT_rich"
     repeat_region   25777. .25802
                     /rpt_family="AT_rich"
     repeat_region   complement(26663. .26756)
                     /rpt_family="L1ME4A"
     repeat_region   26764. .26817
                     /rpt_family="(TAAAAA)n"
     repeat_region   27050. .27082
                     /rpt_family="AT_rich"
     repeat_region   complement(27455. .27626)
                     /rpt_family="L1ME4A"
     repeat_region   30048. .30108
                     /rpt_family="AluJ/FRAM"
     repeat_region   30249. .30269
                     /rpt_family="AT_rich"
     repeat_region   complement(30270. .30568)
                     /rpt_family="AluSq"
     repeat_region   31382. .31690
                     /rpt_family="AluSx"
     repeat_region   complement(32052. .32214)
                     /rpt_family="L2"
     repeat_region   32452. .32743
                     /rpt_family="AluSc"
     repeat_region   32839. .32969
                     /rpt_family="L2"
     repeat_region   complement(33055. .33190)
                     /rpt_family="FLAM_C"
     repeat_region   complement(33266. .33548)
                     /rpt_family="L2"
     repeat_region   33600. .33795
                     /rpt_family="L2"
     repeat_region   33939. .34117
                     /rpt_family="L1MDa"

  Query Match          7.0%;  Score 349.2;  DB 9;  Length 171427;
  Best Local Similarity  65.2%;  Pred. No. 2.8e-60;
  Matches  609;  Conservative    0;  Mismatches  313;  Indels   12;  Gaps    6;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC--AAAAT 58
              || ||||||||| ||||   |||| |||||||||| ||||||||||||||  |||||
Db      74773 AGATCAAGACCATCCTGGCTAACACGGTGAAACCCCGTCTCTACTAAAAATACAAAAAT 74714

Qy         59 TAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGA 118
              |||   |||||||||     |||||||| |||||||||||  |||||| ||||||||||
Db      74713 TAGCAGGGCATGGTGGCGGGCGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGA 74654

Qy        119 ATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCC 177
              ||  |  |||  | | || || |||| |||| | ||| | ||||   |||| ||||||||
Db      74653 ATGGCGGGAA--TCCGGGAGGCGGAGCTTGCAGTGAGCCGAGATAGCGCCACTGCACTCC 74596

Qy        178 AGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTT-----TATA 232
```

```
                |||| || |||       | | |      | |||||||| ||  ||| |     || |
Db      74595 AGCCTGGGGGACAGAGCGAGACTCTGTCTCAAAAAAAAAAAGAAAACATATAAATATAAA 74536

Qy        233 TATATATATATATATATATACACACACACACATATATATATACACATATATATACACA 292
              ||| |||||| |||||||||| |||    | | |||| |||||||   | |||| ||| |||
Db      74535 TATGTATATAAATATATATATACATTTATATACATTTATATATTTATATATAAATATACA 74476

Qy        293 TATATACATATATACACATATATACACATATATATACACATACATATGTACACATATATA 352
              ||||||  | ||||   ||||||    |||||| || | ||   ||| ||   ||||||
Db      74475 TATATATGTGTATATTTATATATGTTTATATATTTATATATGTTTATATATTTATATATG 74416

Qy        353 TACACATATGTATACACATATATACACATATATACACACATATATACACATATATACACA 412
              |  | |||| |||| | ||||  | | || | || | |  ||||| | ||||||| | |
Db      74415 TTTATATATTTATATATATATTAATATATGTTTATATATTTATATATAAATATATATAAA 74356

Qy        413 CATATATACACATATATACACATATATACACATATACACATATATACACATATATACATA 472
              ||| |||| | | ||||| | ||||||| | | ||  | ||||| | | ||| ||| |||
Db      74355 CATTTATATATAAATATATAAATATATATAAACATTTATATATAAATATATAAATATATA 74296

Qy        473 TATACACATATATAT-AATATACACACATATATATACACATATATACACACATATATACA 531
              || |||  ||||||| |||||| | | ||||||| |||  |||||| | | ||||| | |
Db      74295 TAAACATTTATATATAAATATATAAATATATATAAACATTTATATATAAATATATAAATA 74236

Qy        532 CATATATACACATATATATACACATATATACACATATATACATATATACACATATATATA 591
              |||||  |||  |||||||| | ||||||| |||| |||| ||| ||| | | |||||||
Db      74235 TATATAAACATTTATATATAAATATATATAAACATTTATATATAAATATATAAATATATA 74176

Qy        592 CATATATACACATATATACATATATACACATATATACATATATACACACATATATACACA 651
              | | ||  | ||||| | ||||| | | ||| | |  |||||| | | ||||| | | |
Db      74175 TAAACATTTATATATAAATATATAAATATATAAACATTTATATATAAATATATAAATATA 74116

Qy        652 TACATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTA 711
              || | ||  | | | | | ||||| ||||||| |||  ||||||||  ||||||||   |
Db      74115 TAAACATTTATATATAAATATATAAATATATAAACATTTATATATAAATATATATAAACA 74056

Qy        712 TATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAA 771
              | |||||||         || |||         |  | |  |     |    | | |   |
Db      74055 TTTATATATAAATATATATTTATATATAAATATATATAAACATTTATATATAAATATATA 73996

Qy        772 AAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAG 831
              || | |  || | |     ||||    | |   | ||||      |        | |||
Db      73995 AATATATATTTATATATACGTATATATAAATATATATTTATATATACGTATATATAAATA 73936

Qy        832 GGGAAATACATAAATGTGTAATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTAACACTA 890
                | ||  |||||  | ||  || |  ||| | || | |||   |   |  |||   ||
Db      73935 TATATATGTATAAACATATATAAATATATATTTATATATGTATAAACATATATAAATATA 73876

Qy        891 TATATATATATAGTGAAAGCAGTGTAAACCTAAA 924
              ||| |||||||    |||    | ||||  || |
Db      73875 TATTTATATATGTATAAACATATATAAATATATA 73842


RESULT 78
CR759788/c
LOCUS       CR759788              195311 bp    DNA     linear   HTG 07-SEP-2004
DEFINITION  Danio rerio clone CH211-153F13, *** SEQUENCING IN PROGRESS ***, 11
            unordered pieces.
ACCESSION   CR759788
VERSION     CR759788.3  GI:51950464
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 195311)
  AUTHORS   McLay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 9, 2004 this sequence version replaced gi:51890035.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zC153F13
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 191941 bases at least Q40
            Consensus quality: 192830 bases at least Q30
            Consensus quality: 193638 bases at least Q20
            Insert size: 194311; sum-of-contigs
            Insert size: 184045; 2.4% error; agarose-fp
            Quality coverage: 6.41x in Q20 bases; sum-of-contigs Quality
            coverage: 6.89x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     4825: contig of 4825 bp in length
            *     4826     4925: gap of 100 bp
            *     4926    14966: contig of 10041 bp in length
            *    14967    15066: gap of 100 bp
            *    15067    17900: contig of 2834 bp in length
            *    17901    18000: gap of 100 bp
            *    18001    22764: contig of 4764 bp in length
            *    22765    22864: gap of 100 bp
            *    22865    40985: contig of 18121 bp in length
            *    40986    41085: gap of 100 bp
            *    41086    65693: contig of 24608 bp in length
            *    65694    65793: gap of 100 bp
            *    65794   118569: contig of 52776 bp in length
            *   118570   118669: gap of 100 bp
            *   118670   131148: contig of 12479 bp in length
            *   131149   131248: gap of 100 bp
            *   131249   172432: contig of 41184 bp in length
            *   172433   172532: gap of 100 bp
            *   172533   178335: contig of 5803 bp in length
            *   178336   178435: gap of 100 bp
            *   178436   195311: contig of 16876 bp in length.
FEATURES             Location/Qualifiers
     source          1. .195311
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-153F13"
                     /clone_lib="CHORI-211"
     misc_feature    1. .4825
                     /note="assembly_fragment:00162
                     fragment_chain:1
                     clone_end:SP6
                     vector_side:left"
     misc_feature    4926. .14966
                     /note="assembly_fragment:00231
                     fragment_chain:1"
     misc_feature    15067. .17900
                     /note="assembly_fragment:00038
                     fragment_chain:1"
     misc_feature    18001. .22764
                     /note="assembly_fragment:00112
                     fragment_chain:1"
     misc_feature    22865. .40985
                     /note="assembly_fragment:00460
                     fragment_chain:2"
     misc_feature    41086. .65693
                     /note="assembly_fragment:00820
                     fragment_chain:2"
```

```
     misc_feature    65794. .118569
                     /note="assembly_fragment:01583
                     fragment_chain:2"
     misc_feature    118670. .131148
                     /note="assembly_fragment:00325
                     fragment_chain:2"
     misc_feature    131249. .172432
                     /note="assembly_fragment:01060
                     fragment_chain:2"
     misc_feature    172533. .178335
                     /note="assembly_fragment:00062
                     fragment_chain:2"
     misc_feature    178436. .195311
                     /note="assembly_fragment:00627
                     fragment_chain:2"
ORIGIN

  Query Match             7.0%;  Score 349.2;  DB 2;  Length 195311;
  Best Local Similarity   63.3%;  Pred. No. 2.8e-60;
  Matches  475;  Conservative    0;  Mismatches  273;  Indels    2;  Gaps    2;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||||| | | | | | | |||||||||||| | |||||||
Db     178944 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 178885

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | |||||||| | ||| |||| || | |
Db     178884 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 178825

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db     178824 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 178765

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db     178764 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 178705

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db     178704 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 178645

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db     178644 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 178585

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db     178584 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 178525

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||   ||||||
Db     178524 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 178465

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||||          |  ||
Db     178464 TATATATATATATATATATATATATATATATNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 178405

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825

Db     178404 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 178345

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTA 884
                    | ||| ||| || | ||    | |  ||| | || |  ||  ||    |  ||
Db     178344 NNNNNNNNATATATATATATATATATATATATATATATATATATATATATATATATATATA 178285

Qy        885 ACACTATATATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAA 944
                |||||||||||||    | |     | || |  ||      |  |   ||   ||
Db     178284 TATATATATATATATATATATATATATATATATATATATATTCATAGATCTGTGGCGAGAAT 178225

Qy        945 TGTTATAGAAGAATGGCTAAGATGCCAAAG 974
              ||  | | |||   |  |||   |  ||||
Db     178224 TGCAAGAAAAGTTAGAATAACTGGTTAAAG 178195
```

```
RESULT 79
CR762470
LOCUS       CR762470              199230 bp    DNA     linear   HTG 16-SEP-2004
DEFINITION  Danio rerio clone DKEY-12N16, *** SEQUENCING IN PROGRESS ***, 15
            unordered pieces.
ACCESSION   CR762470
VERSION     CR762470.2  GI:52213994
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 199230)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 16, 2004 this sequence version replaced gi:52000316.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK12N16
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 192957 bases at least Q40
            Consensus quality: 193980 bases at least Q30
            Consensus quality: 194701 bases at least Q20
            Insert size: 197830; sum-of-contigs
            Insert size: 204603; 6.7% error; agarose-fp
            Quality coverage: 9.51x in Q20 bases; sum-of-contigs Quality
            coverage: 9.54x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 15 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    22618: contig of 22618 bp in length
            *    22619    22718: gap of 100 bp
            *    22719    28510: contig of 5792 bp in length
            *    28511    28610: gap of 100 bp
            *    28611    61395: contig of 32785 bp in length
            *    61396    61495: gap of 100 bp
            *    61496    65559: contig of 4064 bp in length
            *    65560    65659: gap of 100 bp
            *    65660    72529: contig of 6870 bp in length
            *    72530    72629: gap of 100 bp
            *    72630    92239: contig of 19610 bp in length
            *    92240    92339: gap of 100 bp
            *    92340    98985: contig of 6646 bp in length
            *    98986    99085: gap of 100 bp
            *    99086   146980: contig of 47895 bp in length
            *   146981   147080: gap of 100 bp
            *   147081   152421: contig of 5341 bp in length
            *   152422   152521: gap of 100 bp
            *   152522   157138: contig of 4617 bp in length
            *   157139   157238: gap of 100 bp
            *   157239   167850: contig of 10612 bp in length
            *   167851   167950: gap of 100 bp
            *   167951   170933: contig of 2983 bp in length
            *   170934   171033: gap of 100 bp
            *   171034   186729: contig of 15696 bp in length
            *   186730   186829: gap of 100 bp
```

```
            *   186830   195213: contig of 8384 bp in length
            *   195214   195313: gap of 100 bp
            *   195314   199230: contig of 3917 bp in length.
FEATURES             Location/Qualifiers
     source          1. .199230
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-12N16"
                     /clone_lib="DanioKey"
     misc_feature    1. .22618
                     /note="assembly_fragment:01219
                     fragment_chain:1"
     misc_feature    22719. .28510
                     /note="assembly_fragment:00267
                     fragment_chain:1"
     misc_feature    28611. .61395
                     /note="assembly_fragment:01636
                     fragment_chain:1"
     misc_feature    61496. .65559
                     /note="assembly_fragment:00116
                     fragment_chain:1"
     misc_feature    65660. .72529
                     /note="assembly_fragment:00202
                     fragment_chain:1"
     misc_feature    72630. .92239
                     /note="assembly_fragment:00921
                     fragment_chain:2"
     misc_feature    92340. .98985
                     /note="assembly_fragment:00329
                     fragment_chain:2"
     misc_feature    99086. .146980
                     /note="assembly_fragment:02193
                     fragment_chain:2"
     misc_feature    147081. .152421
                     /note="assembly_fragment:00157
                     fragment_chain:3"
     misc_feature    152522. .157138
                     /note="assembly_fragment:00407
                     fragment_chain:3"
     misc_feature    157239. .167850
                     /note="assembly_fragment:00593
                     fragment_chain:4"
     misc_feature    167951. .170933
                     /note="assembly_fragment:00070
                     fragment_chain:4"
     misc_feature    171034. .186729
                     /note="assembly_fragment:00727
                     fragment_chain:5"
     misc_feature    186830. .195213
                     /note="assembly_fragment:00491
                     fragment_chain:5"
     misc_feature    195314. .199230
                     /note="assembly_fragment:00081.0"
ORIGIN

  Query Match             7.0%;  Score 349.2;  DB 2;  Length 199230;
  Best Local Similarity   68.5%;  Pred. No. 2.7e-60;
  Matches  511;  Conservative    0;  Mismatches  233;  Indels    2;  Gaps    2;

Qy        181 CCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATA 240
              |||| ||| || |   |||  | |  | |  ||| ||| ||   | |  ||||||||||||
Db      26837 CCAGACAAAAAAAAAAAAAAATTAAATAAATAAATAAATAAATAAATAAATATATATATA 26896

Qy        241 TATATATATATACACACACACACATATATATATACACATATATATACACATATATACA 300
              ||||||||||||| | | | | | | |||||||||||| | ||||||||| | ||||||| |
Db      26897 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 26956

Qy        301 TATATACACATATATACACATATATATACACATACATATGTACACATATATATACACATA 360
              |||||| | ||||||| | ||||||||| | ||| |||| || | |||||||||| | |||
Db      26957 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 27016

Qy        361 TGTATACACATATATACACATATATACACACATATATACACATATATACACACATATATA 420
              | |||| | ||||||| | ||||||| | | ||||||| | ||||||| | | |||||||
Db      27017 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 27076

Qy        421 CACATATATACACATATATACACATATACACATATATACACATATATACATATATACACA 480
              | ||||||| | ||||||| | ||||| | ||||||| | ||||||| ||||||| | |
Db      27077 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 27136

Qy        481 TATATATA-ATATACACACATATATATACACATATATACACACATATATACACATATATA 539
              |||||||| ||||| | | ||||||||| | ||||||| | | ||||||| | |||||||
Db      27137 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 27196

Qy        540 CACATATATATACACATATATACACATATATACATATATACACATATATATACATATATA 599
              | ||||||||| | ||||||| | ||||||| ||||||| | |||||||||| |||||||
Db      27197 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 27256

Qy        600 CACATATATACATATATACACATATATACATATATACACACATATATACACATACATATA 659
              | ||||||| ||||||| | ||||||| ||||||| | | ||||||| | ||| |||||
Db      27257 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 27316

Qy        660 CACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATA 719
              | | | ||| ||||| ||||||||| | ||||||||| |||||||||| ||||||||||
Db      27317 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 27376

Qy        720 TGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGT 779
              |          | |||          | |  ||    |    ||| |  | | | |
Db      27377 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 27436

Qy        780 TTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATA 839
              | | |   | |||    | |  | || |    |     | |||      | |||
Db      27437 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 27496

Qy        840 CATAAATGTGTAATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTAACACTATATATATA 898
              ||| || | ||   | |  ||| | || |  ||  ||   |  ||    ||||||||||
Db      27497 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 27556

Qy        899 TATAGTGAAAGCAGTGTAAACCTAAA 924
              ||||  | |    | || |  || |
Db      27557 TATATATATATATATATATATATA 27582


RESULT 80
AC098061/c
LOCUS       AC098061              247615 bp    DNA     linear   HTG 10-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-115G17, WORKING DRAFT SEQUENCE, 2
            unordered pieces.
ACCESSION   AC098061
VERSION     AC098061.7  GI:30522584
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 247615)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
```

Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

TITLE Direct Submission

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 247615)

AUTHORS Worley,K.C.

TITLE Direct Submission

JOURNAL Submitted (23-OCT-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE 3 (bases 1 to 247615)

AUTHORS Rat Genome Sequencing Consortium.

TITLE Direct Submission

JOURNAL Submitted (10-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT On May 10, 2003 this sequence version replaced gi:23096179.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
Center project name: GGEZ
Center clone name: CH230-115G17
---------- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 219440 bases at least Q40
Consensus quality: 221621 bases at least Q30
Consensus quality: 223241 bases at least Q20
Estimated insert size: 234299; sum-of-contigs estimation
Quality coverage: 8x in Q20 bases; sum-of-contigs estimation

```
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     9721: contig of 9721 bp in length
            *    9722     9821: gap of unknown length
            *    9822   247615: contig of 237794 bp in length.
FEATURES             Location/Qualifiers
     source          1. .247615
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-115G17"
ORIGIN

  Query Match             7.0%;  Score 349.2;  DB 2;  Length 247615;
  Best Local Similarity   67.1%;  Pred. No. 2.7e-60;
  Matches  523;  Conservative    0;  Mismatches  254;  Indels    2;  Gaps    2;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||||| | | | | | |||||||||| | |||||||
Db       9080 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 9021

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db       9020 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 8961

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              ||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db       8960 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 8901

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db       8900 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 8841

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db       8840 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 8781

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db       8780 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 8721

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db       8720 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 8661

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db       8660 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 8601

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||          | |||          | |  ||    |    |||
Db       8600 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 8541

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
              |   | | | |  | | |    | |||    | |   | || |      |      | |
Db       8540 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 8481

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTA 884
              ||      | ||| ||| || | ||   | |  ||| | || |  ||  ||    |  ||
Db       8480 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 8421

Qy        885 ACACTATATATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAA 944
                |||||||||||||||  | ||   | || |   | | ||      ||  || || |
Db       8420 TATATATATATATATATAGANAGAGATATATATATAGATATGAATATATAAAAATGTAGA 8361
```

```
Qy        945 TGTTATAGAAGAATGGCTAAGATGCCAAAGCTCAGTGTATGTGGCAGAGGCATGGTGTA 1003
              | | | || |  || | | | |  || ||   |   |      || || | || | |
Db       8360 GATAAAAAAACAGAGGTTTAAAGGGGAAGGCATGGATCAATGACAAGGGGAAAGGGGGA 8302


RESULT 81
CR759820/c
LOCUS       CR759820              162435 bp    DNA     linear   HTG 10-SEP-2004
DEFINITION  Danio rerio clone DKEY-208C7, *** SEQUENCING IN PROGRESS ***, 7
            unordered pieces.
ACCESSION   CR759820
VERSION     CR759820.2  GI:51987786
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 162435)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 10, 2004 this sequence version replaced gi:51890335.
            -------------- Genome Center.
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK208C7
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 159742 bases at least Q40
            Consensus quality: 160382 bases at least Q30
            Consensus quality: 160932 bases at least Q20
            Insert size: 161835; sum-of-contigs
            Insert size: 166033; 3.0% error; agarose-fp
            Quality coverage: 10.96x in Q20 bases; sum-of-contigs Quality
            coverage: 10.78x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     9773: contig of 9773 bp in length
            *     9774     9873: gap of 100 bp
            *     9874    59608: contig of 49735 bp in length
            *    59609    59708: gap of 100 bp
            *    59709    64998: contig of 5290 bp in length
            *    64999    65098: gap of 100 bp
            *    65099    70537: contig of 5439 bp in length
            *    70538    70637: gap of 100 bp
            *    70638    90121: contig of 19484 bp in length
            *    90122    90221: gap of 100 bp
            *    90222    92310: contig of 2089 bp in length
            *    92311    92410: gap of 100 bp
            *    92411   162435: contig of 70025 bp in length.
FEATURES             Location/Qualifiers
     source          1. .162435
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-208C7"
                     /clone_lib="DanioKey"
     misc_feature    1. .9773
                     /note="assembly_fragment:00141
                     fragment_chain:1"
     misc_feature    9874. .59608
                     /note="assembly_fragment:00514
                     fragment_chain:1"
     misc_feature    59709. .64998
                     /note="assembly_fragment:00049
                     fragment_chain:1"
     misc_feature    65099. .70537
                     /note="assembly_fragment:00086
                     fragment_chain:2"
     misc_feature    70638. .90121
                     /note="assembly_fragment:00272
                     fragment_chain:2"
     misc_feature    90222. .92310
                     /note="assembly_fragment:00018.0"
     misc_feature    92411. .162435
                     /note="assembly_fragment:01362"
ORIGIN

  Query Match             7.0%;  Score 349;  DB 2;  Length 162435;
  Best Local Similarity   68.7%;  Pred. No. 3.1e-60;
  Matches  509;  Conservative    0;  Mismatches  230;  Indels    2;  Gaps    2;

Qy        209 AGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACACATA 268
              | |||||||||||    | | |||||||||||||||||||||||| | | | | | |||
Db        850 AAAAAAAAAAAAAAAAAATATATATATATATATATATATATATATATATATATATATA 791

Qy        269 TATATATACACATATATATACACATATATACATATATACACATATATACACATATATATA 328
              |||||||| | ||||||||| | ||||||| ||||||| | ||||||| | |||||||||
Db        790 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 731

Qy        329 CACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACA 388
              | ||| |||| || | ||||||||| | |||| |||| | ||||||| | ||||||| |
Db        730 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 671

Qy        389 CACATATATACACATATATACACACATATATACACATATATACACATATATACACATATA 448
              | ||||||| | ||||||| | | ||||||| | ||||||| | ||||||| | |||||
Db        670 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 611

Qy        449 CACATATATACACATATATACATATATACACATATATATA-ATATACACACATATATATA 507
              | ||||||| | ||||||| ||||||| | ||||||||| ||||| | | |||||||||
Db        610 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 551

Qy        508 CACATATATACACACATATATACACATATATACACATATATATACACATATATACACATA 567
              | ||||||| | | ||||||| | ||||||| | ||||||||| | ||||||| | |||
Db        550 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 491

Qy        568 TATACATATATACACATATATATACATATATACACATATATACATATATACACATATATA 627
              |||| ||||||| | ||||||||| ||||||| | ||||||| ||||||| | |||||||
Db        490 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 431

Qy        628 CATATATACACACATATATACACATACATATACACACACATAGATATACATATATATACA 687
              ||||||| | | ||||||| | ||| ||||| | | | ||| ||||| |||||||||
Db        430 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 371

Qy        688 CATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGT 747
              |||||||||  |||||||||| ||||||||||          | |||         |
Db        370 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 311

Qy        748 TGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACA 807
              |   ||     |   ||| |   | | |  |  | | |    | |||    | |   | |
Db        310 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 251

Qy        808 TTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA 867
              | |     |    | |||     | ||| ||| || | ||   | |  ||| | ||
Db        250 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 191

Qy        868 -AGGTAGTTAAGAAGGTAACACTATATATATATAGTGAAAGCAGTGTAAACCTAAAGG 926
               |  ||  ||   |  ||    |||||||||||||   | |    | || |  || |
Db        190 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 131

Qy        927 ATGGGCCAAGGATTTAAATGT 947
                |  |   ||||| || |
```

```
Db        130 TATAGAGATATATTTATATAT 110


RESULT 82
CR788258
LOCUS       CR788258              165992 bp    DNA     linear   HTG 25-SEP-2004
DEFINITION  Danio rerio clone DKEY-182D21, *** SEQUENCING IN PROGRESS ***, 10
            unordered pieces.
ACCESSION   CR788258
VERSION     CR788258.3  GI:52694485
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 165992)
  AUTHORS   McLay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 25, 2004 this sequence version replaced gi:52353217.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK182D21
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 162106 bases at least Q40
            Consensus quality: 163025 bases at least Q30
            Consensus quality: 163727 bases at least Q20
            Insert size: 165092; sum-of-contigs
            Insert size: 165748; 0.8% error; agarose-fp
            Quality coverage: 7.00x in Q20 bases; sum-of-contigs Quality
            coverage: 7.22x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    12528: contig of 12528 bp in length
            *    12529    12628: gap of 100 bp
            *    12629    31404: contig of 18776 bp in length
            *    31405    31504: gap of 100 bp
            *    31505    37666: contig of 6162 bp in length
            *    37667    37766: gap of 100 bp
            *    37767    62052: contig of 24286 bp in length
            *    62053    62152: gap of 100 bp
            *    62153    79002: contig of 16850 bp in length
            *    79003    79102: gap of 100 bp
            *    79103   113175: contig of 34073 bp in length
            *   113176   113275: gap of 100 bp
            *   113276   122334: contig of 9059 bp in length
            *   122335   122434: gap of 100 bp
            *   122435   137396: contig of 14962 bp in length
            *   137397   137496: gap of 100 bp
            *   137497   147411: contig of 9915 bp in length
            *   147412   147511: gap of 100 bp
            *   147512   165992: contig of 18481 bp in length.
FEATURES             Location/Qualifiers
     source          1. .165992
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-182D21"
                     /clone_lib="DanioKey"
     misc_feature    1. .12528
                     /note="assembly_fragment:00331
                     fragment_chain:1"
     misc_feature    12629. .31404
                     /note="assembly_fragment:00639
                     fragment_chain:1"
     misc_feature    31505. .37666
                     /note="assembly_fragment:00044
                     fragment_chain:1"
     misc_feature    37767. .62052
                     /note="assembly_fragment:01292
                     fragment_chain:2"
     misc_feature    62153. .79002
                     /note="assembly_fragment:00843
                     fragment_chain:2"
     misc_feature    79103. .113175
                     /note="assembly_fragment:01612
                     fragment_chain:2"
     misc_feature    113276. .122334
                     /note="assembly_fragment:00114
                     fragment_chain:3"
     misc_feature    122435. .137396
                     /note="assembly_fragment:00465
                     fragment_chain:3"
     misc_feature    137497. .147411
                     /note="assembly_fragment:00219"
     misc_feature    147512. .165992
                     /note="assembly_fragment:01055.0"
ORIGIN

  Query Match             7.0%;  Score 349;  DB 2;  Length 165992;
  Best Local Similarity   63.5%;  Pred. No. 3.1e-60;
  Matches  455;  Conservative    0;  Mismatches  261;  Indels    1;  Gaps    1;

Qy        209 AGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACACATA 268
              |||||||||||      | ||||||||||||||||||||||||| | | | | | | |||
Db      11982 AAAAAAAAAAAAAAAAATATATATATATATATATATATATATATATATATATATATATA 12041

Qy        269 TATATATACACATATATATACACATATATACATATATACACATATATACACATATATATA 328
              |||||||| | ||||||||| | ||||||| ||||||| | ||||||| | |||||||||
Db      12042 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 12101

Qy        329 CACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACA 388
              | ||| |||| || | ||||||||| | |||| |||| | ||||||| | ||||||| |
Db      12102 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 12161

Qy        389 CACATATATACACATATATACACACATATATACACATATATACACATATATACACATATA 448
              | ||||||| | ||||||| | | ||||||| | ||||||| | ||||||| | |||||
Db      12162 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 12221

Qy        449 CACATATATACACATATATACATATATACACATATATATA-ATATACACACATATATATA 507
              | ||||||| | ||||||| ||||||| | ||||||||| ||||| | | |||||||||
Db      12222 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 12281

Qy        508 CACATATATACACACATATATACACATATATACACATATATATACACATATATACACATA 567
              | ||||||| | | ||||||| | ||||||| | ||||||||| | ||||||| | |||
Db      12282 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 12341

Qy        568 TATACATATATACACATATATATACATATATACACATATATACATATATACACATATATA 627
              |||| ||||||| | ||||||||| ||||||| | ||||||| ||||||| | |||||||
Db      12342 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 12401

Qy        628 CATATATACACACATATATACACATACATATACACACACATAGATATACATATATATACA 687
              ||||||| | | ||||||| | ||| ||||| | | | ||| ||||| |||||||||
Db      12402 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 12461

Qy        688 CATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGT 747
              |||||||||  |||||||| ||||||||||      |  |||       |
Db      12462 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 12521

Qy        748 TGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACA 807
```

```
            |   ||
Db      12522 TATATATNNN
```

```
  Matches  498;  Conservative    0;  Mismatches  230;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||||| | | | | | | |||||||||||| | |||||||
Db      53249 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 53308

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db      53309 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 53368

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db      53369 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 53428

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db      53429 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 53488

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db      53489 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 53548

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db      53549 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 53608

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db      53609 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 53668

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db      53669 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 53728

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||||          | |||          | |   ||    |   |||
Db      53729 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 53788

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
              |   | | | |  | | |   | |||    | |  | || |     |    | | |
Db      53789 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 53848

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885
              ||       | ||| || ||   ||   | |  ||| | ||   ||  ||    |  ||
Db      53849 TATATATATATATATATATATACATATATATATATATATATAATAAATATATATATAT 53908

Qy        886 CACTATATATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAAT 945
                |||||||||||||   | |  | | |||   ||||  ||      |    | |||
Db      53909 ATATATATATATATATATATATATAAATATAATAATAAATAATATAAATATTTATATAAT 53968

Qy        946 GTTATAGAA 954
              |||   ||
Db      53969 ATTAATTAA 53977


RESULT 84
CR812467
LOCUS       CR812467              326441 bp    DNA     linear   HTG 27-SEP-2004
DEFINITION  Danio rerio clone CH211-134M19, *** SEQUENCING IN PROGRESS ***, 23
            unordered pieces.
ACCESSION   CR812467
VERSION     CR812467.2  GI:52748666
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 326441)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 27, 2004 this sequence version replaced gi:52694423.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zC134M19
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 317403 bases at least Q40
            Consensus quality: 319890 bases at least Q30
            Consensus quality: 321680 bases at least Q20
            Insert size: 324241; sum-of-contigs
            Insert size: 167531; 3.4% error; agarose-fp
            Quality coverage: 6.43x in Q20 bases; sum-of-contigs Quality
            coverage: 12.99x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 23 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     9379: contig of 9379 bp in length
            *     9380     9479: gap of 100 bp
            *     9480    17719: contig of 8240 bp in length
            *    17720    17819: gap of 100 bp
            *    17820    22219: contig of 4400 bp in length
            *    22220    22319: gap of 100 bp
            *    22320    33493: contig of 11174 bp in length
            *    33494    33593: gap of 100 bp
            *    33594    57232: contig of 23639 bp in length
            *    57233    57332: gap of 100 bp
            *    57333    63423: contig of 6091 bp in length
            *    63424    63523: gap of 100 bp
            *    63524    67014: contig of 3491 bp in length
            *    67015    67114: gap of 100 bp
            *    67115   106755: contig of 39641 bp in length
            *   106756   106855: gap of 100 bp
            *   106856   121891: contig of 15036 bp in length
            *   121892   121991: gap of 100 bp
            *   121992   135062: contig of 13071 bp in length
            *   135063   135162: gap of 100 bp
            *   135163   153032: contig of 17870 bp in length
            *   153033   153132: gap of 100 bp
            *   153133   190703: contig of 37571 bp in length
            *   190704   190803: gap of 100 bp
            *   190804   193044: contig of 2241 bp in length
            *   193045   193144: gap of 100 bp
            *   193145   198079: contig of 4935 bp in length
            *   198080   198179: gap of 100 bp
            *   198180   207026: contig of 8847 bp in length
            *   207027   207126: gap of 100 bp
            *   207127   214185: contig of 7059 bp in length
            *   214186   214285: gap of 100 bp
            *   214286   235774: contig of 21489 bp in length
            *   235775   235874: gap of 100 bp
            *   235875   268606: contig of 32732 bp in length
            *   268607   268706: gap of 100 bp
            *   268707   270870: contig of 2164 bp in length
            *   270871   270970: gap of 100 bp
            *   270971   278926: contig of 7956 bp in length
            *   278927   279026: gap of 100 bp
            *   279027   294890: contig of 15864 bp in length
            *   294891   294990: gap of 100 bp
            *   294991   302499: contig of 7509 bp in length
            *   302500   302599: gap of 100 bp
            *   302600   326441: contig of 23842 bp in length.
```

```
FEATURES             Location/Qualifiers
     source          1. .326441
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-134M19"
                     /clone_lib="CHORI-211"
     misc_feature    1. .9379
                     /note="assembly_fragment:00388
                     fragment_chain:1"
     misc_feature    9480. .17719
                     /note="assembly_fragment:00332
                     fragment_chain:1"
     misc_feature    17820. .22219
                     /note="assembly_fragment:00211
                     fragment_chain:1"
     misc_feature    22320. .33493
                     /note="assembly_fragment:00785
                     fragment_chain:1"
     misc_feature    33594. .57232
                     /note="assembly_fragment:01360
                     fragment_chain:1"
     misc_feature    57333. .63423
                     /note="assembly_fragment:00160
                     fragment_chain:1"
     misc_feature    63524. .67014
                     /note="assembly_fragment:00138
                     fragment_chain:1"
     misc_feature    67115. .106755
                     /note="assembly_fragment:02559
                     fragment_chain:2"
     misc_feature    106856. .121891
                     /note="assembly_fragment:00914
                     fragment_chain:2"
     misc_feature    121992. .135062
                     /note="assembly_fragment:00663
                     fragment_chain:2"
     misc_feature    135163. .153032
                     /note="assembly_fragment:00542
                     fragment_chain:2"
     misc_feature    153133. .190703
                     /note="assembly_fragment:01930
                     fragment_chain:3"
     misc_feature    190804. .193044
                     /note="assembly_fragment:00089
                     fragment_chain:3"
     misc_feature    193145. .198079
                     /note="assembly_fragment:00185
                     fragment_chain:3"
     misc_feature    198180. .207026
                     /note="assembly_fragment:00288
                     fragment_chain:4"
     misc_feature    207127. .214185
                     /note="assembly_fragment:00248
                     fragment_chain:4"
     misc_feature    214286. .235774
                     /note="assembly_fragment:01110
                     fragment_chain:5"
     misc_feature    235875. .268606
                     /note="assembly_fragment:02953
                     fragment_chain:5"
     misc_feature    268707. .270870
                     /note="assembly_fragment:00100"
     misc_feature    270971. .278926
                     /note="assembly_fragment:00463"
     misc_feature    279027. .294890
                     /note="assembly_fragment:01641.0"
     misc_feature    294991. .302499
                     /note="assembly_fragment:01641.1"
     misc_feature    302600. .326441
                     /note="assembly_fragment:02243"
ORIGIN
```

```
  Query Match             7.0%;  Score 349;  DB 2;  Length 326441;
  Best Local Similarity   64.1%;  Pred. No. 2.8e-60;
  Matches  467;  Conservative    0;  Mismatches  260;  Indels    2;  Gaps    2;

Qy        198 AATGTCATCTCAGAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACAC 257
              ||| | || |    | | | | |      | ||||||||||||||||||||||||| | |
Db     235247 AATTTTATATATATATATATATATATATATATATATATATATATATATATATATATAT 235306

Qy        258 ACACACACATATATATATACACATATATATACACATATATACATATATACACATATATAC 317
              | | | | |||||||||||| | ||||||||| | ||||||| ||||||| | |||||||
Db     235307 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 235366

Qy        318 ACATATATATACACATACATATGTACACATATATATACACATATGTATACACATATATAC 377
              | ||||||||| | ||| |||| || | ||||||||| | |||| |||| | |||||||
Db     235367 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 235426

Qy        378 ACATATATACACACATATATACACATATATACACACATATATACACATATATACACATAT 437
              | ||||||| | | ||||||| | ||||||| | | ||||||| | ||||||| | ||||
Db     235427 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 235486

Qy        438 ATACACATATACACATATATACACATATATACATATATACACATATATATA-ATATACAC 496
              ||| | ||||| | ||||||| | ||||||| ||||||| | ||||||||| ||||| |
Db     235487 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 235546

Qy        497 ACATATATATACACATATATACACACATATATACACATATATACACATATATATACACAT 556
              | ||||||||| | ||||||| | | ||||||| | ||||||| | ||||||||| | ||
Db     235547 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 235606

Qy        557 ATATACACATATATACATATATACACATATATATACATATATACACATATATACATAT 616
              |||||| | ||||||| ||||||| | ||||||||| ||||||| | ||||||| ||||||
Db     235607 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 235666

Qy        617 ACACATATATACATATATACACACATATATACACATACATATACACACACATAGATATAC 676
              | | ||||||| ||||||| | | ||||||| | ||| ||||| | | | ||| ||||
Db     235667 ATATATATATATATATATATATATATATATATATATATATATATATATATATCTATAT 235726

Qy        677 ATATATATACACATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAA 736
              ||||||||| | ||||||||| |||||||||| ||||||||||
Db     235727 ATATATATATATATATATATATGTATATATATATATATATATATATNNNNNNNNNNNN 235786

Qy        737 GAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTG 796

Db     235787 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 235846

Qy        797 TGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAA 856
                                          |||        | ||| ||| || | || | |
Db     235847 NNNNNNNNNNNNNNNNNNNNNNNNNNNATATATATATATATATATATATATTTAT 235906

Qy        857 ACCTATCTGTA-AGGTAGTTAAGAAGGTAACACTATATATATATATAGTGAAAGCAGTGT 915
              | ||| | || |  ||  ||    |  ||    |||||||||||||   | |    | |
Db     235907 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 235966

Qy        916 AAACCTAAA 924
              | |  || |
Db     235967 ATATATATA 235975
```

```
RESULT 85
CR627490
LOCUS       CR627490              216831 bp    DNA     linear   HTG 09-AUG-2004
DEFINITION  Danio rerio clone DKEY-52H23, WORKING DRAFT SEQUENCE, 15 unordered
            pieces.
ACCESSION   CR627490
VERSION     CR627490.3  GI:51090209
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 216831)
  AUTHORS   Burton,J.
  TITLE     Direct Submission
```

```
  JOURNAL   Submitted (08-AUG-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 9, 2004 this sequence version replaced gi:51036472.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK52H23
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 210790 bases at least Q40
            Consensus quality: 211954 bases at least Q30
            Consensus quality: 213128 bases at least Q20
            Insert size: 215431; sum-of-contigs
            Insert size: 201083; 6.6% error; agarose-fp
            Quality coverage: 9.53x in Q20 bases; sum-of-contigs Quality
            coverage: 10.53x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 15 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    10834: contig of 10834 bp in length
            *    10835    10934: gap of 100 bp
            *    10935    15833: contig of 4899 bp in length
            *    15834    15933: gap of 100 bp
            *    15934    43416: contig of 27483 bp in length
            *    43417    43516: gap of 100 bp
            *    43517    50746: contig of 7230 bp in length
            *    50747    50846: gap of 100 bp
            *    50847    56683: contig of 5837 bp in length
            *    56684    56783: gap of 100 bp
            *    56784    61053: contig of 4270 bp in length
            *    61054    61153: gap of 100 bp
            *    61154    65291: contig of 4138 bp in length
            *    65292    65391: gap of 100 bp
            *    65392    93832: contig of 28441 bp in length
            *    93833    93932: gap of 100 bp
            *    93933   118737: contig of 24805 bp in length
            *   118738   118837: gap of 100 bp
            *   118838   122857: contig of 4020 bp in length
            *   122858   122957: gap of 100 bp
            *   122958   175495: contig of 52538 bp in length
            *   175496   175595: gap of 100 bp
            *   175596   203060: contig of 27465 bp in length
            *   203061   203160: gap of 100 bp
            *   203161   206216: contig of 3056 bp in length
            *   206217   206316: gap of 100 bp
            *   206317   208899: contig of 2583 bp in length
            *   208900   208999: gap of 100 bp
            *   209000   216831: contig of 7832 bp in length.
FEATURES             Location/Qualifiers
     source          1. .216831
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-52H23"
                     /clone_lib="DanioKey"
     misc_feature    1. .10834
                     /note="assembly_fragment:00683
                     fragment_chain:1"
     misc_feature    10935. .15833
                     /note="assembly_fragment:00240
                     fragment_chain:1"
     misc_feature    15934. .43416
                     /note="assembly_fragment:01139
                     fragment_chain:1"
     misc_feature    43517. .50746
                     /note="assembly_fragment:00404
                     fragment_chain:1"
     misc_feature    50847. .56683
                     /note="assembly_fragment:00315
                     fragment_chain:1"
     misc_feature    56784. .61053
                     /note="assembly_fragment:00165
                     fragment_chain:1"
     misc_feature    61154. .65291
                     /note="assembly_fragment:00123
                     fragment_chain:1"
     misc_feature    65392. .93832
                     /note="assembly_fragment:02008
                     fragment_chain:1"
     misc_feature    93933. .118737
                     /note="assembly_fragment:00887
                     fragment_chain:1"
     misc_feature    118838. .122857
                     /note="assembly_fragment:00091
                     fragment_chain:1"
     misc_feature    122958. .175495
                     /note="assembly_fragment:02459
                     fragment_chain:2"
     misc_feature    175596. .203060
                     /note="assembly_fragment:01567
                     fragment_chain:2"
     misc_feature    203161. .206216
                     /note="assembly_fragment:00050"
     misc_feature    206317. .208899
                     /note="assembly_fragment:00069"
     misc_feature    209000. .216831
                     /note="assembly_fragment:00508.0"
ORIGIN

  Query Match             7.0%;  Score 348.8;  DB 2;  Length 216831;
  Best Local Similarity   64.1%;  Pred. No. 3.3e-60;
  Matches  448;  Conservative    0;  Mismatches  250;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db      93324 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 93383

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | |||||||||| | ||| |||| || | |
Db      93384 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 93443

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db      93444 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 93503

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db      93504 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 93563

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db      93564 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 93623

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db      93624 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 93683

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db      93684 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 93743

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||| ||
Db      93744 TATATATATATATATATATATATATATATATATATATATATATANATATATATATATTTA 93803
```

```
Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||||          |  ||
Db      93804 TATATATATATATATATATATATATATNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 93863

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825

Db      93864 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 93923

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885
              | ||| | |||| | |       |  |||      |  ||  ||       ||
Db      93924 NNNNNNNNATATATAAAAATATATTTATTTATATATAATATATATATATATTTTTAT 93983

Qy        886 CACTATATATATATATAGTGAAAGCAGTGTAAACCTAAA 924
              ||||||||||||||| | | |    | || |  || |
Db      93984 ATATATATATATATATATTTATATATATATANANATATA 94022


RESULT 86
CR450804/c
LOCUS       CR450804              220888 bp    DNA     linear   HTG 08-OCT-2004
DEFINITION  Danio rerio clone CH211-275D22, *** SEQUENCING IN PROGRESS ***, 28
            unordered pieces.
ACCESSION   CR450804
VERSION     CR450804.4  GI:54019855
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 220888)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 9, 2004 this sequence version replaced gi:53828284.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zC275D22
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 206726 bases at least Q40
            Consensus quality: 211339 bases at least Q30
            Consensus quality: 214628 bases at least Q20
            Insert size: 218188; sum-of-contigs
            Insert size: 163802; 4.5% error; agarose-fp
            Quality coverage: 7.02x in Q20 bases; sum-of-contigs Quality
            coverage: 11.27x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 28 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    25610: contig of 25610 bp in length
            *   25611    25710: gap of 100 bp
            *   25711    32841: contig of 7131 bp in length
            *   32842    32941: gap of 100 bp
            *   32942    47962: contig of 15021 bp in length
            *   47963    48062: gap of 100 bp
            *   48063    99124: contig of 51062 bp in length
            *   99125    99224: gap of 100 bp
            *   99225   126254: contig of 27030 bp in length
            *  126255   126354: gap of 100 bp
            *  126355   132148: contig of 5794 bp in length
            *  132149   132248: gap of 100 bp
            *  132249   135973: contig of 3725 bp in length
            *  135974   136073: gap of 100 bp
            *  136074   160340: contig of 24267 bp in length
            *  160341   160440: gap of 100 bp
            *  160441   163288: contig of 2848 bp in length
            *  163289   163388: gap of 100 bp
            *  163389   165705: contig of 2317 bp in length
            *  165706   165805: gap of 100 bp
            *  165806   168126: contig of 2321 bp in length
            *  168127   168226: gap of 100 bp
            *  168227   173446: contig of 5220 bp in length
            *  173447   173546: gap of 100 bp
            *  173547   175654: contig of 2108 bp in length
            *  175655   175754: gap of 100 bp
            *  175755   178489: contig of 2735 bp in length
            *  178490   178589: gap of 100 bp
            *  178590   181085: contig of 2496 bp in length
            *  181086   181185: gap of 100 bp
            *  181186   185198: contig of 4013 bp in length
            *  185199   185298: gap of 100 bp
            *  185299   188726: contig of 3428 bp in length
            *  188727   188826: gap of 100 bp
            *  188827   191247: contig of 2421 bp in length
            *  191248   191347: gap of 100 bp
            *  191348   194660: contig of 3313 bp in length
            *  194661   194760: gap of 100 bp
            *  194761   197060: contig of 2300 bp in length
            *  197061   197160: gap of 100 bp
            *  197161   199305: contig of 2145 bp in length
            *  199306   199405: gap of 100 bp
            *  199406   202170: contig of 2765 bp in length
            *  202171   202270: gap of 100 bp
            *  202271   204970: contig of 2700 bp in length
            *  204971   205070: gap of 100 bp
            *  205071   207333: contig of 2263 bp in length
            *  207334   207433: gap of 100 bp
            *  207434   210161: contig of 2728 bp in length
            *  210162   210261: gap of 100 bp
            *  210262   213219: contig of 2958 bp in length
            *  213220   213319: gap of 100 bp
            *  213320   217663: contig of 4344 bp in length
            *  217664   217763: gap of 100 bp
            *  217764   220888: contig of 3125 bp in length.
FEATURES             Location/Qualifiers
     source          1. .220888
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-275D22"
                     /clone_lib="CHORI-211"
     misc_feature    1. .25610
                     /note="assembly_fragment:01114
                     fragment_chain:1
                     clone_end:T7
                     vector_side:left"
     misc_feature    25711. .32841
                     /note="assembly_fragment:00426
                     fragment_chain:1"
     misc_feature    32942. .47962
                     /note="assembly_fragment:00502
                     fragment_chain:1"
     misc_feature    48063. .99124
                     /note="assembly_fragment:02105
                     fragment_chain:1"
     misc_feature    99225. .126254
                     /note="assembly_fragment:01573
                     fragment_chain:1"
     misc_feature    126355. .132148
                     /note="assembly_fragment:00373
                     fragment_chain:1"
     misc_feature    132249. .135973
```

```
                     /note="assembly_fragment:00352
                     fragment_chain:1"
     misc_feature    136074. .160340
                     /note="assembly_fragment:00794
                     fragment_chain:1
                     clone_end:SP6
                     vector_side:right"
     misc_feature    160441. .163288
                     /note="assembly_fragment:00075
                     fragment_chain:2"
     misc_feature    163389. .165705
                     /note="assembly_fragment:00116
                     fragment_chain:2"
     misc_feature    165806. .168126
                     /note="assembly_fragment:00155
                     fragment_chain:3"
     misc_feature    168227. .173446
                     /note="assembly_fragment:00298
                     fragment_chain:3"
     misc_feature    173547. .175654
                     /note="assembly_fragment:00176
                     fragment_chain:4"
     misc_feature    175755. .178489
                     /note="assembly_fragment:00200
                     fragment_chain:4"
     misc_feature    178590. .181085
                     /note="assembly_fragment:00192
                     fragment_chain:5"
     misc_feature    181186. .185198
                     /note="assembly_fragment:00276
                     fragment_chain:5"
     misc_feature    185299. .188726
                     /note="assembly_fragment:00244
                     fragment_chain:6"
     misc_feature    188827. .191247
                     /note="assembly_fragment:00254
                     fragment_chain:6"
     misc_feature    191348. .194660
                     /note="assembly_fragment:00311
                     fragment_chain:7"
     misc_feature    194761. .197060
                     /note="assembly_fragment:00225
                     fragment_chain:7"
     misc_feature    197161. .199305
                     /note="assembly_fragment:00131"
     misc_feature    199406. .202170
                     /note="assembly_fragment:00143"
     misc_feature    202271. .204970
                     /note="assembly_fragment:00162"
     misc_feature    205071. .207333
                     /note="assembly_fragment:00169"
     misc_feature    207434. .210161
                     /note="assembly_fragment:00184"
     misc_feature    210262. .213219
                     /note="assembly_fragment:00265"
     misc_feature    213320. .217663
                     /note="assembly_fragment:00324"
     misc_feature    217764. .220888
                     /note="assembly_fragment:00338"
ORIGIN

  Query Match             7.0%;  Score 348.8;  DB 2;  Length 220888;
  Best Local Similarity   65.1%;  Pred. No. 3.2e-60;
  Matches  456;  Conservative    0;  Mismatches  242;  Indels    2;  Gaps    2;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db      26206 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 26147

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db      26146 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 26087
```

```
Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db      26086 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 26027

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db      26026 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 25967

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | |||||||||| ||||| | | |||||||||| | ||||||| | | |||
Db      25966 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 25907

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACACATATATACACATA 585
              |||| | ||||||| | |||||||||| | ||||||| | ||||||| ||||||| | |||
Db      25906 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 25847

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db      25846 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 25787

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db      25786 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 25727

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||
Db      25726 TATATATATATATAAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 25667

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
                                                                        |
Db      25666 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNATTA 25607

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTA 884
              || ||    | ||| ||| || | ||   | |  ||| | || |  ||  |||  |  ||
Db      25606 TAATATATTATATATATATATTTATATATATATATATATATATATATATATAAATATATA 25547

Qy        885 ACACTATATATATATATAGTGAAAGCAGTGTAAACCTAAA 924
                  |||||||||||||   | |    | || |  || |
Db      25546 TATATATATATATATATATATATATATATATATATATA 25507
```

```
RESULT 87
CR812957/c
LOCUS       CR812957              224175 bp    DNA     linear   HTG 28-SEP-2004
DEFINITION  Danio rerio clone DKEY-41L5, *** SEQUENCING IN PROGRESS ***, 13
            unordered pieces.
ACCESSION   CR812957
VERSION     CR812957.2  GI:52839447
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 224175)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 28, 2004 this sequence version replaced gi:52749045.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK41L5
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 219065 bases at least Q40
            Consensus quality: 220415 bases at least Q30
```

```
           Consensus quality: 221446 bases at least Q20
           Insert size: 222975; sum-of-contigs
           Insert size: 218290; 5.1% error; agarose-fp
           Quality coverage: 6.26x in Q20 bases; sum-of-contigs Quality
           coverage: 6.49x in Q20 bases; agarose-fp
           --------------.
           * NOTE: This is a 'working draft' sequence. It currently
           * consists of 13 contigs. The true order of the pieces
           * is not known and their order in this sequence record is
           * arbitrary. Gaps between the contigs are represented as
           * runs of N, but the exact sizes of the gaps are unknown.
           * This record will be updated with the finished sequence
           * as soon as it is available and the accession number will
           * be preserved.
           *        1     2025: contig of 2025 bp in length
           *     2026     2125: gap of 100 bp
           *     2126    41693: contig of 39568 bp in length
           *    41694    41793: gap of 100 bp
           *    41794    50195: contig of 8402 bp in length
           *    50196    50295: gap of 100 bp
           *    50296    54932: contig of 4637 bp in length
           *    54933    55032: gap of 100 bp
           *    55033    75269: contig of 20237 bp in length
           *    75270    75369: gap of 100 bp
           *    75370   105478: contig of 30109 bp in length
           *   105479   105578: gap of 100 bp
           *   105579   128261: contig of 22683 bp in length
           *   128262   128361: gap of 100 bp
           *   128362   130955: contig of 2594 bp in length
           *   130956   131055: gap of 100 bp
           *   131056   177561: contig of 46506 bp in length
           *   177562   177661: gap of 100 bp
           *   177662   185583: contig of 7922 bp in length
           *   185584   185683: gap of 100 bp
           *   185684   195396: contig of 9713 bp in length
           *   195397   195496: gap of 100 bp
           *   195497   208546: contig of 13050 bp in length
           *   208547   208646: gap of 100 bp
           *   208647   224175: contig of 15529 bp in length.
FEATURES             Location/Qualifiers
     source          1. .224175
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-41L5"
                     /clone_lib="DanioKey"
     misc_feature    1. .2025
                     /note="assembly_fragment:00011
                     fragment_chain:1"
     misc_feature    2126. .41693
                     /note="assembly_fragment:01456
                     fragment_chain:1"
     misc_feature    41794. .50195
                     /note="assembly_fragment:00090
                     fragment_chain:1"
     misc_feature    50296. .54932
                     /note="assembly_fragment:00059
                     fragment_chain:1"
     misc_feature    55033. .75269
                     /note="assembly_fragment:00674
                     fragment_chain:1"
     misc_feature    75370. .105478
                     /note="assembly_fragment:01132
                     fragment_chain:2"
     misc_feature    105579. .128261
                     /note="assembly_fragment:00900
                     fragment_chain:2"
     misc_feature    128362. .130955
                     /note="assembly_fragment:00024
                     fragment_chain:2"
     misc_feature    131056. .177561
                     /note="assembly_fragment:01832
                     fragment_chain:2"
     misc_feature    177662. .185583
                     /note="assembly_fragment:00258
                     fragment_chain:2"
     misc_feature    185684. .195396
                     /note="assembly_fragment:00172
                     fragment_chain:3"
     misc_feature    195497. .208546
                     /note="assembly_fragment:00383
                     fragment_chain:3"
     misc_feature    208647. .224175
                     /note="assembly_fragment:00528"
ORIGIN

  Query Match             7.0%;  Score 348.8;  DB 2;  Length 224175;
  Best Local Similarity   60.6%;  Pred. No. 3.2e-60;
  Matches  507;  Conservative    0;  Mismatches  327;  Indels    2;  Gaps    2;

Qy        191 AGAGCGAAATGTCATCTCAGAAAAAAAAAAAGGCATTTTATATATATATATATATATATA 250
              | | | ||  |||  | ||||||||||||  |  ||||||||||||||||||||||
Db       2838 ATAACACAAAACCATAAAAAAAAAAAAAAAAAAAAAAAAAATATATATATATATATATATA 2779

Qy        251 TACACACACACACACATATATATATACACATATATATACACATATATACATATATACACA 310
              || | | | | | ||||||||||| | ||||||||| | ||||||| ||||||| | |
Db       2778 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 2719

Qy        311 TATATACACATATATATACACATACATATGTACACATATATATACACATATGTATACACA 370
              |||||| | ||||||||| | ||| |||| || | ||||||||| | |||| |||| | |
Db       2718 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 2659

Qy        371 TATATACACATATATACACACATATATACACATATATACACACATATATACACATATATA 430
              |||||| | ||||||| | | ||||||| | ||||||| | | ||||||| | |||||||
Db       2658 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 2599

Qy        431 CACATATATACACATATACACATATATACACATATATACATATATACACATATATATAAT 490
              | ||||||| | ||||| | ||||||| | ||||||| ||||||| | |||||||| ||
Db       2598 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT-AT 2540

Qy        491 ATACACACATATATATACACATATATACACACATATATACACATATATACACATATATAT 550
              ||| | | |||||||||| | ||||||| | | ||||||| | ||||||| | ||||||||
Db       2539 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 2480

Qy        551 ACACATATATACACATATATACATATATACACATATATATACATATATACACATATATAC 610
              | | ||||||| | ||||||| ||||||| | |||||||||| ||||||| | |||||||
Db       2479 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 2420

Qy        611 ATATATACACATATATACATATATACACACATATATACACATACATATACACACACATAG 670
              ||||||| | ||||||| ||||||| | | ||||||| | ||| ||||| | | | |||
Db       2419 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 2360

Qy        671 ATATACATATATATACACATATATATACGTATATATATGTATATATATATGCTCCAGAGT 730
              ||||| ||||||||| | |||||||||  |||||||| ||||||||||           |
Db       2359 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 2300

Qy        731 TCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGT 790
              |||          |  |  ||    |   ||| |   | | |  |  | |    |
Db       2299 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 2240

Qy        791 GTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGT 850
              |||    | |   | || |     |     | |||       | ||| ||| || | |
Db       2239 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 2180

Qy        851 AATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTAACACTATATATATATATAGTGAAAG 909
              |   | |  ||| | || |  ||  ||   |  ||    ||||||||||||||
Db       2179 ATATATATATATATATATATATATATATATATATATATATATATATATATATATNNNNNN 2120

Qy        910 CAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTATAGAAGAATGGCTAAGATGC 969

Db       2119 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2060

Qy        970 CAAAGCTCAGTGTATGTGGCAGAGGCATGGTGTAGGGTGTGTCCAGGTTCATATAT 1025
                                                  | | |  |  | ||||||
Db       2059 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNATATATATATATATATATAT 2004
```

```
RESULT 88
AC094041/c
LOCUS       AC094041              236551 bp    DNA     linear   HTG 10-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-18C6, WORKING DRAFT SEQUENCE, 4
            unordered pieces.
ACCESSION   AC094041
VERSION     AC094041.8  GI:30520885
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 236551)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 236551)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 236551)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 10, 2003 this sequence version replaced gi:24942723.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GFNH
                Center clone name: CH230-18C6
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 222835 bases at least Q40
                Consensus quality: 225635 bases at least Q30
                Consensus quality: 228117 bases at least Q20
                Estimated insert size: 236184; sum-of-contigs estimation
                Quality coverage: 9x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1   230804: contig of 230804 bp in length
            *   230805   230904: gap of unknown length
            *   230905   232373: contig of 1469 bp in length
            *   232374   232473: gap of unknown length
            *   232474   233531: contig of 1058 bp in length
            *   233532   233631: gap of unknown length
            *   233632   236551: contig of 2920 bp in length.
FEATURES             Location/Qualifiers
     source          1. .236551
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-18C6"
ORIGIN

  Query Match             7.0%;  Score 348.8;  DB 2;  Length 236551;
  Best Local Similarity   69.6%;  Pred. No. 3.2e-60;
  Matches  487;  Conservative    0;  Mismatches  212;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              |||||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db     234988 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 234929

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db     234928 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 234869

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
```

```
Db     234868 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 234809

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db     234808 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 234749

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db     234748 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 234689

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db     234688 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 234629

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db     234628 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 234569

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db     234568 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 234509

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||||          | |||          | |   ||    |    |||
Db     234508 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 234449

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
              |   | | | |   | | |    | |||    | |   | || |       |    | |
Db     234448 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 234389

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885
              ||       | ||| ||| || | ||   | |. ||| | ||      |  | | | ||
Db     234388 TATATATATATATATATATATATATATATATATATATATATATATATATATATATGATAT 234329

Qy        886 CACTATATATATATAGTGAAAGCAGTGTAAACCTAAAG 925
                ||||||||||||| | | |||  | | || |  || ||
Db     234328 ATATATATATATATAAAATTAAAAAATTATATATATATAG 234289


RESULT 89
BX322798/c
LOCUS       BX322798              109868 bp    DNA     linear   VRT 24-JUN-2004
DEFINITION  Zebrafish DNA sequence from clone CH211-231F6 in linkage group 24,
            complete sequence.
ACCESSION   BX322798
VERSION     BX322798.6  GI:49203424
KEYWORDS    HTG.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 109868)
  AUTHORS   Heath,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JUN-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jun 24, 2004 this sequence version replaced gi:49065270.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep Clone-derived
            Zebrafish pUC subclones occasionally display inconsistency over the
            length of mononucleotide A/T runs and conserved TA repeats. Where
            this is found the longest good quality representation will be
            submitted.
            Repeat names beginning 'Dr' were identified by the Recon repeat
            discovery system (Zhirong Bao and Sean Eddy, submitted), and those
            beginning 'drr' were identified by Rick Waterman (Stephen Johnson
            lab, WashU).  For further information see
            http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml CH211-231F6
            is from a CHORI-211 BAC library
            VECTOR: pTARBAC2.1.
FEATURES             Location/Qualifiers
     source          1. .109868
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-231F6"
                     /clone_lib="CHORI-211"
ORIGIN

  Query Match             7.0%;  Score 348.6;  DB 5;  Length 109868;
  Best Local Similarity   76.5%;  Pred. No. 4e-60;
  Matches  440;  Conservative    0;  Mismatches  134;  Indels    1;  Gaps    1;

Qy        147 TGCGGGGAGCAAGATCATGCCATTGCACTCCAGCCCAGGCAACAAGAGCGAAATGTCATC 206
              || | |||   ||  ||       ||||||  ||    |   ||   | | |  || | |
Db      78151 TGTGTGGAAACAGTCCAAAAACATGCACTATAGGTGAATGAATGGGTTCTTTATTTGACT 78092

Qy        207 TCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATATACACACACACACACA 266
              | | | ||| | |    || ||||| ||| ||||| ||||||||| | | | | | |||
Db      78091 ATATATATAAATATATATATATTATACATACATATACATATATATATATATATATATACA 78032

Qy        267 TATATATATACACATATATATACACATATATACATATATACACATATATACACATATATA 326
              |||||||||||| ||||||||| | ||||| | ||||| | | |||||||||| ||||||||
Db      78031 TATATATATACATATATATATATATATACATATATACATATATATATACATATATATA 77972

Qy        327 TACACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATA 386
              | |||||||||| |||| |||||||||| |||| |||||| ||||||| | ||||||||
Db      77971 CATACATACATATATACATATATATATACATATATATATACATATATATATATATATATA 77912

Qy        387 CACACATATATACACATATATACACACATATATACACATATATACACATATATACACATA 446
              || ||||||| | | |||||||| | |||||||| | ||||| | ||||| | | |||
Db      77911 CATACATATACATATATATATACATATATATATACATACATATATATATACATATATA 77852

Qy        447 TACACATATATACACATATATACATATATACACATATATATA-ATATACACACATATATA 505
              || ||||| ||| |||||||||| ||| ||| | ||||| ||| ||||||| | |||||||
Db      77851 TATACATACATATACATATATATATACATATATATATACATACATATACATATATATATA 77792

Qy        506 TACACATATATACACACATATATACACATATATACACATATATATACACATATATACACA 565
              | | ||||||||| ||||||| | | |||||||| |||||||| ||||| ||| |||
Db      77791 CATATATATATACATACATATACATATATATATACATATATATATATACATACATATACA 77732

Qy        566 TATATACATATATACACATATATATACATATATACACATATATACATATATACACATATA 625
              |||||| ||| ||||| ||||| ||||||||| | | ||||||| ||| ||||| |||||
Db      77731 TATATATATACATACATATATACATACATATACATATATATATATATACATACATATATA 77672

Qy        626 TACATATATACACACATATATACACATACATATACACACACATAGATATACATATATATA 685
              | ||||||| ||| | ||||| ||| |   |||| | | | ||| ||||| |||||||||
Db      77671 CATATATATATACATACATATATACACATGTATATATATATATATATATATATATATA 77612

Qy        686 CACATATATATACGTATATATATGTATATATATAT 720
              |||||| |||||  ||||||||| ||||||||||
```

```
Db      77611 CACATACATATATATATATATATATATATATAT 77577


RESULT 90
CR392338/c
LOCUS       CR392338              132545 bp    DNA     linear   HTG 26-AUG-2004
DEFINITION  Danio rerio clone DKEY-262M9, *** SEQUENCING IN PROGRESS ***, 9
            unordered pieces.
ACCESSION   CR392338
VERSION     CR392338.3  GI:51571696
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 132545)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-AUG-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 26, 2004 this sequence version replaced gi:51490965.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK262M9
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 130047 bases at least Q40
            Consensus quality: 130566 bases at least Q30
            Consensus quality: 130976 bases at least Q20
            Insert size: 131745; sum-of-contigs
            Insert size: 144213; 1.4% error; agarose-fp
            Quality coverage: 7.43x in Q20 bases; sum-of-contigs Quality
            coverage: 6.98x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2980: contig of 2980 bp in length
            *     2981     3080: gap of 100 bp
            *     3081     7869: contig of 4789 bp in length
            *     7870     7969: gap of 100 bp
            *     7970    25024: contig of 17055 bp in length
            *    25025    25124: gap of 100 bp
            *    25125    34501: contig of 9377 bp in length
            *    34502    34601: gap of 100 bp
            *    34602    42063: contig of 7462 bp in length
            *    42064    42163: gap of 100 bp
            *    42164    45417: contig of 3254 bp in length
            *    45418    45517: gap of 100 bp
            *    45518    62732: contig of 17215 bp in length
            *    62733    62832: gap of 100 bp
            *    62833    89992: contig of 27160 bp in length
            *    89993    90092: gap of 100 bp
            *    90093   132545: contig of 42453 bp in length.
FEATURES            Location/Qualifiers
     source         1. .132545
                    /organism="Danio rerio"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:7955"
                    /clone="DKEY-262M9"
                    /clone_lib="DanioKey"
     misc_feature   1. .2980
                    /note="assembly_fragment:00025
                    fragment_chain:1"
     misc_feature   3081. .7869
                    /note="assembly_fragment:00103
                    fragment_chain:1"
     misc_feature   7970. .25024
                    /note="assembly_fragment:00458
                    fragment_chain:1"
     misc_feature   25125. .34501
                    /note="assembly_fragment:00156
                    fragment_chain:1"
     misc_feature   34602. .42063
                    /note="assembly_fragment:00280
                    fragment_chain:1"
     misc_feature   42164. .45417
                    /note="assembly_fragment:00074
                    fragment_chain:2"
     misc_feature   45518. .62732
                    /note="assembly_fragment:00711
                    fragment_chain:2"
     misc_feature   62833. .89992
                    /note="assembly_fragment:00974
                    fragment_chain:2"
     misc_feature   90093. .132545
                    /note="assembly_fragment:01385
                    fragment_chain:2"
ORIGIN

  Query Match             7.0%;  Score 348.6;  DB 2;  Length 132545;
  Best Local Similarity   62.2%;  Pred. No. 3.8e-60;
  Matches  472;  Conservative    0;  Mismatches  284;  Indels    3;  Gaps    1;

Qy        169 TTGCACTCCAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTT 228
              ||  || |   |  |   ||  | |     ||||| |   ||| |||| || |    | |||
Db      63457 TTTTACACACCCTGAAAAAAGTAAAAACCAATGTAAATACAGCAAAACAATATTTTAATT 63398

Qy        229 TATATATATATATATATATATATACACACACACACACATATATATATACACATATATATA 288
              ||||||||||||||||||||||| | | | | | ||||||||||| | |||||||||
Db      63397 AATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 63338

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |||
Db      63337 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 63278

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              |||||| | |||| |||| | ||| |||| | ||||||| | | ||||||| | |||||||
Db      63277 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 63218

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||||
Db      63217 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 63158

Qy        469 CATATATACACATATATATA---ATATACACACATATATATACACATATATACACACATA 525
              |||||||| | |||||||||    ||||| | | |||||||||| | ||||||| | | |||
Db      63157 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 63098

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | |||||||||| | ||||||| | ||||||| ||||||| | |||
Db      63097 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 63038

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| |||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db      63037 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 62978

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db      62977 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 62918

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||          | |||          | |  ||    |   |||
Db      62917 TATATATATATATATATATATATATATATATATATATATATATATATATATATATAAATA 62858
```

```
Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
              |  ||  |  |  |  |  |  |  |     |
Db      62857 AAATAATATATATATATATATATATNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 62798

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885

Db      62797 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 62738

Qy        886 CACTATATATATATATAGTGAAAGCAGTGTAAACCTAAA 924
              ||||||||||||   | |    | || | ||| |
Db      62737 NNNNNTATATATATATATATATATATATATATCTATA 62699
```

RESULT 91
AC004966
LOCUS AC004966 134084 bp DNA linear PRI 03-OCT-2003
DEFINITION Homo sapiens PAC clone RP5-1107K15 from 7, complete sequence.
ACCESSION AC004966
VERSION AC004966.2 GI:5306307
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 134084)
AUTHORS Sulston,J.E. and Wilson,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
PUBMED 9847074
REFERENCE 2 (bases 1 to 134084)
AUTHORS Smith-Craig,Renee., Strowmatt,C. and Wohldmann,P.
TITLE The sequence of Homo sapiens PAC clone RP5-1107K15
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 134084)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (12-JUN-1998) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 134084)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (01-JUL-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 5 (bases 1 to 134084)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (23-OCT-1999) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 6 (bases 1 to 134084)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-1999) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 7 (bases 1 to 134084)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (26-APR-2003) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 8 (bases 1 to 134084)
AUTHORS Wilson,R.
TITLE Direct Submission
JOURNAL Submitted (03-OCT-2003) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On Jul 1, 1999 this sequence version replaced gi:3907492.
---------- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: sapiens@watson.wustl.edu
---------- Summary Statistics
Center project name: H_DJ1107K15
----------.

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see http://www.nhgri.nih.gov/DIR/GTB/CHR7 , send mailto:egreen@nhgri.nih.gov , or see http://genome.wustl.edu

SOURCE INFORMATION:
This clone was derived from human PAC library RPCI-5, prepared by Pieter de Jong and coworkers at http://www.chori.org using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994).
The library is from one male donor.
The clone may be obtained either from Genome Systems, Inc. (http://www.genomesystems.com) or Research Genetics, Inc. (http://www.resgen.com); or from Pieter de Jong.
VECTOR: pCYPAC2

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP5-942I16, 200 bp overlap.
Actual start of this clone is at base position 197 of RP5-1107K15
actual end is at base position 134084 of RP5-1107K15.

FEATURES	Location/Qualifiers
source	1. .134084
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="7"
	/map="7"
	/clone="RP5-1107K15"
	/clone_lib="RPCI-5"
repeat_region	84. .925
	/rpt_family="L1"
repeat_region	1199. .1495
	/rpt_family="Alu"
repeat_region	1692. .2013
	/rpt_family="Alu"
repeat_region	2023. .2340
	/rpt_family="Alu"
repeat_region	3225. .3438
	/rpt_family="CR1"
repeat_region	3459. .3768
	/rpt_family="Alu"
repeat_region	4040. .4346
	/rpt_family="Alu"
repeat_region	4474. .4594
	/rpt_family="MIR"
repeat_region	4854. .5166
	/rpt_family="Alu"
repeat_region	5172. .5337
	/rpt_family="Alu"
repeat_region	5338. .5387
	/rpt_family="(TTA)n"
repeat_region	5388. .5507
	/rpt_family="Alu"

```
   repeat_region   5523. .5800
                   /rpt_family="Alu"
   repeat_region   6624. .6922
                   /rpt_family="Alu"
   repeat_region   7057. .7353
                   /rpt_family="Alu"
   repeat_region   7360. .7642
                   /rpt_family="Alu"
   repeat_region   8314. .8428
                   /rpt_family="L2"
   repeat_region   8464. .8762
                   /rpt_family="Alu"
   repeat_region   8767. .9078
                   /rpt_family="Alu"
   repeat_region   9642. .9950
                   /rpt_family="Alu"
   repeat_region   9959. .10274
                   /rpt_family="Alu"
   repeat_region   10789. .11096
                   /rpt_family="Alu"
   repeat_region   11354. .11489
                   /rpt_family="Alu"
   repeat_region   11526. .11806
                   /rpt_family="Alu"
   repeat_region   11810. .12113
                   /rpt_family="Alu"
   repeat_region   12388. .12679
                   /rpt_family="Alu"
   repeat_region   13254. .13297
                   /rpt_family="(TG)n"
   repeat_region   13298. .13445
                   /rpt_family="Alu"
   repeat_region   13448. .13774
                   /rpt_family="Alu"
   repeat_region   13776. .14066
                   /rpt_family="Alu"
   repeat_region   14091. .14478
                   /rpt_family="ERVL"
   repeat_region   14550. .14612
                   /rpt_family="Alu"
   repeat_region   14618. .14897
                   /rpt_family="Alu"
   repeat_region   14936. .15029
                   /rpt_family="MER1_type"
   repeat_region   15078. .15244
                   /rpt_family="L2"
   repeat_region   15248. .15469
                   /rpt_family="MaLR"
   repeat_region   15485. .15561
                   /rpt_family="MaLR"
   repeat_region   15700. .15955
                   /rpt_family="Alu"
   repeat_region   15997. .16096
                   /rpt_family="L1"
   repeat_region   16097. .16124
                   /rpt_family="(TTTA)n"
   repeat_region   16135. .16422
                   /rpt_family="Alu"
   repeat_region   16877. .17035
                   /rpt_family="Alu"
   repeat_region   17036. .17345
                   /rpt_family="Alu"
   repeat_region   17530. .17671
                   /rpt_family="MIR"
   repeat_region   18281. .18563
                   /rpt_family="Alu"
   repeat_region   18564. .18609
                   /rpt_family="(CA)n"
   repeat_region   18711. .18888

Query Match             7.0%;  Score 348.6;  DB 9;  Length 134084;
Best Local Similarity   83.3%;  Pred. No. 3.8e-60;
Matches  409;  Conservative    0;  Mismatches   79;  Indels    3;  Gaps    1;

Qy        230 ATATATATATATATATATATACACACACACACATATATATATACACATATATATAC 289
              ||||||||| | ||||||||| | | | ||| | ||||| ||||| | | ||||||||
Db     108308 ATATATATACACATATATATACATATATATACATATATATACATATATATACATATATAC 108367

Qy        290 ACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATAT 349
              |||||||||||| ||||| ||| ||||| ||| ||||||| ||||| ||||  | | ||||
Db     108368 ACATATATACACATATATACACATATATACACATATATATACATATATATACATATATAT 108427

Qy        350 ATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATAC 409
              | ||| | |||  |||| ||||| ||| ||||||||| ||| ||||||||| ||||||||
Db     108428 ACATATATATACATATATACATACATATACATATATATACATATATATACATATATATAC 108487

Qy        410 ACACATATATACACATATATACACATATATACACATATACACATATATACACATATATAC 469
              ||| ||||| | | ||||| ||| ||||| ||| ||||||||||||||||| ||||||||
Db     108488 ACATATATACATATATATACACATATATACACATATATACACATATATACATATATATAC 108547

Qy        470 ATATATACACATATATATA---ATATACACACATATATATACACATATATACACACATAT 526
              ||||||| ||||||||||   |||||| | ||| | ||| | | ||||| | ||| ||
Db     108548 ATATATATACATATATATACATATATACATATATACACATATATACATATATATACACAT 108607

Qy        527 ATACACATATATACACATATATATACACATATATACACATATATACATATATACACATAT 586
              ||| ||||||||| ||| ||||||||| ||||||||| ||||||||| ||||| ||||||
Db     108608 ATATACATATATATACACATATATACATATATATACATATATATACACATATATACATAT 108667

Qy        587 ATATACATATATACACATATATACATATATACACATATATACATATATACACACATATAT 646
              |||||||||||||| |||||||||||||| |||||||||||||||| ||| ||||||
Db     108668 ATATACATATATACATATATATACATATATATACATATATACATATATATACATATATAT 108727

Qy        647 ACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATAT 706
              ||| ||| | ||| | | ||||| ||||||||||||||||| ||||||| || |||||||
Db     108728 ACATATATACATATATATACATATATATACATATATATACATATATATACACATATATAT 108787

Qy        707 ATGTATATATA 717
              |   |||||||
Db     108788 ACACATATATA 108798


RESULT 92
CR735106/c
LOCUS       CR735106              170355 bp    DNA     linear   HTG 19-AUG-2004
DEFINITION  Danio rerio clone CH211-232N15, WORKING DRAFT SEQUENCE, 13
            unordered pieces.
ACCESSION   CR735106
VERSION     CR735106.2  GI:51476923
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 170355)
  AUTHORS   Burton,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-AUG-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 20, 2004 this sequence version replaced gi:51241603.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zC232N15
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 167417 bases at least Q40
            Consensus quality: 168324 bases at least Q30
            Consensus quality: 168699 bases at least Q20
            Insert size: 169155; sum-of-contigs
            Insert size: 178984; 3.2% error; agarose-fp
```

```
            Quality coverage: 6.27x in Q20 bases; sum-of-contigs Quality
            coverage: 6.05x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 13 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    11635: contig of 11635 bp in length
            *    11636    11735: gap of 100 bp
            *    11736    22534: contig of 10799 bp in length
            *    22535    22634: gap of 100 bp
            *    22635    25737: contig of 3103 bp in length
            *    25738    25837: gap of 100 bp
            *    25838    30700: contig of 4863 bp in length
            *    30701    30800: gap of 100 bp
            *    30801    47115: contig of 16315 bp in length
            *    47116    47215: gap of 100 bp
            *    47216    67157: contig of 19942 bp in length
            *    67158    67257: gap of 100 bp
            *    67258    93368: contig of 26111 bp in length
            *    93369    93468: gap of 100 bp
            *    93469   103340: contig of 9872 bp in length
            *   103341   103440: gap of 100 bp
            *   103441   113531: contig of 10091 bp in length
            *   113532   113631: gap of 100 bp
            *   113632   125308: contig of 11677 bp in length
            *   125309   125408: gap of 100 bp
            *   125409   138161: contig of 12753 bp in length
            *   138162   138261: gap of 100 bp
            *   138262   143042: contig of 4781 bp in length
            *   143043   143142: gap of 100 bp
            *   143143   170355: contig of 27213 bp in length.
FEATURES             Location/Qualifiers
     source          1. .170355
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-232N15"
                     /clone_lib="CHORI-211"
     misc_feature    1. .11635
                     /note="assembly_fragment:00706
                     fragment_chain:1"
     misc_feature    11736. .22534
                     /note="assembly_fragment:00576
                     fragment_chain:1"
     misc_feature    22635. .25737
                     /note="assembly_fragment:00028
                     fragment_chain:1"
     misc_feature    25838. .30700
                     /note="assembly_fragment:00056
                     fragment_chain:1"
     misc_feature    30801. .47115
                     /note="assembly_fragment:00862
                     fragment_chain:1"
     misc_feature    47216. .67157
                     /note="assembly_fragment:01028
                     fragment_chain:1"
     misc_feature    67258. .93368
                     /note="assembly_fragment:01537
                     fragment_chain:1"
     misc_feature    93469. .103340
                     /note="assembly_fragment:00351
                     fragment_chain:2"
     misc_feature    103441. .113531
                     /note="assembly_fragment:00156
                     fragment_chain:2"
     misc_feature    113632. .125308
                     /note="assembly_fragment:00462
                     fragment_chain:2"
     misc_feature    125409. .138161
                     /note="assembly_fragment:00246
                     fragment_chain:2"
     misc_feature    138262. .143042
                     /note="assembly_fragment:00100.0"
     misc_feature    143143. .170355
                     /note="assembly_fragment:01234"
ORIGIN

  Query Match             7.0%;  Score 348.6;  DB 2;  Length 170355;
  Best Local Similarity   65.6%;  Pred. No. 3.7e-60;
  Matches  451;  Conservative    0;  Mismatches  234;  Indels    2;  Gaps    2;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db     114143 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 114084

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | |||||||||| | ||| |||| || | |
Db     114083 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 114024

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db     114023 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 113964

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db     113963 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 113904

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db     113903 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 113844

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db     113843 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 113784

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db     113783 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 113724

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | ||| ||||| ||||||||| | ||||||||| |||||||
Db     113723 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 113664

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| |||||||||| |  ||       ||
Db     113663 TATATATATATATATATTGGAGTCAGATGAATNNNNNNNNNNNNNNNNNNNNNNNNNN 113604

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825

Db     113603 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 113544

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTA 884
                       || ||| || | ||   | |  ||| | || |  ||  ||   |  ||
Db     113543 NNNNNNNNNNNNTATATATATATATATATATATATATATATATATATATATATATATA 113484

Qy        885 ACACTATATATATATATAGTGAAAGCA 911
                  |||||||||||||| || | ||
Db     113483 TATATATATATATATATAGGGAGATCA 113457


RESULT 93
CR762394
LOCUS       CR762394               233491 bp    DNA     linear   HTG 13-SEP-2004
DEFINITION  Danio rerio clone DKEY-108C6, *** SEQUENCING IN PROGRESS ***, 11
            unordered pieces.
ACCESSION   CR762394
VERSION     CR762394.3  GI:52000387
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 233491)
  AUTHORS   McLay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 13, 2004 this sequence version replaced gi:51988247.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK108C6
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 229390 bases at least Q40
            Consensus quality: 230228 bases at least Q30
            Consensus quality: 230919 bases at least Q20
            Insert size: 232491; sum-of-contigs
            Insert size: 232003; 3.4% error; agarose-fp
            Quality coverage: 6.99x in Q20 bases; sum-of-contigs Quality
            coverage: 7.05x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2489: contig of 2489 bp in length
            *     2490     2589: gap of 100 bp
            *     2590    44861: contig of 42272 bp in length
            *    44862    44961: gap of 100 bp
            *    44962    83642: contig of 38681 bp in length
            *    83643    83742: gap of 100 bp
            *    83743   120587: contig of 36845 bp in length
            *   120588   120687: gap of 100 bp
            *   120688   129304: contig of 8617 bp in length
            *   129305   129404: gap of 100 bp
            *   129405   132761: contig of 3357 bp in length
            *   132762   132861: gap of 100 bp
            *   132862   147991: contig of 15130 bp in length
            *   147992   148091: gap of 100 bp
            *   148092   152438: contig of 4347 bp in length
            *   152439   152538: gap of 100 bp
            *   152539   199358: contig of 46820 bp in length
            *   199359   199458: gap of 100 bp
            *   199459   209632: contig of 10174 bp in length
            *   209633   209732: gap of 100 bp
            *   209733   233491: contig of 23759 bp in length.
FEATURES             Location/Qualifiers
     source          1. .233491
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-108C6"
                     /clone_lib="DanioKey"
     misc_feature    1. .2489
                     /note="assembly_fragment:00019
                     fragment_chain:1"
     misc_feature    2590. .44861
                     /note="assembly_fragment:01734
                     fragment_chain:1"
     misc_feature    44962. .83642
                     /note="assembly_fragment:01269
                     fragment_chain:1"
     misc_feature    83743. .120587
                     /note="assembly_fragment:00842
                     fragment_chain:1"
     misc_feature    120688. .129304
                     /note="assembly_fragment:00086
                     fragment_chain:1"
     misc_feature    129405. .132761
                     /note="assembly_fragment:00035
                     fragment_chain:1"
     misc_feature    132862. .147991
                     /note="assembly_fragment:00295
                     fragment_chain:1"
     misc_feature    148092. .152438
                     /note="assembly_fragment:00060
                     fragment_chain:2"
     misc_feature    152539. .199358
                     /note="assembly_fragment:02231
                     fragment_chain:2"
     misc_feature    199459. .209632
                     /note="assembly_fragment:00184
                     fragment_chain:2"
     misc_feature    209733. .233491
                     /note="assembly_fragment:00499.0"
ORIGIN

  Query Match             7.0%;  Score 348.6;  DB 2;  Length 233491;
  Best Local Similarity   64.2%;  Pred. No. 3.5e-60;
  Matches  449;  Conservative    0;  Mismatches  249;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db      83133 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 83192

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | |||||||||| | ||| |||| || | |
Db      83193 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 83252

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db      83253 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 83312

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db      83313 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 83372

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db      83373 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 83432

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db      83433 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 83492

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db      83493 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 83552

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db      83553 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 83612

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| |||||||||||          |  |||
Db      83613 TATATATATATATATATATATATATATATANNNNNNNNNNNNNNNNNNNNNNNNNNNNN 83672

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825

Db      83673 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 83732

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885
                        || ||| || |    || |   ||| | ||      || | |   ||
Db      83733 NNNNNNNNNNTTTATATATATATAATTTATATATTATATATATTTATATTTATATTATAT 83792

Qy        886 CACTATATATATATATAGTGAAAGCAGTGTAAACCTAAA 924
```

```
                    | ||||||||||||   | |    | || |  || |
Db      83793 ATATTTATATATATATATATATATATATATATATATA 83831


RESULT 94
CR548629
LOCUS       CR548629              131550 bp    DNA     linear   HTG 06-JUL-2004
DEFINITION  Danio rerio clone DKEY-173G4, WORKING DRAFT SEQUENCE, 12 unordered
            pieces.
ACCESSION   CR548629
VERSION     CR548629.2  GI:50057711
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 131550)
  AUTHORS   Burton,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUL-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 8, 2004 this sequence version replaced gi:49658992.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK173G4
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 127208 bases at least Q40
            Consensus quality: 128066 bases at least Q30
            Consensus quality: 128642 bases at least Q20
            Insert size: 130450; sum-of-contigs
            Insert size: 145332; 1.1% error; agarose-fp
            Quality coverage: 7.95x in Q20 bases; sum-of-contigs Quality
            coverage: 7.21x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 12 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    18381: contig of 18381 bp in length
            *    18382    18481: gap of 100 bp
            *    18482    25435: contig of 6954 bp in length
            *    25436    25535: gap of 100 bp
            *    25536    29512: contig of 3977 bp in length
            *    29513    29612: gap of 100 bp
            *    29613    40983: contig of 11371 bp in length
            *    40984    41083: gap of 100 bp
            *    41084    56393: contig of 15310 bp in length
            *    56394    56493: gap of 100 bp
            *    56494    70939: contig of 14446 bp in length
            *    70940    71039: gap of 100 bp
            *    71040    80278: contig of 9239 bp in length
            *    80279    80378: gap of 100 bp
            *    80379    87641: contig of 7263 bp in length
            *    87642    87741: gap of 100 bp
            *    87742   103997: contig of 16256 bp in length
            *   103998   104097: gap of 100 bp
            *   104098   106761: contig of 2664 bp in length
            *   106762   106861: gap of 100 bp
            *   106862   110734: contig of 3873 bp in length
            *   110735   110834: gap of 100 bp
            *   110835   131550: contig of 20716 bp in length.

FEATURES             Location/Qualifiers
     source          1. .131550
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-173G4"
                     /clone_lib="DanioKey"
     misc_feature    1. .18381
                     /note="assembly_fragment:01412
                     fragment_chain:1"
     misc_feature    18482. .25435
                     /note="assembly_fragment:00193
                     fragment_chain:1"
     misc_feature    25536. .29512
                     /note="assembly_fragment:00013
                     fragment_chain:1"
     misc_feature    29613. .40983
                     /note="assembly_fragment:00264
                     fragment_chain:1"
     misc_feature    41084. .56393
                     /note="assembly_fragment:00930
                     fragment_chain:1"
     misc_feature    56494. .70939
                     /note="assembly_fragment:00480
                     fragment_chain:1"
     misc_feature    71040. .80278
                     /note="assembly_fragment:00367
                     fragment_chain:2"
     misc_feature    80379. .87641
                     /note="assembly_fragment:00063
                     fragment_chain:2"
     misc_feature    87742. .103997
                     /note="assembly_fragment:00696
                     fragment_chain:2"
     misc_feature    104098. .106761
                     /note="assembly_fragment:00032"
     misc_feature    106862. .110734
                     /note="assembly_fragment:00122.0"
     misc_feature    110835. .131550
                     /note="assembly_fragment:01164"
ORIGIN

  Query Match             7.0%;  Score 348.4;  DB 2;  Length 131550;
  Best Local Similarity   69.5%;  Pred. No. 4.2e-60;
  Matches  500;  Conservative    0;  Mismatches  217;  Indels    2;  Gaps    2;

Qy        208 CAGAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACACAT 267
              || |||    ||  ||      ||||||||||||||||||||||||||||||||||||
Db      86675 CAAAAAGGCAATAAATACATTTATATATATATATATATATATACACACACACACACAC 86734

Qy        268 ATATATATACACATATATATACACATATATACATATATACACATATATACACATATATAT 327
              | | | |||| |||||||| | ||||||| ||||||| | ||||||| | ||||||||
Db      86735 ACACACATACATATATATATATATATATATATATATATATATATATATATATATATAT 86794

Qy        328 ACACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATAC 387
              | | ||| |||| || | |||||||| | |||| |||| | ||||||| | |||||||
Db      86795 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 86854

Qy        388 ACACATATATACACATATATACACACATATATACACATATATACACATATATACACATAT 447
              | | ||||||| | ||||||| | | ||||||| | ||||||| | ||||||| | ||||
Db      86855 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 86914

Qy        448 ACACATATATACACATATATACATATATACACATATATATA-ATATACACACATATATAT 506
              | | ||||||| | ||||||| ||||||| | ||||||||| ||||| | | ||||||||
Db      86915 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 86974

Qy        507 ACACATATATACACACATATATACACATATATACACATATATATACACATATATACACAT 566
              | | ||||||| | | ||||||| | ||||||| | ||||||||| | ||||||| | ||
Db      86975 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 87034

Qy        567 ATATACATATATACACATATATATACATATATACACATATATACATATATACACATATAT 626
              ||||| ||||||| | ||||||||| ||||||| | ||||||| ||||||| | ||||||
Db      87035 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 87094
```

```
Qy        627 ACATATATACACACATATATACACATACATATACACACACATAGATATACATATATATAC 686
              | ||||||| | | ||||||| | ||| ||||| | | | ||| ||||| |||||||||
Db      87095 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 87154

Qy        687 ACATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAG 746
              | ||||||||| |||||||| ||||||||||          | |||         |
Db      87155 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 87214

Qy        747 TTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGAC 806
              |   ||     |    ||| |   | | | |   | | |     | |||     | |   |
Db      87215 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 87274

Qy        807 ATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGT 866
              || |      |       | |||       | ||| ||| || | ||   | |  ||| | |
Db      87275 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 87334

Qy        867 A-AGGTAGTTAAGAAGGTAACACTATATATATATATAGTGAAAGCAGTGTAAACCTAAA 924
              | |  ||  ||   |  ||    |||||||||||||   | |    | || |  || |
Db      87335 ATATATATATATATATATATATATATATATATATATATATATATATATATATATNTATA 87393


RESULT 95
CR388149/c
LOCUS       CR388149              158085 bp    DNA     linear   HTG 19-AUG-2004
DEFINITION  Danio rerio clone DKEY-223J15, WORKING DRAFT SEQUENCE, 6 unordered
            pieces.
ACCESSION   CR388149
VERSION     CR388149.3  GI:51470665
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 158085)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-AUG-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 20, 2004 this sequence version replaced gi:46408652.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK223J15
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 156562 bases at least Q40
            Consensus quality: 156889 bases at least Q30
            Consensus quality: 157151 bases at least Q20
            Insert size: 157585; sum-of-contigs
            Insert size: 159679; 5.7% error; agarose-fp
            Quality coverage: 8.57x in Q20 bases; sum-of-contigs Quality
            coverage: 8.73x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    64210: contig of 64210 bp in length
            *    64211    64310: gap of 100 bp
            *    64311    82166: contig of 17856 bp in length
            *    82167    82266: gap of 100 bp
            *    82267    87201: contig of 4935 bp in length
            *    87202    87301: gap of 100 bp
            *    87302   120999: contig of 33698 bp in length
            *   121000   121099: gap of 100 bp
            *   121100   131190: contig of 10091 bp in length
            *   131191   131290: gap of 100 bp
            *   131291   158085: contig of 26795 bp in length.
FEATURES             Location/Qualifiers
     source          1. .158085
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-223J15"
                     /clone_lib="DanioKey"
     misc_feature    1. .64210
                     /note="assembly_fragment:01629
                     fragment_chain:1"
     misc_feature    64311. .82166
                     /note="assembly_fragment:00258
                     fragment_chain:1"
     misc_feature    82267. .87201
                     /note="assembly_fragment:00044
                     fragment_chain:1"
     misc_feature    87302. .120999
                     /note="assembly_fragment:00990
                     fragment_chain:1"
     misc_feature    121100. .131190
                     /note="assembly_fragment:00092
                     fragment_chain:1"
     misc_feature    131291. .158085
                     /note="assembly_fragment:00487.0"
ORIGIN

  Query Match             7.0%;  Score 348.4;  DB 2;  Length 158085;
  Best Local Similarity   67.9%;  Pred. No. 4.1e-60;
  Matches  515;  Conservative    0;  Mismatches  241;  Indels    2;  Gaps    2;

Qy        227 TTTATATATATATATATATATATATACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||||| | | | | | |||||||||| | |||||||
Db      65123 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 65064

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db      65063 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 65004

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db      65003 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 64944

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db      64943 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 64884

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db      64883 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 64824

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db      64823 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 64764

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db      64763 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 64704

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| |||||||||| | |||||||||  ||||||
Db      64703 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 64644

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||      | |||       | |   ||    |   |||
Db      64643 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 64584

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
```

```
Db      64583 TATAAAAATATATATATATATATATATATATATATATATATATATATATATATATATATA 64524

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885

Db      64523 TATATATATATATATATATATATATATATATATATTTATATATAATAAATAAATAAAAAA 64464

Qy        886 CACTATATATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAAT 945

Db      64463 ATATATATATATATATTATTTTATTA-TATATATATAAAAAAAAATAAAAATATATAATT 64405

Qy        946 GTTATAGAAGAATGGCTAAGATGCCAAAGCTCAGTGTA 983

Db      64404 TTAAAATATAAATAATTAAAAAAATATATTTAATTATA 64367


RESULT 96
CR774185/c
LOCUS       CR774185              167023 bp    DNA     linear   HTG 01-OCT-2004
DEFINITION  Danio rerio clone CH211-237H18, WORKING DRAFT SEQUENCE, 11
            unordered pieces.
ACCESSION   CR774185
VERSION     CR774185.2  GI:52346261
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 167023)
  AUTHORS   McLaren,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 19, 2004 this sequence version replaced gi:52213970.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zC237H18
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 164787 bases at least Q40
            Consensus quality: 165392 bases at least Q30
            Consensus quality: 165717 bases at least Q20
            Insert size: 166023; sum-of-contigs
            Insert size: 188931; 10.3% error; agarose-fp
            Quality coverage: 9.96x in Q20 bases; sum-of-contigs Quality
            coverage: 8.81x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    21904: contig of 21904 bp in length
            *    21905    22004: gap of 100 bp
            *    22005    25287: contig of 3283 bp in length
            *    25288    25387: gap of 100 bp
            *    25388    69373: contig of 43986 bp in length
            *    69374    69473: gap of 100 bp
            *    69474    71725: contig of 2252 bp in length
            *    71726    71825: gap of 100 bp
            *    71826    74171: contig of 2346 bp in length
            *    74172    74271: gap of 100 bp
            *    74272    81210: contig of 6939 bp in length
            *    81211    81310: gap of 100 bp
            *    81311    91830: contig of 10520 bp in length
            *    91831    91930: gap of 100 bp
            *    91931    99897: contig of 7967 bp in length
            *    99898    99997: gap of 100 bp
            *    99998   111895: contig of 11898 bp in length
            *   111896   111995: gap of 100 bp
            *   111996   149605: contig of 37610 bp in length
            *   149606   149705: gap of 100 bp
            *   149706   167023: contig of 17318 bp in length.
FEATURES             Location/Qualifiers
     source          1. .167023
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-237H18"
                     /clone_lib="CHORI-211"
     misc_feature    1. .21904
                     /note="assembly_fragment:00892
                     clone_end:SP6
                     vector_side:left"
     misc_feature    22005. .25287
                     /note="assembly_fragment:00055
                     fragment_chain:1"
     misc_feature    25388. .69373
                     /note="assembly_fragment:01905
                     fragment_chain:1"
     misc_feature    69474. .71725
                     /note="assembly_fragment:00023"
     misc_feature    71826. .74171
                     /note="assembly_fragment:00031"
     misc_feature    74272. .81210
                     /note="assembly_fragment:00083
                     fragment_chain:2"
     misc_feature    81311. .91830
                     /note="assembly_fragment:00416
                     fragment_chain:2"
     misc_feature    91931. .99897
                     /note="assembly_fragment:00157
                     fragment_chain:2"
     misc_feature    99998. .111895
                     /note="assembly_fragment:00260
                     fragment_chain:2"
     misc_feature    111996. .149605
                     /note="assembly_fragment:01303
                     fragment_chain:2"
     misc_feature    149706. .167023
                     /note="assembly_fragment:00585
                     fragment_chain:2
                     clone_end:T7
                     vector_side:right"
ORIGIN

  Query Match             7.0%;  Score 348.4;  DB 2;  Length 167023;
  Best Local Similarity   64.6%;  Pred. No. 4.1e-60;
  Matches  461;  Conservative    0;  Mismatches  251;  Indels    2;  Gaps    2;

Qy        212 AAAAAAAAAAGGCATTTTATATATATATATATATATATATACACACACACACACATATAT 271

Db     100501 AATATTAAAAAAAATTATATATATATATATATATATATATATATATATATATATATATAT 100442

Qy        272 ATATACACATATATATACACATATATACATATATACACATATATACACATATATATACAC 331

Db     100441 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 100382

Qy        332 ATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACACAC 391

Db     100381 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 100322

Qy        392 ATATATACACATATATACACACATATATACACATATATACACATATATACACATATACAC 451

Db     100321 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 100262
```

```
Qy        452 ATATATACACATATATACATATATACACATATATATAATATACACACATATATATACACA 511
              ||||||| | ||||||| ||||||| | |||||||| ||||| | | ||||||||| | |
Db     100261 ATATATATATATATATATATATATATATATATAT-ATATATATATATATATATATATA 100203

Qy        512 TATATACACACATATATACACATATATACACATATATATACACATATATACACATATATA 571
              |||||| | | ||||||| | ||||||| | ||||||||| | ||||||| | |||||||
Db     100202 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 100143

Qy        572 CATATATACACATATATATACATATATACACATATATACATATATACACATATATACATA 631
              ||||||| | |||||||||| ||||||| | ||||||| ||||||| | ||||||| |||
Db     100142 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 100083

Qy        632 TATACACACATATATACACATACATATACACACACATAGATATACATATATATACACATA 691
              |||| | | ||||||| | ||| ||||| | | | ||| ||||| ||||||||| | |||
Db     100082 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 100023

Qy        692 TATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGAT 751
              ||||||  ||||||||| |||||||
Db     100022 TATATATATATATATATATATATATNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 99963

Qy        752 TACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTC 811

Db      99962 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 99903

Qy        812 AGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA-AGG 870
                  | |||       ||||| ||| || | ||   | |  ||| | || |
Db      99902 NNNNNATATATATATATATATAAATATATATATATATATATATATATATATATATA 99843

Qy        871 TAGTTAAGAAGGTAACACTATATATATATAGTGAAAGCAGTGTAAACCTAAA 924
              ||  |||  |  ||    ||||||||||||||    | |    | || |  || |
Db      99842 TATATAAATATATATATATATATATATATATATATATATATATATATATATA 99789


RESULT 97
CR788285
LOCUS       CR788285              169053 bp    DNA     linear   HTG 03-OCT-2004
DEFINITION  Danio rerio clone DKEYP-99C4, *** SEQUENCING IN PROGRESS ***, 5
            unordered pieces.
ACCESSION   CR788285
VERSION     CR788285.4  GI:53755972
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 169053)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 3, 2004 this sequence version replaced gi:53145965.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zKp99C4
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 167822 bases at least Q40
            Consensus quality: 168197 bases at least Q30
            Consensus quality: 168423 bases at least Q20
            Insert size: 168653; sum-of-contigs
            Insert size: 171627; 2.8% error; agarose-fp
            Quality coverage: 12.96x in Q20 bases; sum-of-contigs Quality
            coverage: 12.74x in Q20 bases; agarose-fp
            --------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     5600: contig of 5600 bp in length
            *     5601     5700: gap of 100 bp
            *     5701    59937: contig of 54237 bp in length
            *    59938    60037: gap of 100 bp
            *    60038    78734: contig of 18697 bp in length
            *    78735    78834: gap of 100 bp
            *    78835   100059: contig of 21225 bp in length
            *   100060   100159: gap of 100 bp
            *   100160   169053: contig of 68894 bp in length.
FEATURES             Location/Qualifiers
     source          1. .169053
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEYP-99C4"
                     /clone_lib="DanioKeypilot"
     misc_feature    1. .5600
                     /note="assembly_fragment:00020
                     fragment_chain:1"
     misc_feature    5701. .59937
                     /note="assembly_fragment:00981
                     fragment_chain:1"
     misc_feature    60038. .78734
                     /note="assembly_fragment:00116
                     fragment_chain:2"
     misc_feature    78835. .100059
                     /note="assembly_fragment:00523
                     fragment_chain:2"
     misc_feature    100160. .169053
                     /note="assembly_fragment:02105.0"
ORIGIN

  Query Match             7.0%;  Score 348.4;  DB 2;  Length 169053;
  Best Local Similarity   64.5%;  Pred. No. 4.1e-60;
  Matches  463;  Conservative    0;  Mismatches  251;  Indels    4;  Gaps    2;

Qy        227 TTTATATATATATATATATATATATACACACACACACATATATATATACACATATATA 286
              |||||||||||||||||||||||||| | | | | | |||||||||| | |||||||
Db       5108 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 5167

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || ||||||||| || ||||| | ||||||| | ||||||||| | ||| |||| || | |
Db       5168 TATACATATATATATGTATATATATATATATATATATATATATATATATATATATATA 5227

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              ||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db       5228 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 5287

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db       5288 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 5347

Qy        467 TACATATATACACATATATATAATATACACACATATATATACACATATATACACACATAT 526
              || ||||||| | |||||||| ||||| | | ||||||||| | ||||||| | | ||||
Db       5348 TATATATATATATATATATAT-ATATATATATATATATATATATATATATATATATAT 5406

Qy        527 ATACACATATATACACATATATATACACATATATACACATATATACACATATATACACATAT 586
              ||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | ||||
Db       5407 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 5466

Qy        587 ATATACATATATACACATATATACATATATACACATATATACATATATACACACATATAT 646
              ||||| |||||||||| |||||||| ||||||| | ||||||| ||||||| | | ||||||
Db       5467 ATATATATATATACATATATATATATATATATATATATATATATATATATATATATAT 5526

Qy        647 ACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATAT 706
              | | ||| ||||| | | | ||| ||||| |||||||||| | |||||||||  |||||||
Db       5527 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 5586
```

```
Qy        707 ATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAG 766
              ||||||||||||||||
Db       5587 ATGTATATATATATNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 5646

Qy        767 AGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAAT 826
                                                                      |||
Db       5647 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNAAAAAA 5706

Qy        827 AGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAAC 886
              |       | ||| ||| || | ||   | |  ||| | ||   ||  ||   |  ||
Db       5707 ATATATATATATATATATATATATATATATATATATATA---TATATATATATATATA 5763

Qy        887 ACTATATATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAA 944
                |||||||||  | |   ||| || || |  |     |||   |   |   ||||||
Db       5764 TATATATATATTAAAAAATAAAACATTGCAGTCTGTTGGGAGATGTGCATCGTTTAAA 5821


RESULT 98
CR749743/c
LOCUS       CR749743              171424 bp    DNA     linear   HTG 10-OCT-2004
DEFINITION  Danio rerio clone DKEY-97M3, WORKING DRAFT SEQUENCE, 9 unordered
            pieces.
ACCESSION   CR749743
VERSION     CR749743.4  GI:54021791
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 171424)
  AUTHORS   McLaren,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 10, 2004 this sequence version replaced gi:52213800.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK97M3
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 165755 bases at least Q40
            Consensus quality: 166566 bases at least Q30
            Consensus quality: 167672 bases at least Q20
            Insert size: 170624; sum-of-contigs
            Insert size: 189327; 3.5% error; agarose-fp
            Quality coverage: 8.39x in Q20 bases; sum-of-contigs Quality
            coverage: 7.61x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    13261: contig of 13261 bp in length
            *    13262    13361: gap of 100 bp
            *    13362    19777: contig of 6416 bp in length
            *    19778    19877: gap of 100 bp
            *    19878    33915: contig of 14038 bp in length
            *    33916    34015: gap of 100 bp
            *    34016    61771: contig of 27756 bp in length
            *    61772    61871: gap of 100 bp
            *    61872    94572: contig of 32701 bp in length
            *    94573    94672: gap of 100 bp
            *    94673   131525: contig of 36853 bp in length
            *   131526   131625: gap of 100 bp
            *   131626   133888: contig of 2263 bp in length
            *   133889   133988: gap of 100 bp
            *   133989   157725: contig of 23737 bp in length
            *   157726   157825: gap of 100 bp
            *   157826   171424: contig of 13599 bp in length.
FEATURES             Location/Qualifiers
     source          1. .171424
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-97M3"
                     /clone_lib="DanioKey"
     misc_feature    1. .13261
                     /note="assembly_fragment:00444
                     fragment_chain:1"
     misc_feature    13362. .19777
                     /note="assembly_fragment:00033
                     fragment_chain:1"
     misc_feature    19878. .33915
                     /note="assembly_fragment:00098
                     fragment_chain:1"
     misc_feature    34016. .61771
                     /note="assembly_fragment:00628
                     fragment_chain:2"
     misc_feature    61872. .94572
                     /note="assembly_fragment:01012
                     fragment_chain:2"
     misc_feature    94673. .131525
                     /note="assembly_fragment:01459
                     fragment_chain:2"
     misc_feature    131626. .133888
                     /note="assembly_fragment:00008
                     fragment_chain:3"
     misc_feature    133989. .157725
                     /note="assembly_fragment:01969
                     fragment_chain:3"
     misc_feature    157826. .171424
                     /note="assembly_fragment:00264.0"
ORIGIN

  Query Match             7.0%;  Score 348.4;  DB 2;  Length 171424;
  Best Local Similarity   67.6%;  Pred. No. 4.1e-60;
  Matches  495;  Conservative    0;  Mismatches  236;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              |||||||||||||||||||||||||| | | | | | | |||||||||| | |||||||
Db     134802 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 134743

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db     134742 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 134683

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db     134682 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 134623

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db     134622 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 134563

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db     134562 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 134503

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db     134502 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 134443

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | |||||||| ||||||| | ||||||| ||||||| | | |||||
```

```
Db     134442 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 134383

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db     134382 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 134323

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||          |  |||          |  |   ||    |   |||
Db     134322 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 134263

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
              |   | | | |  | | |    | |||    | |   | || |     | |   | |
Db     134262 TATATATATATATATATATATATATATATATATATATATATATATATATATAAATATATA 134203

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885
              ||  |    | ||| ||| || | ||   |     ||     |   |   ||  |  ||
Db     134202 TATAAATATATATATATATATATTTATATATTTTAATAATATAANNAAAAAAATATATAN 134143

Qy        886 CACTATATATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAAT 945
              |||||||||||||| | |||    | || |  |     ||     ||  |   | |
Db     134142 ATATATATATATATATATTTAAATATATATATATTTTTTTNATNTTAAAANNANAAANAN 134083

Qy        946 GTTATAGAAGAA 957
                | | || ||
Db     134082 AAAAAATAATAA 134071


RESULT 99
CR759841
LOCUS       CR759841              178975 bp    DNA     linear   HTG 08-SEP-2004
DEFINITION  Danio rerio clone DKEY-192G7, *** SEQUENCING IN PROGRESS ***, 6
            unordered pieces.
ACCESSION   CR759841
VERSION     CR759841.2  GI:51965207
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 178975)
  AUTHORS   Burton,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 9, 2004 this sequence version replaced gi:51890356.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK192G7
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 176373 bases at least Q40
            Consensus quality: 176964 bases at least Q30
            Consensus quality: 177367 bases at least Q20
            Insert size: 178475; sum-of-contigs
            Insert size: 178296; 1.0% error; agarose-fp
            Quality coverage: 11.15x in Q20 bases; sum-of-contigs Quality
            coverage: 11.48x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     3686: contig of 3686 bp in length
            *     3687     3786: gap of 100 bp
            *     3787    62427: contig of 58641 bp in length
            *    62428    62527: gap of 100 bp
            *    62528    84899: contig of 22372 bp in length
            *    84900    84999: gap of 100 bp
            *    85000   136604: contig of 51605 bp in length
            *   136605   136704: gap of 100 bp
            *   136705   158073: contig of 21369 bp in length
            *   158074   158173: gap of 100 bp
            *   158174   178975: contig of 20802 bp in length.
FEATURES             Location/Qualifiers
     source          1. .178975
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-192G7"
                     /clone_lib="DanioKey"
     misc_feature    1. .3686
                     /note="assembly_fragment:00051
                     fragment_chain:1"
     misc_feature    3787. .62427
                     /note="assembly_fragment:01908
                     fragment_chain:1"
     misc_feature    62528. .84899
                     /note="assembly_fragment:00400
                     fragment_chain:1"
     misc_feature    85000. .136604
                     /note="assembly_fragment:01093
                     fragment_chain:2"
     misc_feature    136705. .158073
                     /note="assembly_fragment:00722
                     fragment_chain:2"
     misc_feature    158174. .178975
                     /note="assembly_fragment:00096
                     fragment_chain:2"
ORIGIN

  Query Match             7.0%;  Score 348.4;  DB 2;  Length 178975;
  Best Local Similarity   60.4%;  Pred. No. 4e-60;
  Matches  492;  Conservative    0;  Mismatches  321;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACATATATATATACACATATATA 286
              | |||||||||||||||||||||||| | | | | | |||||||||| | |||||||
Db     135912 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 135971

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| |||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db     135972 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 136031

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db     136032 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 136091

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db     136092 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 136151

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db     136152 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 136211

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | |||||||| | ||||||| | ||||||| ||||||| | |||
Db     136212 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 136271

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db     136272 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 136331

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db     136332 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 136391
```

```
Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||||          | |||          | |   ||     |    |||
Db     136392 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 136451

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
              |   | | | |   | | |     | |||     | |   | || |       |     | |
Db     136452 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 136511

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885
              ||        | ||| ||| || | ||   | |  ||| | ||       |  |  |  ||
Db     136512 TATATATATATATATATATATATATATATATATATATATATATATATATATAATATATAT 136571

Qy        886 CACTATATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAAT 945
                ||||||||||||||   | |     | || |
Db     136572 ATATATATATATATATATATATATATATATANNNNNNNNNNNNNNNNNNNNNNNNNN 136631

Qy        946 GTTATAGAAGAATGGCTAAGATGCCAAAGCTCAGTGTATGTGGCAGAGGCATGGTGTAGG 1005

Db     136632 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 136691

Qy       1006 GTGTGTCCAGGTTCATATATTGCATTAAGTGTGA 1039
                  | | ||  ||||| | | |
Db     136692 NNNNNNNNNNNNNTTTTTTTTTATTAATTATTA 136725
```

RESULT 100
CR792414
LOCUS CR792414 179213 bp DNA linear HTG 25-SEP-2004
DEFINITION Danio rerio clone CH211-95O10, *** SEQUENCING IN PROGRESS ***, 9 unordered pieces.
ACCESSION CR792414
VERSION CR792414.2 GI:52694443
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 179213)
AUTHORS Sims,S.
TITLE Direct Submission
JOURNAL Submitted (24-SEP-2004) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT On Sep 25, 2004 this sequence version replaced gi:52626621.
---------- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfish-help@sanger.ac.uk
---------- Project Information
Center project name: zC95O10
---------- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 176437 bases at least Q40
Consensus quality: 176944 bases at least Q30
Consensus quality: 177295 bases at least Q20
Insert size: 178413; sum-of-contigs
Insert size: 195379; 9.0% error; agarose-fp
Quality coverage: 8.87x in Q20 bases; sum-of-contigs Quality coverage: 8.31x in Q20 bases; agarose-fp
----------.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 4542: contig of 4542 bp in length
* 4543 4642: gap of 100 bp
* 4643 41427: contig of 36785 bp in length
* 41428 41527: gap of 100 bp
* 41528 50062: contig of 8535 bp in length
* 50063 50162: gap of 100 bp
* 50163 59923: contig of 9761 bp in length
* 59924 60023: gap of 100 bp
* 60024 103384: contig of 43361 bp in length
* 103385 103484: gap of 100 bp
* 103485 121057: contig of 17573 bp in length
* 121058 121157: gap of 100 bp
* 121158 148007: contig of 26850 bp in length
* 148008 148107: gap of 100 bp
* 148108 169047: contig of 20940 bp in length
* 169048 169147: gap of 100 bp
* 169148 179213: contig of 10066 bp in length.

FEATURES Location/Qualifiers
source 1. .179213
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="CH211-95O10"
/clone_lib="CHORI-211"
misc_feature 1. .4542
/note="assembly_fragment:00045
fragment_chain:1
clone_end:T7
vector_side:left"
misc_feature 4643. .41427
/note="assembly_fragment:01381
fragment_chain:1"
misc_feature 41528. .50062
/note="assembly_fragment:00108
fragment_chain:1"
misc_feature 50163. .59923
/note="assembly_fragment:00205
fragment_chain:1"
misc_feature 60024. .103384
/note="assembly_fragment:01959"
misc_feature 103485. .121057
/note="assembly_fragment:00463
fragment_chain:2"
misc_feature 121158. .148007
/note="assembly_fragment:00984
fragment_chain:2"
misc_feature 148108. .169047
/note="assembly_fragment:00719
fragment_chain:2"
misc_feature 169148. .179213
/note="assembly_fragment:00316
fragment_chain:2
clone_end:SP6
vector_side:right"

ORIGIN

Query Match 7.0%; Score 348.4; DB 2; Length 179213;
Best Local Similarity 66.8%; Pred. No. 4e-60;
Matches 524; Conservative 0; Mismatches 258; Indels 2; Gaps 2;

```
Qy        191 AGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATA 250
              |||  |  || ||  |   ||  ||||||  |   |  ||||||||||||||||||||||
Db     120224 AGACAGGCATTTCGACAGTGATTAAAAAACAACAACATTATATATATATATATATATATA 120283

Qy        251 TACACACACACACACATATATATATACACATATATATACACATATATACATATATACACA 310
              || | | | | | | |||||||||||| | ||||||||| | ||||||| ||||||| | |
Db     120284 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 120343

Qy        311 TATATACACATATATATACACATACATATGTACACATATATATACACATATGTATACACA 370
              |||||| | ||||||||| | ||| |||| || | ||||||||| | |||| |||| | |
Db     120344 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 120403

Qy        371 TATATACACATATATACACACATATATACACATATATACACACATATATACACATATATA 430
              |||||| | ||||||| | | ||||||| | ||||||| | | ||||||| | |||||||
```

```
Db     120404 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 120463

Qy        431 CACATATATACACATATACACATATATACACATATATACATATATACACATATATATAAT 490
              | ||||||| | ||||| | ||||||| | ||||||| ||||||| | |||||||| ||
Db     120464 TATATATATATATATATATATATATATATATATATATATATATATATATATATATAT-AT 120522

Qy        491 ATACACACATATATATACACATATATACACACATATATACACATATATACACATATATAT 550
              ||| | | |||||||||| | ||||||| | | ||||||| | ||||||| | ||||||||
Db     120523 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 120582

Qy        551 ACACATATATACACATATATACATATATACACATATATATACATATATACACATATATAC 610
              | | ||||||| | ||||||| ||||||| | ||||||||| ||||||| | |||||||
Db     120583 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 120642

Qy        611 ATATATACACATATATACATATATACACACATATATACACATACATATACACACACATAG 670
              ||||||| | ||||||| ||||||| | | ||||||| | ||| ||||| | | | |||
Db     120643 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 120702

Qy        671 ATATACATATATATACACATATATATACGTATATATATGTATATATATATGCTCCAGAGT 730
              ||||| ||||||||| | ||||||||| ||||||||| |||||||||| |
Db     120703 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 120762

Qy        731 TCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGT 790
              |||         | | ||    |   ||| | | | | | | | | |   |
Db     120763 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 120822

Qy        791 GTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGT 850
              |||   | |  | || |    |      | |||       | ||| ||| || | |
Db     120823 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 120882

Qy        851 AATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTAACACTATATATATATATAGTGAAAG 909
              |   | |  ||| | || |  ||  ||   |  ||   ||||||||||||||   |||
Db     120883 ATATATATATATATATATATATATATATATATATATATATATATATATATATATAAAT 120942

Qy        910 CAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTATAGAAGAATGGCTAAGATGC 969
              | || |  || |   |     |   || || || |     || | |   ||  |
Db     120943 ATATATATATATATATATTTATATATATATATATATATATATAAAAANTAAATATAAATA 121002

Qy        970 CAAA 973
              |||
Db     121003 NAAA 121006


Search completed: April 21, 2005, 17:03:21
Job time : 13946.5 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 20, 2005, 22:49:00 ; Search time 1505.38 Seconds
(without alignments)
18171.562 Million cell updates/sec

Title: US-09-939-209A-3_COPY_1_4621
Perfect score: 4621
Sequence: 1 agttcaagaccagcctgagc.....ttgctttttagtcctaaaaa 4621

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_16Dec04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	4621	100.0	20300	6	ABK47337	Abk47337 Genomic n
c	2	356.6	7.7	121600	6	ABT10748	Abt10748 Human bre
c	3	356.6	7.7	121600	12	ADQ18307	Adq18307 Human sof
c	4	327.4	7.1	188971	12	ADL08108	Adl08108 Human gen
	5	323.6	7.0	163382	13	ABD32659	Abd32659 Human can
c	6	322.2	7.0	26997	4	AAS46748	Aas46748 Tumour su
c	7	311.2	6.7	26997	4	AAS46747	Aas46747 Tumour su
	8	308	6.7	1990	10	ADC87572	Adc87572 Human GPC
	9	296	6.4	15418	3	AAA63785	Aaa63785 Nucleotid
	10	296	6.4	15418	6	AAL38601	Aal38601 Human TER
	11	296	6.4	15418	6	ABS54997	Abs54997 Lambda cl
	12	296	6.4	15418	6	AAL38595	Aal38595 DNA of pl
	13	296	6.4	15418	10	ADC21253	Adc21253 Lambda cl
c	14	294.8	6.4	1486	10	ADC87552	Adc87552 Human GPC
	15	293.6	6.4	27067	10	AAD50021	Aad50021 Human sec
c	16	291.6	6.3	1672	10	ADC87570	Adc87570 Human GPC
	17	288	6.2	30393	4	AAK67239	Aak67239 Human imm
c	18	286.6	6.2	44147	6	ABK84481	Abk84481 Human cDN
c	19	286.6	6.2	44147	10	ADD14691	Add14691 Human src
	20	280.8	6.1	32206	4	AAK89374	Aak89374 Human dig
	21	270.4	5.9	20420	4	AAK73165	Aak73165 Human imm
	22	270.4	5.9	20420	6	ABK69933	Abk69933 Human sec
c	23	268.6	5.8	110000	13	ABD32806_4	Continuation (5 of
c	24	268.6	5.8	191127	10	ADH13788	Adh13788 Prostate
	25	267.6	5.8	110000	13	ABD32806_4	Continuation (5 of
	26	267.4	5.8	19965	4	AAK73166	Aak73166 Human imm
	27	267.4	5.8	19965	6	ABK69932	Abk69932 Human sec
c	28	266.4	5.8	6971	6	ABL33237	Abl33237 Human imm
c	29	263.6	5.7	2522	11	ADM01858	Adm01858 Human cDN
c	30	263	5.7	117962	8	AAD54480	Aad54480 Human CIP
	31	263	5.7	165199	6	ABK83460	Abk83460 Human cDN
c	32	263	5.7	227968	6	ABK83497	Abk83497 Human cDN
c	33	263	5.7	227968	12	ADQ18538	Adq18538 Human sof
c	34	261.4	5.7	12069	6	ABK39930	Abk39930 Human che
c	35	260.6	5.6	20420	4	AAK73165	Aak73165 Human imm
c	36	260.6	5.6	20420	6	ABK69933	Abk69933 Human sec
	37	258.8	5.6	1489	10	ADC87554	Adc87554 Human GPC
c	38	258.8	5.6	2761	12	ADO15781	Ado15781 4 synthes
	39	258.8	5.6	201766	10	ADL13771	Adl13771 Osteoarth
c	40	258.6	5.6	163382	13	ABD32659	Abd32659 Human can
	41	258.4	5.6	82938	6	ABV72623	Abv72623 Human tra
c	42	258	5.6	19965	4	AAK73166	Aak73166 Human imm
c	43	258	5.6	19965	6	ABK69932	Abk69932 Human sec
	44	257.6	5.6	12069	6	ABK39930	Abk39930 Human che
	45	256.4	5.5	1486	10	ADC87552	Adc87552 Human GPC

ALIGNMENTS

RESULT 1
ABK47337
ID ABK47337 standard; DNA; 20300 BP.
XX
AC ABK47337;
XX
DT 18-JUN-2002 (first entry)
XX
DE Genomic nucleotide sequence encoding human RGS-4 protein.
XX
KW RGS-4; schizophrenia; human; regulator of G protein signalling 4;
KW neuroleptic; polymorphism; schizotypy; schizoid; chromosome 1q21-22;
KW gene; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT variation replace(4121,T)
FT /*tag= a
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace(4123,A)
FT /*tag= b
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace(4368,C)
FT /*tag= c
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace(4621,C)
FT /*tag= d
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace(4790,T)
FT /*tag= e
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace(4816,T)
FT /*tag= f
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace(4970,T)
FT /*tag= g
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace(5055,G)
FT /*tag= h
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace(5295,A)
FT /*tag= i

```
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(5695,A)
FT                   /*tag= j
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(7375,T)
FT                   /*tag= k
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(7759,A)
FT                   /*tag= l
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(8596,A)
FT                   /*tag= m
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(9602. .9610,CA)
FT                   /*tag= n
FT                   /note= "Deletion polymorphism"
FT   variation       replace(9892,A)
FT                   /*tag= o
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(9963,A)
FT                   /*tag= p
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(10132,A)
FT                   /*tag= q
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(11056,C)
FT                   /*tag= r
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(11091,T)
FT                   /*tag= s
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(11106,A)
FT                   /*tag= t
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(11774,T)
FT                   /*tag= u
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(12143,A)
FT                   /*tag= v
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(12145,T)
FT                   /*tag= w
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(14367,G)
FT                   /*tag= x
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(17027. .17029,GA)
FT                   /*tag= y
FT                   /note= "Deletion polymorphism"
FT   variation       replace(17630,T)
FT                   /*tag= Z
FT                   /standard_name= "Single-nucleotide polymorphism"
XX
PN   WO200216653-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US026622.
XX
PR   24-AUG-2000; 2000US-0228021P.
XX
PA   (UYPI-) UNIV PITTSBURGH.
XX
PI   Levitt PR,  Mirnics K,  Kodavali VC,  Nimgaonkar VL;
XX
DR   WPI; 2002-292070/33.
XX
PT   Diagnosing, assessing susceptibility and treating schizophrenia, involves
PT   observing regulator of G-protein signaling 4, RGS4 levels in a subject.
XX
PS   Claim 1; Page 20-33; 112pp; English.
XX
CC   This invention relates to a novel method for diagnosing schizophrenia or
CC   determining susceptibility to schizophrenia in a human. The method
CC   comprises obtaining from a patient a DNA sample and detecting variations
CC   in the regulator of G-protein signalling 4 (RGS4) gene. Alternatively,
CC   the method involves measuring RGS4 mRNA or protein levels in a tissue
CC   sample from the patient and determining if there is a reduced level. The
CC   method of the invention is useful for diagnosing and determining
CC   susceptibility to schizophrenia. The invention also comprises a method
CC   that is useful for treating schizophrenia which includes a prophylactic
CC   treatment. The method of genotyping polymorphic variants in the RGS-4
CC   gene is applied to diagnosing pathologies of the schizophrenic spectrum,
CC   such as in particular schizotypy, schizoid individuals, etc. This method
CC   offers the possibility of diagnosing schizophrenia by a biological test
CC   and no longer exclusively by clinical evaluations. The present sequence
CC   represents the genomic DNA encoding the human regulator of G-protein
CC   signalling 4 (RGS4) protein used in the method of the invention. The gene
CC   for the RGS4 protein is located on human chromosome 1q21-22
XX
SQ   Sequence 20300 BP; 6157 A; 4102 C; 3775 G; 6266 T; 0 U; 0 Other;

  Query Match           100.0%;  Score 4621;  DB 6;  Length 20300;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 4621;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAATTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAATTA 60

Qy         61 GACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAAT 120

Qy        121 CACTTGAACCTGCTGGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACTTGAACCTGCTGGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAGC 180

Qy        181 CCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATA 240

Qy        241 TATATATATATACACACACACACATATATATATACACATATATATACACATATATACA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TATATATATATACACACACACACATATATATATACACATATATATACACATATATACA 300

Qy        301 TATATACACATATATACACATATATATACACATACATATGTACACATATATATACACATA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TATATACACATATATACACATATATATACACATACATATGTACACATATATATACACATA 360

Qy        361 TGTATACACATATATACACATATATACACACATATATACACATATATACACACATATATA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGTATACACATATATACACATATATACACACATATATACACATATATACACACATATATA 420

Qy        421 CACATATATACACATATATACACATATACACATATATACACATATATACATATATACACA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CACATATATACACATATATACACATATACACATATATACACATATATACATATATACACA 480

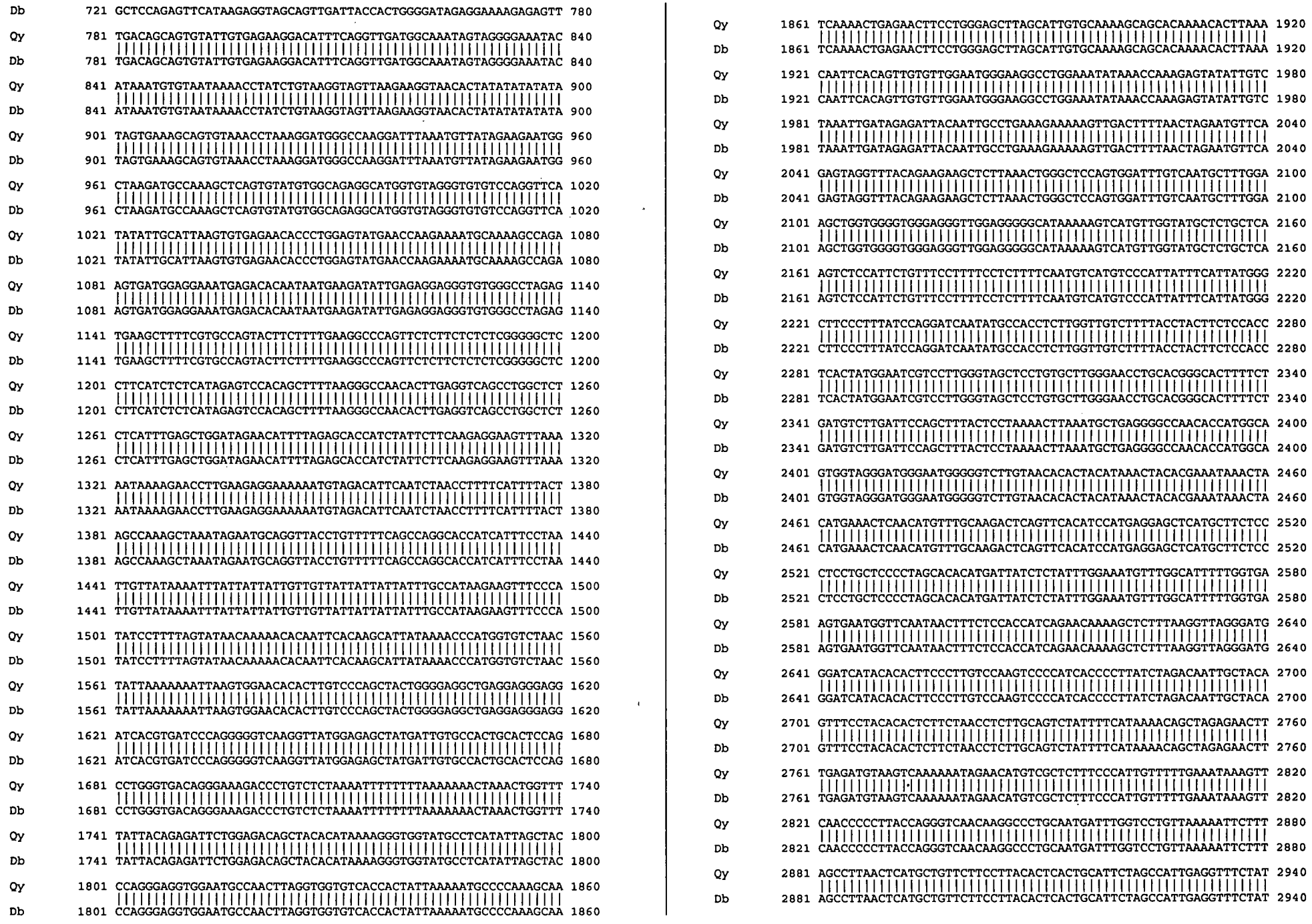
Qy        481 TATATATAATATACACACATATATATACACATATATACACACATATATACACATATATAC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TATATATAATATACACACATATATATACACATATATACACACATATATACACATATATAC 540

Qy        541 ACATATATATACACATATATACACATATATACATATATACACATATATATACATATATAC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ACATATATATACACATATATACACATATATACATATATACACATATATATACATATATAC 600

Qy        601 ACATATATACATATATACACATATATACATATATACACACATATATACACATACATATAC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ACATATATACATATATACACATATATACATATATACACACATATATACACATACATATAC 660

Qy        661 ACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATAT 720

Qy        721 GCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```


Db 721 GCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTT 780
Qy 781 TGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATAC 840
Db 781 TGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATAC 840
Qy 841 ATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATATATATA 900
Db 841 ATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATATATATA 900
Qy 901 TAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTATAGAAGAATGG 960
Db 901 TAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTATAGAAGAATGG 960
Qy 961 CTAAGATGCCAAAGCTCAGTGTATGTGGCAGAGGCATGGTGTAGGGTGTGTCCAGGTTCA 1020
Db 961 CTAAGATGCCAAAGCTCAGTGTATGTGGCAGAGGCATGGTGTAGGGTGTGTCCAGGTTCA 1020
Qy 1021 TATATTGCATTAAGTGTGAGAACACCCTGGAGTATGAACCAAGAAAATGCAAAAGCCAGA 1080
Db 1021 TATATTGCATTAAGTGTGAGAACACCCTGGAGTATGAACCAAGAAAATGCAAAAGCCAGA 1080
Qy 1081 AGTGATGGAGGAAATGAGACACAATAATGAAGATATTGAGAGGAGGGTGTGGGCCTAGAG 1140
Db 1081 AGTGATGGAGGAAATGAGACACAATAATGAAGATATTGAGAGGAGGGTGTGGGCCTAGAG 1140
Qy 1141 TGAAGCTTTTCGTGCCAGTACTTCTTTTGAAGGCCCAGTTCTCTTCTCTCTCGGGGGCTC 1200
Db 1141 TGAAGCTTTTCGTGCCAGTACTTCTTTTGAAGGCCCAGTTCTCTTCTCTCTCGGGGGCTC 1200
Qy 1201 CTTCATCTCTCATAGAGTCCACAGCTTTTAAGGGCCAACACTTGAGGTCAGCCTGGCTCT 1260
Db 1201 CTTCATCTCTCATAGAGTCCACAGCTTTTAAGGGCCAACACTTGAGGTCAGCCTGGCTCT 1260
Qy 1261 CTCATTTGAGCTGGATAGAACATTTTAGAGCACCATCTATTCTTCAAGAGGAAGTTTAAA 1320
Db 1261 CTCATTTGAGCTGGATAGAACATTTTAGAGCACCATCTATTCTTCAAGAGGAAGTTTAAA 1320
Qy 1321 AATAAAAGAACCTTGAAGAGGAAAAAATGTAGACATTCAATCTAACCTTTTCATTTTACT 1380
Db 1321 AATAAAAGAACCTTGAAGAGGAAAAAATGTAGACATTCAATCTAACCTTTTCATTTTACT 1380
Qy 1381 AGCCAAAGCTAAATAGAATGCAGGTTACCTGTTTTTCAGCCAGGCACCATCATTTCCTAA 1440
Db 1381 AGCCAAAGCTAAATAGAATGCAGGTTACCTGTTTTTCAGCCAGGCACCATCATTTCCTAA 1440
Qy 1441 TTGTTATAAAATTTATTATTATTGTTGTTATTATTATTATTTGCCATAAGAAGTTTCCCA 1500
Db 1441 TTGTTATAAAATTTATTATTATTGTTGTTATTATTATTATTTGCCATAAGAAGTTTCCCA 1500
Qy 1501 TATCCTTTTAGTATAACAAAAACACAATTCACAAGCATTATAAAACCCATGGTGTCTAAC 1560
Db 1501 TATCCTTTTAGTATAACAAAAACACAATTCACAAGCATTATAAAACCCATGGTGTCTAAC 1560
Qy 1561 TATTAAAAAAAATTAAGTGGAACACACTTGTCCCAGCTACTGGGGAGGCTGAGGAGGGAGG 1620
Db 1561 TATTAAAAAAAATTAAGTGGAACACACTTGTCCCAGCTACTGGGGAGGCTGAGGAGGGAGG 1620
Qy 1621 ATCACGTGATCCCAGGGGGTCAAGGTTATGGAGAGCTATGATTGTGCCACTGCACTCCAG 1680
Db 1621 ATCACGTGATCCCAGGGGGTCAAGGTTATGGAGAGCTATGATTGTGCCACTGCACTCCAG 1680
Qy 1681 CCTGGGTGACAGGGAAAGACCCTGTCTCTAAAATTTTTTTTAAAAAAACTAAACTGGTTT 1740
Db 1681 CCTGGGTGACAGGGAAAGACCCTGTCTCTAAAATTTTTTTTAAAAAAACTAAACTGGTTT 1740
Qy 1741 TATTACAGAGATTCTGGAGACAGCTACACATAAAAGGGTGGTATGCCTCATATTAGCTAC 1800
Db 1741 TATTACAGAGATTCTGGAGACAGCTACACATAAAAGGGTGGTATGCCTCATATTAGCTAC 1800
Qy 1801 CCAGGGAGGTGGAATGCCAACTTAGGTGGTGTCACCACTATTAAAAATGCCCCAAAGCAA 1860
Db 1801 CCAGGGAGGTGGAATGCCAACTTAGGTGGTGTCACCACTATTAAAAATGCCCCAAAGCAA 1860
Qy 1861 TCAAAACTGAGAACTTCCTGGGAGCTTAGCATTGTGCAAAAGCAGCACAAAACACTTAAA 1920
Db 1861 TCAAAACTGAGAACTTCCTGGGAGCTTAGCATTGTGCAAAAGCAGCACAAAACACTTAAA 1920
Qy 1921 CAATTCACAGTTGTGTTGGAATGGGAAGGCCTGGAAATATAAACCAAAGAGTATATTGTC 1980
Db 1921 CAATTCACAGTTGTGTTGGAATGGGAAGGCCTGGAAATATAAACCAAAGAGTATATTGTC 1980
Qy 1981 TAAATTGATAGAGATTACAATTGCCTGAAAGAAAAAGTTGACTTTTAACTAGAATGTTCA 2040
Db 1981 TAAATTGATAGAGATTACAATTGCCTGAAAGAAAAAGTTGACTTTTAACTAGAATGTTCA 2040
Qy 2041 GAGTAGGTTTACAGAAGAAGCTCTTAAACTGGGCTCCAGTGGATTTGTCAATGCTTTGGA 2100
Db 2041 GAGTAGGTTTACAGAAGAAGCTCTTAAACTGGGCTCCAGTGGATTTGTCAATGCTTTGGA 2100
Qy 2101 AGCTGGTGGGGTGGGAGGGTTGGAGGGGGCATAAAAAGTCATGTTGGTATGCTCTGCTCA 2160
Db 2101 AGCTGGTGGGGTGGGAGGGTTGGAGGGGGCATAAAAAGTCATGTTGGTATGCTCTGCTCA 2160
Qy 2161 AGTCTCCATTCTGTTTCCTTTTCCTCTTTTCAATGTCATGTCCCATTATTTCATTATGGG 2220
Db 2161 AGTCTCCATTCTGTTTCCTTTTCCTCTTTTCAATGTCATGTCCCATTATTTCATTATGGG 2220
Qy 2221 CTTCCCTTTATCCAGGATCAATATGCCACCTCTTGGTTGTCTTTTACCTACTTCTCCACC 2280
Db 2221 CTTCCCTTTATCCAGGATCAATATGCCACCTCTTGGTTGTCTTTTACCTACTTCTCCACC 2280
Qy 2281 TCACTATGGAATCGTCCTTGGGTAGCTCCTGTGCTTGGGAACCTGCACGGGCACTTTTCT 2340
Db 2281 TCACTATGGAATCGTCCTTGGGTAGCTCCTGTGCTTGGGAACCTGCACGGGCACTTTTCT 2340
Qy 2341 GATGTCTTGATTCCAGCTTTACTCCTAAAACTTAAATGCTGAGGGGCCAACACCATGGCA 2400
Db 2341 GATGTCTTGATTCCAGCTTTACTCCTAAAACTTAAATGCTGAGGGGCCAACACCATGGCA 2400
Qy 2401 GTGGTAGGGATGGGAATGGGGGTCTTGTAACACACTACATAAACTACACGAAATAAACTA 2460
Db 2401 GTGGTAGGGATGGGAATGGGGGTCTTGTAACACACTACATAAACTACACGAAATAAACTA 2460
Qy 2461 CATGAAACTCAACATGTTTGCAAGACTCAGTTCACATCCATGAGGAGCTCATGCTTCTCC 2520
Db 2461 CATGAAACTCAACATGTTTGCAAGACTCAGTTCACATCCATGAGGAGCTCATGCTTCTCC 2520
Qy 2521 CTCCTGCTCCCCTAGCACACATGATTATCTCTATTTGGAAATGTTTGGCATTTTTGGTGA 2580
Db 2521 CTCCTGCTCCCCTAGCACACATGATTATCTCTATTTGGAAATGTTTGGCATTTTTGGTGA 2580
Qy 2581 AGTGAATGGTTCAATAACTTTCTCCACCATCAGAACAAAAGCTCTTTAAGGTTAGGGATG 2640
Db 2581 AGTGAATGGTTCAATAACTTTCTCCACCATCAGAACAAAAGCTCTTTAAGGTTAGGGATG 2640
Qy 2641 GGATCATACACACTTCCCTTGTCCAAGTCCCCATCACCCCTTATCTAGACAATTGCTACA 2700
Db 2641 GGATCATACACACTTCCCTTGTCCAAGTCCCCATCACCCCTTATCTAGACAATTGCTACA 2700
Qy 2701 GTTTCCTACACACTCTTCTAACCTCTTGCAGTCTATTTTCATAAAACAGCTAGAGAACTT 2760
Db 2701 GTTTCCTACACACTCTTCTAACCTCTTGCAGTCTATTTTCATAAAACAGCTAGAGAACTT 2760
Qy 2761 TGAGATGTAAGTCAAAAAATAGAACATGTCGCTCTTTCCCATTGTTTTTGAAATAAAGTT 2820
Db 2761 TGAGATGTAAGTCAAAAAATAGAACATGTCGCTCTTTCCCATTGTTTTTGAAATAAAGTT 2820
Qy 2821 CAACCCCCTTACCAGGGTCAACAAGGCCCTGCAATGATTTGGTCCTGTTAAAAATTCTTT 2880
Db 2821 CAACCCCCTTACCAGGGTCAACAAGGCCCTGCAATGATTTGGTCCTGTTAAAAATTCTTT 2880
Qy 2881 AGCCTTAACTCATGCTGTTCTTCCTTACACTCACTGCATTCTAGCCATTGAGGTTTCTAT 2940
Db 2881 AGCCTTAACTCATGCTGTTCTTCCTTACACTCACTGCATTCTAGCCATTGAGGTTTCTAT 2940

```
Qy       2941 GCATCAAACTTTTTTTGGTCCCAGCACTGTGCACATCCTTCTGGGTAGAATGCCCCTTGA 3000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2941 GCATCAAACTTTTTTTGGTCCCAGCACTGTGCACATCCTTCTGGGTAGAATGCCCCTTGA 3000

Qy       3001 TTTGTATAATTAGCACCTCCTTCATCATTTAGGTCTTAGTATAACTACTACCTTCTTAGA 3060
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3001 TTTGTATAATTAGCACCTCCTTCATCATTTAGGTCTTAGTATAACTACTACCTTCTTAGA 3060

Qy       3061 GAAGCTCTGCTTCTTCATCCTATAAAAAAGTAAAATTCCTTACCCTGTTATTTTTTAAGT 3120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3061 GAAGCTCTGCTTCTTCATCCTATAAAAAAGTAAAATTCCTTACCCTGTTATTTTTTAAGT 3120

Qy       3121 CATCCGTGTTTCATTCTGTTAAAGTTCTTATCACAATTTATCATTATTTTATTTACAGTC 3180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3121 CATCCGTGTTTCATTCTGTTAAAGTTCTTATCACAATTTATCATTATTTTATTTACAGTC 3180

Qy       3181 ATGTGCCACATAACAATGTTTCAGTCAGGGATAGAACACAAATGTATCTGGCCCCATAAT 3240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3181 ATGTGCCACATAACAATGTTTCAGTCAGGGATAGAACACAAATGTATCTGGCCCCATAAT 3240

Qy       3241 ATTATAAGCTGAGAAATTTCTATTAACTAGTGATATCGCAGCCATCATAAGTGTAATGCA 3300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3241 ATTATAAGCTGAGAAATTTCTATTAACTAGTGATATCGCAGCCATCATAAGTGTAATGCA 3300

Qy       3301 GGACATTACCTTTTCTATGTTTAGATATGTTAGATACACAAATATATTTCATTGTGTTAT 3360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3301 GGACATTACCTTTTCTATGTTTAGATATGTTAGATACACAAATATATTTCATTGTGTTAT 3360

Qy       3361 AATTTCCTACAGTATTCAGTACAGTAACATGCTGTACAGGTTTGTAACCTAGGAGTAATA 3420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3361 AATTTCCTACAGTATTCAGTACAGTAACATGCTGTACAGGTTTGTAACCTAGGAGTAATA 3420

Qy       3421 GGCTATACCATACAGCTTAGGTGTGTAGTAGGCTATAACCATCTAGGTTTGTGTAAGTAC 3480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3421 GGCTATACCATACAGCTTAGGTGTGTAGTAGGCTATAACCATCTAGGTTTGTGTAAGTAC 3480

Qy       3481 ATTCTATGATATTCCCACAATGATGAAATCACCTAACTACACATTTCTCAGAATGTTTCA 3540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3481 ATTCTATGATATTCCCACAATGATGAAATCACCTAACTACACATTTCTCAGAATGTTTCA 3540

Qy       3541 CTGTTGTGAAGTGACCCATGACTATATTTTCCTATATACTTGATATTTTTGTGCATCTGC 3600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3541 CTGTTGTGAAGTGACCCATGACTATATTTTCCTATATACTTGATATTTTTGTGCATCTGC 3600

Qy       3601 CCATGAGAATGTAGTGTAAGATCAAAGGATGCAAGAATGGGTTCTATCCAGTATAGTACC 3660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3601 CCATGAGAATGTAGTGTAAGATCAAAGGATGCAAGAATGGGTTCTATCCAGTATAGTACC 3660

Qy       3661 CACTACACTGGTGGATGTCAATATGTATTTGTTAGATTAATATCTCAAGAATGAGCACCT 3720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3661 CACTACACTGGTGGATGTCAATATGTATTTGTTAGATTAATATCTCAAGAATGAGCACCT 3720

Qy       3721 TTCTCAGACACATAAAAGATGCTCAATATAAAAGTTTGTTGAACTGAACGTTATTGGCAA 3780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3721 TTCTCAGACACATAAAAGATGCTCAATATAAAAGTTTGTTGAACTGAACGTTATTGGCAA 3780

Qy       3781 ATGTAACATGATCGGATTTAAAGAGGAGCGAAACAGAGGTCTGGCTCAAACACCATACTT 3840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3781 ATGTAACATGATCGGATTTAAAGAGGAGCGAAACAGAGGTCTGGCTCAAACACCATACTT 3840

Qy       3841 CTAGAGTGCATAAGAGGTAGCAGTTGATTACCACTGGCGACAGGAGAAAAAAGAGCTTGA 3900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3841 CTAGAGTGCATAAGAGGTAGCAGTTGATTACCACTGGCGACAGGAGAAAAAAGAGCTTGA 3900

Qy       3901 CCGCAGGGTACTGTGAAGACATTTCAGGTTGATGGCACAGAACAGGGGAAATACATAAAT 3960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3901 CCGCAGGGTACTGTGAAGACATTTCAGGTTGATGGCACAGAACAGGGGAAATACATAAAT 3960

Qy       3961 GTGTGGGAATATTCAGTGGTCTGGGATGACTACATAGTAGAATATAATGAAGAAAAGAGT 4020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3961 GTGTGGGAATATTCAGTGGTCTGGGATGACTACATAGTAGAATATAATGAAGAAAAGAGT 4020

Qy       4021 GGAAGGGAAAGATGAAAAGTTGGAATGGGGATGAATTATGAAAGTACCAGAATGTTATGC 4080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4021 GGAAGGGAAAGATGAAAAGTTGGAATGGGGATGAATTATGAAAGTACCAGAATGTTATGC 4080

Qy       4081 TAAGGAATCTAGATTTTAAAATGTGAGGGCAAATTGAAGTCCTGGGCACGTTACAAAACT 4140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4081 TAAGGAATCTAGATTTTAAAATGTGAGGGCAAATTGAAGTCCTGGGCACGTTACAAAACT 4140

Qy       4141 AGAGGTCATAAAGTTTACCCTAATTTACCAAGATTTCCTAGAGGATCTATAATTGGAATC 4200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4141 AGAGGTCATAAAGTTTACCCTAATTTACCAAGATTTCCTAGAGGATCTATAATTGGAATC 4200

Qy       4201 CAGATCTGCCTCTCTGTAAAGTTCAAGCACTTTCCATGACACCATACTGTTTCTTTCCAC 4260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4201 CAGATCTGCCTCTCTGTAAAGTTCAAGCACTTTCCATGACACCATACTGTTTCTTTCCAC 4260

Qy       4261 CTGCACAATGCAAATGAACTCTTATGAAACTGCTGTTTCTATCCTGGGCTAAATGTTGCA 4320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4261 CTGCACAATGCAAATGAACTCTTATGAAACTGCTGTTTCTATCCTGGGCTAAATGTTGCA 4320

Qy       4321 GAAAAAAGATTTAATCTTTGGGATAAGGCTATTTTGGGTTTTCTCCTACTTCTTGGGAAA 4380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4321 GAAAAAAGATTTAATCTTTGGGATAAGGCTATTTTGGGTTTTCTCCTACTTCTTGGGAAA 4380

Qy       4381 CAAGGTTTTCTTCCCCTGGCTAATTAAGTGTGGTATTGTTCTTCCAGGGAAATCAGTGAT 4440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4381 CAAGGTTTTCTTCCCCTGGCTAATTAAGTGTGGTATTGTTCTTCCAGGGAAATCAGTGAT 4440

Qy       4441 GCATCACCTGCTGCTATCAAATGTCAGGGTTGGAGTTCCTGATTTATTGCATGTGCCCAC 4500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4441 GCATCACCTGCTGCTATCAAATGTCAGGGTTGGAGTTCCTGATTTATTGCATGTGCCCAC 4500

Qy       4501 AAAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCT 4560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4501 AAAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCT 4560

Qy       4561 GTTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAA 4620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4561 GTTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAA 4620

Qy       4621 A 4621
              |
Db       4621 A 4621
```

RESULT 2
ABT10748/c
ID ABT10748 standard; cDNA; 121600 BP.
XX
AC ABT10748;
XX
DT 04-DEC-2002 (first entry)
XX
DE Human breast cancer associated coding sequence SEQ ID NO: 882.
XX
KW Human; breast specific gene; breast cancer; differential expression;
KW cytostatic; gene therapy; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200259271-A2.
XX
PD 01-AUG-2002.
XX
PF 25-JAN-2002; 2002WO-US002176.
XX
PR 25-JAN-2001; 2001US-0263757P.
PR 25-APR-2001; 2001US-0286090P.
PR 23-MAY-2001; 2001US-0292517P.
XX
PA (GENE-) GENE LOGIC INC.
XX
PI Orr MS, Nation M, Diggans JC, Zeng W;
XX

```
DR   WPI; 2002-674803/72.
XX
PT   Diagnosing breast cancer in a patient comprises detecting the level of
PT   gene expression in cell or tissue samples, where a differential gene
PT   expression is indicative of breast cancer.
XX
PS   Claim 1; SEQ ID NO 882; 260pp + Sequence Listing; English.
XX
CC   The present invention relates to methods of diagnosing breast cancer in a
CC   patient, which comprise detecting the level of expression in a tissue
CC   sample of two or more genes selected from those shown in ABT09867-
CC   ABT11112, where a differential expression of the genes indicates breast
CC   cancer. The methods are useful in diagnosing, treating, detecting the
CC   progression, and in monitoring treatment of breast cancer in patients.
CC   The methods are also useful as a screening tool for agents that modulate
CC   the onset or progression of breast cancer. The breast cancer genes may be
CC   used as diagnostic markers for the prediction or identification of the
CC   malignant state of breast tissue, for confirming the type and progression
CC   of cancer, and for drug screening and assays. The present sequence is a
CC   coding sequence of the invention. Note: The sequence data for this patent
CC   did not form part of the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub.published_pct_sequences
XX
SQ   Sequence 121600 BP; 38372 A; 24652 C; 24048 G; 34528 T; 0 U; 0 Other;

  Query Match             7.7%;  Score 356.6;  DB 6;  Length 121600;
  Best Local Similarity   74.7%;  Pred. No. 8.6e-52;
  Matches  543;  Conservative    0;  Mismatches  169;  Indels   15;  Gaps    7;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAA--- 57
              || || |||||| ||||   ||||  |||||||||| |||||||||||||| |||||
Db      44776 AGATCGAGACCATCCTGGCTAACACAGTGAAACCCCGTCTCTACTAAAAATCCAAAAAAA 44717

Qy         58 --TTAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGG 115
                |||| | ||| |||||     | ||||||| ||||||||||| |||||| ||||||||
Db      44716 AGTTAGCCGGGCGTGGTGGCGGGCACCTGTAGTCCCAGCTACTCTGGAGGCTGAGGCAGG 44657

Qy        116 AGAATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCAC 174
              ||||| | ||||| ||  || || |||| |||| | ||| | ||||   |||| |||||
Db      44656 AGAATGGCGTGAACTTG--GGAGGCGGAGCTTGCAGTGAGCCGAGATAGCGCCACTGCAC 44599

Qy        175 TCCAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATA 234
              ||||| |  ||| || ||||||| |   ||||||| |||||||| |      | ||||||
Db      44598 TCCAGTCTGGGCGAC-AGAGCGAGACTCCATCTCAAAAAAAAAATA----TATATATATA 44544

Qy        235 TATATATATATATATACACACACACACATATATATATACACATATATATACACATA 294
              || ||||||||| ||||||| | | | | | ||||||| ||| | | ||||||| | |||
Db      44543 TAGATATATATAGATATACAGATATATAGATATATATAGATATATAGATATATATAGATA 44484

Qy        295 TATACATATATACACATATATACACATATATATACACATACATATGTACACATATATATA 354
              |||| ||||||| | ||||||| | ||||||| | | ||| |||| || | |||||| |
Db      44483 TATAGATATATATAGATATATAGATATATATAGATATATAGATATATATAGATATATAGA 44424

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              | ||||  ||| | | ||||| | | ||||| | | ||||| | | ||||| | | | |
Db      44423 TATATATAGATATATAGATATATATAGATATATAGATATATAGATATATATAGATATATA 44364

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              ||||| | | ||||| | ||||||| | ||||| | ||||||| | ||||||| |||||
Db      44363 GATATATATAGATATATAGATATATATAGATATATAGATATATAGATATATATAGATATA 44304

Qy        475 TACACATATATATAATATACACACATATATATACACATATATACACACATATATACACAT 534
              || |  |||| || ||||| | | ||||| ||| | ||| ||| | | ||||||| | ||
Db      44303 TAGATGTATAGAT-ATATATAGATATATAGATATATATAGATATATAGATATATAGATAT 44245

Qy        535 ATATACACATATATATACACATATATACACATATATACATATATACACATATATATACAT 594
              ||||| | ||||| ||| | ||| ||| | ||||| | ||||| | | ||| |||||  | :
Db      44244 ATATAGATATATAGATATATATAGATATATATATAGATATATAGATATATAGATATATAG 44185

Qy        595 ATATACACATATATACATATATACA-CATATATACATATATACACACATATATACACATA 653
              ||||| | |||||||| |||||||  | ||||||| ||||||| | | ||||| | | |||
Db      44184 ATATATAGATATATAGATATATAGATGATATATAGATATATATAGATATATAGATATATA 44125

Qy        654 CATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATA 713
              ||||| | | | ||||||||| ||| ||||| | ||||||| |  |||| |||| ||||
Db      44124 GATATATAGATATATAGATATATATAGATATATAGATATATAGATATATAGATATATATA 44065

Qy        714 TATATAT 720
              |||||||
Db      44064 GATATAT 44058


RESULT 3
ADQ18307/c
ID   ADQ18307 standard; DNA; 121600 BP.
XX
AC   ADQ18307;
XX
DT   26-AUG-2004  (first entry)
XX
DE   Human soft tissue sarcoma-upregulated DNA - SEQ ID 1125.
XX
KW   soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO2004048938-A2.
XX
PD   10-JUN-2004.
XX
PF   26-NOV-2003; 2003WO-US038193.
XX
PR   26-NOV-2002; 2002US-0429739P.
XX
PA   (PROT-) PROTEIN DESIGN LABS INC.
XX
PI   Aziz N,  Ginsburg WM,  Zlotnik A;
XX
DR   WPI; 2004-441208/41.
XX
PT   Early detection of soft tissue sarcoma comprises determining expression
PT   of a gene in a first soft tissue sample and a normal soft tissue sample
PT   and comparing the gene expression, also useful in treating soft tissue
PT   sarcoma.
XX
PS   Example 2; SEQ ID NO 1125; 210pp; English.
XX
CC   The invention relates to a novel method for detecting soft tissue sarcoma
CC   which comprises obtaining a first soft tissue sample from an individual
CC   and a normal soft tissue sample from the same or different individual,
CC   determining the expression of a gene in both samples and comparing the
CC   expression of the gene in both soft tissue samples, where a higher level
CC   of protein expression in the first soft tissue sample indicates the
CC   presence of soft tissue sarcoma. The method of the invention has
CC   cytostatic applications and may be useful for detecting soft tissue
CC   sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC   acid sequences may be useful in diagnostic and screening applications.
CC   The current sequence is that of a human soft tissue sarcoma-upregulated
CC   DNA of the invention. The current sequence is not shown within the
CC   specification per se but was submitted in CD format by the inventor.
XX
SQ   Sequence 121600 BP; 38372 A; 24652 C; 24048 G; 34528 T; 0 U; 0 Other;

  Query Match             7.7%;  Score 356.6;  DB 12;  Length 121600;
  Best Local Similarity   74.7%;  Pred. No. 8.6e-52;
  Matches  543;  Conservative    0;  Mismatches  169;  Indels   15;  Gaps    7;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAA--- 57
              || || |||||| ||||   ||||  |||||||||| |||||||||||||| |||||
Db      44776 AGATCGAGACCATCCTGGCTAACACAGTGAAACCCCGTCTCTACTAAAAATCCAAAAAAA 44717

Qy         58 --TTAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGG 115
                |||| | ||| |||||     | ||||||| ||||||||||| |||||| ||||||||
Db      44716 AGTTAGCCGGGCGTGGTGGCGGGCACCTGTAGTCCCAGCTACTCTGGAGGCTGAGGCAGG 44657
```

```
Qy        116 AGAATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCAC 174
              |||||  | |||||  ||  || || |||| |||| | ||| | ||||   |||| |||||
Db      44656 AGAATGGCGTGAACTTG--GGAGGCGGAGCTTGCAGTGAGCCGAGATAGCGCCACTGCAC 44599

Qy        175 TCCAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATA 234
              |||||  |  ||| || |||||||  |   ||||||| ||||||||  |      | ||||||
Db      44598 TCCAGTCTGGGCGAC-AGAGCGAGACTCCATCTCAAAAAAAAAAATA----TATATATATA 44544

Qy        235 TATATATATATATATATACACACACACACACATATATATATACACATATATATACACATA 294
              || |||||||| |||||||  | | | | |||||||  ||| | | |||||||  | |||
Db      44543 TAGATATATATAGATATACAGATATATAGATATATATAGATATATAGATATATATAGATA 44484

Qy        295 TATACATATATACACATATATACACATATATATACACATACATATGTACACATATATATA 354
              ||||  |||||||  | |||||||  | |||||||  | | |||  ||||  ||  | |||||||  |
Db      44483 TATAGATATATATAGATATATAGATATATATAGATATATAGATATATATAGATATATAGA 44424

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              | ||||  |||  | | |||||  | |  |||||  | |  |||||  | |  |||||  | |  | |
Db      44423 TATATATAGATATATAGATATATATAGATATATAGATATATAGATATATATAGATATATA 44364

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              |||||  | |  |||||  | |||||||  | |||||  | |||||||  | |||||||  |||||
Db      44363 GATATATATAGATATATAGATATATATAGATATATAGATATATAGATATATATAGATATA 44304

Qy        475 TACACATATATATAATATACACACATATATATACACATATATACACACATATATACACAT 534
              ||  |  ||||  ||  |||||  | |  |||||  |||  | |||  |||  | |  |||||||  | ||
Db      44303 TAGATGTATAGAT-ATATATAGATATATAGATATATATAGATATATAGATATATAGATAT 44245

Qy        535 ATATACACATATATATACACATATATACACATATATACATATATACACATATATATACAT 594
              |||||  |  |||||  |||  |  |||  |||  |  |||||  |  |||||  |  |||  |||||  |
Db      44244 ATATAGATATATAGATATATATAGATATATATATAGATATATAGATATATAGATATATAG 44185

Qy        595 ATATACACATATATACATATATACA-CATATATACATATATACACACATATATACACATA 653
              |||||  |  |||||||  |||||||  |  ||||||| |||||||  | |  |||||  | |  |||
Db      44184 ATATATAGATATATAGATATATAGATGATATATAGATATATATAGATATATAGATATATA 44125

Qy        654 CATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATA 713
              |||||  | | |  |||||||||  |||  |||||  |  |||||||  |   ||||  ||||  ||||
Db      44124 GATATATAGATATATAGATATATATAGATATATAGATATATAGATATATAGATATATATA 44065

Qy        714 TATATAT 720
              |||||||
Db      44064 GATATAT 44058


RESULT 4
ADL08108/c
ID   ADL08108 standard; DNA; 188971 BP.
XX
AC   ADL08108;
XX
DT   20-MAY-2004  (first entry)
XX
DE   Human gene associated with low HDL-C APOA1.
XX
KW   Human; ds; SNP; single nucleotide polymorphism;
KW   high density lipoprotein-C; HDL-C; vascular disease; metabolic disease;
KW   coronary artery disease; gene.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       replace(123408,T)
FT                   /*tag=  a
FT                   /standard_name= "Single nucleotide polymorphism"
XX
PN   US2004043389-A1.
XX
PD   04-MAR-2004.
XX
PF   04-SEP-2002; 2002US-00235192.
XX
PR   04-SEP-2002; 2002US-00235192.
XX
PA   (VITI-) VITIVITY INC.
XX
PI   Mccarthy J;
XX
DR   WPI; 2004-214170/20.
XX
PT   Determining whether a subject has, or is at risk of developing, an
PT   abnormally low high density lipoprotein-C (HDL-C) level comprises
PT   detecting an allelic variant of a polymorphic region from any of a set of
PT   27 genes.
XX
PS   Disclosure; SEQ ID NO 27; 37pp; English.
XX
CC   The invention relates to determining whether a subject has, or is at risk
CC   of developing, an abnormally low high density lipoprotein-C (HDL-C) level
CC   comprises determining whether the subject has an allelic variant of a
CC   polymorphic region from any of 27 genes (alleles listed in Table 5 of the
CC   specification). Also included are determining whether a male subject has,
CC   or is at risk of developing, an abnormally low HDL-C level, comprising
CC   determining whether the male subject has an allelic variant of a
CC   polymorphic region listed in Table 5 which is associated with abnormally
CC   low HDL-C levels in males, and determining whether a female subject has,
CC   or is at risk of developing, an abnormally low HDL-C level, comprising
CC   determining whether the female subject has an allelic variant of a
CC   polymorphic region listed in Table 5 which is associated with abnormally
CC   low HDL-C levels in females. The allelic variant in determining whether a
CC   subject has, or is at risk of developing, an abnormally low HDL-C level
CC   is APOA 1 CC, CD14 1 CT, COL5A2 1 GG, EDNRB 1 AG or AA, FABP3 1 CT, GBE1
CC   1 AG or GG, LIPC 5 AA, MTHFR 1 CC, VWF 2 GG, or their complements. The
CC   allelic variant in determining whether a male subject has, or is at risk
CC   of developing, an abnormally low HDL-C level, LRP1 3 CC or CT, PAI2 4 GG,
CC   or PPARG 1 CG, or their complements. The allelic variants are also COL5A2
CC   1 GG, CD14 1 CT or CC, and FABP3 1 CT, in combination, or their
CC   complements. The methods are useful for diagnosing (a predisposition to)
CC   abnormally low levels of low high density lipoprotein-C (HDL-C) in a
CC   subject, The methods are useful in diagnosing (a predisposition to) or
CC   prognosticating diseases and disorders associated with abnormal lipid
CC   levels such as vascular and metabolic diseases, e.g., coronary artery
CC   disease. The present sequence is a human gene containing a SNP (single
CC   nucleotide polymorphism associated with low high density lipoprotein-C
CC   (HDL-C) levels.
XX
SQ   Sequence 188971 BP; 53091 A; 44581 C; 42653 G; 48646 T; 0 U; 0 Other;

  Query Match             7.1%;  Score 327.4;  DB 12;  Length 188971;
  Best Local Similarity   67.6%;  Pred. No. 8.8e-47;
  Matches  601;  Conservative    0;  Mismatches  261;  Indels   27;  Gaps    9;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC--AAAAT 58
              || ||||||||| ||||  |||||||||||||| || ||||||||||||||  |||||
Db     104585 AGATCAAGACCATCCTGGCCAACATGGTGAAACTCCGTCTCTACTAAAAATACAAAAAAT 104526

Qy         59 TAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGA 118
              |||   ||| ||||| |   | |||||||||||||||||| |||||| || ||||||||
Db     104525 TAGCTGGGCGTGGTGGTGGGCTCCTGTAATCCCAGCTACTTGGGAGGCTGAAGCAGGAGA 104466

Qy        119 ATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTG---CAC 174
              ||| ||||||||  | || || |||||||||  | ||| |||||||  |||||||   |||
Db     104465 ATCGCTTGAACC--CAGGAGGCGGAGGTTGCAGTGAGCCAAGATCGCGCCATTGCACCAC 104408

Qy        175 TCCAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATA 234
              |||||||  |||||  |||||  ||||      |||||  ||||||||||  |   | | |
Db     104407 TCCAGCCTGGGCAAAAAGAGAGAAACTCTGTCTCAAAAAAAAAAAAAAAAAAAAAAAAAA 104348

Qy        235 TATATATATATATATATACACACACACACACATATATATATACACATATATATACACATA 294
              ||||||||||||||||  | | | | |||||  | |||||||  ||||||||       |
Db     104347 AATATATATATATATATATATATATACACACACATATATATTCATATATA-------A 104295

Qy        295 TATACATATATACACATATATACACATATATATACACATACATATGTACACATATATATA 354
              |  | |||||||  ||||||||   ||  |||||  | ||  |||||  ||  ||||||||||
Db     104294 TTCATATATATATTCATATATA---ATTCATATATATATTCATATATATTCATATATATA 104238
```

```
Qy        355 CACATATGT-ATACACATATATACACATATATACACACATATATACACATATATACACAC 413
              ||||| | || || |||||||  ||||||||  || |||||||  |||||||| |   |
Db     104237 TTCATATATAATTCATATATATATTCATATATATTCATATATATATTCATATATATTC 104178

Qy        414 ATATATACACATATATACACATATATACACATATACACATATATACACATATATACATAT 473
              |||||||     ||  | | ||||||  || |||||  ||||||||  |   ||||| ||||
Db     104177 ATATATA-----ATTCATATATATATTCATATATATTCATATATATATTCATATATATAT 104123

Qy        474 ATACACATATATATAATATACACACATATATATACACATATATACACACATATATACACA 533
              | | |||  |   ||||| |  |||||||| || |||||||  || |||||| || |
Db     104122 TCATATATAATTCATATATATATTCATATATATTCATATATATATTCATATATATTCATA 104063

Qy        534 TATATACACATATATATACACATATATACACATATATACATATATACACATATATATACA 593
              ||||||  |||||||||  ||||||||  |||||||| ||  ||| | ||||  ||| |
Db     104062 TATATATTCATATATATATTCATATATATTCATATATATATTCATATATATATTCATATA 104003

Qy        594 TATATACACATATATACATATATACACATATATACATATATACACACATATATACACATA 653
              |||||  || |||||||   |||||  | ||  ||| ||||||  || | ||||  | | |||
Db     104002 TATATTCATATATATATTCATATATATATTCATATATATATTCATATATATTCATATATA 103943

Qy        654 CATATACACACACATAGATATACA---TATATATACACATATATATACGTATATATATGT 710
              ||  | | | | ||  |||||  |    |||||||  |   |||||||  |  ||||||
Db     103942 TATTCATATATATATTCATATATAATTCATATATATATTCATATATATATTCATATATAA 103883

Qy        711 ATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGA 770
              ||  ||||||  | || |  |      |  |||   || ||    | |   |||||
Db     103882 ATTCATATATAATTCATATATATATTCATATAGATATTCATATATAATTCATATAGATAT 103823

Qy        771 AAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTA 830
              | |||  || | |    |  ||  || |   | || |     |      || ||| |
Db     103822 TCATAGATATTCATATATATTCATATAGATATTCATATATTCATATATATTCATATATTC 103763

Qy        831 GGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGA 879
              | | || |||| |  | |    | |   || |  |  |   || | |
Db     103762 ATGTATATTCATATAGATATTCATATATAGATATTCATAGATATTCATA 103714
```

RESULT 5
ABD32659
ID ABD32659 standard; DNA; 163382 BP.
XX
AC ABD32659;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human cancer-associated genomic DNA HD13-065.
XX
KW Human; ds; cancer-associated protein; gene; cytostatic; cancer;
KW leukaemia; lymphoma; CAP.
XX
OS Homo sapiens.
XX
PN WO2004074320-A2.
XX
PD 02-SEP-2004.
XX
PF 17-FEB-2004; 2004WO-US004730.
XX
PR 14-FEB-2003; 2003US-00367094.
PR 14-MAR-2003; 2003US-00388838.
PR 15-APR-2003; 2003US-00417375.
PR 13-JUN-2003; 2003US-00461862.
PR 15-SEP-2003; 2003US-00663431.
PR 15-DEC-2003; 2003US-00737318.
XX
PA (SAGR-) SAGRES DISCOVERY INC.
XX
PI Morris DW, Morris DW, Malandro MS;
XX
DR WPI; 2004-652914/63.
XX
PT New isolated cancer-associated polynucleotides and polypeptides useful
PT for diagnosing, preventing or treating cancers, especially lymphoma and
PT leukemia, or in screening for agents that modulate cancer.
XX
PS claim 16; seqid 199; 310pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising at least 10
CC contiguous nucleotides of any of the 233 polynucleotide sequences given
CC in the specification, or its complement. The nucleic acids encode cancer-
CC associated proteins. Also included are an expression vector comprising
CC the isolated nucleic acid cited above, a host cell comprising the above
CC recombinant nucleic acid or expression vector, a microarray for detecting
CC a cancer-associated (CA) nucleic acid comprising at least one probe
CC comprising at least 10 contiguous nucleotides of any of the above-
CC mentioned nucleotide sequences, an isolated polypeptide (encoded within
CC an open reading frame of a CA sequence selected from any of the 95
CC polynucleotide sequences as mentioned in the specification, or its
CC complement), an isolated antibody, (or its antigen binding fragment) that
CC binds to the above polypeptide, a hybridoma that produces the above
CC monoclonal antibody, a pharmaceutical composition comprising the above
CC antibody and a pharmaceutical excipient, a kit for detecting cancer
CC cells(comprising the antibody cited above, methods for diagnosing cancer
CC or for detecting the presence or absence of cancer cells in an
CC individual, a method for inhibiting growth of cancer cells in an
CC individual, a method for delivering a therapeutic agent to cancer cells
CC in an individual, an electronic library comprising the above
CC polynucleotide or polypeptide (or their fragments), methods of screening
CC for anticancer activity or for a bioactive agent capable of modulating
CC the activity of a CA protein (CAP), methods for detecting cancer
CC associated with expression of a polypeptide in a test cell sample, a
CC method for treating cancers and a method for inhibiting the expression of
CC CA gene in a cell. The composition and methods are useful for detecting,
CC diagnosing, preventing and treating cancers, especially lymphoma and
CC leukaemia. These may also be used in screening for agents that modulate
CC cancer. The present sequence is a human CAP genomic sequence. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 163382 BP; 50073 A; 25805 C; 27837 G; 53184 T; 0 U; 6483 Other;

Query Match 7.0%; Score 323.6; DB 13; Length 163382;
Best Local Similarity 68.2%; Pred. No. 3.9e-46;
Matches 495; Conservative 0; Mismatches 224; Indels 7; Gaps 3;

```
Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              |||||||||||||||||  ||||||||||||||||||||||||||||||||| ||||||
Db     144405 AGTTCAAGACCAGCCTGGCCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAAATT 144464

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              || ||||| |||||  ||  |||||||||||||||||   |||||| ||||||||||||
Db     144465 AGCCAGGCGTGGTGGCACGTGCCTGTAATCCCAGCTACCCAGGAGGCTGAGGCAGGAGAA 144524

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              |  |||||| |    | ||  | || |   | |  | ||||||||||||||||||||
Db     144525 TTGCTTGAAGC-CAAGAGGCAGAAGTTGCAGTGAGCCGAGATCATGCCATTGCACTCCAG 144583

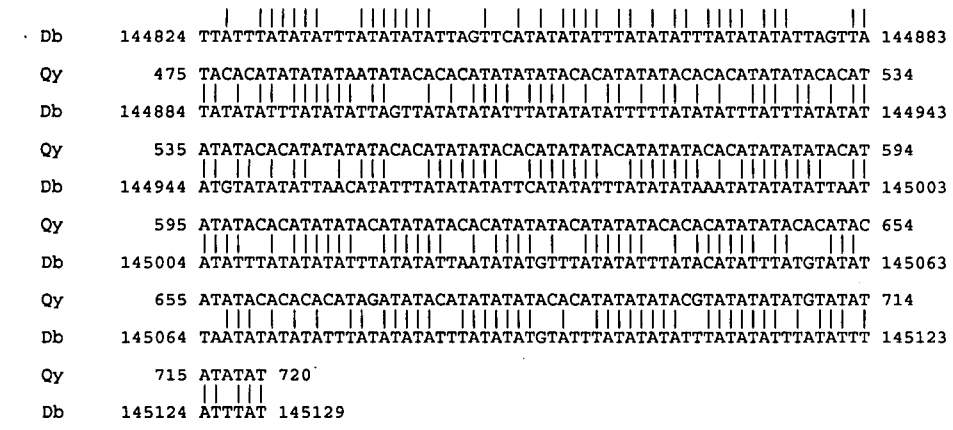
Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  ||  |||||| ||| |  |||| ||| ||||||||||       |  |   |   |
Db     144584 CCTGGGGGACAAGAACGAGACTTCATGTCAAAAAAAAAAAATACACAGTCTTCCTTGGCT 144643

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
                      | | | | | | || ||| |||| | |||| |||| | || |||
Db     144644 TAGCTGTATGGGAAATATATATATATTTATTTATATATATATTTATATATATTTATTT 144703

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACAT 359
              |||||||   || |||| | || ||| || | |||  ||| || | || ||| || | |
Db     144704 ATATATATTTATTTATATATATTTATTTATATATATTTATTTATATATTTATTTATATTT 144763

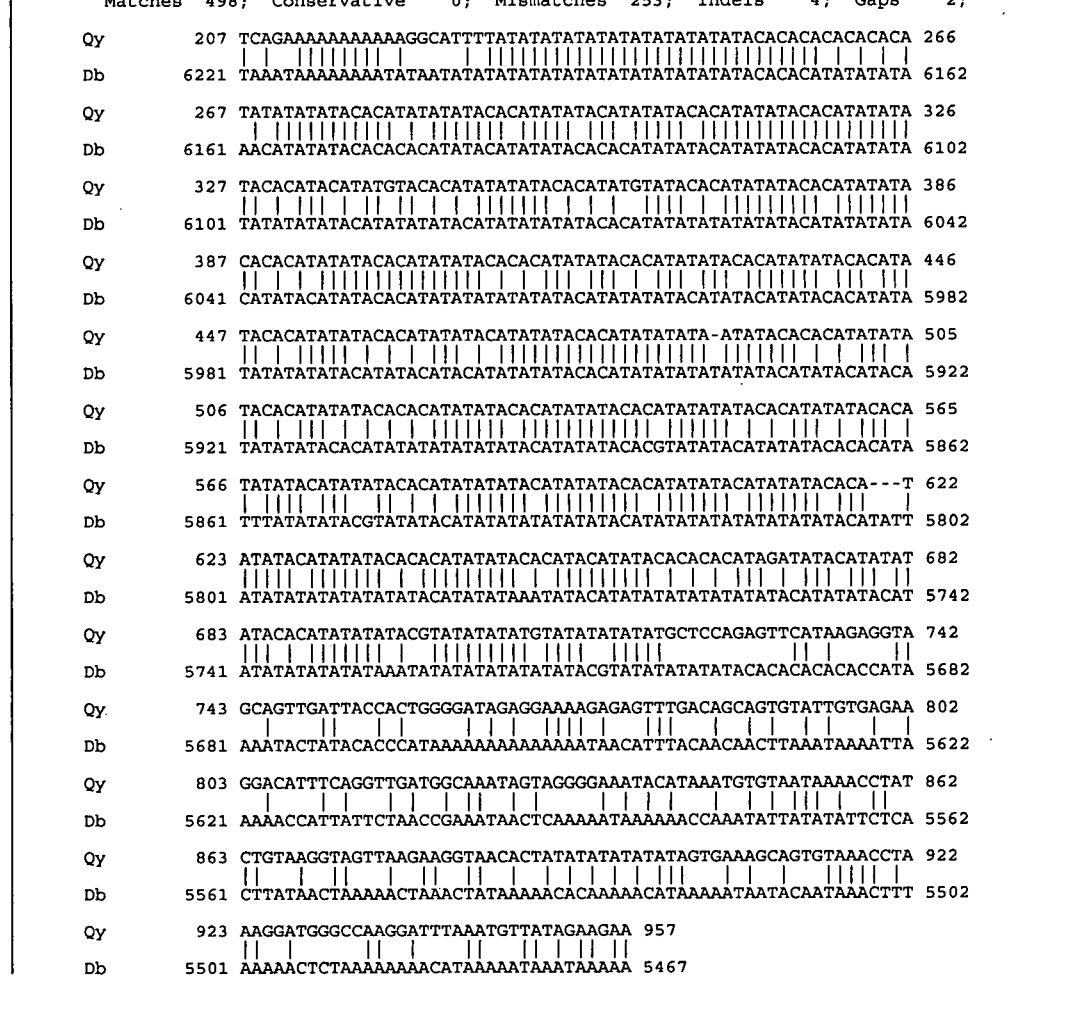
Qy        360 ATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATAT 419
              || | || | || |||  | ||||||    | |||| ||   ||||||  |   ||||||
Db     144764 ATTTTTATATATTTATTTATATATATTTGTATATATTTATTTATATATTTATTTATATAT 144823

Qy        420 ACACATATATACACATATATACACATATACACATATATACACATATATACATA-----TA 474
```



RESULT 6
AAS46748/c
ID AAS46748 standard; DNA; 26997 BP.
XX
AC AAS46748;
XX
DT 18-DEC-2001 (first entry)
XX
DE Tumour suppressor gene derived chemically modified sequence #472.
XX
KW Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer;
KW tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW cytosine methylation; ds.
XX
OS Homo sapiens.
XX
PN WO200168912-A2.
XX
PD 20-SEP-2001.
XX
PF 15-MAR-2001; 2001WO-EP002955.
XX
PR 15-MAR-2000; 2000DE-01013847.
PR 06-APR-2000; 2000DE-01019058.
PR 07-APR-2000; 2000DE-01019173.
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2001-602752/68.
XX
PT Fragments of chemically modified genes associated with tumor suppressor
PT genes and oncogenes, useful in designing primers and probes for analyzing
PT diseases associated with cytosine methylation state e.g. cancer.
XX
PS Claim 1; SEQ ID NO 472; 27pp; English.
XX
CC The invention relates to a nucleic acid comprising a sequence of 18
CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC bisulphite, of genes associated with tumour suppression and oncogenes
CC having a sequence taken from 536 (actually 533 since numbers 408, 458 and
CC 500 are missing from the sequence listing) sequences (Ss) and sequences
CC complementary to (Ss). The nucleic acid may be a peptide nucleic acid-
CC oligomer (PNA) of at least 9 nucleotides and may form part of a set of
CC probes for detecting the cytosine methylation state and/or single
CC nucleotide polymorphisms and also to be used in an array for analysing
CC diseases associated with CpG dinucleotides e.g. cancers and tumours. The
CC probes can also be used in a method for ascertaining genetic and/or
CC epigenetic parameters for the diagnosis and/or therapy of existing
CC diseases or the predisposition to specific diseases, by analysing
CC cytosine methylations. The parameters may be compared to another set of
CC genetic and/or epigenetic parameters, the differences serving as basis
CC for diagnosis and/or prognosis events which are disadvantageous to
CC patients. The present sequence is one of the 533 genomic sequences
CC derived from tumour suppressor genes and oncogenes. Sequences with even
CC numbered Seq ID numbers are the complementary sequence of the
CC corresponding odd numbered sequence (e.g. ID 2 and ID1, ID 536 and ID
CC 535, except for those whose partner sequence is missing). Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 26997 BP; 7650 A; 354 C; 5924 G; 13069 T; 0 U; 0 Other;

Query Match 7.0%; Score 322.2; DB 4; Length 26997;
Best Local Similarity 66.0%; Pred. No. 6.3e-46;
Matches 498; Conservative 0; Mismatches 253; Indels 4; Gaps 2;



```
RESULT 7
AAS46747/c
ID   AAS46747 standard; DNA; 26997 BP.
XX
AC   AAS46747;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Tumour suppressor gene derived chemically modified sequence #471.
XX
KW   Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer;
KW   tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW   cytosine methylation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200168912-A2.
XX
PD   20-SEP-2001.
XX
PF   15-MAR-2001; 2001WO-EP002955.
XX
PR   15-MAR-2000; 2000DE-01013847.
PR   06-APR-2000; 2000DE-01019058.
PR   07-APR-2000; 2000DE-01019173.
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2001-602752/68.
XX
PT   Fragments of chemically modified genes associated with tumor suppressor
PT   genes and oncogenes, useful in designing primers and probes for analyzing
PT   diseases associated with cytosine methylation state e.g. cancer.
XX
PS   Claim 1; SEQ ID NO 471; 27pp; English.
XX
CC   The invention relates to a nucleic acid comprising a sequence of 18
CC   bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC   bisulphite, of genes associated with tumour suppression and oncogenes
CC   having a sequence taken from 536 (actually 533 since numbers 408, 458 and
CC   500 are missing from the sequence listing) sequences (Ss) and sequences
CC   complementary to (Ss). The nucleic acid may be a peptide nucleic acid-
CC   oligomer (PNA) of at least 9 nucleotides and may form part of a set of
CC   probes for detecting the cytosine methylation state and/or single
CC   nucleotide polymorphisms and also to be used in an array for analysing
CC   diseases associated with CpG dinucleotides e.g. cancers and tumours. The
CC   probes can also be used in a method for ascertaining genetic and/or
CC   epigenetic parameters for the diagnosis and/or therapy of existing
CC   diseases or the predisposition to specific diseases, by analysing
CC   cytosine methylations. The parameters may be compared to another set of
CC   genetic and/or epigenetic parameters, the differences serving as basis
CC   for diagnosis and/or prognosis events which are disadvantageous to
CC   patients. The present sequence is one of the 533 genomic sequences
CC   derived from tumour suppressor genes and oncogenes. Note: The sequence
CC   data for this patent did not form part of the printed specification, but
CC   was obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 26997 BP; 7233 A; 354 C; 6190 G; 13220 T; 0 U; 0 Other;

  Query Match             6.7%;  Score 311.2;  DB 4;  Length 26997;
  Best Local Similarity   76.9%;  Pred. No. 4.8e-44;
  Matches  409;  Conservative    0;  Mismatches  113;  Indels   10;  Gaps    2;

Qy        203 CATCTCAGAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACA 262
              | | ||| ||| | | ||       | ||||||||||||||||||| | | || || | | |
Db      21339 CTTTTCATAAATACATAATATTCCATAATATATATATATATATACACACACGCATATATA 21280

Qy        263 CACATATATATATACACATATATATACACATATATACATATATACACATATATACACATA 322
              ||||||||   | ||| ||| ||||| | |||||||||||||| ||| ||||||| |  ||
Db      21279 CACATATACCCACACATATACATATATATATATATACATATACACATATATATATCTA 21220

Qy        323 TATATACACATACATATGT-----ACACATATATATACACATATGTATACACATATATAC 377
              |||||||||| | |||  |      | | ||| | ||| |||||| |||| ||| ||||||
Db      21219 CATATACACACATATACATATATAATATATACACATATACATATATATATACACATATAC 21160

Qy        378 ACATATATACACACATATATACACATATATACACACATATATACACATATATACACATAT 437
              ||||||| | || || ||||| ||||||||| | |||||||||||| |||| ||| |||
Db      21159 ACATATACATACGCACATATAAACATATATATATACATATATACACGTATACACATATAC 21100

Qy        438 ATACACATATACACATATATACACATATATACATATATACACATATATATA-----ATAT 492
              ||| ||| ||| | ||||||||| ||||| ||||||| | | ||| |||||      |||
Db      21099 ATATACACATATATATATATACATATATACACATATACATATATACATATATATATAC 21040

Qy        493 ACACACATATATATACACATATATACACACATATATACACATATATACACATATATATAC 552
              | | | | | ||||||| ||||| | | | ||||||| ||||||||||||||||| |||
Db      21039 ATATATACACATATACATATATACATATATATATATATACATATATACACATATACATAT 20980

Qy        553 ACATATATACACATATATACATATATACACATATATATACATATATACACATATATACAT 612
              | | ||||| | ||||||||||||||||||||||| ||| ||| ||| | ||||||| |
Db      20979 ATACATATATATATATATACATATATACACATATACATATATACATATATATATATAC 20920

Qy        613 ATATACACATATATACATATATACACACATATATACACATACATATACACACACATAGAT 672
              ||||| ||| |||||||||||||||| | ||||||| | ||| ||||| | | | ||| ||
Db      20919 ATATATACACATATACATATATACATATATATATATATATATATATATATATATACAT 20860

Qy        673 ATACATATATATACACATATATATACGTATATATATGTATATATATATGCTC 724
              ||| |||||||||| |||||| ||  ||||||||| |||||||||| | |
Db      20859 ATATATATATATACATATATATCTATATATATATATATATATATATATACAC 20808

RESULT 8
ADC87572
ID   ADC87572 standard; DNA; 1990 BP.
XX
AC   ADC87572;
XX
DT   01-JAN-2004  (first entry)
XX
DE   Human GPCR gene SEQ ID NO:2025.
XX
KW   ds; gene; human; GPCR;
KW   guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1270724-A2.
XX
PD   02-JAN-2003.
XX
PF   18-JUN-2002; 2002EP-00013517.
XX
PR   18-JUN-2001; 2001JP-00246789.
XX
PA   (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA   (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI   Suwa M,  Asai K,  Akiyama Y,  Aburatani H;
XX
DR   WPI; 2003-315783/31.
DR   P-PSDB; ADC87573.
XX
PT   New polynucleotide, useful for preparing a composition for treating a
PT   patient in need of increased or suppressed activity or expression of the
PT   guanosine triphosphate-binding protein coupled receptor.
XX
PS   Claim 1; SEQ ID NO 2025; 28pp; English.
XX
CC   The invention relates to a novel polynucleotide encoding a guanosine
CC   triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
```

```
CC   the invention may have a use in gene therapy. The polynucleotide and
CC   polypeptide are useful for preparing a composition for treating a patient
CC   in need of increased or suppressed activity or expression of the
CC   guanosine triphosphate-binding protein coupled receptor. The
CC   polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC   invention.
XX
SQ   Sequence 1990 BP; 833 A; 366 C; 180 G; 611 T; 0 U; 0 Other;

  Query Match             6.7%;  Score 308;  DB 10;  Length 1990;
  Best Local Similarity   80.2%;  Pred. No. 1.5e-43;
  Matches  401;  Conservative    0;  Mismatches   90;  Indels    9;  Gaps    3;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | || | | |||| |||||||| | | | | ||||||| || ||||||||||||||||||
Db        826 TATACACACATATGTATATATACACATATATACACACATATGTATATATACACATATATA 885

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
               |||||||| || ||||| ||| ||||| |||||||| |||| | | ||||||| |||||||
Db        886 CACACATATGTATATATACACATATATACACACATATGTATATATACACATATATACACA 945

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||| || | |||   ||| | | |  || |  |||||| ||||||||||||||||| |
Db        946 CATATGTATATATACACATATATACACGTATATGTATATATACACATATATACACATACA 1005

Qy        407 TACACA----CATATATACACATATATACACATATATACACATATACACATATATACACA 462
              ||||||     ||||| ||||||||||||||||| ||||||||||||  ||||||||||
Db       1006 TACACATATGTATATACACACATATATACACATACATACACATATACGTATATATACACA 1065

Qy        463 TATATACATATATACACATATATATAATATACACACATATATATACACATATATACACAC 522
              |||| ||||| |||||||||||  | ||||| |||||||||| ||| ||   || | ||
Db       1066 TATACACATACATACACATATACGT-ATATATACACATATATACACATATGCGTATATAC 1124

Qy        523 ATATATACACATATATACACATATATATACACATATATACACATATATACATATATACAC 582
              ||||||||||||||||||||    ||||||  |||||||||||||||||||| ||||||
Db       1125 ATATATACACATATATACAC----ATATACGTATATATACACATATATACACATATACGT 1180

Qy        583 ATATATATACATATATACACATATATACATATATACACATATATACATATATACACACAT 642
              ||||||| |||||||||||||||||||||| |||||  ||||| |||||||||||||| ||
Db       1181 ATATATACACATATATACACATATATACACATATATGTATATACACATATATACACATAT 1240

Qy        643 ATATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTAT 702
              |||| | | ||||||| ||||| |||  |||| ||| ||||| |||  ||| ||  |||
Db       1241 GTATATATACACATATATACACATATATGTATATATACATATATACACGTATGTATATAT 1300

Qy        703 ATATATGTATATATATATGC 722
              | | || ||| ||||||| |
Db       1301 ACACATATATGTATATATAC 1320


RESULT 9
AAA63785
ID   AAA63785 standard; DNA; 15418 BP.
XX
AC   AAA63785;
XX
DT   04-DEC-2000  (first entry)
XX
DE   Nucleotide sequence of the human TERT promoter and gene from pGRN142.
XX
KW   Telomerase reverse transcriptase; TERT; promoter; telomerase complex;
KW   telomere length; hyperplasic disease; cancer; oncolytic virus;
KW   cis-acting transcriptional control sequence; viral replication;
KW   cell proliferation; aging; immunological disorder; infertility; ss.
XX
OS   Homo sapiens.
XX
PN   WO200046355-A2.
XX
PD   10-AUG-2000.
XX
PF   04-FEB-2000; 2000WO-US003104.
XX
PR   04-FEB-1999;   99US-00244438.
XX
PA   (GERO-) GERON CORP.
XX
PI   Morin GB,  Lichtsteiner S,  Vasserot A,  Adams R,  Cardoza LM;
PI   Lebkowski JS;
XX
DR   WPI; 2000-532898/48.
XX
PT   New polynucleotides comprising cis-acting transcriptional control
PT   sequences, e.g. promoter sequence, of telomerase reverse transcriptase
PT   genes, useful in the treatment of cancer.
XX
PS   Claim 9; Page 58-61; 63pp; English.
XX
CC   The present sequence represents the human telomerase reverse
CC   transcriptase (TERT) promoter and gene. TERT is part of the telomerase
CC   complex responsible for maintaining telomere length and increasing
CC   replicative capacity of progenitor cells. Telomerase activity is turned
CC   off in mature differentiated cells, but is turned back on again in
CC   hyperplasic diseases, including many cancers. The polynucleotide
CC   comprises cis-acting transcriptional control sequences, e.g. promoter
CC   sequence. These promoter sequences are used to produce oncolytic viruses,
CC   in which a toxin or genetic element essential for viral replication is
CC   placed under the control of a TERT promoter. As a result, the virus
CC   replicates preferentially in cells expressing TERT, and selectively lyses
CC   cancer cells. The oncolytic viruses are useful for treating cancer in
CC   humans or animals. The TERT promoter sequences are useful in the
CC   treatment of cancer and other diseases of cell proliferation such as
CC   degenerative and aging processes and diseases of aging, immunological
CC   disorders, or infertility
XX
SQ   Sequence 15418 BP; 4518 A; 3797 C; 3765 G; 3338 T; 0 U; 0 Other;

  Query Match             6.4%;  Score 296;  DB 3;  Length 15418;
  Best Local Similarity   69.2%;  Pred. No. 1.9e-41;
  Matches  479;  Conservative    0;  Mismatches  200;  Indels   13;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||| |||||||||||   ||| || | |||||||||||| ||||||||| || ||||||
Db        646 AGTTCGAGACCAGCCTGGCCAATATAGCGAAACCCCATCTGTACTAAAAACACAAAAATT 705

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   ||  |||||| ||| |||||||||||||| ||||  ||||||  ||||||||||
Db        706 AGCTGGGTGTGGTGACACATGCCTGTAATCCCAGGTACTCAGGAGGCTAAGGCAGGAGAA 765

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              |  |||||||  |  ||  | || |||     |  ||||||    ||| ||||||||||
Db        766 TTGCTTGAACTGGGAGGCAGAGGTGGT-----GAGCCAAGATTGCACCAGTGCACTCCAG 820

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  ||  || |||| ||||    ||||||| ||| ||| |||   | |  ||||| |||
Db        821 CCTTGGTGAC-AGAGTGAAACTCCATCTCAAAAACAAACAAACAAAATACATATACATAA 879

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              ||||||||  | | | | | ||| | ||| ||||||| ||| ||||||| |  |||||||
Db        880 ATATATATGCACATATATATACATATATAAATATATATACACATATATAAATCTATATAC 939

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACAT-----ATATATA 354
              |||||||| ||||| ||| ||||| ||  | |||||||| || ||||     ||||||
Db        940 ATATATACATATATACACATATATAAATCTATATACATATATATACATATATAATATATT 999

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              ||||||  || | | ||| ||| | | |||||||  | ||| |||||  ||||| | | |||
Db       1000 TACATATATAAATATATACATATATAAATATACATATATAAATACATATATAAATATACA 1059

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              |||||| | ||| ||| | | |||||||  ||||| | |||||  | | ||| ||| | |||
Db       1060 TATATAAATATACATATATAAATATACATATATAAATATATACATATATAAATATACATA 1119

Qy        475 TACACATATATATA-ATATACACACATATATATACACATATATACACACATATATACACA 533
              || | |||||||||  |||||  | | |||||  ||| |||  ||||  ||| |||||||||  |
Db       1120 TATAAATATATATACATATATAAATATATAAATATACAAGTATATACAAATATATACATA 1179
```

```
Qy        534 TATATACACATATATATACACATATATACACATATATACATATATACACATATATATACA 593
              |||| |   |||||  || | | ||||| | ||| ||| ||| | | || | |   |
Db       1180 TATAAATGTATATACGTATATACATATATATATAAATATATAAAAAAACTTTTGGCTGGG 1239

Qy        594 TATATACACATATATACATATATACACATATATACATATATACACACATATATACACATA 653
              | |   || || |      | | | |  | |     ||  | || ||| | | |||||
Db       1240 CACCTTTCCAAATCTCATGGCACATATAAGTCTCATGGTAACCTCAAATAAAAAAACATA 1299

Qy        654 CATATACACACACATAGATATACATATATATA 685
              |          || | ||| | |   ||| |
Db       1300 TAACAGATACACCAAAAATAAAAACCAATAAA 1331
```

RESULT 10
AAL38601
ID AAL38601 standard; DNA; 15418 BP.
XX
AC AAL38601;
XX
DT 16-AUG-2002 (first entry)
XX
DE Human TERT promoter and upstream sequence.
XX
KW Cytostatic; glycosyltransferase; tumour; cell-surface carbohydrate;
KW tissue specific; transcriptional control element; cancer cell;
KW gene therapy; human TERT; telomerase reverse transcriptase; ds.
XX
OS Homo sapiens.
XX
PN WO200242468-A2.
XX
PD 30-MAY-2002.
XX
PF 26-NOV-2001; 2001WO-US044306.
XX
PR 27-NOV-2000; 2000US-0253395P.
XX
PA (GERO-) GERON CORP.
XX
PI Schiff MJ;
XX
DR WPI; 2002-479954/51.
XX
PT New polynucleotide encoding glycosyltransferase enzymes including histo
PT blood group transferase useful for treating conditions associated with
PT hyperproliferation, such as cancers and other neoplasias.
XX
PS Claim 9; Page 17-20; 49pp; English.
XX
CC The invention relates to a polynucleotide comprising an encoding sequence
CC for a glycosyltransferase, under control of a heterologous tumour
CC specific or tissue specific transcriptional control element, where
CC expression of the polynucleotide in a human cell causes the cell to
CC express a cell-surface carbohydrate determinant to which some or all
CC humans have a naturally occurring antibody. The polynucleotide of the
CC invention is useful for killing cancer cells and for preparing a
CC medicament for the treatment of cancer. The polynucleotide sequence of
CC the invention can be used to treat disorders by gene therapy. This
CC polynucleotide sequence represents the human TERT (telomerase reverse
CC transcriptase) promoter and upstream sequence related to the invention
XX
SQ Sequence 15418 BP; 4518 A; 3797 C; 3765 G; 3338 T; 0 U; 0 Other;

```
  Query Match             6.4%;  Score 296;  DB 6;  Length 15418;
  Best Local Similarity   69.2%;  Pred. No. 1.9e-41;
  Matches  479;  Conservative    0;  Mismatches  200;  Indels   13;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||| |||||||||||  ||| || | |||||||||||| |||||||| || ||||||
Db        646 AGTTCGAGACCAGCCTGGCCAATATAGCGAAACCCCATCTGTACTAAAAACACAAAAATT 705

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   ||  |||||| ||| |||||||||||||| ||||  ||||||  ||||||||||
Db        706 AGCTGGGTGTGGTGACACATGCCTGTAATCCCAGGTACTCAGGAGGCTAAGGCAGGAGAA 765

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              | |||||||  |  ||  | || |||     | .||||||    ||| ||||||||||
Db        766 TTGCTTGAACTGGGAGGCAGAGGTGGT-----GAGCCAAGATTGCACCAGTGCACTCCAG 820

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  ||  || |||| ||||   ||||||| ||| ||| |||   | |  ||||| |||
Db        821 CCTTGGTGAC-AGAGTGAAACTCCATCTCAAAAACAAACAAACAAAATACATATACATAA 879

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              ||||||||  | | | | ||| | ||| |||||||| ||| |||||||  |  |||||||
Db        880 ATATATATGCACATATATATACATATATAAATATATATACACATATATAAATCTATATAC 939

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACAT-----ATATATA 354
              |||||||| ||||| ||| ||||| ||  | |||||||| || ||||      ||||||
Db        940 ATATATACATATATACACATATATAAATCTATATACATATATATACATATATAATATATT 999

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              |||||| || | | ||| ||| | | |||||||  | ||| ||||| ||||| | | |||
Db       1000 TACATATATAAATATATACATATATAAATATACATATATAAATACATATATAAATATACA 1059

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              |||||| | ||| ||| | | |||||||  ||||| | ||||| | | ||| ||| | |||
Db       1060 TATATAAATATACATATATAAATATACATATATAAATATATACATATATAAATATACATA 1119

Qy        475 TACACATATATATA-ATATACACACATATATATACACATATATACACACATATATACACA 533
              || | |||||||||  ||||| | | ||||| ||| |||  |||| ||| |||||||||  |
Db       1120 TATAAATATATATACATATATAAATATATAAATATACAAGTATATACAAATATATACATA 1179

Qy        534 TATATACACATATATATACACATATATACACATATATACATATATACACATATATATACA 593
              |||| |   |||||  || | | ||||| | ||| ||| ||| | | || | |   |
Db       1180 TATAAATGTATATACGTATATACATATATATATAAATATATAAAAAAACTTTTGGCTGGG 1239

Qy        594 TATATACACATATATACATATATACACATATATACATATATACACACATATATACACATA 653
              | |   || || |      | | | |  | |     ||  | || ||| | | |||||
Db       1240 CACCTTTCCAAATCTCATGGCACATATAAGTCTCATGGTAACCTCAAATAAAAAAACATA 1299

Qy        654 CATATACACACACATAGATATACATATATATA 685
              |          || | ||| | |   ||| |
Db       1300 TAACAGATACACCAAAAATAAAAACCAATAAA 1331
```

RESULT 11
ABS54997
ID ABS54997 standard; DNA; 15418 BP.
XX
AC ABS54997;
XX
DT 10-DEC-2002 (first entry)
XX
DE Lambda clone containing human TERT genomic insert.
XX
KW Telomerase reverse transcriptase; TERT; replication-conditional virus;
KW adenovirus replication gene; cancer cell; lung; pancreatic cancer;
KW medulloblastoma; cervical carcinoma; fibrosarcoma; osteosarcoma;
KW cytolysis; replication defective adenovirus vector; congenital defect;
KW proinflammatory; antiinflammatory; heterologous effector gene;
KW cancer therapy; cytostatic; gene therapy; lambda clone; human; ds.
XX
OS Bacteriophage lambda.
OS Homo sapiens.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT misc_feature 1. .43
FT /*tag= a
FT /note= "From Lambda clone"
FT misc_feature 44. .15375
FT /*tag= b
FT /note= "Human telomerase reverse transcriptase (hTERT)
FT sequence including the hTERT promoter and upstream

```
FT                   sequence"
FT   misc_feature    15376. .15418
FT                   /*tag=  c
FT                   /note= "From Lambda clone"
XX
PN   WO200253760-A2.
XX
PD   11-JUL-2002.
XX
PF   17-DEC-2001; 2001WO-US048785.
XX
PR   18-DEC-2000; 2000US-0256418P.
XX
PA   (GERO-) GERON CORP.
XX
PI   Irving JM,  Lebkowski JS;
XX
DR   WPI; 2002-723123/78.
XX
PT   Novel replication-conditional virus useful for cytolysis of target cells
PT   e.g. cancer cells and preparing a medicament for treating cancer,
PT   comprises heterologous replication element in an adenovirus-based
PT   construct.
XX
PS   Claim 11; Page 26-29; 32pp; English.
XX
CC   The present invention relates to a new replication-conditional virus with
CC   a genome comprising adenovirus replication genes and one or more
CC   heterologous gene(s) that functionally replaces one or more adenovirus
CC   gene(s) required for replication or assembly of the virus. The invention
CC   is useful for killing a cancer cell (such as lung, pancreatic cancer,
CC   medulloblastoma, cervical carcinoma, fibrosarcoma or osteosarcoma),
CC   killing a cell expressing TERT (telomerase reverse transcriptase), and in
CC   preparing a medicament for treating cancer and a condition associated
CC   with increased expression of TERT in affected cells, in a subject. The
CC   invention is also useful for cytolysis of specific target cells. The
CC   invention is further useful for producing replication defective
CC   adenovirus vector which is useful for transient expression of a
CC   heterologous therapeutic gene to correct a congenital defect, introducing
CC   proinflammatory or antiinflammatory activity, enhancing telomerase
CC   function, and delivering heterologous effector genes that induce killing
CC   of the transduced cells. The invention is more safe for use in cancer
CC   therapy. The present nucleic acid sequence represents the human TERT
CC   sequence contained within a lambda clone sequence of the invention
XX
SQ   Sequence 15418 BP; 4518 A; 3797 C; 3765 G; 3338 T; 0 U; 0 Other;

  Query Match          6.4%;  Score 296;  DB 6;  Length 15418;
  Best Local Similarity   69.2%;  Pred. No. 1.9e-41;
  Matches  479;  Conservative    0;  Mismatches  200;  Indels   13;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||| ||||||||||  ||| || | ||||||||||| |||||||| || ||||||
Db        646 AGTTCGAGACCAGCCTGGCCAATATAGCGAAACCCCATCTGTACTAAAAACACAAAAATT 705

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   ||  |||||| ||| ||||||||||||| ||||  ||||||  ||||||||||
Db        706 AGCTGGGTGTGGTGACACATGCCTGTAATCCCAGGTACTCAGGAGGCTAAGGCAGGAGAA 765

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              |  |||||||  |  ||  | ||  |||     |   ||||||    ||| ||||||||||
Db        766 TTGCTTGAACTGGGAGGCAGAGGTGGT-----GAGCCAAGATTGCACCAGTGCACTCCAG 820

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  ||  || |||| ||||   ||||||| ||| ||| |||    | |  ||||| |||
Db        821 CCTTGGTGAC-AGAGTGAAACTCCATCTCAAAAACAAACAAACAAAATACATATACATAA 879

Qy        240 ATATATATATATACACACACACACACATATATATATACACATATATATACACATATATAC 299
              ||||||||  | | | | | ||| | ||| ||||||| ||| ||||||| |  |||||||
Db        880 ATATATATGCACATATATATACATATATAAATATATATACACATATATAAATCTATATAC 939

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACAT-----ATATATA 354
              ||||||||| ||||| ||| ||||| ||  | |||||||| || ||||     ||||||

Db        940 ATATATACATATATACACATATATAAATCTATATACATATATATACATATATAATATATT 999

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              |||||| || | | ||| ||| | | ||||||| | ||| ||||| ||||| | | |||
Db       1000 TACATATATAAATATATACATATATAAATATACATATATAAATACATATATAAATATACA 1059

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              |||||| | ||| ||| | | ||||||| ||||| | ||||| | | ||| ||| | |||
Db       1060 TATATAAATATACATATATAAATATACATATATAAATATATACATATATAAATATACATA 1119

Qy        475 TACACATATATATA-ATATACACACATATATATACACATATATACACACATATATACACA 533
              || | |||||||||| ||||| | | ||||| ||| |||  |||| ||| |||||||||| |
Db       1120 TATAAATATATATACATATATAAATATATAAATATACAAGTATATACAAATATATACATA 1179

Qy        534 TATATACACATATATATACACATATATACACATATATACATATATACACATATATATACA 593
              |||| |   |||||  || | | ||||| | ||| ||| ||| | | || | |   |
Db       1180 TATAAATGTATATACGTATATACATATATATATAAATATATAAAAAAACTTTTGGCTGGG 1239

Qy        594 TATATACACATATATACATATATACACATATATACATATATACACACATATATACACATA 653
              | |   || || |       | | | |  | |     ||  | || ||| | | |||||
Db       1240 CACCTTTCCAAATCTCATGGCACATATAAGTCTCATGGTAACCTCAAATAAAAAAACATA 1299

Qy        654 CATATACACACACATAGATATACATATATATA 685
              |           || | ||| | |   ||| |
Db       1300 TAACAGATACACCAAAAATAAAAACCAATAAA 1331


RESULT 12
AAL38595
ID   AAL38595 standard; DNA; 15418 BP.
XX
AC   AAL38595;
XX
DT   16-AUG-2002  (first entry)
XX
DE   DNA of plasmid pGRN144 containing human hTERT gene.
XX
KW   Anticonvulsant; cerebroprotective; vasotropic; nootropic; stem cell;
KW   neuroprotective; antiparkinsonian; antiinflammatory; undifferentiated;
KW   cardiant; transcriptional control element; human embryonic; tissue graft;
KW   regenerative medicine; tissue reconstitution; neural progenitor cell;
KW   nerve system; epilepsy; stroke; ischaemia; Huntington's disease;
KW   Parkinson's disease; multiple sclerosis; leukodystrophy; neuritis;
KW   hepatocyte; liver damage; cardiomyocyte; cryoinjury; cardiac muscle;
KW   pGRN144; human hTERT gene; chimeric; ds.
XX
OS   Homo sapiens.
OS   Unidentified.
OS   Chimeric.
XX
PN   WO200242445-A2.
XX
PD   30-MAY-2002.
XX
PF   26-NOV-2001; 2001WO-US044309.
XX
PR   27-NOV-2000; 2000US-0253357P.
PR   27-NOV-2000; 2000US-0253443P.
PR   13-FEB-2001; 2001US-00783203.
XX
PA   (GERO-) GERON CORP.
XX
PI   Gold JD,  Lebrowski JS;
XX
DR   WPI; 2002-479952/51.
XX
PT   Depleting a cell e.g., human embryonic stem cell population of
PT   undifferentiated stem cells (UC) for use in regenerative medicine
PT   comprises genetically altering UC in a population to express nucleic acid
PT   encoding a lethal product.
XX
PS   Example 9; Page 62-66; 67pp; English.
XX
```

CC The invention relates to a system for depleting a cell population of
CC undifferentiated stem cells, by introducing nucleic acids of structure P-
CC X, where X is a nucleic acid encoding product that is lethal to a cell in
CC which it is expressed; or renders a cell in which it is expressed
CC susceptible to the lethal effect of an external agent, and P is a
CC transcriptional control element causing X to be preferentially expressed
CC in the undifferentiated stem cells. The system is used for depleting a
CC cell population (preferably, human embryonic stem cells) of
CC undifferentiated stem cells. A population of differentiated cells is
CC useful in regenerative medicine, and for preparing antibodies and cDNA
CC libraries that are specific for a differentiated phenotype. The cell
CC populations are also useful for drug screening and therapeutic
CC applications. The differentiated cells are useful for tissue
CC reconstitution or regeneration in a human patient in need of treatment.
CC The cells are administered in a manner that permits to graft to the
CC intended tissue side and reconstitute or regenerate the functionally
CC deficient area. The neural progenitor cells are useful for treating acute
CC or chronic damage to the nerve system e.g. epilepsy, stroke, ischaemia,
CC Huntington's disease, Parkinson's disease, multiple sclerosis,
CC leukodystrophies, neuritis, etc. The hepatocytes and hepatocyte
CC precursors are useful for assessing animal models for ability to repair
CC liver damage. The cardiomyocyte population is useful for assessing animal
CC models for cryoinjury, regenerating cardiac muscle and to treat
CC insufficient cardiac function. This polynucleotide sequence represents
CC the DNA of plasmid pGRN144 containing the human hTERT gene relating to
CC the invention
XX
SQ Sequence 15418 BP; 4519 A; 3797 C; 3764 G; 3338 T; 0 U; 0 Other;

Query Match 6.4%; Score 296; DB 6; Length 15418;
Best Local Similarity 69.2%; Pred. No. 1.9e-41;
Matches 479; Conservative 0; Mismatches 200; Indels 13; Gaps 5;

```
Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||| ||||||||||  ||| || | ||||||||||||| ||||||||| || ||||||
Db        646 AGTTCGAGACCAGCCTGGCCAATATAGCGAAACCCCATCTGTACTAAAAACACAAAAATT 705

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   ||  |||||| ||| ||||||||||||||| ||||  ||||||  ||||||||||
Db        706 AGCTGGGTGTGGTGACACATGCCTGTAATCCCAGGTACTCAGGAGGCTAAGGCAGGAGAA 765

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              |  |||||||  |  ||  | || |||      |   ||||||    ||| ||||||||||
Db        766 TTGCTTGAACTGGGAGGCAGAGGTGGT-----GAGCCAAGATTGCACCAGTGCACTCCAG 820

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  ||  || |||| ||||   ||||||| ||| ||| |||   | |  ||||| |||
Db        821 CCTTGGTGAC-AGAGTGAAACTCCATCTCAAAAACAAACAAACAAAATACATATACATAA 879

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              ||||||||  | | | | | ||| | ||| ||||||| ||| ||||||| |  |||||||
Db        880 ATATATATGCACATATATATACATATATAAATATATATACACATATATAAATCTATATAC 939

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACAT-----ATATATA 354
              |||||||||  ||||| ||| ||||| ||  | |||||||| || ||||      ||||||
Db        940 ATATATACATATATACACATATATAAATCTATATACATATATATACATATATAATATATT 999

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              ||||||  || | | ||| ||| | | ||||||| | ||| ||||| ||||| | | |||
Db       1000 TACATATATAAATATATACATATATAAATATACATATATAAATACATATATAAATATACA 1059

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              |||||| | ||| ||| | | ||||||| ||||| | ||||| | | ||| ||| | |||
Db       1060 TATATAAATATACATATATAAATATACATATATAAATATATACATATATAAATATACATA 1119

Qy        475 TACACATATATATA-ATATACACACATATATATACACATATATACACACATATATACACA 533
              || | |||||||||  |||| | | |||| ||| |||  |||| ||| |||||||||  |
Db       1120 TATAAATATATATACATATATAAATATATAAATATACAAGTATATACAAATATATACATA 1179

Qy        534 TATATACACATATATATACACATATATACACATATATACATATATACACATATATATACA 593
              |||| |   |||||  || | | |||| | ||| ||| ||| | | || | |   |
Db       1180 TATAAATGTATATACGTATATACATATATATATAAATATATAAAAAAACTTTTGGCTGGG 1239
```

```
Qy         594 TATATACACATATATACATATATACACATATATACATATATACACACATATATACACATA 653
               |  |   || || |      | | | | | |     ||  | || ||| | | |||||
Db        1240 CACCTTTCCAAATCTCATGGCACATATAAGTCTCATGGTAACCTCAAATAAAAAAACATA 1299

Qy         654 CATATACACACACATAGATATACATATATATA 685
               |          || | ||| | |   |||. |
Db        1300 TAACAGATACACCAAAAATAAAAACCAATAAA 1331
```

RESULT 13
ADC21253
ID ADC21253 standard; DNA; 15418 BP.
XX
AC ADC21253;
XX
DT 18-DEC-2003 (first entry)
XX
DE Lambda clone (lambdaGphi5) containing human TERT DNA.
XX
KW Viral vector; heterologous control element; gene expression;
KW human telomerase reverse transcriptase; hTERT; tumour specific gene;
KW cell death; transcriptional control element;
KW human telomerase RNA component; hTR; cancer cell; liver cancer;
KW prostate cancer; muscle cancer; neural cell cancer; lung cancer;
KW pancreatic cancer; medulloblastoma; cervical carcinoma; fibrosarcoma;
KW osteosarcoma; lambda clone; lambdaGphi5; human; ds.
XX
OS Synthetic.
OS Homo sapiens.
OS Bacteriophage lambda.
XX
PN US2003099616-A1.
XX
PD 29-MAY-2003.
XX
PF 25-JUL-2002; 2002US-00206447.
XX
PR 25-JUL-2001; 2001US-0308029P.
XX
PA (IRVI/) IRVING J M.
PA (KARP/) KARPF D B.
PA (SCHI/) SCHIFF J M.
XX
PI Irving JM, Karpf DB, Schiff JM;
XX
DR WPI; 2003-730140/69.
XX
PT New dual specificity vectors driven by the telomerase promoter, useful
PT for killing or slowing the growth of tumor cells, or for treating cancer,
PT e.g. liver cancer, prostate cancer, lung cancer, or pancreatic cancer.
XX
PS Example 1; Page 13-20; 25pp; English.
XX
CC The present invention relates to a viral vector comprising first and
CC second genes controlled by heterologous control elements, where the first
CC gene is preferentially expressed in cells expressing human telomerase
CC reverse transcriptase (hTERT), and another gene under the control of a
CC heterologous transcriptional control element for a tissue or tumour
CC specific gene other than TERT, and where transduction of the vector into
CC a mammalian cell expressing TERT causes the death of the cell or its
CC progeny. In particular, the second gene may be under the control of a
CC transcriptional control element for a tissue specific gene selected from
CC albumin, alpha-fetoprotein, prostate-specific antigen, mitochondrial
CC creatine kinase, myelin basic protein, glial fibrillary acidic protein,
CC and neuron-specific enolase. The second gene may be under the control of
CC a transcriptional control element for a human telomerase RNA component
CC (hTR). The vector of the invention is useful for killing a cancer cell,
CC or treating a subject for a condition associated with increased
CC expression of TERT in affected cells. The vector is also useful in the
CC preparation of a medicament for treatment of a condition associated with
CC increased expression of TERT, particularly cancer, and especially liver
CC cancer, prostate cancer, muscle cancer, neural cell cancer, lung cancer,
CC pancreatic cancer, medulloblastoma, cervical carcinoma, fibrosarcoma, and

```
CC   osteosarcoma. The present sequence represents a lambda clone
CC   (lambdaGphi5) containing human TERT DNA.
XX
SQ   Sequence 15418 BP; 4518 A; 3797 C; 3765 G; 3338 T; 0 U; 0 Other;

  Query Match             6.4%;  Score 296;  DB 10;  Length 15418;
  Best Local Similarity   69.2%;  Pred. No. 1.9e-41;
  Matches  479;  Conservative    0;  Mismatches  200;  Indels   13;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||| |||||||||||  ||| || | ||||||||||||| |||||||| || ||||||
Db        646 AGTTCGAGACCAGCCTGGCCAATATAGCGAAACCCCATCTGTACTAAAAACACAAAAATT 705

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   ||  |||||| ||| ||||||||||||| ||||  ||||||  ||||||||||
Db        706 AGCTGGGTGTGGTGACACATGCCTGTAATCCCAGGTACTCAGGAGGCTAAGGCAGGAGAA 765

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              |  |||||||  |  ||  | || |||     |   ||||||    ||| ||||||||||
Db        766 TTGCTTGAACTGGGAGGCAGAGGTGGT-----GAGCCAAGATTGCACCAGTGCACTCCAG 820

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  ||  || |||| ||||   ||||||| ||| ||| |||   | |  ||||| |||
Db        821 CCTTGGTGAC-AGAGTGAAACTCCATCTCAAAAACAAACAAACAAAATACATATACATAA 879

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              ||||||||  | | | | | ||| | ||| ||||||| ||| ||||||| |  |||||||
Db        880 ATATATATGCACATATATATACATATATAAATATATATACACATATATAAATCTATATAC 939

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACAT-----ATATATA 354
              |||||||| ||||| ||| ||||| ||  | |||||||| || ||||     ||||||
Db        940 ATATATACATATATACACATATATAAATCTATATACATATATATACATATATAAATATATT 999

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              ||||.|| || | | ||| ||| | | ||||||| | ||| ||||| ||||| | | |||
Db       1000 TACATATATAAATATATACATATATAAATATACATATATAAATACATATATAAATATACA 1059

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              |||||| | ||| ||| | | ||||||| ||||| | ||||| | | ||| ||| | |||
Db       1060 TATATAAATATACATATATAAATATACATATATAAATATATACATATATAAATATACATA 1119

Qy        475 TACACATATATATA-ATATACACACATATATATACACATATATACACACATATATACACA 533
              || | ||||||||| ||||| | | ||||| ||| |||  |||| ||| |||||||| |
Db       1120 TATAAATATATATACATATATAAATATATAAATATACAAGTATATACAAATATATACATA 1179

Qy        534 TATATACACATATATATACACATATATACACATATATACATATATACACATATATATACA 593
              |||| |   |||||  || | | ||||| | ||| ||| ||| | | || | |   |
Db       1180 TATAAATGTATATACGTATATACATATATATATAAATATATAAAAAAACTTTTGGCTGGG 1239

Qy        594 TATATACACATATATACATATATACACATATATACATATATACACACATATATACACATA 653
              | |   || ||  |      | | | |  | |      ||  | || ||| | | |||||
Db       1240 CACCTTTCCAAATCTCATGGCACATATAAGTCTCATGGTAACCTCAAATAAAAAAACATA 1299

Qy        654 CATATACACACACATAGATATACATATATATA 685
              |  .    || | ||| | |   ||| |
Db       1300 TAACAGATACACCAAAAATAAAAACCAATAAA 1331


RESULT 14
ADC87552/c
ID   ADC87552 standard; DNA; 1486 BP.
XX
AC   ADC87552;
XX
DT   01-JAN-2004  (first entry)
XX
DE   Human GPCR gene SEQ ID NO:2005.
XX
KW   ds; gene; human; GPCR;
KW   guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1270724-A2.
XX
PD   02-JAN-2003.
XX
PF   18-JUN-2002; 2002EP-00013517.
XX
PR   18-JUN-2001; 2001JP-00246789.
XX
PA   (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA   (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI   Suwa M,  Asai K,  Akiyama Y,  Aburatani H;
XX
DR   WPI; 2003-315783/31.
DR   P-PSDB; ADC87553.
XX
PT   New polynucleotide, useful for preparing a composition for treating a
PT   patient in need of increased or suppressed activity or expression of the
PT   guanosine triphosphate-binding protein coupled receptor.
XX
PS   Claim 1; SEQ ID NO 2005; 28pp; English.
XX
CC   The invention relates to a novel polynucleotide encoding a guanosine
CC   triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC   the invention may have a use in gene therapy. The polynucleotide and
CC   polypeptide are useful for preparing a composition for treating a patient
CC   in need of increased or suppressed activity or expression of the
CC   guanosine triphosphate-binding protein coupled receptor. The
CC   polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC   invention.
XX
SQ   Sequence 1486 BP; 517 A; 149 C; 242 G; 578 T; 0 U; 0 Other;

  Query Match             6.4%;  Score 294.8;  DB 10;  Length 1486;
  Best Local Similarity   74.9%;  Pred. No. 2.8e-41;
  Matches  397;  Conservative    0;  Mismatches  127;  Indels    6;  Gaps    2;

Qy        197 AAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACA 256
              | || |   | || | || | | |      || || | | ||  | ||||||| |||||||
Db        999 ATATATACACACATATAATATATACATATTTATACACACATGCACATATATAAATACACA 940

Qy        257 CACACACACATATATATATACACATATATATACACATATATACATATATACACATATATA 316
              | | ||| ||||| | | | | || |||||| ||||||||||||||||||||||||| ||
Db        939 TATATACATATATACACACATATATGTATATATACATATATACACATATGTA 880

Qy        317 CACATATATATACACATACATATGTACACATATATATACACATATGTATACACATATATA 376
              | ||| ||||| ||||| |||   | |  |||||||||||||||||||||| | | |||
Db        879 TATATACATATATACATATATACACATATGTATATATACACATATGTATATATACACATA 820

Qy        377 CACATATATACACACATATATACACATATATACACACATATATACACATATATACACATA 436
              ||||||| || | | | |||||||||| || | | | |||||||||| || | |||
Db        819 CACATATGTATATATACACATACACATATGTATATATACACATACACATATGTATATATA 760

Qy        437 TATACACATATACACATATATACACATATATACATATATACACATATATATA-----ATA 491
              | | ||| |||   |||||||||||| ||  |||||||||||||||| ||     | |
Db        759 CACATACACATATGTATATATACACATACATGTATATATACACATATATGTATATATACA 700

Qy        492 TACACACATATATATACACATATATACACACATATATACACATATATACACATATATATA 551
              | |||||| |||| ||| | | ||| |   |||||||||||| | | | ||| | |||
Db        699 CATACACATGTATACACATACACATATATGTATATATACACATACACATATATACACATA 640

Qy        552 CACATATATACACATATATACATATATACACATATATATACATATATACACATATATACA 611
              |||  |||||||||||||||| ||||||||||||||   |||||||||||||| |||
Db        639 TACACATATACACATATATACACATATACACATATATAAGTATATATACACATATACACA 580

Qy        612 TATATACACATATATACATATATACACACATATATACACATACATAT-ACACACACATAG 670
              ||||   | ||||| |||||||||||| |||  || | ||||| || | ||||| |||
Db        579 TATACGTATATATACACATATATACACATATACGTATATATACACATAATACACATATAC 520

Qy        671 ATATACATATATATACACATATATATACGTATATATATGTATATATATAT 720
              |||| ||| | ||| | | | |||||||||||||||| ||||||| | ||
Db        519 GTATAAATACACATATATACACATATACGTATATATACGTATATACACAT 470
```

```
RESULT 15
AAD50021
ID   AAD50021 standard; DNA; 27067 BP.
XX
AC   AAD50021;
XX
DT   24-MAR-2003  (first entry)
XX
DE   Human secreted protein encoding gene.
XX
KW   Human; secreted protein; cancer; gene therapy; cytostatic; gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       replace(921,T)
FT                   /*tag= a
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(1022,T)
FT                   /*tag= ba
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(1781,T)
FT                   /*tag= b
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(1850,A)
FT                   /*tag= c
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(1882,T)
FT                   /*tag= bb
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(1951,A)
FT                   /*tag= bc
FT                   /standard_name= "Single nucleotide polymorphism"
FT   CDS             2031. .25941
FT                   /*tag= d
FT                   /product= "Human secreted protein"
FT   exon            2031. .2094
FT                   /*tag= e
FT   intron          2095. .5569
FT                   /*tag= f
FT   exon            2132. .2195
FT                   /*tag= bd
FT   intron          2196. .5670
FT                   /*tag= be
FT   variation       replace(2839,G)
FT                   /*tag= g
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(2940,G)
FT                   /*tag= bf
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(3730,A)
FT                   /*tag= h
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(3831,A)
FT                   /*tag= bg
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            5570. .5762
FT                   /*tag= i
FT   exon            5671. .5863
FT                   /*tag= bh
FT   intron          5763. .7571
FT                   /*tag= j
FT   intron          5864. .7672
FT                   /*tag= bi
FT   variation       replace(6631,A)
FT                   /*tag= k
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(6732,A)
FT                   /*tag= bj
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(6945,A)
FT                   /*tag= l
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(6952,T)
FT                   /*tag= m
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(7457,A)
FT                   /*tag= n
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(7558,A)
FT                   /*tag= bk
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            7572. .7741
FT                   /*tag= o
FT   exon            7673. .7761
FT                   /*tag= bl
FT   intron          7742. .10000
FT                   /*tag= p
FT   intron          7762. .9149
FT                   /*tag= bm
FT   variation       replace(7830,A)
FT                   /*tag= q
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(7931,A)
FT                   /*tag= bn
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(8089,C)
FT                   /*tag= r
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(8190,C)
FT                   /*tag= bo
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(8551,T)
FT                   /*tag= s
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(8652,T)
FT                   /*tag= bp
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            9150. .9302
FT                   /*tag= bq
FT   variation       replace(9269,C)
FT                   /*tag= t
FT                   /standard_name= "Single nucleotide polymorphism"
FT   intron          9303. .10101
FT                   /*tag= br
FT   variation       replace(9362,T)
FT                   /*tag= u
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(9370,C)
FT                   /*tag= bs
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(9463,T)
FT                   /*tag= bt
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(9782,T)
FT                   /*tag= v
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(9883,T)
FT                   /*tag= bu
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            10001. .10173
FT                   /*tag= w
FT   exon            10102. .10274
FT                   /*tag= bv
FT   intron          10174. .10298
FT                   /*tag= x
FT   intron          10275. .10399
FT                   /*tag= bw
FT   exon            10299. .10485
FT                   /*tag= y
FT   exon            10400. .10586
FT                   /*tag= bx
FT   intron          10486. .12027
FT                   /*tag= z
FT   intron          10587. .12128
```

```
FT                   /*tag=  by
FT   variation       replace(11493,W)
FT                   /*tag=  aa
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(11594,W)
FT                   /*tag=  bz
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            12028. .12193
FT                   /*tag=  ab
FT   exon            12129. .12294
FT                   /*tag=  ca
FT   intron          12194. .25821
FT                   /*tag=  ac
FT   variation       replace(12260,G)
FT                   /*tag=  ad
FT                   /standard_name= "Single nucleotide polymorphism"
FT   intron          12295. .25922
FT                   /*tag=  cb
FT   variation       replace(12361,G)
FT                   /*tag=  cc
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(13086,C)
FT                   /*tag=  ae
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(13183,C)
FT                   /*tag=  af
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(13187,C)
FT                   /*tag=  cd
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(13284,C)
FT                   /*tag=  ce
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(21240,G)
FT                   /*tag=  ag
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(21341,G)
FT                   /*tag=  cf
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(21695,G)
FT                   /*tag=  ah
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(21796,G)
FT                   /*tag=  cg
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(22058,T)
FT                   /*tag=  ai
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(22159,T)
FT                   /*tag=  ch
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(22233,A)
FT                   /*tag=  aj
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(22244. .22245,A)
FT                   /*tag=  ak
FT                   /standard_name= "Single nucleotide polymorphism"

  Query Match             6.4%;  Score 293.6;  DB 10;  Length 27067;
  Best Local Similarity   75.8%;  Pred. No. 5e-41;
  Matches  391;  Conservative    0;  Mismatches  119;  Indels    6;  Gaps    2;

Qy        230 ATATATATATATATATATATATACACACACACACACATATAT-ATATACACATATATATA 288
              || ||||  |  | |  |  ||| |     | ||| |||||| | ||| | ||||| | |
Db      26509 ATGTATAACTTAAAAATTTCATATATGTGTATACATATATATGAAATATATATATACAGA 26568

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              |||| ||||| || |||||| |||||||||||||||||||  | | | |||| ||||||| 
Db      26569 CACACATATATATGTATACATATATATACACATATATATGTATACATATATACACACATA 26628

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              ||| || |||||| |||||||| |||||||||||||||||||| ||| | | ||| | |
Db      26629 TATGTATACATATATATACACACATATACACATATATACACACACATACATATATACACA 26688

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              |||| ||||| |||||||||||||| ||  |||||||| |  |||| | | ||| | |
Db      26689 CACATATATACACACATATATACACACATGCACACATATATATGTATACATATATACACA 26748

Qy        469 CATATATACACATATATATA-----ATATACACACATATATATACACATATATACACACA 523
              ||| |||||  ||||||| |     ||||||||||||||||||||||| ||||||||||
Db      26749 CATGTATACGTATATATACACACATATATACACACATATATATACACACATATACACACA 26808

Qy        524 TATATACACATATATACACATATATATACACATATATACACATATATACATATATACACA 583
              || | |||||||||||||| |||||||||| ||||| ||| | ||| |||| || |||
Db      26809 TACACACACATATATACACACATATATACACACATATATACACACATATATATGTATACA 26868

Qy        584 TATATATACATATATACACATATATACATATATACACATATATACATATATACACACATA 643
              |||||||||| | ||| | | | ||||| | ||||| ||||| ||||||| ||| | | |
Db      26869 TATATATACACACATATATACACATACACACATACATATATACACATATACACATATACA 26928

Qy        644 TATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATA 703
              | ||| ||||| | |   | ||||| ||||||| | || || |||||| |||||  | |
Db      26929 CACACATATACACACATGTATACATATATATACACACATGTATACATATGTATACACACA 26988

Qy        704 TATATGTATATATATATGCTCCAGAGTTCATAAGAG 739
              |||| | |||| ||||    |     | |||| | |
Db      26989 CATATATGTATACATATATACACACATACATATGTG 27024
```

```
RESULT 16
ADC87570/c
ID   ADC87570 standard; DNA; 1672 BP.
XX
AC   ADC87570;
XX
DT   01-JAN-2004  (first entry)
XX
DE   Human GPCR gene SEQ ID NO:2023.
XX
KW   ds; gene; human; GPCR;
KW   guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1270724-A2.
XX
PD   02-JAN-2003.
XX
PF   18-JUN-2002; 2002EP-00013517.
XX
PR   18-JUN-2001; 2001JP-00246789.
XX
PA   (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA   (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI   Suwa M,  Asai K,  Akiyama Y,  Aburatani H;
XX
DR   WPI; 2003-315783/31.
DR   P-PSDB; ADC87571.
XX
PT   New polynucleotide, useful for preparing a composition for treating a
PT   patient in need of increased or suppressed activity or expression of the
PT   guanosine triphosphate-binding protein coupled receptor.
XX
PS   Claim 1; SEQ ID NO 2023; 28pp; English.
XX
CC   The invention relates to a novel polynucleotide encoding a guanosine
CC   triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC   the invention may have a use in gene therapy. The polynucleotide and
CC   polypeptide are useful for preparing a composition for treating a patient
CC   in need of increased or suppressed activity or expression of the
CC   guanosine triphosphate-binding protein coupled receptor. The
CC   polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC   invention.
XX
SQ   Sequence 1672 BP; 499 A; 171 C; 300 G; 602 T; 0 U; 100 Other;
```

```
  Query Match             6.3%;  Score 291.6;  DB 10;  Length 1672;
  Best Local Similarity   75.5%;  Pred. No. 9.8e-41;
  Matches  376;  Conservative    0;  Mismatches  119;  Indels    3;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACATATATATATACACATATATA 286
              |||||||||||||||||||||||||| | | ||||||||| | ||||||||||||||||||
Db       1390 TATATATATATATATATATATATATATATACACACACACATATATACACATATATA 1331

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              |||||| | | | ||||| ||| ||||| ||||||||||||||||||||| |||   | | |
Db       1330 TACACACACACATATATACACATATATACACACATATATATACACATATATACACATATA 1271

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||| | | | ||||  | | | |||| |||||||||||||||||||| |||| || ||||
Db       1270 TATACACATATATATACACATATATATGTACACATATATACACACATGTATATACGTATA 1211

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              |    |||||||||||||| | ||| |  ||| |  | ||||||||||||||  | |||||
Db       1210 TGTGTACATATATACACACACATATATGTATGTGTATATATACACATATATGTATATATA 1151

Qy        467 TACATATATACACATATATATA---ATATACACACATATATATACACATATATACACACA 523
              |||| | | | |||| ||||||    |||| | ||| | ||| | | || |||| |
Db       1150 TACACACACAGACATGTATATATGTGTATATATACACACATAGATATATGTATATATGTG 1091

Qy        524 TATATACACATATATACACATATATATACACATATATACACATATATACATATATACACA 583
              |||||||||| | ||| |  | ||||||||| |||||    |||||||| | | | | |
Db       1090 TATATACACACACATATATGTGTATATACACACATATATGTGTATATACACACACATATA 1031

Qy        584 TATATATACATATATACACATATATACATATATACACATATATACATATATACACACATA 643
              | | |||| | | | ||| |||| |  ||||| ||||| |||||  | |||| ||||| |
Db       1030 TGTGTATATACACACACATATATGTGTATATACACACACATATATGTGTATATACACACA 971

Qy        644 TATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATA 703
               ||| |  |  |||||||||||||||  |||||  | |||||||||||||| ||  |||
Db        970 CATATATGTGTATATACACACACATGTATATATGTGTATATACACATATATGTATATATG 911

Qy        704 TATATGTATATATATATG 721
              | ||| || | |||||||
Db        910 TGTATATACACATATATG 893


RESULT 17
AAK67239
ID   AAK67239 standard; DNA; 30393 BP.
XX
AC   AAK67239;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:22051.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001354.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0235484P.
PR   27-SEP-2000; 2000US-0235834P.
PR   27-SEP-2000; 2000US-0235836P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   13-OCT-2000; 2000US-0239937P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241221P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
```

PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS Disclosure; SEQ ID NO 22051; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention
XX
SQ Sequence 30393 BP; 8271 A; 7315 C; 8076 G; 6731 T; 0 U; 0 Other;

Query Match 6.2%; Score 288; DB 4; Length 30393;
Best Local Similarity 76.1%; Pred. No. 4.6e-40;
Matches 382; Conservative 0; Mismatches 115; Indels 5; Gaps 2;

```
Qy      227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
            | |||||||| | ||||| ||||||| | | ||||||||| | | | | ||| |||||||
Db    14755 TATATATATACACATATACATATATATATAGACACACACACACACACACACATATATATA 14814

Qy      287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
            ||||| | |||||||||||||||| ||||||| ||||| | ||| |||||||| |||| |
Db    14815 CACACACACATACATATATACACACATATACATATATACACACATATACATATATACATA 14874

Qy      347 TATA----TATACACATATGTATACACATATATACACATATATACACACATATATACACA 402
            ||||    | |||| ||||     | | | ||||| ||| ||||||| | ||||| |||||
Db    14875 TATACACGTGTACATATATACGCATATACATATATACACATATACATATATATACACACA 14934

Qy      403 TATATACACACATATATACACATATATACACATATATACACATATACACATATATACACA 462
            |||| | | | | ||||| | | | ||  |||||||||| |||||||| ||||||||||||
Db    14935 TATACATATATACATATATATACACATGTACATATATACGCATATACATATATATACACA 14994

Qy      463 TATATACATATATACACATATATATA-ATATACACACATATATATACACATATATACACA 521
            |||| | |||||||||||  |||| | ||||||| | ||| | ||  |||||||||| ||
Db    14995 TATACATATATATACACACCTATACATATATACACATATATACACATGTACATATATACGCA 15054

Qy      522 CATATATACACATATATACACATATATATACACATATATACACATATATACATATATACA 581
            ||| ||| | | | | | | | || ||||||| ||||| |||||||||| |  |||| | |
Db    15055 TATACATATATACACACATATACATGTATACATATATACACACATATACATGTATACATA 15114

Qy      582 CATATATATACATATATACACATATATACATATATACACATATATACATATATACACACA 641
            ||| | ||  | ||||| ||  ||||||||||||||||||||||| | ||||| ||||| |
Db    15115 TATACACATGTACATATATACGCATATACATATATACACATATACATATATACACACATA 15174

Qy      642 TATATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTA 701
            |||||  ||||| ||| |||   ||||| |||  |||||| ||| | | | |||||| ||
Db    15175 CATATATACATATATACACATGTACATATATACGCATATACATATATACACATATACATA 15234

Qy      702 TATATATGTATATATATATGCT 723
            |||| |  | || |||||| ||
Db    15235 TATACACATGTACATATATACT 15256
```

RESULT 18
ABK84481/c
ID ABK84481 standard; cDNA; 44147 BP.
XX
AC ABK84481;
XX
DT 14-AUG-2002 (first entry)
XX
DE Human cDNA differentially expressed in granulocytic cells #1052.
XX
KW Human; ss; granulocytic cell; DNA chip; bacterial infection;
KW viral infection; parasitic infection; protozoal infection;
KW fungal infection; sterile inflammatory disease; psoriasis;
KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KW cardiac reperfusion injury; renal reperfusion injury; ARDS;
KW adult respiratory distress syndrome; inflammatory bowel disease;
KW Crohn's disease; ulcerative colitis; periodontal disease;
KW granulocyte activation; chronic inflammation; allergy.
XX
OS Homo sapiens.

XX
PN WO200228999-A2.
XX
PD 11-APR-2002.
XX
PF 03-OCT-2001; 2001WO-US030821.
XX
PR 03-OCT-2000; 2000US-0237189P.
XX
PA (GENE-) GENE LOGIC INC.
XX
PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
XX
DR WPI; 2002-435328/46.
XX
PT Detecting granulocyte activation by detecting differential expression of
PT genes associated with granulocyte activation, which serves as diagnostic
PT markers that is useful for monitoring disease states and drug toxicity.
XX
PS Claim 1; SEQ ID NO 1052; 114pp; English.
XX
CC The invention relates to detecting (M1) granulocyte (GC) activation
CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
CC DNA chip analysis as given in the specification, and comparing the
CC expression level to an expression level in an unactivated GC, where
CC differential expression of Gs is indicative of GCA. Also included are
CC modulating (M2) GA by contacting GC with an agent that alters the
CC expression of at least one gene in Gs; (2) screening (M3) for an agent
CC capable of modulating GCA or an inflammation (especially chronic) in a
CC tissue, an allergic response in a subject, exposure of a subject to a
CC pathogen or sterile inflammatory disease using the gene expression
CC profile; (3) detecting (M4) an inflammation (especially chronic) in a
CC tissue, an allergic response in a subject, exposure of a subject to a
CC pathogen or sterile inflammatory disease, by detecting the level of
CC expression in a sample of the tissue of gene(s) from Gs, where the level
CC of expression of the gene is indicative of inflammation; (4) treating
CC (M5) an inflammation (especially chronic) or in a tissue, an allergic
CC response in a subject, exposure of a subject to a pathogen or sterile
CC inflammatory disease, by contacting a tissue having inflammation with an
CC agent that modulates the expression of gene(s) from Gs in the tissue. M1
CC is useful for detecting GCA; M2 is useful for modulating GA; M3 is useful
CC for screening an agent capable of modulating GCA preferably in an
CC inflammation in a tissue; M4 is useful for detecting an inflammation
CC (especially chronic) in a tissue, an allergic response in a subject,
CC exposure of a subject to a pathogen or sterile inflammatory disease (e.g.
CC psoriasis, rheumatoid arthritis, glomerulonephritis, asthma, thrombosis,
CC cardiac reperfusion injury, renal reperfusion injury, ARDS, adult
CC respiratory distress syndrome, inflammatory bowel disease, Crohn's
CC disease, ulcerative colitis, periodontal disease; also bacterial
CC infection, viral infection, parasitic infection, protozoal infection,
CC fungal infection and M5 is useful for treating one of the above
CC conditions. The present sequence represents a gene differentially
CC expressed in granulocytes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 44147 BP; 9813 A; 11709 C; 11789 G; 10833 T; 0 U; 3 Other;

Query Match 6.2%; Score 286.6; DB 6; Length 44147;
Best Local Similarity 72.9%; Pred. No. 8.1e-40;
Matches 453; Conservative 0; Mismatches 154; Indels 14; Gaps 6;

```
Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC---AAAA 57
              ||||  ||||||||||| |||||||| |||||||| |||||| |||||||||   ||||
Db      10942 AGTTTGAGACCAGCCTGACCAACATGGAGAAACCCCGTCTCTATTAAAAATACAAAAAAA 10883

Qy         58 TTAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAG 117
              |||| ||||| |||||  ||| ||||||||||||||||||||| | |||| || |||||||
Db      10882 TTAGCCAGGCGTGGTGGCACATGCCTGTAATCCCAGCTACTTGGAAGGCTGAAGCAGGAG 10823

Qy        118 AATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTC 176
              || | ||||||||  | || || ||||||| ||| ||| | ||||||||||||||||||
Db      10822 AAGCGCTTGAACC--CGGGAGGCGGAGGTTTCGGTGAGCCCAGATCATGCCATTGCACTG 10765

Qy        177 CAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATA 236
              |||||  |||||||||||| ||||       || |   || ||| |   |||| |||||||
Db      10764 CAGCCTGGGCAACAAGAGCAAAAT------TCCGTTTAATAAATAATTATTTAATATATA 10711

Qy        237 TATATATATATATATACACACACACACACATATATATATACACATATATATACACATATA 296
              ||||||| | | | || | | | | |||| ||||||||||  ||||||||||||   ||
Db      10710 TATATATGTGTGTGTATATATATATACACGTATATATATACGTATATATATACACGCGTA 10651

Qy        297 TACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACA 356
              || |||||   | ||||||||||   |||||| || || |||| ||| | ||||||||| |
Db      10650 TATATATACGTATATATATACACGCGTATATATACGTATATATATACGCGTATATATATA 10591

Qy        357 CATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATA 416
              | ||| |||| |||  || | | |  |||| | | |||  || | ||||| ||  |||||
Db      10590 CGTATATATATACACGTACATATACGTATATATATATACGTATATATATACACGTACATA 10531

Qy        417 TATACACATATATACACATATATACACATATACACATATATACACATATATACATATATA 476
              ||      ||||||| ||| ||||||  ||||| | | | ||| | |  |||| ||| | |
Db      10530 TACGTGTATATATATACACATATACGTATATATATACACATATATACGTATATATACACA 10471

Qy        477 CACATATATATAATATACACACATATATATACACATATATACACACATATATACACATAT 536
              | ||||| |  ||||  | |||| |||||||| ||||||| |   ||   |  || |||
Db      10470 TATATATACA-CATATGTATACATCTATATACATATATATAAAATAATTATTTTACGTAT 10412

Qy        537 ATACACATATATATACACATATATA-CACATATATACATATATACACATATATATACATA 595
              |||          |   | |||  ||  | |  ||||| | |||| | |||||||| |||
Db      10411 ATAATTTACGTAAATAATATACGTATAATAATTATACGTGTATATATATATATATATATA 10352

Qy        596 TATACACATATATACATATAT 616
              | ||||| ||||| ||||||
Db      10351 CACACACACATATATATATAT 10331
```

RESULT 19
ADD14691/c
ID ADD14691 standard; cDNA; 44147 BP.
XX
AC ADD14691;
XX
DT 01-JAN-2004 (first entry)
XX
DE Human src biomarker polynucleotide SEQ ID NO:85.
XX
KW predictor set; protein tyrosine kinase activity modulator;
KW protein tyrosine kinase pathway; protein tyrosine kinase; cytostatic;
KW gene therapy; drug sensitivity; genetic profile; cancer; human; gene; ss.
XX
OS Homo sapiens.
XX
PN WO2003062395-A2.
XX
PD 31-JUL-2003.
XX
PF 17-JAN-2003; 2003WO-US001981.
XX
PR 18-JAN-2002; 2002US-0350061P.
XX
PA (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
PI Huang F, Fairchild CR, Lee FY, Shaw P;
XX
DR WPI; 2003-636735/60.
DR P-PSDB; ADD14095.
XX
PT New polynucleotides and polypeptides for predicting the activity of
PT compounds that interact with protein tyrosine kinases and/or protein
PT tyrosine kinase pathways.
XX
PS Claim 2; SEQ ID NO 85; 139pp; English.
XX
CC The present invention describes a predictor set comprising a plurality of

```
CC   polynucleotides or polypeptides whose expression pattern is predictive of
CC   the response of cells to treatment with a compound that modulates protein
CC   tyrosine kinase activity or members of the protein tyrosine kinase
CC   pathway. Also described: (1) predicting whether a compound is capable of
CC   modulating the activity of cells, comprising obtaining a sample of cells,
CC   determining whether the cells express a plurality of markers, and
CC   correlating the expression of the markers to the compound's ability to
CC   modulate the activity of the cells; (2) a plurality of cell lines for
CC   identifying polynucleotides and polypeptides whose expression levels
CC   correlate with compound sensitivity or resistance of cells associated
CC   with a disease state; and (3) identifying polynucleotides and
CC   polypeptides that predict compound sensitivity or resistance of cells
CC   associated with a disease state, comprising subjecting the plurality of
CC   cell lines to one or more compounds, analysing the expression pattern of
CC   a microarray of polynucleotides or polypeptides, and selecting
CC   polynucleotides or polypeptides that predict the sensitivity or
CC   resistance of cells associated with a disease state by using the
CC   expression pattern of the microarray. The polynucleotides and
CC   polypeptides have cytostatic activities, and can be used in gene therapy.
CC   The polynucleotides and polypeptides are useful in predicting the
CC   activity of compounds that interact with protein tyrosine kinases and/or
CC   protein tyrosine kinase pathways. These may be used in determining drug
CC   sensitivity in patients to allow the development of individualized
CC   genetic profiles which aid in treating diseases and disorders (e.g.
CC   cancer) based on patient response at a molecular level. The present
CC   sequence is used in the exemplification of the present invention.
XX
SQ   Sequence 44147 BP; 9813 A; 11709 C; 11789 G; 10833 T; 0 U; 3 Other;

  Query Match             6.2%;  Score 286.6;  DB 10;  Length 44147;
  Best Local Similarity   72.9%;  Pred. No. 8.1e-40;
  Matches  453;  Conservative    0;  Mismatches  154;  Indels   14;  Gaps    6;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC---AAAA 57
              |||| |||||||||||| |||||||| |||||||| |||||| |||||||||   ||||
Db      10942 AGTTTGAGACCAGCCTGACCAACATGGAGAAACCCCGTCTCTATTAAAAATACAAAAAAA 10883

Qy         58 TTAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAG 117
              |||| ||||| |||||  ||| |||||||||||||||||||| | |||| || |||||||
Db      10882 TTAGCCAGGCGTGGTGGCACATGCCTGTAATCCCAGCTACTTGGAAGGCTGAAGCAGGAG 10823

Qy        118 AATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTC 176
              || | ||||||||  | || || ||||||| ||| ||| | ||||||||||||||||||
Db      10822 AAGCGCTTGAACC--CGGGAGGCGGAGGTTTCGGTGAGCCCAGATCATGCCATTGCACTG 10765

Qy        177 CAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATA 236
              ||||| |||||||||||| ||||      || |   || ||| |   |||| |||||||
Db      10764 CAGCCTGGGCAACAAGAGCAAAAT------TCCGTTTAATAAATAATTATTTAATATATA 10711

Qy        237 TATATATATATATATACACACACACACATATATATATACACATATATATACACATATA 296
              ||||||| | | || | | | | |||| |||||||||||  ||||||||||||   ||
Db      10710 TATATATGTGTGTGTATATATATATACACGTATATATATACGTATATATATACACGCGTA 10651

Qy        297 TACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACA 356
              || |||||   | ||||||||||   |||||| || || |||| ||| | |||||||| |
Db      10650 TATATATACGTATATATATATACACGCGTATATATACGTATATATATATACGCGTATATATATA 10591

Qy        357 CATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATA 416
              | ||| |||| ||| || | | |  |||| | | |||  || | ||||| ||  |||||
Db      10590 CGTATATATATACACGTACATATACGTATATATATATACGTATATATATACACGTACATA 10531

Qy        417 TATACACATATATACACATATATACACATATACACATATATACACATATATACATATATA 476
              ||      ||||||| ||| ||||||  ||||| | | | ||| | |  |||| ||| | |
Db      10530 TACGTGTATATATATACACATATACGTATATATATACACATATATACGTATATATACACA 10471

Qy        477 CACATATATATAATATACACACATATATATACACATATATACACACATATATACACATAT 536
              | ||||| |  ||||  | |||| |||||||| |||||||  |   ||   |  || |||
Db      10470 TATATATACA-CATATGTATACATCTATATACATATATATAAAATAATTATTTTACGTAT 10412

Qy        537 ATACACATATATATACACATATATA-CACATATATACATATATACACATATATATACATA 595
              |||          |   | |||  ||  | |  ||||| | |||| | |||||||| |||
Db      10411 ATAATTTACGTAAATAATATACGTATAATAATTATACGTGTATATATATATATATATATA 10352

Qy        596 TATACACATATATACATATAT 616
              | ||||| ||||| ||||||
Db      10351 CACACACACATATATATATAT 10331


RESULT 20
AAK89374
ID   AAK89374 standard; DNA; 32206 BP.
XX
AC   AAK89374;
XX
DT   05-NOV-2001  (first entry)
XX
DE   Human digestive system antigen genomic sequence SEQ ID NO: 2950.
XX
KW   Human; digestive system antigen; gene therapy; cancer; appendicitis;
KW   ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
KW   digestive system disorder; Meckel's diverticulum; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155314-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001324.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
```

PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-502630/55.
XX
PT Polynucleotides encoding digestive system antigens, useful for
PT diagnosing, treating, preventing and/or prognozing disorders of the
PT digestive system, particularly cancer and cancer metastases.
XX
PS Disclosure; SEQ ID NO 2950; 986pp; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human digestive system antigens. These can be used in the
CC diagnosis, treatment and prevention of digestive system disorders,
CC including cancer, Meckel's diverticulum, bacterial or parasitic
CC infections, appendicitis, Hirschsprung's disease, chronic colitis or
CC ulcerative colitis. The present sequence is a genomic DNA fragment
CC encoding a digestive system antigen of the invention
XX
SQ Sequence 32206 BP; 9503 A; 5480 C; 7017 G; 10206 T; 0 U; 0 Other;

Query Match 6.1%; Score 280.8; DB 4; Length 32206;
Best Local Similarity 74.4%; Pred. No. 7.9e-39;
Matches 369; Conservative 0; Mismatches 122; Indels 5; Gaps 1;

```
Qy      229 TATATATATATATATATATATACACACACACACATATATATATACACATATATATA 288
            ||||| |||||||| | ||| | | | | ||| | |  |||||||| ||| | | |||||
Db    21803 TATATGTATATATACACATACACATATATACATATATGTATATATACACATACACATATA 21862

Qy      289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
            ||||||||  |||||||||||||| | | |  ||| ||||||  | | ||||  ||| ||
Db    21863 TACATATATGTATATATACACATACACATATGTATGTATACATACATATATACACATATG 21922

Qy      349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
            ||||||  |||||  ||||  ||| || |  | |||||  |  ||  |||||||||||  | | ||
Db    21923 TATATATACATACATATATACACATGTGTATATATACATGCAGATATATACACGTGTGTA 21982

Qy      409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
            | ||| ||||  | ||| || |||||||| |||| | || | ||| | | |||  ||||
Db    21983 TATACACATATGTATATACATGCACATATACACACGTGTATATATACACATACACGTATA 22042

Qy      469 CATATATACACATATATATA-----ATATACACACATATATATACACATATATACACACA 523
            ||| ||| ||| |||||||     | ||| |||||| |||||| ||||| | | | |||
Db    22043 TATACATATACACATATATACATACACATATACACATGTATATATACATACACATATACA 22102

Qy      524 TATATACACATATATACACATATATATACACATATATACACATATATACATATATACACA 583
            || || | ||| ||||| ||||| |||||  |||||| ||||| ||| ||| |||||
Db    22103 CATGTATATATACATACATATATACATACATGTATATATACATACATATATACATACATG 22162

Qy      584 TATATATACATATATACACATATATACATATATACACATATATACATATATACACACATA 643
            |||||||||||| ||| | | ||| |  |||||| ||||| |||||||||||||  |||
Db    22163 TATATATACATACATATATACATACATGTATATATACATACATACATATATACACGTATA 22222

Qy      644 TATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATA 703
```

```
              |||||| | | |||||||||||||||| ||| | | |||||| || | || ||||  || |
Db      22223 TATACATACATATATACACACACATATATACATACATATATGCATACATGTATATATACA 22282

Qy        704 TATATGTATATATATA 719
              || |  |||||| |||
Db      22283 TACAGATATATACATA 22298


RESULT 21
AAK73165
ID   AAK73165 standard; DNA; 20420 BP.
XX
AC   AAK73165;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:27977.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001354.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0235484P.
PR   27-SEP-2000; 2000US-0235834P.
PR   27-SEP-2000; 2000US-0235836P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   13-OCT-2000; 2000US-0239937P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241221P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   20-OCT-2000; 2000US-0241826P.
PR   01-NOV-2000; 2000US-0244617P.
PR   08-NOV-2000; 2000US-0246474P.
PR   08-NOV-2000; 2000US-0246475P.
PR   08-NOV-2000; 2000US-0246476P.
PR   08-NOV-2000; 2000US-0246477P.
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246526P.
PR   08-NOV-2000; 2000US-0246527P.
PR   08-NOV-2000; 2000US-0246528P.
PR   08-NOV-2000; 2000US-0246532P.
PR   08-NOV-2000; 2000US-0246609P.
PR   08-NOV-2000; 2000US-0246610P.
PR   08-NOV-2000; 2000US-0246611P.
PR   08-NOV-2000; 2000US-0246613P.
PR   17-NOV-2000; 2000US-0249207P.
PR   17-NOV-2000; 2000US-0249208P.
PR   17-NOV-2000; 2000US-0249209P.
PR   17-NOV-2000; 2000US-0249210P.
PR   17-NOV-2000; 2000US-0249211P.
PR   17-NOV-2000; 2000US-0249212P.
PR   17-NOV-2000; 2000US-0249213P.
PR   17-NOV-2000; 2000US-0249214P.
PR   17-NOV-2000; 2000US-0249215P.
PR   17-NOV-2000; 2000US-0249216P.
PR   17-NOV-2000; 2000US-0249217P.
PR   17-NOV-2000; 2000US-0249218P.
PR   17-NOV-2000; 2000US-0249244P.
```

```
PR   17-NOV-2000; 2000US-0249245P.
PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250160P.
PR   01-DEC-2000; 2000US-0250391P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS   Disclosure; SEQ ID NO 27977; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting the
CC   nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention
XX
SQ   Sequence 20420 BP; 4834 A; 5403 C; 5630 G; 4553 T; 0 U; 0 Other;

  Query Match             5.9%;  Score 270.4;  DB 4;  Length 20420;
  Best Local Similarity   61.4%;  Pred. No. 4.7e-37;
  Matches  467;  Conservative    0;  Mismatches  291;  Indels    2;  Gaps    2;

Qy        144 GGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAGCCCAGGCAACAAGAGCGAAATGTC 203
Db       4055 GCTGGAGTGCAGTAGCATGATCACAGCTCACTGCAGCCATATATATGTATATATATATAT 4114

Qy        204 ATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATATACACACACACAC 263
Db       4115 ATATGAAAAAATATATATATATATATATATATATGAAAAAATATATATAAATATATATATGA 4174

Qy        264 ACATATATATATACACATATATATACACATATATACATATATACACAT-ATATACACATA 322
Db       4175 AAATATATATATAAATATATATATGAAAATATATATAAATATATATATAATATATATAAA 4234

Qy        323 TATATACACATACATATGTACACATATATATACACATATGTATACACATATATACACATA 382
Db       4235 TATATATATATATAAATATATATATGAAAATATATATAAATATATATATAAATATATATA 4294

Qy        383 TATACACACATATATACACATATATACACACATATATACACATATATACACATATATACA 442
```

```
Db       4295 TGAAAATATATATATGAAAATATATATGAAAATATATATATAAATATATATATATGAAAA 4354

Qy        443 CATATACACATATATACACATATATACATATATACACATATATATAATATACACACATAT 502
Db       4355 TATATATATAAATATATATATGAAAATATATATATAAATATATAT-ATATGAAAATATAT 4413

Qy        503 ATATACACATATATACACACATATATACACATATATACACATATATATACACATATATAC 562
Db       4414 ATATAAATATATATATATGAAAATATATATATAAATATATATATATGAAAATATATATAT 4473

Qy        563 ACATATATACATATATACACATATATATACATATATACACATATATACATATATACACAT 622
Db       4474 AAATATATATATATGAAAATATATATATAAATATATATATATGAAAATATATATATAAAT 4533

Qy        623 ATATACATATATACACACATATATACACATACATATACACACACATAGATATACATATAT 682
Db       4534 ATATATATATGAAAATATATATATAAATATATATATATGAAAATATATATATAAATATAT 4593

Qy        683 ATACACATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTA 742
Db       4594 ATATGAAAATATATATATAAATATATATGAAAATATATATATAAATATATATATATGA 4653

Qy        743 GCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAA 802
Db       4654 AAATATATATATAAATATATATATATGAAAATATATATAAATATATATATGAAAATATAT 4713

Qy        803 GGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTAT 862
Db       4714 ATATATATATATATGAAAATATATATATATGAAAATATATATATAAATATATATATGAAA 4773

Qy        863 CTGTAAGGTAGTTAAGAAGGTAACACTATATATATATATA 902
Db       4774 ATAATATATATAAATATATATAAATATATATAAAAATATA 4813
```

```
RESULT 22
ABK69933
ID   ABK69933 standard; DNA; 20420 BP.
XX
AC   ABK69933;
XX
DT   15-JUL-2002  (first entry)
XX
DE   Human secreted protein gene 68 genomic DNA fragment #21.
XX
KW   Human; ds; secreted protein; gene therapy; immunosuppressive;
KW   antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;
KW   vasotropic; cerebroprotective; nootropic; neuroprotective; antibacterial;
KW   virucide; fungicide; opthalmalogical; autoimmune disease; neoplasm;
KW   rheumatoid arthritis; hyperproliferative disorder; cardiac arrest;
KW   cardiovascular disorder; cerebrovascular disorder; cerebral ischaemia;
KW   angiogenesis; nervous system disorder; Alzheimer's disease; infection;
KW   ocular disorder; corneal infection; wound healing; skin aging;
KW   epithelial cell proliferation; food additive.
XX
OS   Homo sapiens.
XX
PN   WO200226931-A2.
XX
PD   04-APR-2002.
XX
PF   24-SEP-2001; 2001WO-US029871.
XX
PR   25-SEP-2000; 2000US-0234925P.
PR   12-JAN-2001; 2001WO-US000911.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Komatsoulis G,  Duan DR,  Rosen CA,  Moore PA,  Shi Y;
PI   Lafleur DW,  Olsen H,  Brewer LA,  Florence KA,  Young PE,  Soppet DR;
PI   Endress GA,  Mucenski M,  Ebner R;
XX
DR   WPI; 2002-362489/39.
```

```
XX
PT   Novel 71 isolated secreted polypeptides and polynucleotides encoding the
PT   polypeptides, useful for treating Huntington's disease, sepsis,
PT   meningitis, thrombocytopenia, hemolytic anemia, rheumatoid arthritis,
PT   asthma.
XX
PS   Example 2; Page 1440-1445; 1478pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule (or its
CC   fragment, homologue complement or allelic variant) encoding a human
CC   secreted protein (and its fragment, domain, epitope, variant, secreted
CC   form and species variant). Also included are a recombinant vector
CC   comprising the nucleic acid, a recombinant host cell comprising the
CC   vector, an antibody against the secreted protein, a recombinant host cell
CC   that expresses the secreted protein and a method of identifying a binding
CC   partner of the secreted protein. The nucleic acid and protein are used to
CC   prevent, diagnose, treat or ameliorate a medical condition in e.g.
CC   humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep for
CC   example autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
CC   disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
CC   e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC   angiogenesis, nervous system disorders e.g. Alzheimer's disease,
CC   infections caused by bacteria, viruses and fungi and ocular disorders
CC   e.g. corneal infection. Many other diseases and disorders are listed in
CC   the specification. The polypeptides can also be used to aid wound healing
CC   an epithelial cell proliferation, to prevent skin aging due to sunburn,
CC   to maintain organs before transplantation, for supporting cell culture of
CC   primary tissues, to regenerate tissues and in chemotaxis. The
CC   polypeptides can also be used as a food additive or preservative to
CC   increase or decrease storage capabilities. The present sequence
CC   represents a ds DNA fragment of the gene for a novel human secreted
CC   protein of the invention
XX
SQ   Sequence 20420 BP; 4834 A; 5403 C; 5630 G; 4553 T; 0 U; 0 Other;

  Query Match             5.9%;  Score 270.4;  DB 6;  Length 20420;
  Best Local Similarity   61.4%;  Pred. No. 4.7e-37;
  Matches  467;  Conservative    0;  Mismatches  291;  Indels    2;  Gaps    2;

Qy        144 GGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAGCCCAGGCAACAAGAGCGAAATGTC 203
Db       4055 GCTGGAGTGCAGTAGCATGATCACAGCTCACTGCAGCCATATATATGTATATATATATAT 4114

Qy        204 ATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATATACACACACACAC 263
Db       4115 ATATGAAAAAATATATATATATATATATATATATGAAAAAATATATATAAATATATATATGA 4174

Qy        264 ACATATATATATACACATATATATACACATATATACATATATACACAT-ATATACACATA 322
Db       4175 AAATATATATATAAATATATATATGAAAATATATATAAATATATATATAATATATATAAA 4234

Qy        323 TATATACACATACATATGTACACATATATATACACATATGTATACACATATATACACATA 382
Db       4235 TATATATATATATAAATATATATATGAAAATATATATAAATATATATATAAATATATATA 4294

Qy        383 TATACACACATATATACACATATATACACACATATATACACATATATACACATATATACA 442
Db       4295 TGAAAATATATATATGAAAATATATATGAAAATATATATATAAATATATATATATGAAAA 4354

Qy        443 CATATACACATATATACACATATATACATATATACACATATATATAATATACACACATAT 502
Db       4355 TATATATATAAATATATATATGAAAATATATATATAAATATATAT-ATATGAAAATATAT 4413

Qy        503 ATATACACATATATACACACATATATACACATATATACACATATATATACACATATATAC 562
Db       4414 ATATAAATATATATATATGAAAATATATATATAAATATATATATATGAAAATATATATAT 4473

Qy        563 ACATATATACATATATACACATATATATACATATATACACATATATACATATATACACAT 622
Db       4474 AAATATATATATATGAAAATATATATATAAATATATATATATGAAAATATATATAAAT 4533

Qy        623 ATATACATATATACACACATATATACACATACATATACACACACATAGATATACATATAT 682
```

```
Db       4534 ATATATATATGAAAATATATATATAAATATATATATATGAAAATATATATATAAATATAT 4593

Qy        683 ATACACATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTA 742
Db       4594 ATATGAAAATATATATATAAATATATATATGAAAATATATATATAAATATATATATGA 4653

Qy        743 GCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAA 802
Db       4654 AAATATATATATAAATATATATATATGAAAATATATATAAATATATATATGAAAATATAT 4713

Qy        803 GGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTAT 862
Db       4714 ATATATATATATATGAAAATATATATATATGAAAATATATATAAATATATATATGAAA 4773

Qy        863 CTGTAAGGTAGTTAAGAAGGTAACACTATATATATATATA 902
Db       4774 ATAATATATATAAATATATATAAATATATATAAAAATATA 4813


RESULT 23
ABD32806_4/c
Continuation (5 of 7) of ABD32806 from base 400001 (Human cancer-associated genomic DNA
WP Sequence split into 7 fragments  LOCUS ABD32806 Accession Abd32806
WP      Fragment Name      Begin      End
WP      ABD32806_0             1    110000
WP      ABD32806_1        100001    210000
WP      ABD32806_2        200001    310000
WP      ABD32806_3        300001    410000
WP      ABD32806_4        400001    510000
WP      ABD32806_5        500001    610000
WP      ABD32806_6        600001    653458

  Query Match             5.8%;  Score 268.6;  DB 13;  Length 110000;
  Best Local Similarity   69.2%;  Pred. No. 1e-36;
  Matches  441;  Conservative    0;  Mismatches  184;  Indels   12;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC--AAAAT 58
Db      95587 AGTTTGAGACCAGCCTGGCCAACATGGTAAAACCCCATCTCTATTTAAAAAAACAAAAAT 95528

Qy         59 TAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGA 118
Db      95527 TAACCAGGTGTGATGGTACACATCTGTACTCCCGGCTACTCAGGAGGCTGAGGCACGAGA 95468

Qy        119 ATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCA-AGATCATGCCATTGCACTCC 177
Db      95467 ATTGTTTGAACCTG--GAAAGTGGAGGTTGTAGTGAGCAGAGATCGCAACACTGTACTCC 95410

Qy        178 AGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAGGCATTTTATATATAT 237
Db      95409 AGCCTGGGTGACAGAGCAAGACTGTC------TGAAAAAAAAAATCAACAAAATAAACAA 95356

Qy        238 ATATATATATATATACACACACACACATATATATATACACATATATATACACATATAT 297
Db      95355 AGAGGTAATGACTGACATATATATATATACATATATATTCCCATATATATACACATAT 95296

Qy        298 ACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACAC 357
Db      95295 ACGTATATATACGTATATACACGTATATACACGTATACGTGTATATACGTGTATACACGT 95236

Qy        358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417
Db      95235 GTATACACGTATATGTGTATATATACGTATATACATGTATATACGTATATATACATATAT 95176

Qy        418 ATACACATATATACACATATATACACATATACACATATATACACATATATACATATATAC 477
Db      95175 ACATATTCATATACGTATATATACGTATATACATATTCATATATGTATATATACGTGTAC 95116

Qy        478 ACATATATATAATATACACACATATATATACACATATATACACACATATATACACATATA 537
Db      95115 ATATATATGT-ATATATACATATATACGTACACGTATATATGTATATATACATATATACG 95057

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATACATATA 597
```

```
Db      95056 TACACGTATATATGTATATATACATATATACGTACACGTATATATGTATATATATACACA 94997

Qy        598 TACACATATATACATATATACACATATATACATATAT 634
              ||| || | |   |||||| | ||||||| || | |
Db      94996 CACATATGTGTGTATATATATATATATATATTTTT 94960


RESULT 24
ADH13788/c
ID   ADH13788 standard; DNA; 191127 BP.
XX
AC   ADH13788;
XX
DT   11-MAR-2004  (first entry)
XX
DE   Prostate leucine zipper regulatory element/PrLZ gene SEQ ID NO:2.
XX
KW   prostate leucine zipper; PrLZ; prostate; cytostatic; cancer; dysplasia;
KW   ds; human; gene.
XX
OS   Chimeric.
OS   Unidentified.
OS   Homo sapiens.
XX
PN   WO2003087311-A2.
XX
PD   23-OCT-2003.
XX
PF   07-APR-2003; 2003WO-US010536.
XX
PR   05-APR-2002; 2002US-0370557P.
XX
PA   (UYEM-) UNIV EMORY.
XX
PI   Wang R,  Chung LWK;
XX
DR   WPI; 2003-845313/78.
DR   P-PSDB; ADH13786.
XX
PT   New isolated prostate leucine zipper (PrLZ) nucleic acid molecule, useful
PT   in diagnosing, prognosticating and/or treating cancers or high grade
PT   dysplasias affecting PrLZ-positive tissues, such as those in the
PT   prostate.
XX
PS   Claim 23; SEQ ID NO 2; 103pp; English.
XX
CC   The invention relates to a novel isolated nucleic acid molecule that
CC   regulats the expression of prostate-expressed genes, specifically the
CC   prostate leucine zipper (PrLZ) gene. A polypeptide of the invention has
CC   cytostatic activity, and may act as a leucine zipper antagonist. The
CC   agent or vector that inhibits the expression or activity of PrLZ is
CC   useful for the treatment of cancer or high grade dysplasia, and for the
CC   preparation of a medicament for the treatment of cancer or high grade
CC   dysplasia, where the cancer or dysplasia affects prostate tissue, breast
CC   tissue, renal tissue, tissue within the gastrointestinal tract, or any
CC   other PrLZ-positive tissue. The methods and compositions of the present
CC   invention are also useful for the diagnosis and prognosis of such
CC   cancers. The present sequence represents a genomic sequence that
CC   regulates the expression of PrLZ, an additional 5' flanking sequence, the
CC   PrLZ gene, and 3' flanking sequence.
XX
SQ   Sequence 191127 BP; 51194 A; 37977 C; 39502 G; 62454 T; 0 U; 0 Other;

  Query Match             5.8%;  Score 268.6;  DB 10;  Length 191127;
  Best Local Similarity   64.4%;  Pred. No. 1.1e-36;
  Matches  467;  Conservative    0;  Mismatches  249;  Indels    9;  Gaps    4;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAATTA 60
              || || |||||| |||| |||||||||||||||          ||||    || ||||
Db     122784 AGATCGAGACCATCCTGGCCAACATGGTGAAACCCGTCTCAACTAAAAACACAAAGATTA 122725

Qy         61 GACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAAT 120
              |   ||| |||||      | |||||| ||||| |||||   ||||| ||||||||||||||
Db     122724 GCTGGGCGTGGTGGCGTGCACCTGTAGTCCCAACTACTCATGAGGCTGAGGCAGGAGAAT 122665

Qy        121 CACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGC-AAGATCATGCCATTGCACTCCAG 179
              | | |||||   | || ||  |||||||| | ||||  ||||   ||||   ||||||||
Db     122664 CTCCTGAAC--TCGGGAGGCAGAGGTTGCAGTGAGCTGAGATTGCGCCACCACACTCCAG 122607

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  |  |  |||| ||   ||     | |||| | | | |    | | | | | | |
Db     122606 CCTGGCGACAGAGAGAGA---CTCCATCTAAAAAACATATATGTGTGTGTGTGTGTGTGT 122550

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              | | | | | ||        | | |||| || | ||| | ||   | || | | || |||
Db     122549 GTGTGTGTGTGTATGTGTGTATATACATGTACACATAGATATGCGTGTATATACATGTAC 122490

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACAT 359
              | ||| | |    | || | | || || ||| | | ||  ||  | ||| || |||||||
Db     122489 ACATAGATATGCGTGTATATACATGTACACATAGATATGCGTGTATATACATGTACACAT 122430

Qy        360 ATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATAT 419
              | ||| |   |||||||| || | | | | ||   | | ||| ||  ||||||| ||
Db     122429 AGATATGCGTGTATATACATGTACACATAGATATGCGTGTATATACATGTACACATAGAT 122370

Qy        420 ACACATATATACACATATATACACATATACACATATATACACATATATACATATATACAC 479
              | | | |||| |||| || ||| | |||  | | |||| |||| || ||||| ||| |
Db     122369 ATGCGTGTATATACATGTACACATAGATATGCGTGTATATACATGTACACATAGATATGC 122310

Qy        480 ATATATATA---ATATACACACATATATATACACATATATACACACATATATACACATAT 536
              | ||||||    || ||| | ||||   |  | ||| ||  ||||||| |||  | | |
Db     122309 GTGTATATACATGTACACATAGATATGCGTGTATATACATGTACACATAGATATGCGTGT 122250

Qy        537 ATACACATATATATACACATATATACACATATATACATATATACACATATATATACATAT 596
              ||| |||| || | | | || | |  | ||| ||  | | ||| | || | | || |||
Db     122249 ATATACATGTACACAGATATGTGTGTATATACATGTACACATAGATATGTGTGTATATAC 122190

Qy        597 ATACACATATATACATATATACACATATATACATATATACACACATATATACACATACAT 656
              | ||| ||||| |  ||||  |||| ||||| | ||||| | ||| | | | | ||||||
Db     122189 ACACATATATACATCTATACGCACACATATATACATATATATACACACACATATATACAT 122130

Qy        657 ATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATAT 716
              ||| ||||||| | ||||| ||| | ||||||| | | | || ||||||||| ||||||
Db     122129 ATATACACACACACATATATATACACATACACACACACACACATATATATATATATAA 122070

Qy        717 ATATG 721
              |  ||
Db     122069 ACTTG 122065


RESULT 25
ABD32806_4
Continuation (5 of 7) of ABD32806 from base 400001 (Human cancer-associated genomic DNA F
WP Sequence split into 7 fragments  LOCUS ABD32806 Accession Abd32806
WP     Fragment Name      Begin     End
WP     ABD32806_0             1     110000
WP     ABD32806_1        100001     210000
WP     ABD32806_2        200001     310000
WP     ABD32806_3        300001     410000
WP     ABD32806_4        400001     510000
WP     ABD32806_5        500001     610000
WP     ABD32806_6        600001     653458

  Query Match             5.8%;  Score 267.6;  DB 13;  Length 110000;
  Best Local Similarity   69.9%;  Pred. No. 1.5e-36;
  Matches  432;  Conservative    0;  Mismatches  179;  Indels    7;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||| |||||||||||  |||||||| ||||||||||||||||| ||||| || ||||||
Db      94745 AGTTCGAGACCAGCCTGGCCAACATGGCGAAACCCCATCTCTACAAAAAACACAAAAATT 94804

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              || ||||| |||||  | |  ||| || |||||| ||||| |||||| ||||| ||||||
Db      94805 AGCCAGGCGTGGTGGCATATACCTATAGTCCCAG-TACTTGGGAGGCTGAGGCGGGAGAA 94863
```

```
Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              ||||||||||     || |      |   |||   | ||||||||||| || |||||||
Db      94864 TCACTTGAACCCAGGAGGTGAAAGATTGCAGGGAGCCGAGATCATGCCACTGAACTCCAG 94923

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  ||  || ||||| | |     ||||| |||||||||||     | ||||||||||
Db      94924 CCTGGGTGAC-AGAGCCAGACTCTGTCTCAAAAAAAAAAAA---TATATATATATATAT 94979

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              ||| | | | |||         | | ||||| ||||| || | ||  || | || ||||
Db      94980 ATACACACACATATGTGTGTGTATATATATACATATATACGTGTACGTATATATGTATAT 95039

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACAT 359
              ||| ||| | |  | |||  |||||| || | |||||||| |||   ||  |||| | |
Db      95040 ATACATATATACGTGTACGTATATATGTATATATACATATATACGTGTACGTATATATGT 95099

Qy        360 ATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATAT 419
              || |||||| ||||||  |||  |||| | ||||||||  | || |||| ||  ||||||
Db      95100 ATATATACATATATATGTACACGTATATATACATATATGAATATGTATATACGTATATAT 95159

Qy        420 ACACATATATACACATATATACACATATACACATATATACACATATATACATATATACAC 479
              ||  ||||  | |  ||||||   |||| || ||||||||  |||||| |||||| ||
Db      95160 ACGTATATGAATATGTATATATGTATATATACGTATATACATGTATATACGTATATATAC 95219

Qy        480 ATATATA-TAATATACACACATATATATACACATATATACACACATATATACACATATAT 538
              | |||||    |||||||   |||| |   | ||| |   | || | |||| || | |||
Db      95220 ACATATACGTGTATACACGTGTATACACGTATATACACGTATACGTGTATATACGTGTAT 95279

Qy        539 ACACATATATATACACATATATACACATATATACATATATACACATATATATACATATAT 598
              | || |||||||||  |||| |   | ||||||    |||| | || ||||||  ||||||
Db      95280 ATACGTATATATACGTATATGTGTATATATATGGGAATATATATATGTATATATATATAT 95339

Qy        599 ACACATATATACATATAT 616
                  |   ||| | | |
Db      95340 GTCAGTCATTACCTCTTT 95357
```

RESULT 26
AAK73166
ID AAK73166 standard; DNA; 19965 BP.
XX
AC AAK73166;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:27978.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001354.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.

```
PR   08-NOV-2000; 2000US-0246474P.
PR   08-NOV-2000; 2000US-0246475P.
PR   08-NOV-2000; 2000US-0246476P.
PR   08-NOV-2000; 2000US-0246477P.
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246526P.
PR   08-NOV-2000; 2000US-0246527P.
PR   08-NOV-2000; 2000US-0246528P.
PR   08-NOV-2000; 2000US-0246532P.
PR   08-NOV-2000; 2000US-0246609P.
PR   08-NOV-2000; 2000US-0246610P.
PR   08-NOV-2000; 2000US-0246611P.
PR   08-NOV-2000; 2000US-0246613P.
PR   17-NOV-2000; 2000US-0249207P.
PR   17-NOV-2000; 2000US-0249208P.
PR   17-NOV-2000; 2000US-0249209P.
PR   17-NOV-2000; 2000US-0249210P.
PR   17-NOV-2000; 2000US-0249211P.
PR   17-NOV-2000; 2000US-0249212P.
PR   17-NOV-2000; 2000US-0249213P.
PR   17-NOV-2000; 2000US-0249214P.
PR   17-NOV-2000; 2000US-0249215P.
PR   17-NOV-2000; 2000US-0249216P.
PR   17-NOV-2000; 2000US-0249217P.
PR   17-NOV-2000; 2000US-0249218P.
PR   17-NOV-2000; 2000US-0249244P.
PR   17-NOV-2000; 2000US-0249245P.
PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250160P.
PR   01-DEC-2000; 2000US-0250391P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS   Disclosure; SEQ ID NO 27978; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting the
CC   nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention
XX
SQ   Sequence 19965 BP; 4699 A; 5296 C; 5505 G; 4465 T; 0 U; 0 Other;

  Query Match             5.8%;  Score 267.4;  DB 4;  Length 19965;
  Best Local Similarity   63.9%;  Pred. No. 1.5e-36;
  Matches  453;  Conservative    0;  Mismatches  251;  Indels    5;  Gaps    3;

Qy        215 AAAAAAAGGCATTTTATATATATATATATATATATACACACACACACATATATATA 274
              ||||||      |  ||||||||||  | |||||||||| | | | | | |   | |||||
Db       4148 AAAAAATATATATAAATATATATATGAAAATATATATATAAATATATATATGAAAATATA 4207

Qy        275 TACACATATATATACACATATATACATATATACACATATATACACATATATATACACATA 334
              || | |||||||||    | ||||| ||||  | | ||||||| | |||||||||  | |||
Db       4208 TATAAATATATATATGAAAATATATATATGAAAATATATATAAATATATATATGAAAATA 4267

Qy        335 CATATGTACACATATATATACACATATGTATACACATATATACACATATATACACACATA 394
              ||||| | | ||||||||  | |||| |||  | |||||||  | |  |||||  | |  ||
Db       4268 TATATGAAAATATATATATGAAAATATATATGAAAATATATATATAAATATATATATATG 4327

Qy        395 TATACACATATATACACACATATATACACATATATACACATATATACACATATACACATA 454
              | | | |||||||  | |  ||||||  | |||||||  | |  |||||  | ||  | | |||
Db       4328 AAAATATATATATAAATATATATATGAAAATATATATATAAATATATATATGAAAATATA 4387

Qy        455 TATACACATATATACATATATACACATATATAT-AATATACACACATATATATACACATA 513
              ||||  | |||||||  ||||  | |  ||||||||  ||||||  | |  ||  | |||  | |||
Db       4388 TATATAAATATATATATATGAAAATATATATATAAATATATATATATGAAAATATATATA 4447

Qy        514 TATACACACATATATACACATATATACACATATATATACACATATATACACATATATACA 573
              || | | |  ||||||  | |||||||  | |  |||||||  | ||  | | |  |||||||  |
Db       4448 TAAATATATATATATGAAAATATATATATAAATATATATATATGAAAATATATATATAAA 4507

Qy        574 TATATACACATATATATACATATATACACATATATACATATATACACATATATACATATA 633
              ||||||  |   | |||||  |||||  | |  |||||||    | |||  | |||||  | |||||
Db       4508 TATATATATGAAAATATATATATAAATATATATATATGAAAATATATATATAAATATATA 4567

Qy        634 TACACACATATATACACATACATATACACACACATAGATATACATATATATACACATATA 693
              ||  |   | |||||  | | |  |||||  |   |  |||  |||||  ||||||  | |  |||||
Db       4568 TATATGAAAATATATATAAATATATATATGAATATATATATATATATATGAAAATATATA 4627

Qy        694 TATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTA 753
              ||||  | |  ||||||  ||||  ||||||      | | |  |||  |       |  |||  ||
Db       4628 TATATGAAAATATATATATAAATATATATATGAAAATATATATAAATATATATATGAATA 4687

Qy        754 CCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAG 813
                  |   |||  |  |||||  |  |   |  |      | |  |||     | |   | || | |
Db       4688 TATATATATATATATGAAAATATATATATATGAAAATATATATATAAATATATATATGAA 4747

Qy        814 GTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAG 873
              | ||   | |||  |   |  ||| ||| || |  ||  | |  ||| | |||   |
Db       4748 AATAAT---ATATATAAATATATATAAATATATATAAAAATATAAATATATATAA-AAAT 4803

Qy        874 TTAAGAAGGTAACACTATATATATATATAGTGAAAGCAGTGTAAACCTA 922
              ||  ||  ||    |||| | |||||||   |||    | || |  ||
Db       4804 ATATAAATATATATATATAAACATATATATATAAATATATATATATATA 4852


RESULT 27
ABK69932
ID   ABK69932 standard; DNA; 19965 BP.
XX
AC   ABK69932;
XX
DT   15-JUL-2002  (first entry)
XX
DE   Human secreted protein gene 68 genomic DNA fragment #20.
XX
KW   Human; ds; secreted protein; gene therapy; immunosuppressive;
KW   antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;
KW   vasotropic; cerebroprotective; nootropic; neuroprotective; antibacterial;
```

```
KW   virucide; fungicide; opthalmalogical; autoimmune disease; neoplasm;
KW   rheumatoid arthritis; hyperproliferative disorder; cardiac arrest;
KW   cardiovascular disorder; cerebrovascular disorder; cerebral ischaemia;
KW   angiogenesis; nervous system disorder; Alzheimer's disease; infection;
KW   ocular disorder; corneal infection; wound healing; skin aging;
KW   epithelial cell proliferation; food additive.
XX
OS   Homo sapiens.
XX
PN   WO200226931-A2.
XX
PD   04-APR-2002.
XX
PF   24-SEP-2001; 2001WO-US029871.
XX
PR   25-SEP-2000; 2000US-0234925P.
PR   12-JAN-2001; 2001WO-US000911.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Komatsoulis G,  Duan DR,  Rosen CA,  Moore PA,  Shi Y;
PI   Lafleur DW,  Olsen H,  Brewer LA,  Florence KA,  Young PE,  Soppet DR;
PI   Endress GA,  Mucenski M,  Ebner R;
XX
DR   WPI; 2002-362489/39.
XX
PT   Novel 71 isolated secreted polypeptides and polynucleotides encoding the
PT   polypeptides, useful for treating Huntington's disease, sepsis,
PT   meningitis, thrombocytopenia, hemolytic anemia, rheumatoid arthritis,
PT   asthma.
XX
PS   Example 2; Page 1434-1440; 1478pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule (or its
CC   fragment, homologue complement or allelic variant) encoding a human
CC   secreted protein (and its fragment, domain, epitope, variant, secreted
CC   form and species variant). Also included are a recombinant vector
CC   comprising the nucleic acid, a recombinant host cell comprising the
CC   vector, an antibody against the secreted protein, a recombinant host cell
CC   that expresses the secreted protein and a method of identifying a binding
CC   partner of the secreted protein. The nucleic acid and protein are used to
CC   prevent, diagnose, treat or ameliorate a medical condition in e.g.
CC   humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep for
CC   example autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
CC   disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
CC   e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC   angiogenesis, nervous system disorders e.g. Alzheimer's disease,
CC   infections caused by bacteria, viruses and fungi and ocular disorders
CC   e.g. corneal infection. Many other diseases and disorders are listed in
CC   the specification. The polypeptides can also be used to aid wound healing
CC   an epithelial cell proliferation, to prevent skin aging due to sunburn,
CC   to maintain organs before transplantation, for supporting cell culture of
CC   primary tissues, to regenerate tissues and in chemotaxis. The
CC   polypeptides can also be used as a food additive or preservative to
CC   increase or decrease storage capabilities. The present sequence
CC   represents a ds DNA fragment of the gene for a novel human secreted
CC   protein of the invention
XX
SQ   Sequence 19965 BP; 4699 A; 5296 C; 5505 G; 4465 T; 0 U; 0 Other;

  Query Match             5.8%;  Score 267.4;  DB 6;  Length 19965;
  Best Local Similarity   63.9%;  Pred. No. 1.5e-36;
  Matches  453;  Conservative    0;  Mismatches  251;  Indels    5;  Gaps    3;

Qy        215 AAAAAAAGGCATTTTATATATATATATATATATATACACACACACACATATATATA 274
              ||||||      |  ||||||||||  | |||||||||| | | | | | |   | |||||
Db       4148 AAAAAATATATATAAATATATATATGAAAATATATATATAAATATATATATGAAAATATA 4207

Qy        275 TACACATATATATACACATATATACATATATACACATATATACACATATATATACACATA 334
              || | |||||||||   | |||||  ||||  | | |||||||  | ||||||||  | |||
Db       4208 TATAAATATATATATGAAAATATATATATGAAAATATATATAAATATATATATGAAAATA 4267

Qy        335 CATATGTACACATATATATACACATATGTATACACATATATACACATATATACACACATA 394
```

```
              |||||  | |  ||||||||  | ||||  |||  | |||||||  | |  |||||  | |  ||
Db       4268 TATATGAAAATATATATATGAAAATATATATGAAAATATATATATAAATATATATATG 4327

Qy        395 TATACACATATATACACACATATATACACATATATACACATATATACACATATACACATA 454
              | | |  |||||||  | |  ||||||   | |||||||  | |  |||||  | ||   | |  |||
Db       4328 AAAATATATATATAAATATATATATGAAAATATATATATAAATATATATATGAAAATATA 4387

Qy        455 TATACACATATATACATATATACACATATATAT-AATATACACACATATATATACACATA 513
              ||||  | |||||||  ||||   | |  ||||||||  ||||||  | |  ||   |  |||  | |||
Db       4388 TATATAAATATATATATATGAAAATATATATATAAATATATATATATGAAAATATATATA 4447

Qy        514 TATACACACATATATACACATATATACACATATATATACACATATATACACATATATACA 573
              ||  | |  |  ||||||   |  |||||||  | |  |||||||  | ||   | |  |  |||||||  |
Db       4448 TAAATATATATATATGAAAATATATATATAAATATATATATATGAAAATATATATATAAA 4507

Qy        574 TATATACACATATATATACATATATACACATATATACATATATACACATATATACATATA 633
              ||||||  |   | |||||  |||||  | |  |||||||   | |||  | |||||  | |||||
Db       4508 TATATATATGAAAATATATATATAAATATATATATATGAAAATATATATATAAATATATA 4567

Qy        634 TACACACATATATACACATACATATACACACACATAGATATACATATATATACACATATA 693
              ||  |   | |||||  | |  |  |||||  |   | |||  |||||  ||||||   | |  |||||
Db       4568 TATATGAAAATATATATAAATATATATATGAATATATATATATATATATGAAAATATATA 4627

Qy        694 TATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTA 753
              ||||  | |  ||||||  ||||  ||||||     | |  |  |||  |       |  |||  ||
Db       4628 TATATGAAAATATATATATAAATATATATATGAAAATATATATAAATATATATATGAATA 4687

Qy        754 CCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAG 813
                  |    ||| |  |  |  |  |    | |  |||    | |  |  || | | |
Db       4688 TATATATATATATATGAAAATATATATATATGAAAATATATATATAAATATATATATGAA 4747

Qy        814 GTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAG 873
                | ||   | |||  |    | |||  |||  ||  |  ||  |  |  |||  | |||   |
Db       4748 AATAAT---ATATATAAATATATATAAATATATATAAAAATATAAATATATATAA-AAAT 4803

Qy        874 TTAAGAAGGTAACACTATATATATATATAGTGAAAGCAGTGTAAACCTA 922
              ||  ||  ||    ||||  |  |||||||   |||    |  ||  |  ||
Db       4804 ATATAAATATATATATATAAACATATATATATAAATATATATATATATA 4852
```

RESULT 28
ABL33237/c
ID   ABL33237 standard; DNA; 6971 BP.
XX
AC   ABL33237;
XX
DT   26-MAR-2002 (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 1210.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP007537.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A, Piepenbrock C, Berlin K;

XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 1210; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 6971 BP; 1665 A; 242 C; 1831 G; 3232 T; 0 U; 1 Other;

Query Match 5.8%; Score 266.4; DB 6; Length 6971;
Best Local Similarity 71.1%; Pred. No. 2.2e-36;
Matches 367; Conservative 0; Mismatches 146; Indels 3; Gaps 1;

```
Qy       212 AAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACACATATAT 271
Db      6472 ATACATATATACCTATATATATATACCTATATATATAAATACACATATATATATATAT 6413

Qy       272 ATATACACATATATATACACATATATACATATATACACATATATACACATATATATACAC 331
Db      6412 ATACCAATATATATATATACACGTATATATATATACACACTTATATACATATATACCTAT 6353

Qy       332 ATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACACAC 391
Db      6352 ATATATAAATACACATATATATATATATACCTATATATATAAATACACATTTATATATAT 6293

Qy       392 ATATATACACATATATACACACATATATACACATATATACACATATATACACATATACAC 451
Db      6292 AAATACACATTTATATATATATACGCCTATATAAATACACATTTATATATATATATACCT 6233

Qy       452 ATATATACACATATATACATATATACACATATATATAATATACACACATATATATACACA 511
Db      6232 ATATATATACCTATATATATAAATACCTATATATATAAAACACATTTATATATATATATA 6173

Qy       512 TATATACACACATATATACACATATATACACATATATATACACATATATACACATATATA 571
Db      6172 CCTATATATAAATACACATTTATATATATACCTCTATATATATACCTATATATAAAAATA 6113

Qy       572 CATATATACACATATATATACATATATACACATATATACATATATACACATATATACATA 631
Db      6112 CACATTTATATATACGCCTATATATAAAAATACACATTTATATATACGCCTATATATATA 6053

Qy       632 TATACACACATATATACACATACATATACACACACATAGATATACATATATAT---ACAC 688
Db      6052 AATACACATTTATATATATATACGCCTATATATATAAATACCTATATATATATAAAACAC 5993

Qy       689 ATATATATACGTATATATATGTATATATATATGCTC 724
Db      5992 ATTTATATATATATATATACCTATATATAAATACAC 5957
```

RESULT 29
ADM01858/c
ID ADM01858 standard; cDNA; 2522 BP.
XX
AC ADM01858;
XX
DT 20-MAY-2004 (first entry)
XX
DE Human cDNA of the invention SEQ ID NO:543.
XX
KW ss; gene; human; gene therapy; diagnostic marker; pharmaceutical.
XX
OS Homo sapiens.
XX
PN EP1347046-A1.
XX
PD 24-SEP-2003.
XX
PF 12-APR-2002; 2002EP-00008400.
XX
PR 22-MAR-2002; 2002JP-00137785.
XX
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX
DR WPI; 2003-723558/69.
DR P-PSDB; ADM04301.
XX
PT New polynucleotides and polypeptides are useful in gene therapy, for
PT developing a diagnostic marker or medicines for regulating their
PT expression and activity, or as a target of gene therapy.
XX
PS Claim 1; SEQ ID NO 543; 305pp; English.
XX
CC The invention relates to a novel human polynucleotide and the encoded
CC polypeptide. A polynucleotide of the invention may have a use in gene
CC therapy. An oligonucleotide of the invention ADM06202-ADM06773 is useful
CC as a primer for synthesizing the polynucleotide or as a probe for
CC detecting the polynucleotide. The polynucleotides ADM01316-ADM03758 are
CC useful in gene therapy, for developing a diagnostic marker or medicines
CC for regulating their expression and activity, or as a target of gene
CC therapy. The proteins ADM03759-ADM06201 encoded by the polynucleotides
CC are useful as pharmaceutical agents. The present sequence represents a
CC cDNA sequence of the invention.
XX
SQ Sequence 2522 BP; 789 A; 399 C; 547 G; 787 T; 0 U; 0 Other;

Query Match 5.7%; Score 263.6; DB 11; Length 2522;
Best Local Similarity 68.9%; Pred. No. 6.3e-36;
Matches 376; Conservative 0; Mismatches 169; Indels 1; Gaps 1;

```
Qy       236 ATATATATATATATATACACACACACACACATATATATATACACATATATATACACATAT 295
Db      2217 ATTCAACTGTGTATATATATATACACACATACGTATATATGTGTATATATATACACACGT 2158

Qy       296 ATACATATATACACATATATACACATATATATACACATACATATGTACACATATATATAC 355
Db      2157 ATATATACACATATATGTATACATACACACGCATATATACACATATATGTATACATACAC 2098

Qy       356 ACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACAT 415
Db      2097 ACGTATATATACACATATATGTATACATACACACGCATATATACACATATATGTATACAT 2038

Qy       416 ATATACACATATATACACATATATACACATATACACATATATACACATATATACATATAT 475
Db      2037 ACACACGTATATATACACATATATGCATACATACACACGCATATATACACATATATGCAT 1978

Qy       476 ACACATATATATAATATACACACATATATATACACATATATACACACATATATACACATA 535
Db      1977 ACATACATACGT-ATATATACACATATATGTATACATACATACGTATACATACACATATA 1919

Qy       536 TATACACATATATATACACATATATACACATATATACATATATACACATATATATACATA 595
Db      1918 TGTATACATACATGCGCATATATATACACATATATGTATACATACATACGTATATATACACA 1859

Qy       596 TATACACATATATACATATATACACATATATACATATATACACACATATATACACATACA 655
Db      1858 TATATGTATACATACATACGTATACATACACATATGTATACATACATACGTATATATACA 1799

Qy       656 TATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATA 715
Db      1798 CATATATGTATACATACATGCGTATATATACACATATATGTATACATACATGCGTATATA 1739
```

```
Qy        716 TATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGA 775
              || |       |  | ||||      | ||  |    ||    || | ||| |   | | |
Db       1738 TACACATATATGCATACATACATATGAACACATATATATGTGTGTGTATATATACATATA 1679

Qy        776 GAGTTT 781
                 |||
Db       1678 AGATTT 1673
```

RESULT 30
AAD54480/c
ID AAD54480 standard; DNA; 117962 BP.
XX
AC AAD54480;
XX
DT 26-JUN-2003 (first entry)
XX
DE Human CIP DNA #1.
XX
KW Human; p53 pathway; chloride cotransporter interactor protein; CIP;
KW angiogenic disorder; cell proliferation disorder; apoptotic disorder;
KW breast cancer; gene therapy; ds.
XX
OS Homo sapiens.
XX
PN WO200299055-A2.
XX
PD 12-DEC-2002.
XX
PF 03-JUN-2002; 2002WO-US017473.
XX
PR 05-JUN-2001; 2001US-0296076P.
PR 10-OCT-2001; 2001US-0328605P.
PR 15-FEB-2002; 2002US-0357253P.
XX
PA (EXEL-) EXELIXIS INC.
XX
PI Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP;
XX
DR WPI; 2003-175140/17.
XX
PT Identifying p53 pathway modulators for treating or diagnosing disorders
PT with defective p53 function e.g. breast cancer, by providing an assay
PT system having a purified cotransporter interactor protein (CIP)
PT polypeptide or nucleic acid.
XX
PS Disclosure; Page 38-101; 123pp; English.
XX
CC The invention relates to a method of Identifying p53 pathway modulating
CC agent. The method involves contacting a test agent with an assay system
CC comprising a purified cation Cl- cotransporter interactor protein (CIP)
CC polypeptide or polynucleotide, or their functionally active fragment or
CC derivative. The method is useful for identifying modulators of the p53
CC pathway particularly for identifying agents for treating disorders (e.g.
CC breast cancer) associated with defective p53 function. Modulators of the
CC invention are useful as targets for novel therapeutics. CIP sequences are
CC useful as modifiers of the p53 pathway, and as therapeutic targets for
CC disorders associated with defective p53 function e.g. angiogenic,
CC apoptotic or cell proliferation disorders. The invention is useful in
CC gene therapy. The present sequence is human CIP DNA
XX
SQ Sequence 117962 BP; 27840 A; 32096 C; 30624 G; 27402 T; 0 U; 0 Other;

Query Match 5.7%; Score 263; DB 8; Length 117962;
Best Local Similarity 71.6%; Pred. No. 9.4e-36;
Matches 373; Conservative 0; Mismatches 145; Indels 3; Gaps 2;

```
Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||  ||||||||||  ||||||||||||||||||||||||||||||||||| ||||||
Db     107030 AGTTTGAGACCAGCCTGGCCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAAATT 106971

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
                || |||||||||||  ||| |||||||||||||||||| | |||||| || | |||||||
Db     106970 AGCCAGGCATGGTGGCACATGCCTGTAATCCCAGCTACCTGGGAGGCTGAAGTAGGAGAA 106911

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              || || ||||| |  || ||  |||| ||| | |||| |         | | |   ||
Db     106910 TCTCTGGAACCAG--GGAGGCAGAGGCTGCAGTGAGCCAAGATCATTCCTGGGTGTCAGA 106853

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              || |  ||      |       |  | || ||||||||||    |     | | | | |
Db     106852 GAGAGACTCCATCCCTGCCCGTCCCCCGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 106793

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              | | | ||||||| | | | | | | |||||||||||| | ||||| | ||||| | |||
Db     106792 AAAAAAATATATATATATATATATATATATATATATATATATATACACACACACACATAT 106733

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACAT 359
              ||||| | | ||||| |||||||||||||||| ||| |||  || | ||| ||||| | ||
Db     106732 ATATACATATATATACACACATATATATACATATATATACATATATATACATATATATAT 106673

Qy        360 ATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATAT 419
              | ||||| | ||| ||| | ||||| ||| | ||||| ||| |||||||||| | ||||||
Db     106672 ACGTATATATATACATATATATATACACATATATATACACATATATATACATATATATAT 106613

Qy        420 ACACATATATACACATATATACACATATACACATATATACACATATATACATATATACAC 479
              |||||||||| ||| ||||| | | ||| | ||||| ||| ||||||||||||||| |
Db     106612 ACACATATATATACACATATATATACATATATATATACACATATATATACATATATATAT 106553

Qy        480 ATATATATAATATACACACATATATATACACATATATACAC 520
              || |||||  |||  | |       || |    || || |||
Db     106552 ATGTATATTTTATTGAGATCACATTAAATGTGTACATTCAC 106512
```

RESULT 31
ABK83460
ID ABK83460 standard; cDNA; 165199 BP.
XX
AC ABK83460;
XX
DT 14-AUG-2002 (first entry)
XX
DE Human cDNA differentially expressed in granulocytic cells #31.
XX
KW Human; ss; granulocytic cell; DNA chip; bacterial infection;
KW viral infection; parasitic infection; protozoal infection;
KW fungal infection; sterile inflammatory disease; psoriasis;
KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KW cardiac reperfusion injury; renal reperfusion injury; ARDS;
KW adult respiratory distress syndrome; inflammatory bowel disease;
KW Crohn's disease; ulcerative colitis; periodontal disease;
KW granulocyte activation; chronic inflammation; allergy.
XX
OS Homo sapiens.
XX
PN WO200228999-A2.
XX
PD 11-APR-2002.
XX
PF 03-OCT-2001; 2001WO-US030821.
XX
PR 03-OCT-2000; 2000US-0237189P.
XX
PA (GENE-) GENE LOGIC INC.
XX
PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
XX
DR WPI; 2002-435328/46.
XX
PT Detecting granulocyte activation by detecting differential expression of
PT genes associated with granulocyte activation, which serves as diagnostic
PT markers that is useful for monitoring disease states and drug toxicity.
XX
PS Claim 1; SEQ ID NO 31; 114pp; English.
XX

```
CC   The invention relates to detecting (M1) granulocyte (GC) activation
CC   (GCA), by detecting the level of expression of gene(s) (Gs) identified by
CC   DNA chip analysis as given in the specification, and comparing the
CC   expression level to an expression level in an unactivated GC, where
CC   differential expression of Gs is indicative of GCA. Also included are
CC   modulating (M2) GA by contacting GC with an agent that alters the
CC   expression of at least one gene in Gs; (2) screening (M3) for an agent
CC   capable of modulating GCA or an inflammation (especially chronic) in a
CC   tissue, an allergic response in a subject, exposure of a subject to a
CC   pathogen or sterile inflammatory disease using the gene expression
CC   profile; (3) detecting (M4) an inflammation (especially chronic) in a
CC   tissue, an allergic response in a subject, exposure of a subject to a
CC   pathogen or sterile inflammatory disease, by detecting the level of
CC   expression in a sample of the tissue of gene(s) from Gs, where the level
CC   of expression of the gene is indicative of inflammation; (4) treating
CC   (M5) an inflammation (especially chronic) or in a tissue, an allergic
CC   response in a subject, exposure of a subject to a pathogen or sterile
CC   inflammatory disease, by contacting a tissue having inflammation with an
CC   agent that modulates the expression of gene(s) from Gs in the tissue. M1
CC   is useful for detecting GCA; M2 is useful for modulating GA; M3 is useful
CC   for screening an agent capable of modulating GCA preferably in an
CC   inflammation in a tissue; M4 is useful for detecting an inflammation
CC   (especially chronic) in a tissue, an allergic response in a subject,
CC   exposure of a subject to a pathogen or sterile inflammatory disease (e.g.
CC   psoriasis, rheumatoid arthritis, glomerulonephritis, asthma, thrombosis,
CC   cardiac reperfusion injury, renal reperfusion injury, ARDS, adult
CC   respiratory distress syndrome, inflammatory bowel disease, Crohn's
CC   disease, ulcerative colitis, periodontal disease; also bacterial
CC   infection, viral infection, parasitic infection, protozoal infection,
CC   fungal infection and M5 is useful for treating one of the above
CC   conditions. The present sequence represents a gene differentially
CC   expressed in granulocytes. Note: The sequence data for this patent did
CC   not form part of the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 165199 BP; 48510 A; 33223 C; 34406 G; 49060 T; 0 U; 0 Other;

  Query Match             5.7%;  Score 263;  DB 6;  Length 165199;
  Best Local Similarity   69.3%;  Pred. No. 9.5e-36;
  Matches  472;  Conservative    0;  Mismatches  200;  Indels    9;  Gaps    8;

Qy         46 AAAAATACAAAATTAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGG 105
              | ||| | | |   |   | |||||||| | ||| ||||||||||||||||||   |||||
Db     138669 AGAAAGAAAGATGGATGGATGCATGGTGGTGCACACCTGTAATCCCAGCTACACAGGAGG 138728

Qy        106 CCGAGGCAGGAGAATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGA-GCAAGATCAT 164
              | | ||||  | ||||||||||||||||  || ||| ||| | ||   ||  ||||||  |
Db     138729 CTGGGGCATTAAAATCACTTGAACCTG--GGAGGTAGAGTTAGCAATGACACAAGATTGT 138786

Qy        165 GCCATTGCACTCCAGCCCAGGCAACAAGAGCGAAATGTC-ATCTCAGAAAAAAAAAAAGG 223
              ||   ||||||||||||  ||   ||  |    | ||||  ||||| | |   |   |
Db     138787 GCTGCTGCACTCCAGCCTGGGTGGCAGAATGAGACTGTCTCTCTCTGTATCTAGCTATCT 138846

Qy        224 CATTTTATATATATATATATATATATACACACACA-CACACATATATATATACACATA 282
                |  |||||||||  ||||| | || | | | || | ||||  ||||| | |||
Db     138847 ATCCATCCATATATATATCCATATATGTCCATATATATCATATATATCCATATATAGATA 138906

Qy        283 TATATACACATATATACATATATACACATATATACACATATATATACACATACATATGT- 341
              ||||||   |||||||  |||||| | ||  ||| | |||| |||| | ||  |||| |
Db     138907 TATATATCTATATATATCTATATATA-ATCCATATATATATCTATATATATCTATATATA 138965

Qy        342 ACACATATATATACACATATGTATACACATATATACACATATATACACACATATATACAC 401
              |  ||||||||||  |||| ||| || ||||| | | |||||| || | ||||  | |
Db     138966 ATCCATATATATCCATATATATCCATATATAGATATATATCCATATATATATCCATAT 139025

Qy        402 ATATATACACACATATATACACATATATACACATATATACA-CATATACACATATATACA 460
              ||| ||| | | ||  ||| | |||| || | || |||| |  ||  | | |||||| ||
Db     139026 ATAGATATATATATCCATATATATATCTATATATCTATATATAATCAATATATATATCCA 139085

Qy        461 CATATATACATATATACACATATATA-TAATATACACACATATATATACACATATATACA 519
              |||||  |  ||||||  |||||||| | |||||  |  ||||||||||  |||||||| |
Db     139086 TATATAGATGTATATATCCATATATAGTTATATATATCCATATATATATCCATATATAGA 139145

Qy        520 CACATATATACACATATATACACATATATATACACATATATACACATATATACATATATA 579
              | |||||| || |||||||  ||||||||  | | ||||||  || |||||||   |||||
Db     139146 TATATATATCCATATATATATCCATATAGATATATATATCCATATATATATCCATATA 139205

Qy        580 CACATATATATACATATATACACATATATACATATATACACATATATACATATATACACA 639
              | |||||||||   |||||  | ||  ||| ||| ||| | ||||  | |||||||  ||
Db     139206 TAGATATATATATCCATATATATATCCATATATAGATATATATATCCATATATATATCCA 139265

Qy        640 CATATATACACATACATATATACACACACATAGATATACATATATATACACATATATATACG 699
              |||||  |  |||   |||  |  |  |   | |||||   |||||||  | ||||||||  |
Db     139266 TATATAGATATATATCCATATATATATCCATATATATCCATATATAGATATATATATCCA 139325

Qy        700 TATATATATGTATATATATAT 720
              |||||||||  |||||||  ||
Db     139326 TATATATATCCATATATAGAT 139346


RESULT 32
ABK83497/c
ID   ABK83497 standard; cDNA; 227968 BP.
XX
AC   ABK83497;
XX
DT   14-AUG-2002  (first entry)
XX
DE   Human cDNA differentially expressed in granulocytic cells #68.
XX
KW   Human; ss; granulocytic cell; DNA chip; bacterial infection;
KW   viral infection; parasitic infection; protozoal infection;
KW   fungal infection; sterile inflammatory disease; psoriasis;
KW   rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KW   cardiac reperfusion injury; renal reperfusion injury; ARDS;
KW   adult respiratory distress syndrome; inflammatory bowel disease;
KW   Crohn's disease; ulcerative colitis; periodontal disease;
KW   granulocyte activation; chronic inflammation; allergy.
XX
OS   Homo sapiens.
XX
PN   WO200228999-A2.
XX
PD   11-APR-2002.
XX
PF   03-OCT-2001; 2001WO-US030821.
XX
PR   03-OCT-2000; 2000US-0237189P.
XX
PA   (GENE-) GENE LOGIC INC.
XX
PI   Beazer-Barclay Y,  Weissman SM,  Yamaga S,  Vockley J;
XX
DR   WPI; 2002-435328/46.
XX
PT   Detecting granulocyte activation by detecting differential expression of
PT   genes associated with granulocyte activation, which serves as diagnostic
PT   markers that is useful for monitoring disease states and drug toxicity.
XX
PS   Claim 1; SEQ ID NO 68; 114pp; English.
XX
CC   The invention relates to detecting (M1) granulocyte (GC) activation
CC   (GCA), by detecting the level of expression of gene(s) (Gs) identified by
CC   DNA chip analysis as given in the specification, and comparing the
CC   expression level to an expression level in an unactivated GC, where
CC   differential expression of Gs is indicative of GCA. Also included are
CC   modulating (M2) GA by contacting GC with an agent that alters the
CC   expression of at least one gene in Gs; (2) screening (M3) for an agent
CC   capable of modulating GCA or an inflammation (especially chronic) in a
CC   tissue, an allergic response in a subject, exposure of a subject to a
CC   pathogen or sterile inflammatory disease using the gene expression
CC   profile; (3) detecting (M4) an inflammation (especially chronic) in a
CC   tissue, an allergic response in a subject, exposure of a subject to a
CC   pathogen or sterile inflammatory disease, by detecting the level of
CC   expression in a sample of the tissue of gene(s) from Gs, where the level
```

```
CC   of expression of the gene is indicative of inflammation; (4) treating
CC   (M5) an inflammation (especially chronic) or in a tissue, an allergic
CC   response in a subject, exposure of a subject to a pathogen or sterile
CC   inflammatory disease, by contacting a tissue having inflammation with an
CC   agent that modulates the expression of gene(s) from Gs in the tissue. M1
CC   is useful for detecting GCA; M2 is useful for modulating GA; M3 is useful
CC   for screening an agent capable of modulating GCA preferably in an
CC   inflammation in a tissue; M4 is useful for detecting an inflammation
CC   (especially chronic) in a tissue, an allergic response in a subject,
CC   exposure of a subject to a pathogen or sterile inflammatory disease (e.g.
CC   psoriasis, rheumatoid arthritis, glomerulonephritis, asthma, thrombosis,
CC   cardiac reperfusion injury, renal reperfusion injury, ARDS, adult
CC   respiratory distress syndrome, inflammatory bowel disease, Crohn's
CC   disease, ulcerative colitis, periodontal disease; also bacterial
CC   infection, viral infection, parasitic infection, protozoal infection,
CC   fungal infection and M5 is useful for treating one of the above
CC   conditions. The present sequence represents a gene differentially
CC   expressed in granulocytes. Note: The sequence data for this patent did
CC   not form part of the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 227968 BP; 53185 A; 60440 C; 61985 G; 52358 T; 0 U; 0 Other;

  Query Match           5.7%;  Score 263;  DB 6;  Length 227968;
  Best Local Similarity   71.6%;  Pred. No. 9.6e-36;
  Matches  373;  Conservative    0;  Mismatches  145;  Indels    3;  Gaps    2;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||  ||||||||||  ||||||||||||||||||||||||||||||||| ||||||
Db       5103 AGTTTGAGACCAGCCTGGCCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAAATT 5044

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||  ||||||||||  ||| |||||||||||||||||| | |||||| || | |||||||
Db       5043 AGCCAGGCATGGTGGCACATGCCTGTAATCCCAGCTACCTGGGAGGCTGAAGTAGGAGAA 4984

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              || || |||||  |  || ||  |||| ||| | |||| |         | | |   ||
Db       4983 TCTCTGGAACCAG--GGAGGCAGAGGCTGCAGTGAGCCAAGATCATTCCTGGGTGTCAGA 4926

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
                || |  ||      |       |  | || ||||||||||||   |    | | | | |
Db       4925 GAGAGACTCCATCCCTGCCCGTCCCCCGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4866

Qy        240 ATATATATATATACACACACACACACATATATATATACACATATATATACACATATATAC 299
              | | | ||||||| | | | | | | |||||||||||| | ||||| | ||||| | |||
Db       4865 AAAAAAATATATATATATATATATATATATATATATATATATATACACACACACACATAT 4806

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACAT 359
              ||||| | | ||||| |||||||||||||||| ||| |||  || | ||| ||||| | ||
Db       4805 ATATACATATATATACACACATATATATACATATATATACATATATATACATATATATAT 4746

Qy        360 ATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATAT 419
              | ||||| | ||| ||| | ||||| ||| | ||||| ||| |||||||| | ||||||
Db       4745 ACGTATATATATACATATATATATACACATATATATACACATATATATACATATATATAT 4686

Qy        420 ACACATATATACACATATATACACATATACACATATATACACATATATACATATATACAC 479
              |||||||||| ||| ||||| | | ||| | ||||| ||| ||||||||||||||| |
Db       4685 ACACATATATATACACATATATATACATATATATATACACATATATATACATATATATAT 4626

Qy        480 ATATATATAATATACACACATATATATACACATATATACAC 520
              || |||||  |||  | |       || |    || || |||
Db       4625 ATGTATATTTTATTGAGATCACATTAAATGTGTACATTCAC 4585


RESULT 33
ADQ18538/c
ID   ADQ18538 standard; DNA; 227968 BP.
XX
AC   ADQ18538;
XX
DT   26-AUG-2004  (first entry)
XX
DE   Human soft tissue sarcoma-upregulated DNA - SEQ ID 1357.
XX
KW   soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO2004048938-A2.
XX
PD   10-JUN-2004.
XX
PF   26-NOV-2003; 2003WO-US038193.
XX
PR   26-NOV-2002; 2002US-0429739P.
XX
PA   (PROT-) PROTEIN DESIGN LABS INC.
XX
PI   Aziz N,  Ginsburg WM,  Zlotnik A;
XX
DR   WPI; 2004-441208/41.
XX
PT   Early detection of soft tissue sarcoma comprises determining expression
PT   of a gene in a first soft tissue sample and a normal soft tissue sample
PT   and comparing the gene expression, also useful in treating soft tissue
PT   sarcoma.
XX
PS   Example 2; SEQ ID NO 1357; 210pp; English.
XX
CC   The invention relates to a novel method for detecting soft tissue sarcoma
CC   which comprises obtaining a first soft tissue sample from an individual
CC   and a normal soft tissue sample from the same or different individual,
CC   determining the expression of a gene in both samples and comparing the
CC   expression of the gene in both soft tissue samples, where a higher level
CC   of protein expression in the first soft tissue sample indicates the
CC   presence of soft tissue sarcoma. The method of the invention has
CC   cytostatic applications and may be useful for detecting soft tissue
CC   sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC   acid sequences may be useful in diagnostic and screening applications.
CC   The current sequence is that of a human soft tissue sarcoma-upregulated
CC   DNA of the invention. The current sequence is not shown within the
CC   specification per se but was submitted in CD format by the inventor.
XX
SQ   Sequence 227968 BP; 53185 A; 60440 C; 61985 G; 52358 T; 0 U; 0 Other;

  Query Match           5.7%;  Score 263;  DB 12;  Length 227968;
  Best Local Similarity   71.6%;  Pred. No. 9.6e-36;
  Matches  373;  Conservative    0;  Mismatches  145;  Indels    3;  Gaps    2;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||  ||||||||||  ||||||||||||||||||||||||||||||||| ||||||
Db       5103 AGTTTGAGACCAGCCTGGCCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAAATT 5044

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||  ||||||||||  ||| |||||||||||||||||| | |||||| || | |||||||
Db       5043 AGCCAGGCATGGTGGCACATGCCTGTAATCCCAGCTACCTGGGAGGCTGAAGTAGGAGAA 4984

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              || || |||||  |  || ||  |||| ||| | |||| |         | | |   ||
Db       4983 TCTCTGGAACCAG--GGAGGCAGAGGCTGCAGTGAGCCAAGATCATTCCTGGGTGTCAGA 4926

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
                || |  ||      |       |  | || ||||||||||||   |    | | | | |
Db       4925 GAGAGACTCCATCCCTGCCCGTCCCCCGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4866

Qy        240 ATATATATATATACACACACACACACATATATATATACACATATATATACACATATATAC 299
              | | | ||||||| | | | | | | |||||||||||| | ||||| | ||||| | |||
Db       4865 AAAAAAATATATATATATATATATATATATATATATATATATATACACACACACACATAT 4806

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACAT 359
              ||||| | | ||||| |||||||||||||||| ||| |||  || | ||| ||||| | ||
Db       4805 ATATACATATATATACACACATATATATACATATATATACATATATATACATATATATAT 4746

Qy        360 ATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATAT 419
```

```
                 | ||||| | ||| ||| | ||||| ||| | ||||| ||| |||||||||| | ||||||
Db       4745 ACGTATATATATACATATATATATACACATATATATACACATATATATACATATATATAT 4686

Qy        420 ACACATATATACACATATATACACATATACACATATATACACATATATACATATATACAC 479
              |||||||||| ||| ||||| | | ||| | ||||| ||| ||||||||||||||| |
Db       4685 ACACATATATATACACATATATATACATATATATATACACATATATATACATATATATAT 4626

Qy        480 ATATATATAATATACACACATATATATACACATATATACAC 520
              || |||||  |||  | |      || |    || || |||
Db       4625 ATGTATATTTTATTGAGATCACATTAAATGTGTACATTCAC 4585
```

RESULT 34
ABK39930/c
ID ABK39930 standard; DNA; 12069 BP.
XX
AC ABK39930;
XX
DT 21-MAY-2002 (first entry)
XX
DE Human chemically pretreated gene sequence #6 strand 1.
XX
KW Human; ds; bisulphite treatment; CpG; DNA methylation; cancer; tumour;
KW cytostatic; ALDH6; CYP11A; CYP11B1; CYP3A3; DPYD; EPHX2; OCLN; TXNRD1;
KW UGT8; MRP; pharmacogenomics; SNP; single nucleotide polymorphism.
XX
OS Homo sapiens.
XX
PN WO200202806-A2.
XX
PD 10-JAN-2002.
XX
PF 29-JUN-2001; 2001WO-EP007470.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-154757/20.
XX
PT New nucleic acid, oligonucleotides and peptide nucleic acid-oligomers,
PT useful for detecting cytosine methylation state of genes associated with
PT pharmacogenomics and for therapy of diseases e.g. cancer.
XX
PS Claim 1; SEQ ID NO 11; 24pp; English.
XX
CC The invention relates to a nucleic acid comprising a sequence at least 18
CC bases in length of a segment of the chemically pretreated DNA of genes
CC associated with pharmacogenomics according to one of the sequences of the
CC genes ALDH6 (NM_000693), CYP11A (NM_000781), CYP11B1 (NM_000497), CYP3A3
CC (NM_000776 and NM_017460), DPYD (NM_000110), EPHX2 (NM_001979), OCLN
CC (NM_002538), TXNRD1 (NM_003330), UGT8 (NM_003360), MRP (NM_004996,
CC NM_019900, NM_019901, NM_019902, NM_019862, NM_019898, NM_019899) and
CC their complementary sequences, or a sequence (S1) chosen from 87
CC sequences and their complements. The chemical pretreatment is bisulphite
CC treatment to convert cytosines (but not methyl-cytosines) into uracils.
CC Also included are an oligomer (II) in particular an oligonucleotide or a
CC peptide nucleic acid (PNA)-oligomer, comprising in each case at least one
CC base sequence having a length of 9 nucleotides which hybridises to or is
CC identical to a chemically pretreated DNA of genes associated with
CC pharmacogenomics and their complements, arranged in an array for
CC analysing diseases associated with the methylation state (CpG) and/or
CC detecting SNPs (single nucleotide polymorphisms) of the 87 sequences. The
CC oligomers may also be used as PCR primers. The set of 87 nucleic acids
CC and their complements is useful for diagnosis and therapy of solid
CC tumours and cancer. The present sequence represents one the 87 DNA
CC sequences or its complement. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 12069 BP; 3210 A; 185 C; 2849 G; 5823 T; 0 U; 2 Other;

Query Match 5.7%; Score 261.4; DB 6; Length 12069;
Best Local Similarity 70.5%; Pred. No. 1.6e-35;
Matches 363; Conservative 0; Mismatches 151; Indels 1; Gaps 1;

```
Qy        204 ATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATATACACACACACAC 263
              || ||   | | | | ||      | |||||||||||| |||| |||||| | | | | |
Db       9846 ATATCTATATATATATAAATATATCTATATATATATAAATATCTATATATATATAAATAT 9787

Qy        264 ACATATATATATACACATATATATACACATATATACATATATACACATATATACACATAT 323
              |||||  |||| | || |||||| | ||| |||  ||||| ||| |||||  | ||||
Db       9786 CTATATACGTATAAATATCTATATATATATAAATATCTATATCCACGTATATCTATATAT 9727

Qy        324 ATATACACATACATATGTACACATATATATACACATATGTATACACATATATACACATAT 383
              ||||| | ||  |||| || | | |||||||  |   | |||||   ||||||| | | ||
Db       9726 ATATAAATATCTATATATATATAAATATATATATCCACGTATATCTATATATATATAAAT 9667

Qy        384 ATACACACATATATACACATATATACACACATATATACACATATATACACATATATACAC 443
              ||  | | ||||||| | |||||||  ||| |||||  | ||||||| | |||| |||
Db       9666 ATCTATATATATATAAATATATATATCCACGTATATCTATATATATATAAATATCTACGT 9607

Qy        444 ATATACACATATATACACATATATACATATATACACATATATATAATATACACACATATA 503
              ||| | | |||||| ||| |||||  ||||||| | | |||| | |  || | | |||||
Db       9606 ATAAATATATATATCCACGTATATCTATATATATATAAATATCT-ACGTATAAATATATA 9548

Qy        504 TATACACATATATACACACATATATACACATATATACACATATATATACACATATATACA 563
              ||| ||| |||||  | | ||||||| | || |||| | |||| ||| | ||||  ||
Db       9547 TATCCACGTATATCTATATATATATAAATATCTATATATATATAAATATATATATCCACG 9488

Qy        564 CATATATACATATATACACATATATATACATATATACACATATATACATATATACACATA 623
              |||| || ||||||| | |||| |||| ||||||| | || | || ||||||| | |||
Db       9487 TATATCTATATATATATAAATATCTATATATATATAAATATCTCTATATATATAAATATA 9428

Qy        624 TATACATATATACACACATATATACACATACATATACACACACATAGATATACATATATA 683
              ||||   | ||| |   ||||||| | | | || || | | | ||| |||| |||||||
Db       9427 TATATCCACATATATCTATATATATATAAATATCTATATATATATAAATATCTATATATA 9368

Qy        684 TACACATATATATACGTATATATATGTATATATAT 718
              || | | |||| ||  |||||||  ||| |||||
Db       9367 TATATAAATATCTATATATATATAAATATCTATAT 9333
```

RESULT 35
AAK73165/c
ID AAK73165 standard; DNA; 20420 BP.
XX
AC AAK73165;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:27977.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001354.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.

PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS Disclosure; SEQ ID NO 27977; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased

```
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting the
CC   nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention
XX
SQ   Sequence 20420 BP; 4834 A; 5403 C; 5630 G; 4553 T; 0 U; 0 Other;

  Query Match             5.6%;  Score 260.6;  DB 4;  Length 20420;
  Best Local Similarity   71.7%;  Pred. No. 2.2e-35;
  Matches  355;  Conservative    0;  Mismatches  139;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              |||  ||||||||||||||||||||    || | | | |  |||||||||   ||||||||
Db       4732 TTTTCATATATATATATATATATTTTCATATATATATTTATATATATTTTCATATATA 4673

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || |  |||||| |||| | || ||||||| |  |||||||| |  | ||||| || |
Db       4672 TATATTTATATATATATTTTCATATATATATATTTATATATATATTTTCATATATATATT 4613

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| |  |   |||| | || |||| | |||| | || | |||||||   |||||
Db       4612 TATATATATATTTTCATATATATATTTATATATATATTTTCATATATATATATTTATATA 4553

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || |     ||||| | |||| || | ||||||   |||||| | |||| || | |||||
Db       4552 TATATTTTCATATATATATATTTATATATATATTTTCATATATATATATTTATATATATA 4493

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              |    ||||| | |||| |||| |||||    |||||||||| |  |||||| | |  |
Db       4492 TTTTCATATATATATATTTATATATATATTTTCATATATATATATTTATATATATATTTT 4433

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              ||| | ||||||  | ||||||||   ||||||||  | || |||| ||||||   ||||
Db       4432 CATATATATATATTTATATATATATTTTCATATATATATATTTATATATATATTTTCATA 4373

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              ||||||  |||||| | || |  | |||||||  |  |||||| |||| | || | ||||
Db       4372 TATATATTTATATATATATTTTCATATATATATATTTATATATATATTTTCATATATATT 4313

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              | || ||| ||||   || | ||| || || ||||||||  | |||||| | | ||||||
Db       4312 TTCATATATATATTTTCATATATATATTTATATATATATTTATATATATTTTCATATATA 4253

Qy        706 TATGTATATATATAT 720
              ||| ||||||||||||
Db       4252 TATTTATATATATAT 4238


RESULT 36
ABK69933/c
ID   ABK69933 standard; DNA; 20420 BP.
XX
AC   ABK69933;
XX
DT   15-JUL-2002  (first entry)
XX
DE   Human secreted protein gene 68 genomic DNA fragment #21.
XX
KW   Human; ds; secreted protein; gene therapy; immunosuppressive;
KW   antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;
KW   vasotropic; cerebroprotective; nootropic; neuroprotective; antibacterial;
KW   virucide; fungicide; opthalmalogical; autoimmune disease; neoplasm;
KW   rheumatoid arthritis; hyperproliferative disorder; cardiac arrest;
KW   cardiovascular disorder; cerebrovascular disorder; cerebral ischaemia;
KW   angiogenesis; nervous system disorder; Alzheimer's disease; infection;
KW   ocular disorder; corneal infection; wound healing; skin aging;
KW   epithelial cell proliferation; food additive.
XX
OS   Homo sapiens.
XX
PN   WO200226931-A2.
XX
PD   04-APR-2002.
XX
PF   24-SEP-2001; 2001WO-US029871.
XX
PR   25-SEP-2000; 2000US-0234925P.
PR   12-JAN-2001; 2001WO-US000911.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Komatsoulis G,  Duan DR,  Rosen CA,  Moore PA,  Shi Y;
PI   Lafleur DW,  Olsen H,  Brewer LA,  Florence KA,  Young PE,  Soppet DR;
PI   Endress GA,  Mucenski M,  Ebner R;
XX
DR   WPI; 2002-362489/39.
XX
PT   Novel 71 isolated secreted polypeptides and polynucleotides encoding the
PT   polypeptides, useful for treating Huntington's disease, sepsis,
PT   meningitis, thrombocytopenia, hemolytic anemia, rheumatoid arthritis,
PT   asthma.
XX
PS   Example 2; Page 1440-1445; 1478pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule (or its
CC   fragment, homologue complement or allelic variant) encoding a human
CC   secreted protein (and its fragment, domain, epitope, variant, secreted
CC   form and species variant). Also included are a recombinant vector
CC   comprising the nucleic acid, a recombinant host cell comprising the
CC   vector, an antibody against the secreted protein, a recombinant host cell
CC   that expresses the secreted protein and a method of identifying a binding
CC   partner of the secreted protein. The nucleic acid and protein are used to
CC   prevent, diagnose, treat or ameliorate a medical condition in e.g.
CC   humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep for
CC   example autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
CC   disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
CC   e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC   angiogenesis, nervous system disorders e.g. Alzheimer's disease,
CC   infections caused by bacteria, viruses and fungi and ocular disorders
CC   e.g. corneal infection. Many other diseases and disorders are listed in
CC   the specification. The polypeptides can also be used to aid wound healing
CC   an epithelial cell proliferation, to prevent skin aging due to sunburn,
CC   to maintain organs before transplantation, for supporting cell culture of
CC   primary tissues, to regenerate tissues and in chemotaxis. The
CC   polypeptides can also be used as a food additive or preservative to
CC   increase or decrease storage capabilities. The present sequence
CC   represents a ds DNA fragment of the gene for a novel human secreted
CC   protein of the invention
XX
SQ   Sequence 20420 BP; 4834 A; 5403 C; 5630 G; 4553 T; 0 U; 0 Other;

  Query Match             5.6%;  Score 260.6;  DB 6;  Length 20420;
  Best Local Similarity   71.7%;  Pred. No. 2.2e-35;
  Matches  355;  Conservative    0;  Mismatches  139;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              |||  ||||||||||||||||||||    || | | | |  |||||||||   ||||||||
Db       4732 TTTTCATATATATATATATATATTTTCATATATATATTTATATATATTTTCATATATA 4673

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || |  |||||| |||| | || ||||||| |  |||||||| |  | ||||| || |
Db       4672 TATATTTATATATATATTTTCATATATATATATTTATATATATATTTTCATATATATATT 4613

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| |  |   |||| | || |||| | |||| | || | |||||||   |||||
Db       4612 TATATATATATTTTCATATATATATTTATATATATATTTTCATATATATATATTTATATA 4553

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || |     ||||| | |||| || | ||||||   |||||| | |||| || | |||||
```

```
Db       4552 TATATTTTCATATATATATATTTATATATATATTTTCATATATATATATTTATATATATA 4493

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              |    ||||| | |||| |||| |||||     |||||||||| |  |||||| | |   |
Db       4492 TTTTCATATATATATATTTATATATATATTTTCATATATATATATTTATATATATATTTT 4433

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              ||| | ||||||   | ||||||||    |||||||| | || |||| ||||||    ||||
Db       4432 CATATATATATATTTATATATATATTTTCATATATATATATTTATATATATATTTTCATA 4373

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              ||||||  |||||| | || |  | |||||||  |  |||||| |||| | || | ||||
Db       4372 TATATATTTATATATATATTTTCATATATATATATTTATATATATATTTTCATATATATT 4313

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              | || ||| ||||   || | ||| || || ||||||||  | |||||| | | ||||||
Db       4312 TTCATATATATATTTTCATATATATATTTATATATATATTTATATATATTTTCATATATA 4253

Qy        706 TATGTATATATATAT 720
              ||| ||||||||||
Db       4252 TATTTATATATATAT 4238
```

RESULT 37
ADC87554
ID ADC87554 standard; DNA; 1489 BP.
XX
AC ADC87554;
XX
DT 01-JAN-2004 (first entry)
XX
DE Human GPCR gene SEQ ID NO:2007.
XX
KW ds; gene; human; GPCR;
KW guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1270724-A2.
XX
PD 02-JAN-2003.
XX
PF 18-JUN-2002; 2002EP-00013517.
XX
PR 18-JUN-2001; 2001JP-00246789.
XX
PA (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI Suwa M, Asai K, Akiyama Y, Aburatani H;
XX
DR WPI; 2003-315783/31.
DR P-PSDB; ADC87555.
XX
PT New polynucleotide, useful for preparing a composition for treating a
PT patient in need of increased or suppressed activity or expression of the
PT guanosine triphosphate-binding protein coupled receptor.
XX
PS Claim 1; SEQ ID NO 2007; 28pp; English.
XX
CC The invention relates to a novel polynucleotide encoding a guanosine
CC triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC the invention may have a use in gene therapy. The polynucleotide and
CC polypeptide are useful for preparing a composition for treating a patient
CC in need of increased or suppressed activity or expression of the
CC guanosine triphosphate-binding protein coupled receptor. The
CC polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC invention.
XX
SQ Sequence 1489 BP; 553 A; 208 C; 160 G; 568 T; 0 U; 0 Other;

Query Match 5.6%; Score 258.8; DB 10; Length 1489;
Best Local Similarity 72.9%; Pred. No. 4.1e-35;

Matches 360; Conservative 0; Mismatches 132; Indels 2; Gaps 2;

```
Qy        229 TATATATATATATATATATATATACACACACACACATATATATATACACATATATATA 288
              | |||||||||| ||| |||||||| ||| | | |||||||| |||||| || ||   |||
Db        519 TGTATATATATGTATGTATATATATACATATATACACATATGTATATATACGTACC-ATA 577

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              |  |||||   |||||| || |||||| |||| |||||| || || ||||  ||| |||
Db        578 TATGTATATGTGTATATATACGTATATATACATGTATATATACGTATATATACACATATA 637

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              |||||  |||| | |||| || |||||| || |||| | | | ||| |    | |||||
Db        638 CATATATACATTTATATATACGTATATATACGTATACATATATATACACGTATATATACA 697

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              |||  |||||  ||   |||||| || |||| | | ||| ||   |||||| || |||| |
Db        698 TACATATATACACGTGTATATATACGTATACATATATACACGTGTATATATACGTATACA 757

Qy        469 CATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATATA 527
              ||||||||||  | ||||||  |   || ||| |||||  |      ||||||||  | | |||
Db        758 TATATATACACGTGTATATATACGTATACATATATACACGTGTATATATACGTATACATA 817

Qy        528 TACACATATATACACATATATATACACATATATACACATATATACATATATACACATATA 587
              || |||  | || | |||  |||||| ||||| ||    |||||| |  |||||| |||||
Db        818 TATACACGTGTATATATACGTATACATATATACACGTGTATATATACGTATACATATATA 877

Qy        588 TATACATATATACACATATATACATATATACACATATATACATATATACACACATATATA 647
              |    |||||||| | ||||| | |   |  | ||| |||  |||||||||    ||||||
Db        878 CACGTGTATATACATACATATATACACGTGTATATACATATGTATATACACGTGTATATA 937

Qy        648 CACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATA 707
              |||  |||||| | || || |||||||| |||||||| |||||||||| ||||||
Db        938 TGTATATGTATACATATACGTATATATACATGTATATACACGTATATATACGTGTATATA 997

Qy        708 TGTATATATATATG 721
              |||||| ||||| |
Db        998 TGTATACATATACG 1011
```

RESULT 38
ADO15781/c
ID ADO15781 standard; DNA; 2761 BP.
XX
AC ADO15781;
XX
DT 29-JUL-2004 (first entry)
XX
DE 4 synthesis-period of neuroblastoma related DNA, SEQ ID 43.
XX
KW Human; 4 synthesis-period; neuroblastoma; stage 4S; ds.
XX
OS Homo sapiens.
XX
PN WO2004039975-A1.
XX
PD 13-MAY-2004.
XX
PF 30-OCT-2003; 2003WO-JP013932.
XX
PR 30-OCT-2002; 2002JP-00316586.
XX
PA (HISM) HISAMITSU PHARM CO LTD.
PA (CHIB-) CHIBA PREFECTURE.
XX
PI Nakagawara A, Ohira M;
XX
DR WPI; 2004-390323/36.
XX
PT Novel nucleic acid obtained from 4 synthesis-period of neuroblastoma
PT cells useful for prognosing and determining progress stage of
PT neuroblastomas.
XX
PS Claim 1; SEQ ID NO 43; 455pp; Japanese.

```
XX
CC   The present invention relates to human nucleic acid sequences (I;
CC   ADO15739-ADO15912) obtained from 4 synthesis-period (stage 4S) of
CC   neuroblastoma cell. (I) is useful for prognosing and determining the
CC   progress stage of 4 synthesis-period of neuroblastoma.
XX
SQ   Sequence 2761 BP; 847 A; 475 C; 598 G; 841 T; 0 U; 0 Other;

  Query Match             5.6%;  Score 258.8;  DB 12;  Length 2761;
  Best Local Similarity   71.0%;  Pred. No. 4.2e-35;
  Matches  375;  Conservative    0;  Mismatches  142;  Indels   11;  Gaps    2;

Qy        229 TATATATATATATATATATATATACACACACACACACATATATATATACACATATATATA 288
              | | |||||||||||| |||||||| | | |     ||||||||| |||| ||||||||
Db       2444 TGTGTATATATATATACATATATACGTATATATGTGTATATATATACACACGTATATATA 2385

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              |||||||| ||| ||||||| |||| | | | |||||   |||||||   || ||||
Db       2384 CACATATATGTATACATACACACGTATATATACACATATATGTATACATACACACGCATA 2325

Qy        349 TATATACACATATGTATACA--------CATATATACACATATATACACACATATATACA 400
              |||| ||| ||||||||||         ||||||||||||||||  | ||||| | ||
Db       2324 TATACACATATATGTATACATACACACGTATATATACACATATATGTATACATACACACG 2265

Qy        401 CATATATACACACATATATACACATATATACACATATATACACATATACACATATATACA 460
              |||||||||| |||| || ||||| ||||  ||||||||||||||||   ||| |||||
Db       2264 TATATATACACATATATGTATACATACATACGTATATATACACATATATGTATACATACA 2205

Qy        461 CATATATACATATATACACATATATATAATATACACACATATATATACACATATATACAC 520
              |  |||| ||| | | | || |||   | ||||| || ||||||||||||||||  |
Db       2204 TACGTATATATACACATATATGTAT---ACATACATACGTATATATACACATATATGTAT 2148

Qy        521 ACATATATACACATATATACACATATATATACACATATATACACATATATACATATATAC 580
              ||||| ||||  ||||||||||||||| || ||||| ||||  ||||||||| ||||
Db       2147 ACATACATACGTATATATACACATATATGTATACATACATACGTATATATACACATATAT 2088

Qy        581 ACATATATATACATATATACACATATATACATATATACACATATATACATATATACACAC 640
              ||| ||| |    |||| | | | ||| || |||||| | ||    ||||||||||
Db       2087 GTATACATACATGCGTATATATACACATATATGTATACATACATGCGTATATATACACAT 2028

Qy        641 ATATATACACATACATATACACACACATAGATATACATATATATACACATATATATACGT 700
              |||| || ||||| ||||  ||||| ||| || |   | |||| | | |||||      |
Db       2027 ATATGTATACATATATATGAACACATATATATGTGTGTGTATACATATATATAAGATTTT 1968

Qy        701 ATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTT 748
              |      |  ||    ||| |   |||| ||  || | ||| | ||
Db       1967 TCTTGACCATGAATCACCATGTTTGGGAGTACAGTAGGGTTAGGAATT 1920


RESULT 39
ADL13771
ID   ADL13771 standard; DNA; 201766 BP.
XX
AC   ADL13771;
XX
DT   06-MAY-2004  (first entry)
XX
DE   Osteoarthritis-associated polymorphic nucleotide #303.
XX
KW   ds; gene; osteopathic; antiinflammatory; antiarthritic; gene therapy;
KW   joint space narrowing; osteophyte development; joint pain;
KW   osteoarthritis; SNP; single nucleotide polymorphism.
XX
OS   Homo sapiens.
XX
PN   WO2003054166-A2.
XX
PD   03-JUL-2003.
XX
PF   19-DEC-2002; 2002WO-US041225.
XX
PR   20-DEC-2001; 2001US-0342603P.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Jones KA,  Schafer A;
XX
DR   WPI; 2003-559141/52.
XX
PT   Determining susceptibility of an individual to joint space narrowing,
PT   osteophyte development and/or joint pain comprises identifying whether
PT   the individual has at least one polymorphism in a polynucleotide encoding
PT   a protein.
XX
PS   Disclosure; SEQ ID NO 303; 297pp; English.
XX
CC   The invention relates to a method of determining susceptibility of an
CC   individual to joint space narrowing and/or osteophyte development and/or
CC   joint pain comprising identifying whether the individual has at least one
CC   polymorphism in a polynucleotide encoding at least one of the protein
CC   listed in the specification. The methods, composition and agent are
CC   useful for modulating the susceptibility of an individual to joint space
CC   narrowing and/or osteophyte development and/or joint pain that is
CC   associated with a disease, preferably osteoarthritis. The cell line and
CC   the non-human animal are useful for screening for an agent for diagnosing
CC   an individual having susceptibility to joint space narrowing and/or
CC   osteophyte development and/or joint pain. This sequence corresponds to
CC   the polynucleotide encoding a protein listed in the specification. (Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification but was obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences).
XX
SQ   Sequence 201766 BP; 57593 A; 35575 C; 36821 G; 54688 T; 0 U; 17089 Other;

  Query Match             5.6%;  Score 258.8;  DB 10;  Length 201766;
  Best Local Similarity   72.9%;  Pred. No. 5e-35;
  Matches  360;  Conservative    0;  Mismatches  132;  Indels    2;  Gaps    2;

Qy        229 TATATATATATATATATATATATACACACACACACACATATATATATACACATATATATA 288
              | |||||||||| ||| |||||||| ||| | | |||||||| |||||| || ||   |||
Db     188638 TGTATATATATGTATGTATATATATACATATATACACATATGTATATATACGTACC-ATA 188696

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              |  |||||   |||||| || |||||| |||| |||||| || || ||||  ||| |||
Db     188697 TATGTATATGTGTATATATACGTATATATACATGTATATATACGTATATATACACATATA 188756

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              ||||| |||| | |||| || |||||| || |||| | | | ||| |   | |||||
Db     188757 CATATATACATTTATATATACGTATATATACGTATACATATATATACACGTATATATACA 188816

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              ||| ||||| ||   |||||| || |||| | | ||| ||   ||||||  || |||| |
Db     188817 TACATATATACACGTGTATATATACGTATACATATATACACGTGTATATATACGTATACA 188876

Qy        469 CATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATATA 527
              |||||||||| | |||||| |  || ||| ||||| |     ||||||||  | | |||
Db     188877 TATATATACACGTGTATATATACGTATACATATATACACGTGTATATATACGTATACATA 188936

Qy        528 TACACATATATACACATATATATACACATATATACACATATATACATATATACACATATA 587
              || |||  | || | |||  |||||| ||||| ||   ||||||  |  |||||| |||||
Db     188937 TATACACGTGTATATATACGTATACATATATACACGTGTATATATACGTATACATATATA 188996

Qy        588 TATACATATATACACATATATACATATATACACATATATACATATATACACACATATATA 647
              |    |||||||| | ||||| | |   |  ||| |||  |||||||||    ||||||
Db     188997 CACGTGTATATACATACATATATACACGTGTATATACATATGTATATACACGTGTATATA 189056

Qy        648 CACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATA 707
                  |||  |||||| | || || |||||||| |||||||| |||||||||| ||||||
Db     189057 TGTATATGTATACATATACGTATATATACATGTATATACACGTATATATACGTGTATATA 189116

Qy        708 TGTATATATATATG 721
              |||||| ||||| |
Db     189117 TGTATACATATACG 189130


RESULT 40
```

```
ABD32659/c
ID   ABD32659 standard; DNA; 163382 BP.
XX
AC   ABD32659;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Human cancer-associated genomic DNA HD13-065.
XX
KW   Human; ds; cancer-associated protein; gene; cytostatic; cancer;
KW   leukaemia; lymphoma; CAP.
XX
OS   Homo sapiens.
XX
PN   WO2004074320-A2.
XX
PD   02-SEP-2004.
XX
PF   17-FEB-2004; 2004WO-US004730.
XX
PR   14-FEB-2003; 2003US-00367094.
PR   14-MAR-2003; 2003US-00388838.
PR   15-APR-2003; 2003US-00417375.
PR   13-JUN-2003; 2003US-00461862.
PR   15-SEP-2003; 2003US-00663431.
PR   15-DEC-2003; 2003US-00737318.
XX
PA   (SAGR-) SAGRES DISCOVERY INC.
XX
PI   Morris DW,  Morris DW,  Malandro MS;
XX
DR   WPI; 2004-652914/63.
XX
PT   New isolated cancer-associated polynucleotides and polypeptides useful
PT   for diagnosing, preventing or treating cancers, especially lymphoma and
PT   leukemia, or in screening for agents that modulate cancer.
XX
PS   claim 16; seqid 199; 310pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising at least 10
CC   contiguous nucleotides of any of the 233 polynucleotide sequences given
CC   in the specification, or its complement. The nucleic acids encode cancer-
CC   associated proteins. Also included are an expression vector comprising
CC   the isolated nucleic acid cited above, a host cell comprising the above
CC   recombinant nucleic acid or expression vector, a microarray for detecting
CC   a cancer-associated (CA) nucleic acid comprising at least one probe
CC   comprising at least 10 contiguous nucleotides of any of the above-
CC   mentioned nucleotide sequences, an isolated polypeptide (encoded within
CC   an open reading frame of a CA sequence selected from any of the 95
CC   polynucleotide sequences as mentioned in the specification, or its
CC   complement), an isolated antibody, (or its antigen binding fragment) that
CC   binds to the above polypeptide, a hybridoma that produces the above
CC   monoclonal antibody, a pharmaceutical composition comprising the above
CC   antibody and a pharmaceutical excipient, a kit for detecting cancer
CC   cells(comprising the antibody cited above, methods for diagnosing cancer
CC   or for detecting the presence or absence of cancer cells in an
CC   individual, a method for inhibiting growth of cancer cells in an
CC   individual, a method for delivering a therapeutic agent to cancer cells
CC   in an individual, an electronic library comprising the above
CC   polynucleotide or polypeptide (or their fragments), methods of screening
CC   for anticancer activity or for a bioactive agent capable of modulating
CC   the activity of a CA protein (CAP), methods for detecting cancer
CC   associated with expression of a polypeptide in a test cell sample, a
CC   method for treating cancers and a method for inhibiting the expression of
CC   CA gene in a cell. The composition and methods are useful for detecting,
CC   diagnosing, preventing and treating cancers, especially lymphoma and
CC   leukaemia. These may also be used in screening for agents that modulate
CC   cancer. The present sequence is a human CAP genomic sequence. Note: The
CC   sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 163382 BP; 50073 A; 25805 C; 27837 G; 53184 T; 0 U; 6483 Other;
```

```
  Query Match             5.6%;  Score 258.6;  DB 13;  Length 163382;
  Best Local Similarity   70.0%;  Pred. No. 5.4e-35;
  Matches  362;  Conservative    0;  Mismatches  154;  Indels    1;  Gaps    1;

Qy        204 ATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACAC 263
              || | |  | | | |||        |||| ||||||| |||| |||| | ||| | | |
Db     145185 ATATAAATATATATAAATAAATAAATATAAATATATAAATATGTATAAATACATATATAA 145126

Qy        264 ACATATATATATACACATATATATACACATATATACATATATACACATATATACACATAT 323
              | | ||||| ||| | | ||||||| | ||| ||| ||| ||| | ||| ||| | ||||
Db     145125 ATAAATATAAATATATAAATATATATAAATACATATATAAATATATATAAATATATATAT 145066

Qy        324 ATATACACATACATATGTACACATATATATACACATATGTATACACATATATACACATAT 383
                 ||| ||||| ||||||| | |||||| ||| ||||  ||| | | ||||| | | ||
Db     145065 TAATATACATAAATATGTATAAATATATAAACATATATTAATATATAAATATATATAAAT 145006

Qy        384 ATACACACATATATACACATATATACACACATATATACACATATATACACATATATACAC 443
              |||   | |||||||   ||||||| | | ||  ||| | ||| |||     ||||| ||
Db     145005 ATATTAATATATATATTTATATATAAATATATGAATATATATAAATATGTTAATATATAC 144946

Qy        444 ATATACACATATATACACATATATACATATATACACATATATATAATATACACACATATA 503
              ||||| | | ||| | | ||| | | ||||||| | |||||||    ||| | | |||||
Db     144945 ATATATATAAATAAATATATAAAAATATATATAAATATATATAACTAATATATAAATATA 144886

Qy        504 TATACACATATATACACACATATATACACATATATACACATATATATACACATATATACA 563
              |||| ||  ||||| | | ||||||| | ||||||  ||  ||||||| | ||||||| |
Db     144885 TATA-ACTAATATATATAAATATATAAATATATATGAACTAATATATATAAATATATAAA 144827

Qy        564 CATATATACATATATACACATATATATACATATATACACATATATACATATATACACATA 623
              | ||||| | ||| | | ||| ||| | ||||| | | ||||||| ||| ||| | | |
Db     144826 TAAATATATAAATAAATATATAAATAAATATATACAAATATATATAAATAAATATATAAA 144767

Qy        624 TATACATATATACACACATATATACACATACATATACACACACATAGATATACATATATA 683
              ||| ||||| | | | ||||| | | ||| |||||  | |  ||| ||| | | |||||
Db     144766 AATAAATATAAATAAATATATAAATAAATATATATAAATAAATATATATAAATAAATATA 144707

Qy        684 TACACATATATATACGTATATATATGTATATATATAT 720
              || | ||| |||||  || |||||| |||| ||| ||
Db     144706 TATAAATAAATATATATAAATATATATATAAATAAAT 144670


RESULT 41
ABV72623
ID   ABV72623 standard; DNA; 82938 BP.
XX
AC   ABV72623;
XX
DT   27-NOV-2002  (first entry)
XX
DE   Human transporter protein encoding gene.
XX
KW   Human; transporter protein; hTP; gamma-aminobutyric-acid; GABA;
KW   gene therapy; protein therapy; vaccine; gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1120. .81932
FT                   /*tag=  a
FT                   /product= "Transporter protein"
FT                   /note= "Contains 8 introns"
FT   exon            1120. .1223
FT                   /*tag=  b
FT                   /number= 1
FT   intron          1224. .27733
FT                   /*tag=  c
FT                   /number= 1
FT   variation       replace(2419,C)
FT                   /*tag=  d
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(3191,A)
FT                   /*tag=  e
```

```
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(4195,T)
FT                   /*tag= f
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(5910,C)
FT                   /*tag= g
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(7119,T)
FT                   /*tag= h
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(7529,G)
FT                   /*tag= h
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(8855,C)
FT                   /*tag= i
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(8956,A)
FT                   /*tag= j
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(10220,G)
FT                   /*tag= k
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(11628,C)
FT                   /*tag= l
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(11670,A)
FT                   /*tag= m
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(11671,G)
FT                   /*tag= n
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(11931,C)
FT                   /*tag= o
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(15984,G)
FT                   /*tag= p
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(17735,A)
FT                   /*tag= q
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(17910,A)
FT                   /*tag= r
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(18130,C)
FT                   /*tag= s
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19293. .19294,TWA)
FT                   /*tag= t
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19294. .19295,ATT)
FT                   /*tag= u
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19296. .19297,CYT)
FT                   /*tag= v
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19303. .19304,TYC)
FT                   /*tag= w
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19304. .19305,CYC)
FT                   /*tag= x
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19307. .19308,TMC)
FT                   /*tag= y
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19308. .19309,CCC)
FT                   /*tag= z
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19313. .19314,TMA)
FT                   /*tag= aa
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19314. .19315,AWT)
FT                   /*tag= ab
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19315. .19316,TCC)
FT                   /*tag= ac
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19316. .19317,CYT)
FT                   /*tag= ad
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19317. .19318,TWA)
FT                   /*tag= ae
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19318. .19319,ATT)
FT                   /*tag= af
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19860,G)
FT                   /*tag= ag
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(20843,G)
FT                   /*tag= ah
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(21144,T)
FT                   /*tag= ai
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(21371,G)
FT                   /*tag= aj
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(22061. .22062,A)
FT                   /*tag= ak
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(22321,T)
FT                   /*tag= al
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(22470. .22471,A)
FT                   /*tag= am
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(22472. .22473,A)
FT                   /*tag= an
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(24990,C)
FT                   /*tag= ao
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(25284,A)
FT                   /*tag= ap
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(25934,G)
FT                   /*tag= aq
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(27509. .27510,TAT)
FT                   /*tag= ar
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(27513. .27514,TTT)
FT                   /*tag= as
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            27734. .27882
FT                   /*tag= at
FT                   /number= 2
FT   intron          27883. .41045
FT                   /*tag= au
FT                   /number= 2
FT   variation       replace(28608,G)
FT                   /*tag= av
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(28627,C)
FT                   /*tag= aw
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(29709,T)
FT                   /*tag= ax
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(29788,T)
FT                   /*tag= ay
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(29878,T)
FT                   /*tag= az
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(30294,T)
```

```
FT                   /*tag=  ba
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(30896. .30897,T)
FT                   /*tag=  bb
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(31815,G)
FT                   /*tag=  bc
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(31999,TTT)
FT                   /*tag=  bd
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(32026,T)
FT                   /*tag=  be
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(32146,A)
FT                   /*tag=  bf
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(32184,S)
FT                   /*tag=  bg
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(32214. .32215,T)
FT                   /*tag=  bh
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(32215,C)
FT                   /*tag=  bi

  Query Match          5.6%;  Score 258.4;  DB 6;  Length 82938;
  Best Local Similarity   70.5%;  Pred. No. 5.7e-35;
  Matches  361;  Conservative    0;  Mismatches  146;  Indels    5;  Gaps    1;

Qy        214 AAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACATATATAT 273
              | | | | |     | |||| | | | |||| |||| |||||||||    | ||||||  |||
Db       6946 ATACATATGTATATATATACACACATATGTATACATACACACATGTATACATATGTAT 7005

Qy        274 ATACACATATATATACACATATATACATATATACACATATATACACATATATATACACAT 333
              |||   ||||| |||||||| ||||||||| || | || |||| | ||| | ||  | |
Db       7006 ATATGTATATAGATACACATGTATACATATGTATATATGTATATAGATACACATGTATAC 7065

Qy        334 ACATATGTACACATATATATACACATATGTATACACATATATACACATATATACACACAT 393
              | || | || |  |||||| | ||| ||||||||| || |||| |  |||||| | |||
Db       7066 ATATGTATATATGTATATAGATACACATGTATACATATGTATATATGTATATAGATACAC 7125

Qy        394 ATATACACATATATACACACATATATACACATATATACACATATATACACATATACACAT 453
              || || |||||| || | |  |||||| | | | ||  | | |||| | ||||  | ||
Db       7126 ATGTATACATATGTATATATGTATATAGATACACATGTATACATATGTATATATGTATAT 7185

Qy        454 ATATACACATATATACATATATACACATATATATA-----ATATACACACATATATATAC 508
              | |||||||| ||||||||| || | || ||||||     | ||  | |||||| ||||
Db       7186 AGATACACATGTATACATATGTATATATGTATATAGATACACATGTATACATATGTATAT 7245

Qy        509 ACATATATACACACATATATACACATATATACACATATATATACACATATATACACATAT 568
              |  |||||| | ||| || || |||||| || | || ||||||||  ||| || | || |
Db       7246 ATGTATATAGATACACATGTATACATATGTATATATGTATATACATGTATGTATATATGT 7305

Qy        569 ATACATATATACACATATATATACATATATACACATATATACATATATACACATATATAC 628
              |||  ||||||||  ||| |||| | ||||   ||||||  |||| || | | || ||
Db       7306 ATATGTATATACATGTATGTATATGTGTATATGTATATATGTATATGTATATACATGTAT 7365

Qy        629 ATATATACACACATATATACACATACATATACACACACATAGATATACATATATATACAC 688
               |||||| |||  |||||||  |  ||||| | |||  || |||   ||||| ||| |
Db       7366 GTATATATACAGCTATATACATGTGTATATATATACAGCTATATACGTATATACATATAT 7425

Qy        689 ATATATATACGTATATATATGTATATATATAT 720
              || |||||| ||||| |||||| |||||||||
Db       7426 ATGTATATATGTATACATATGTGTATATATAT 7457


RESULT 42
AAK73166/c
ID   AAK73166 standard; DNA; 19965 BP.
XX
AC   AAK73166;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:27978.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001354.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
```

PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS Disclosure; SEQ ID NO 27978; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention
XX
SQ Sequence 19965 BP; 4699 A; 5296 C; 5505 G; 4465 T; 0 U; 0 Other;

Query Match 5.6%; Score 258; DB 4; Length 19965;
Best Local Similarity 64.7%; Pred. No. 6.2e-35;
Matches 431; Conservative 0; Mismatches 230; Indels 5; Gaps 3;

```
Qy        59 TAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGA 118
             | | ||||   ||||   ||||||| ||||||||||   || ||||||||||||||||
Db      4999 TGGCCAGGAGCGGTGGCTCACGCCTATAATCCCAGCACGTTGGGAGGCCGAGGCAGGGAG 4940

Qy       119 ATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCA 178
               |  | | | | | |  ||          |||| |   ||  ||| ||   |
Db      4939 GATCGCTTGAGAT-CAAGAGCTCCAGACCAGCCTAGGCAACACGGTGAAATTCCATCTCT 4881

Qy       179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATA 238
               |  |   || | |  |||    | |   | | | | |      | ||||| | ||
Db      4880 ACAAAAAAAATACAAAAAAAAAAAAAAAATATATATATATATTTATATATATATGTTTA 4821

Qy       239 TATATATATATATACACACACACACATATATATATACACATATATATACACATATATA 298
             |||||||||||  || | |      | | |||| |||||    ||||||||  | ||||||
Db      4820 TATATATATATTTATATATTTTTATATATATTTATATTTTTATATATATTTATATATATT 4761

Qy       299 CATATATA-CACATATATACACATATATATACACATACATATGTACACATATATATACAC 357
               |||||||  |  |  |||  | |||| |||| | |||  |   || | ||||||  | ||
Db      4760 TATATATATTATTTTCATATATATATTTATATATATATTTTCATATATATATATTTTCAT 4701

Qy       358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417
             |||| |||| | ||||  | | ||||||  | | |||| | || |||||||   | ||||
Db      4700 ATATATATATATATATTCATATATATATTTATATATATTTTCATATATATATTTATATAT 4641

Qy       418 ATACACATATATACACATATATACACATATACACATATATACACATATATACACATATATAC 477
             |||     ||||| | |||| | || ||||| | |||||||  ||||||||  || ||||
Db      4640 ATATTTTCATATATATATATTTTCATATATATATATATATATTCATATATATATTTATAT 4581

Qy       478 ACATATATATA---ATATACACACATATATATACACATATATACACACATATATACACAT 534
             | || |  |||   |||||   | ||||||||  ||||||||  | |  |||||  | ||
Db      4580 ATATTTTCATATATATATATTTATATATATATTTTCATATATATATATTTATATATATAT 4521
```

```
Qy        535 ATATACACATATATATACACATATATACACATATATACATATATACACATATATATACAT 594
              |  | | |||||| || | ||||||   |||||||| |||| || | |||||| | |||
Db       4520 TTTCATATATATATTTATATATATATTTTCATATATATATATTTATATATATATTTTCAT 4461

Qy        595 ATATACACATATATACATATATACACATATATACATATATACACACATATATACACATAC 654
              ||||| | || |||| ||||||   |||||||| |||| || | | |||| | || |||
Db       4460 ATATATATATTTATATATATATTTTCATATATATATATTTATATATATATTTTCATATAT 4401

Qy        655 ATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATAT 714
              |||||   | | ||| || | ||||||||||   |||||||||  |  |||||| ||| |
Db       4400 ATATATTTATATATATATTTTCATATATATATTTATATATATATTTTCATATATATATTT 4341

Qy        715 ATATAT 720
              ||||||
Db       4340 ATATAT 4335
```

RESULT 43
ABK69932/c
ID ABK69932 standard; DNA; 19965 BP.
XX
AC ABK69932;
XX
DT 15-JUL-2002 (first entry)
XX
DE Human secreted protein gene 68 genomic DNA fragment #20.
XX
KW Human; ds; secreted protein; gene therapy; immunosuppressive;
KW antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;
KW vasotropic; cerebroprotective; nootropic; neuroprotective; antibacterial;
KW virucide; fungicide; opthalmalogical; autoimmune disease; neoplasm;
KW rheumatoid arthritis; hyperproliferative disorder; cardiac arrest;
KW cardiovascular disorder; cerebrovascular disorder; cerebral ischaemia;
KW angiogenesis; nervous system disorder; Alzheimer's disease; infection;
KW ocular disorder; corneal infection; wound healing; skin aging;
KW epithelial cell proliferation; food additive.
XX
OS Homo sapiens.
XX
PN WO200226931-A2.
XX
PD 04-APR-2002.
XX
PF 24-SEP-2001; 2001WO-US029871.
XX
PR 25-SEP-2000; 2000US-0234925P.
PR 12-JAN-2001; 2001WO-US000911.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Komatsoulis G, Duan DR, Rosen CA, Moore PA, Shi Y;
PI Lafleur DW, Olsen H, Brewer LA, Florence KA, Young PE, Soppet DR;
PI Endress GA, Mucenski M, Ebner R;
XX
DR WPI; 2002-362489/39.
XX
PT Novel 71 isolated secreted polypeptides and polynucleotides encoding the
PT polypeptides, useful for treating Huntington's disease, sepsis,
PT meningitis, thrombocytopenia, hemolytic anemia, rheumatoid arthritis,
PT asthma.
XX
PS Example 2; Page 1434-1440; 1478pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (or its
CC fragment, homologue complement or allelic variant) encoding a human
CC secreted protein (and its fragment, domain, epitope, variant, secreted
CC form and species variant). Also included are a recombinant vector
CC comprising the nucleic acid, a recombinant host cell comprising the
CC vector, an antibody against the secreted protein, a recombinant host cell
CC that expresses the secreted protein and a method of identifying a binding
CC partner of the secreted protein. The nucleic acid and protein are used to
CC prevent, diagnose, treat or ameliorate a medical condition in e.g.
CC humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep for
CC example autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC angiogenesis, nervous system disorders e.g. Alzheimer's disease,
CC infections caused by bacteria, viruses and fungi and ocular disorders
CC e.g. corneal infection. Many other diseases and disorders are listed in
CC the specification. The polypeptides can also be used to aid wound healing
CC an epithelial cell proliferation, to prevent skin aging due to sunburn,
CC to maintain organs before transplantation, for supporting cell culture of
CC primary tissues, to regenerate tissues and in chemotaxis. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities. The present sequence
CC represents a ds DNA fragment of the gene for a novel human secreted
CC protein of the invention
XX
SQ Sequence 19965 BP; 4699 A; 5296 C; 5505 G; 4465 T; 0 U; 0 Other;

Query Match 5.6%; Score 258; DB 6; Length 19965;
Best Local Similarity 64.7%; Pred. No. 6.2e-35;
Matches 431; Conservative 0; Mismatches 230; Indels 5; Gaps 3;

```
Qy         59 TAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGA 118
              | | ||||   ||||   ||||||| ||||||||||   || ||||||||||||||||
Db       4999 TGGCCAGGAGCGGTGGCTCACGCCTATAATCCCAGCACGTTGGGAGGCCGAGGCAGGGAG 4940

Qy        119 ATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCA 178
                |  | | | | | ||          |||| |   ||  ||| ||   |
Db       4939 GATCGCTTGAGAT-CAAGAGCTCCAGACCAGCCTAGGCAACACGGTGAAATTCCATCTCT 4881

Qy        179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATA 238
              |  |   || | |   |||    | |   | | | | |     | ||||| | ||
Db       4880 ACAAAAAAAATACAAAAAAAAAAAAAAAATATATATATATATATTTATATATATATGTTTA 4821

Qy        239 TATATATATATATACACACACACACATATATATATACACATATATATACACATATATA 298
              |||||||||| || | |     | | |||| |||||    ||||||||  | ||||||
Db       4820 TATATATATATTTATATATTTTTATATATATTTATATTTTTATATATATTTATATATATT 4761

Qy        299 CATATATA-CACATATATACACATATATATACACATACATATGTACACATATATATACAC 357
              |||||||  |  |  ||| | |||| |||| | |||  |   || | |||||| | ||
Db       4760 TATATATATTATTTTCATATATATATTTATATATATATTTTCATATATATATATTTTCAT 4701

Qy        358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417
              |||| |||| | ||||  | | ||||||  | | ||||  | || |||||||   | ||||
Db       4700 ATATATATATATATATTCATATATATATTTATATATATTTTCATATATATATTTATATAT 4641

Qy        418 ATACACATATATACACATATATACACATATACACATATATACACATATATACATATATAC 477
              |||     |||||  | |||| | || ||||| | |||||||  ||||||||  || ||||
Db       4640 ATATTTTCATATATATATATTTTCATATATATATATATATATTCATATATATATTTATAT 4581

Qy        478 ACATATATATA---ATATACACACATATATATACACATATATACACACATATATACACAT 534
              | || |  |||   |||||   | ||||||||   ||||||||  | |  ||||||  | ||
Db       4580 ATATTTTCATATATATATATTTATATATATATTTTCATATATATATATTTATATATATAT 4521

Qy        535 ATATACACATATATATACACATATATACACATATATACATATATACACATATATATACAT 594
              | | | |||||| || | ||||||   |||||||| |||| || | |||||| | |||
Db       4520 TTTCATATATATATTTATATATATATTTTCATATATATATATTTATATATATATTTTCAT 4461

Qy        595 ATATACACATATATACATATATACACATATATACATATATACACACATATATACACATAC 654
              ||||| | || |||| ||||||   |||||||| |||| || | | |||| | || |||
Db       4460 ATATATATATTTATATATATATTTTCATATATATATATTTATATATATATTTTCATATAT 4401

Qy        655 ATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATAT 714
              |||||   | | ||| || | ||||||||||   |||||||||  |  |||||| ||| |
Db       4400 ATATATTTATATATATATTTTCATATATATATTTATATATATATTTTCATATATATATTT 4341

Qy        715 ATATAT 720
              ||||||
Db       4340 ATATAT 4335
```

RESULT 44
ABK39930
ID ABK39930 standard; DNA; 12069 BP.

```
XX
AC   ABK39930;
XX
DT   21-MAY-2002  (first entry)
XX
DE   Human chemically pretreated gene sequence #6 strand 1.
XX
KW   Human; ds; bisulphite treatment; CpG; DNA methylation; cancer; tumour;
KW   cytostatic; ALDH6; CYP11A; CYP11B1; CYP3A3; DPYD; EPHX2; OCLN; TXNRD1;
KW   UGT8; MRP; pharmacogenomics; SNP; single nucleotide polymorphism.
XX
OS   Homo sapiens.
XX
PN   WO200202806-A2.
XX
PD   10-JAN-2002.
XX
PF   29-JUN-2001; 2001WO-EP007470.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-154757/20.
XX
PT   New nucleic acid, oligonucleotides and peptide nucleic acid-oligomers,
PT   useful for detecting cytosine methylation state of genes associated with
PT   pharmacogenomics and for therapy of diseases e.g. cancer.
XX
PS   Claim 1; SEQ ID NO 11; 24pp; English.
XX
CC   The invention relates to a nucleic acid comprising a sequence at least 18
CC   bases in length of a segment of the chemically pretreated DNA of genes
CC   associated with pharmacogenomics according to one of the sequences of the
CC   genes ALDH6 (NM_000693), CYP11A (NM_000781), CYP11B1 (NM_000497), CYP3A3
CC   (NM_000776 and NM_017460), DPYD (NM_000110), EPHX2 (NM_001979), OCLN
CC   (NM_002538), TXNRD1 (NM_003330), UGT8 (NM_003360), MRP (NM_004996,
CC   NM_019900, NM_019901, NM_019902, NM_019862, NM_019898, NM_019899) and
CC   their complementary sequences, or a sequence (S1) chosen from 87
CC   sequences and their complements. The chemical pretreatment is bisulphite
CC   treatment to convert cytosines (but not methyl-cytosines) into uracils.
CC   Also included are an oligomer (II) in particular an oligonucleotide or a
CC   peptide nucleic acid (PNA)-oligomer, comprising in each case at least one
CC   base sequence having a length of 9 nucleotides which hybridises to or is
CC   identical to a chemically pretreated DNA of genes associated with
CC   pharmacogenomics and their complements, arranged in an array for
CC   analysing diseases associated with the methylation state (CpG) and/or
CC   detecting SNPs (single nucleotide polymorphisms) of the 87 sequences. The
CC   oligomers may also be used as PCR primers. The set of 87 nucleic acids
CC   and their complements is useful for diagnosis and therapy of solid
CC   tumours and cancer. The present sequence represents one the 87 DNA
CC   sequences or its complement. Note: The sequence data for this patent did
CC   not form part of the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 12069 BP; 3210 A; 185 C; 2849 G; 5823 T; 0 U; 2 Other;

  Query Match             5.6%;  Score 257.6;  DB 6;  Length 12069;
  Best Local Similarity   62.9%;  Pred. No. 7.2e-35;
  Matches  433;  Conservative    0;  Mismatches  249;  Indels    6;  Gaps    2;

Qy        226 TTTTATATATATATATATATATATATACACACACACACATATATATATACACATATAT 285
              || |||||||||| |||| ||||||||| | | | | ||    |||||||||   ||||||
Db       9239 TTATATATATATAGATATTTATATATATATAGATATACGTGGATATATATATTTATATAT 9298

Qy        286 ATACACATATATACATATATACACATATATACACATATATATACACATACATATGTACAC 345
              ||| | |||| || ||||||| | |||||     |||||  |||   ||| |||| || |
Db       9299 ATATAGATATTTATATATATATAGATATACGTGGATATAGATATTTATATATATATAGAT 9358

Qy        346 ATATATATACACATATGTATACACATATATACACATATATACACACATATATACACATAT 405
              || |||||| | ||||  |||   ||||||| | | ||||  | | ||||||| | ||||
Db       9359 ATTTATATATATATATAGATATTTATATATATATAGATATTTATATATATATAGATATAT 9418

Qy        406 ATACACACATATATACACATATATACACATATATACACATATACACATATATACACATAT 465
              |  | | ||||||  | ||||||| | |||| || | ||||| | |||| || | ||||
Db       9419 GTGGATATATATATTTATATATATAGAGATATTTATATATATATAGATATTTATATATAT 9478

Qy        466 ATACATATATACACATATATATA-----ATATACACACATATATATACACATATATACAC 520
              ||| |||||      |||||||||      ||||| | | |||| |||| | ||||| | |
Db       9479 ATAGATATACGTGGATATATATATTTATATATATATAGATATTTATATATATATAGATAT 9538

Qy        521 ACATATATACACATATATACACATATATATACACATATATACACATATATACATATATAC 580
              || |  ||| | |||| || || || ||||  | |||||| | |||||     |||||
Db       9539 ACGTGGATATATATATTTATACGTAGATATTTATATATATATAGATATACGTGGATATAT 9598

Qy        581 ACATATATATACATATATACACATATATACATATATACACATATATACATATATACACAC 640
              | || ||||   | ||||  | ||||||| | ||||||    ||||| |||| || | |
Db       9599 ATATTTATACGTAGATATTTATATATATATAGATATACGTGGATATATATATTTATATAT 9658

Qy        641 ATATATACACATACATATACACACACATAGATATACATATATATACACATATATATACGT 700
              ||||| | |  || ||||| | | | |||  |  | |||||||  | |||||||||  |
Db       9659 ATATAGATATTTATATATATATAGATATACGTGGATATATATATTTATATATATATAGAT 9718

Qy        701 ATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGG 760
              || ||||| |||||| |||| |       |  |||         |  |   ||    |
Db       9719 ATTTATATATATATAGATATACGTGGATATAGATATTTATATATATATAGATATTTATAC 9778

Qy        761 GGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATG 820
              | ||| ||  |   | |  |  | ||   | |||    | |  || || |     |
Db       9779 GTATATAGATATTTATATATATATAGATATTTATATATATATAGATATATTTATATATAT 9838

Qy        821 GCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAA 880
              | |||     ||| ||| |||  | | | ||| |   ||| | ||       |  | |
Db       9839 ATAGATATATATGGATATATATATTTATAT-ATATATAATATTTATATATAGATATATAT 9897

Qy        881 GGTAACACTATATATATATATAGTGAAA 908
              ||  | | ||| ||||||||||  || |
Db       9898 GGATATAGTATTTATATATATATAGATA 9925


RESULT 45
ADC87552
ID   ADC87552 standard; DNA; 1486 BP.
XX
AC   ADC87552;
XX
DT   01-JAN-2004  (first entry)
XX
DE   Human GPCR gene SEQ ID NO:2005.
XX
KW   ds; gene; human; GPCR;
KW   guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1270724-A2.
XX
PD   02-JAN-2003.
XX
PF   18-JUN-2002; 2002EP-00013517.
XX
PR   18-JUN-2001; 2001JP-00246789.
XX
PA   (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA   (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI   Suwa M,  Asai K,  Akiyama Y,  Aburatani H;
XX
DR   WPI; 2003-315783/31.
DR   P-PSDB; ADC87553.
XX
PT   New polynucleotide, useful for preparing a composition for treating a
```

```
PT   patient in need of increased or suppressed activity or expression of the
PT   guanosine triphosphate-binding protein coupled receptor.
XX
PS   Claim 1; SEQ ID NO 2005; 28pp; English.
XX
CC   The invention relates to a novel polynucleotide encoding a guanosine
CC   triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC   the invention may have a use in gene therapy. The polynucleotide and
CC   polypeptide are useful for preparing a composition for treating a patient
CC   in need of increased or suppressed activity or expression of the
CC   guanosine triphosphate-binding protein coupled receptor. The
CC   polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC   invention.
XX
SQ   Sequence 1486 BP; 517 A; 149 C; 242 G; 578 T; 0 U; 0 Other;

  Query Match             5.5%;  Score 256.4;  DB 10;  Length 1486;
  Best Local Similarity   70.2%;  Pred. No. 1.1e-34;
  Matches  344;  Conservative    0;  Mismatches  146;  Indels    0;  Gaps    0;

Qy        227 TTTATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | |||||||| || ||| | |||||| ||| |     |  | |||||||||| || | ||
Db        718 TGTATATATACATGTATGTGTATATATACATATGTGTATGTGTATATATACATATGTGTA 777

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              |    |||||||||||| |   | |||| | | |||| |  |   |  |||| |||| |
Db        778 TGTGTATATATACATATGTGTATGTGTATATATACATATGTGTATGTGTATATATACATA 837

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              | | |||| | | || | |  | |||| || | || |||| | |||||| |  | ||||
Db        838 TGTGTATATATACATATGTGTATATATGTATATATGTATATATACATATGTGTATATATG 897

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | |  |||||| ||||||||     ||||||   |||||    ||| || | || |
Db        898 TATATATGTATATATACATATATGTGTGTATATATGTATATATGTGTATTTATATATGTG 957

Qy        467 TACATATATACACATATATATAATATACACACATATATATACACATATATACACACATAT 526
              |  | | || | |||| ||||   || |    | |||||| |||||||| |   ||||
Db        958 CATGTGTGTATAAATATGTATATATTATATGTGTGTATATATACATATATGTATGTATAT 1017

Qy        527 ATACACATATATACACATATATATACACATATATACACATATATACATATATACACATAT 586
              ||||| ||||| | |||||||||| | | ||| | ||| |||   |||||||| ||| ||||
Db       1018 ATACATATATACATACATATATACATATATACACACATATACGCACATATAGACATATAT 1077

Qy        587 ATATACATATATACACATATATACATATATACACATATATACATATATACACACATATAT 646
              || || ||||| |||||||||||||||||||   |||||||| | ||| | ||||||||
Db       1078 ATGTATATATACACACATATATACATATATATGTATATATACACACATATATACATATAT 1137

Qy        647 ACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATAT 706
              |   ||| ||| ||||| | ||| |||||  ||||||||||||||||||||   ||||||
Db       1138 ATGTATATATACACACATATATACATATATGTATATATACACATATATATGTATATATAC 1197

Qy        707 ATGTATATAT 716
              |  |||||||
Db       1198 ACATATATAT 1207


RESULT 46
AAQ39248/c
ID   AAQ39248 standard; DNA; 870 BP.
XX
AC   AAQ39248;
XX
DT   25-MAR-2003  (revised)
DT   10-AUG-1993  (first entry)
XX
DE   Sequence of cosmid clone 4F7 isolated from a chromosome 11q-specific
DE   cosmid library genomic polymorphic locus (GPL).
XX
KW   Genomic polymorphic locus; restriction site; chromosome 11 q13; ss.
XX
OS   Synthetic.
XX
PN   WO9307166-A1.
XX
PD   15-APR-1993.
XX
PF   07-OCT-1992;   92WO-US008539.
XX
PR   09-OCT-1991;   91US-00773099.
XX
PA   (SALK ) SALK INST BIOLOGICAL STUDIES.
XX
PI   Evans GA,  Selleri L,  Eubanks JH;
XX
DR   WPI; 1993-134379/16.
XX
PT   Oligo-nucleotide(s) useful as primers - for identifying individuals on
PT   basis of presence of polymorphic locus, allows specific detection.
XX
PS   Disclosure; Page 33-34; 51pp; English.
XX
CC   Two cosmids, 4F7 and 7H12, were identified by screening cosmid libraries
CC   with a sequence from the 2nd transmembrane domain of a rat D2 dopamine
CC   recptor cDNA. The cosmids had potential structural homology to G-protein
CC   receptor gene family as well as several loci on chromosome 11. A highly
CC   repetitive element was also detected in the region of of overlap of the 2
CC   cosmids and was subcloned into Bluescript and its DNA sequence
CC   determined. It revealed a highly degenerate repetative hexamer motif
CC   spanning 438 bp consisting of 73 repeats. (Updated on 25-MAR-2003 to
CC   correct PN field.)
XX
SQ   Sequence 870 BP; 185 A; 127 C; 247 G; 311 T; 0 U; 0 Other;

  Query Match             5.5%;  Score 256;  DB 2;  Length 870;
  Best Local Similarity   73.9%;  Pred. No. 1.2e-34;
  Matches  340;  Conservative    0;  Mismatches  115;  Indels    5;  Gaps    1;

Qy        238 ATATATATATATATACACACACACACATATATATATACACATATATATACACATATAT 297
              ||| ||| | ||||||| ||||| ||||||||| ||| ||| ||| ||||||| | | ||
Db        810 ATACATACACATATACATACACATACACATATACATACACATATACATATACATACAT 751

Qy        298 ACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACAC 357
              | | ||||||| ||| | | ||||||| ||| ||||||| || |||| ||| ||||||||
Db        750 ATACATATACATATACACATACATATACATATACATACACATATACATATACATATACAC 691

Qy        358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417
              |||  |||||| ||| ||||||||||| | | ||||||| ||| | ||||||| | | ||
Db        690 ATACATATACATATACATACACATATACATATACATATACACATACATATACATATACAT 631

Qy        418 ATACACATATATACACATATATACACATATACACATATATACACATATATACATATATAC 477
              ||||| | | ||| ||| ||| ||| | ||| ||||||| | ||| ||| ||||||| |
Db        630 ATACATACACATATACACATACACATACATATACATATACATACACATACACATATACAT 571

Qy        478 ACATATATATAATATACACACATATATATACACATATATACACACATATATACACATATA 537
              | | |      ||| | | ||||||| |||||||||| | | ||||| | | ||| |||
Db        570 ATACA-----CATACATATACATATACATACACATATACATATACATACACATACACATA 516

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATACATATA 597
              |||| | | ||| ||| | | ||||||||| | ||| | |||||| || |  ||||  ||
Db        515 TACATACACATACACATACACATACACATACACACACACATACACGTACACGTACACGTA 456

Qy        598 TACACATATATACATATATACACATATATACATATATACACACATATATACACATACATA 657
              ||  | |  ||||  |||||| ||| ||| | ||||||| | | ||||||| |||||||
Db        455 CACGTACACGTACACGTATACATATACATATACATATACATATACATATACATATACATA 396

Qy        658 TACACACACATAGATATACATATATATACACATATATATA 697
              |||| | ||||| | | | | ||| |    | |   |||
Db        395 TACATATACATACACACATACACATACAAGAAGAAGGATA 356


RESULT 47
ADC87554/c
ID   ADC87554 standard; DNA; 1489 BP.
XX
AC   ADC87554;
```

```
XX
DT   01-JAN-2004  (first entry)
XX
DE   Human GPCR gene SEQ ID NO:2007.
XX
KW   ds; gene; human; GPCR;
KW   guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1270724-A2.
XX
PD   02-JAN-2003.
XX
PF   18-JUN-2002; 2002EP-00013517.
XX
PR   18-JUN-2001; 2001JP-00246789.
XX
PA   (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA   (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI   Suwa M,  Asai K,  Akiyama Y,  Aburatani H;
XX
DR   WPI; 2003-315783/31.
DR   P-PSDB; ADC87555.
XX
PT   New polynucleotide, useful for preparing a composition for treating a
PT   patient in need of increased or suppressed activity or expression of the
PT   guanosine triphosphate-binding protein coupled receptor.
XX
PS   Claim 1; SEQ ID NO 2007; 28pp; English.
XX
CC   The invention relates to a novel polynucleotide encoding a guanosine
CC   triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC   the invention may have a use in gene therapy. The polynucleotide and
CC   polypeptide are useful for preparing a composition for treating a patient
CC   in need of increased or suppressed activity or expression of the
CC   guanosine triphosphate-binding protein coupled receptor. The
CC   polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC   invention.
XX
SQ   Sequence 1489 BP; 553 A; 208 C; 160 G; 568 T; 0 U; 0 Other;

  Query Match           5.5%;  Score 255.4;  DB 10;  Length 1489;
  Best Local Similarity   63.8%;  Pred. No. 1.6e-34;
  Matches  438;  Conservative    0;  Mismatches  241;  Indels    8;  Gaps    3;

Qy        236 ATATATATATATATATACACACACACACATATATATATACACATATATATACACATAT 295
              | || | |||||||||||| | | ||  | |||| | |||| | | ||| |||   ||||
Db       1197 ACATGTGTATATATATACATATATACGTATATATGTGTATATATACATACATAGGTATAT 1138

Qy        296 ATACATATATACACATATATACACATATATATACACATACATATGTACACATATATATAC 355
              | | ||||| ||| | ||||| || |||||||  | |||  ||| || |||| ||||||
Db       1137 ACATATATACACATACATATATACGTATATATGTATATATGTATATATACATGTATATAT 1078

Qy        356 ACATATG-TATACACATATATACACATATATACACACATATATACACATATATACACACA 414
               ||||| | || ||||||||||||| |||||| ||  |||| || ||||||||||||   |
Db       1077 GTATATGTTGTATACATATATACACGTATATATACGTATATGTATACATATATACACGTA 1018

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              |||||||  |||| || |||||||||||| |||| | |   | || | | || || |||||
Db       1017 TATATACGTATATGTATACATATATACACGTATATATACGTGTATATACATGTATATATA 958

Qy        475 TACACATATATATAATATACACACATATATATACACATATATACACACATATATACACAT 534
              | || |     ||| | | ||||||||| ||   |||||||| |  |||||||| |
Db        957 CGTATATGT-----ATACATATACATATATACACGTGTATATACATATGTATATACACGT 903

Qy        535 ATATACACATATATATACACATATATACACATATATACATATATACACATATATATACAT 594
              |||| |  ||| |||| |||  | || | || ||||| |||||| |||  | |||| ||
Db        902 GTATATATGTATGTATATACACGTGTATATATGTATACGTATATATACACGTGTATATAT 843

Qy        595 ATATACACATATATACATATATACACATATATACATATATACACACATATATACACATAC 654
               |||||  |||||||||  | || | || ||||| |||||| |||| | |||| |  ||
Db        842 GTATACGTATATATACACGTGTATATATGTATACGTATATATACACGTGTATATATGTAT 783

Qy        655 ATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATAT 714
              |  || | | |||    |||| |||| |||||  |||||||| |||| ||||||||||||
Db        782 ACGTATATATACACGTGTATATATATGTATACGTATATATACACGTGTATATATGTATAC 723

Qy        715 ATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAG 774
              |||||  |  | ||  ||| |        |  |   |     |     || | | | |
Db        722 GTATATATACACGTGTATATATGTATGTATATATACGTGTATATATATGTATACGTATAT 663

Qy        775 AGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGG 834
              | |  |  | |  | |||||    |      |  | |     |      | |||   |
Db        662 ATACGTATATATAAATGTAT--ATATGTATATGTGTATATATACGTATATATACATGTAT 605

Qy        835 AAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATA 894
              | |||| || || |  |    | || ||| ||  | |||  ||   |  |     |||||
Db        604 ATATACGTATATATACACATATACATATATGGTACGTATATATACATATGTGTATATATG 545

Qy        895 TATATATAGTGAAAGCAGTGTAAACCT 921
              ||||||||  | |    | || || |
Db        544 TATATATATACATACATATATATACAT 518


RESULT 48
ABD33148/c
ID   ABD33148 standard; DNA; 104399 BP.
XX
AC   ABD33148;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Murine cancer-associated (CA) gene MD07-019.
XX
KW   Mouse; cancer-associated protein; CAP; cancer-associated gene; CA; gene;
KW   ds; cancer; cytostatic.
XX
OS   Mus musculus.
XX
PN   WO2004058146-A2.
XX
PD   15-JUL-2004.
XX
PF   15-DEC-2003; 2003WO-US040081.
XX
PR   17-DEC-2002; 2002US-00322281.
XX
PA   (SAGR-) SAGRES DISCOVERY INC.
XX
PI   Morris DW,  Malandro MS;
XX
DR   WPI; 2004-499109/47.
XX
PT   Novel human cancer associated protein encoded within open reading frame
PT   of cancer associated gene, useful as targets for diagnosing cancer.
XX
PS   Disclosure; SEQ ID NO 123; 182pp; English.
XX
CC   The invention relates to cancer-associated proteins (CAP) and the cancer-
CC   associated (CA) nucleic acids encoding them. The invention also relates
CC   to a method for treating cancers involving administering to a patient an
CC   inhibitor of CAP, and a method of screening for anticancer activity in a
CC   potential drug involving providing a cell that expresses a CA gene,
CC   contacting a tissue sample derived from a cancer cell with an anticancer
CC   drug candidate and monitoring the effect of the anticancer drug candidate
CC   on expression of the CA gene. The CAP proteins are useful for detecting
CC   cancer associated with expression of a CAP protein in a test cell sample
CC   and for screening for a bioactive agent capable of modulating the
CC   activity of a CAP protein. The CA nucleic acids are useful for diagnosing
CC   cancer, involving determining the expression of a CA nucleic acid in a
CC   tissue. This sequence represents a murine CA gene of the invention. Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
```

```
XX
SQ   Sequence 104399 BP; 25540 A; 20083 C; 20918 G; 29473 T; 0 U; 8385 Other;

  Query Match             5.5%;  Score 255.4;  DB 13;  Length 104399;
  Best Local Similarity   73.3%;  Pred. No. 1.9e-34;
  Matches  367;  Conservative    0;  Mismatches  131;  Indels    3;  Gaps    3;

Qy        230 ATATATATATATATATATATATACACACACACACACATATATATATACACATATATATAC 289
              ||| ||| ||| ||||||| |||| ||| | | ||||| ||| |||| | ||||||||
Db      13721 ATACATACATACATATATACATACGTACATATATACATACATACATACGTACATATATAC 13662

Qy        290 ACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATAT 349
              | |  || | ||||| | ||||| ||||| ||||| ||||  | | ||||  | |||||
Db      13661 ATACGTACATATATACATACATACATACATATATACATACGTACATATATACATACATAC 13602

Qy        350 ATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATAC 409
              |||   |||||| || | || || ||| | | ||| | ||| | ||||| |||||  |||
Db      13601 ATACGTACATATATACATACGTACATATATACATACATACATACATATATACATACGTAC 13542

Qy        410 ACACATATATACACATATATACACATATATACACATATACACATATATACACATATATAC 469
              | | ||| ||||| | |||   |||||||||||| | ||||| | ||||| ||||| ||||
Db      13541 ATATATACATACATACATACGTACATATATACATACATACATACATATATACATACATAC 13482

Qy        470 ATATATACACATATATATAATATACACACATATATATACACATATATACACACATATATA 529
              ||||||||| | ||| |||||| | | | | ||  ||| | |  ||| | |||  ||||| |
Db      13481 ATATATACATACATACATAATACATATATACATTCATATATACATA-ATACATATATACA 13423

Qy        530 CACATATATACACATATATAT-ACACATATATACACATAT-ATACATATATACACATATA 587
              |||||||||||| | ||| || | |||||||||||| | || ||||||| || || |||||
Db      13422 TACATATATACATACATACATAATACATATATACATACATAATACATACATGCATATATA 13363

Qy        588 TATACATATATACACATATATACATATATACACATATATACATATATACACACATATATA 647
              || |||| ||| | | ||| ||| | ||||| | ||| | | ||||| ||| |   || |
Db      13362 CATGCATACATATATACATACACACACATACATACATACATATATATACATACGTACA 13303

Qy        648 CACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATA 707
              | |||||| | ||| || || ||||| | ||| ||||| ||||| |||| || |||||
Db      13302 TATATACATACATACATACGTACATATATACATACATACATATATACATACATACATATA 13243

Qy        708 TGTATATATATATGCTCCAGA 728
              |  ||| ||| ||  | || |
Db      13242 TACATACATACATAATACATA 13222


RESULT 49
ADC86154/c
ID   ADC86154 standard; DNA; 84510 BP.
XX
AC   ADC86154;
XX
DT   01-JAN-2004  (first entry)
XX
DE   Human GPCR gene SEQ ID NO:607.
XX
KW   ds; gene; human; GPCR;
KW   guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1270724-A2.
XX
PD   02-JAN-2003.
XX
PF   18-JUN-2002; 2002EP-00013517.
XX
PR   18-JUN-2001; 2001JP-00246789.
XX
PA   (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA   (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI   Suwa M,  Asai K,  Akiyama Y,  Aburatani H;
XX
DR   WPI; 2003-315783/31.
DR   P-PSDB; ADC86155.
XX
PT   New polynucleotide, useful for preparing a composition for treating a
PT   patient in need of increased or suppressed activity or expression of the
PT   guanosine triphosphate-binding protein coupled receptor.
XX
PS   Claim 1; SEQ ID NO 607; 28pp; English.
XX
CC   The invention relates to a novel polynucleotide encoding a guanosine
CC   triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC   the invention may have a use in gene therapy. The polynucleotide and
CC   polypeptide are useful for preparing a composition for treating a patient
CC   in need of increased or suppressed activity or expression of the
CC   guanosine triphosphate-binding protein coupled receptor. The
CC   polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC   invention.
XX
SQ   Sequence 84510 BP; 21253 A; 19989 C; 19504 G; 23764 T; 0 U; 0 Other;

  Query Match             5.5%;  Score 255;  DB 10;  Length 84510;
  Best Local Similarity   72.6%;  Pred. No. 2.2e-34;
  Matches  358;  Conservative    0;  Mismatches  130;  Indels    5;  Gaps    2;

Qy        232 ATATATATATATATATATATACACACACACACATATATATATACACATATATATACAC 291
              ||||||||| |||| |||| | | ||||| || |||||||||| ||  ||||| ||
Db      80292 ATATATATACATATTCTTATATATATATACACATATGTATATATACATATGAATATATAC 80233

Qy        292 ATATATACATATATACACATATATACACATATATATACACATACATATGTACACATATAT 351
              | ||||  |||||||||| ||  ||| | | | ||||||| ||  |||| ||||||||||
Db      80232 ACATATGTATATATACATATGAATATATACACATATACATATGAATATATACACATATAC 80173

Qy        352 ATACACATATGTATACACATATATACACATATATACACACATATATACACATATATACAC 411
              |||   |||| || | | ||| |||   |||||||||||| |||||   |   ||||| ||
Db      80172 ATATGAATATATACATATATACATATGAATATATACACATATATACGTATGAATATATAC 80113

Qy        412 ACATATATAC----ACATATATACACATATATACACATATACACATATATACACATATAT 467
              ||||||||||     ||||||| ||| ||||| || |||  | ||||||||||||||||
Db      80112 ACATATATACGTATGAATATATATACACATATATACGTATGAATATATATACACATATAT 80053

Qy        468 ACATATATACACATATATATAATATACACACATATATATACACATATATACACACATATA 527
              || |||  | | ||||| |  ||||||  |  |||||||  ||  ||||||  |   |||
Db      80052 ACGTATGAATATATATACA-CATATACGTATGAATATATATACCCATATACGTATGAATA 79994

Qy        528 TACACATATATACACATATATATACACATATATACACATATATACATATATACACATATA 587
              |||||||||||| || ||  |||| | ||| | | | |||  ||| |||||| | | | |
Db      79993 TACACATATATATACGTACGTATATATATACACATATATACGTACGTATATATATACACA 79934

Qy        588 TATACATATATACACATATATACATATATACACATATATACATATATACACACATATATA 647
              |||| |  ||   | ||||||||| ||||| || ||   | ||||||||||| |||||
Db      79933 TATATACGTACGTATATATATACACATATATACGTACGAATATATATACACATATATACG 79874

Qy        648 CACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATA 707
              ||  | ||||| ||||| ||| |  |||  ||||||| ||| ||||||||||   |||
Db      79873 TACGAATATATATACACATATATACGTACGAATATATATACACATATATACGTACGAATA 79814

Qy        708 TGTATATATATAT 720
              | |||| | ||||
Db      79813 TATATACACATAT 79801


RESULT 50
ADL13771/c
ID   ADL13771 standard; DNA; 201766 BP.
XX
AC   ADL13771;
XX
DT   06-MAY-2004  (first entry)
XX
DE   Osteoarthritis-associated polymorphic nucleotide #303.
XX
KW   ds; gene; osteopathic; antiinflammatory; antiarthritic; gene therapy;
KW   joint space narrowing; osteophyte development; joint pain;
KW   osteoarthritis; SNP; single nucleotide polymorphism.
```

XX
OS Homo sapiens.
XX
PN WO2003054166-A2.
XX
PD 03-JUL-2003.
XX
PF 19-DEC-2002; 2002WO-US041225.
XX
PR 20-DEC-2001; 2001US-0342603P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Jones KA, Schafer A;
XX
DR WPI; 2003-559141/52.
XX
PT Determining susceptibility of an individual to joint space narrowing,
PT osteophyte development and/or joint pain comprises identifying whether
PT the individual has at least one polymorphism in a polynucleotide encoding
PT a protein.
XX
PS Disclosure; SEQ ID NO 303; 297pp; English.
XX
CC The invention relates to a method of determining susceptibility of an
CC individual to joint space narrowing and/or osteophyte development and/or
CC joint pain comprising identifying whether the individual has at least one
CC polymorphism in a polynucleotide encoding at least one of the protein
CC listed in the specification. The methods, composition and agent are
CC useful for modulating the susceptibility of an individual to joint space
CC narrowing and/or osteophyte development and/or joint pain that is
CC associated with a disease, preferably osteoarthritis. The cell line and
CC the non-human animal are useful for screening for an agent for diagnosing
CC an individual having susceptibility to joint space narrowing and/or
CC osteophyte development and/or joint pain. This sequence corresponds to
CC the polynucleotide encoding a protein listed in the specification. (Note:
CC The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences).
XX
SQ Sequence 201766 BP; 57593 A; 35575 C; 36821 G; 54688 T; 0 U; 17089 Other;

Query Match 5.5%; Score 255; DB 10; Length 201766;
Best Local Similarity 63.6%; Pred. No. 2.3e-34;
Matches 437; Conservative 0; Mismatches 242; Indels 8; Gaps 3;

Qy 236 ATATATATATATATATACACACACACACATATATATATACACATATATATACACATAT 295
Db 189316 ACATGTGTATATATATACATATATACGTATATATGTGTATATATACATACATAGGTATAT 189257

Qy 296 ATACATATATACACATATATACACATATATATACACATACATATGTACACATATATATAC 355
Db 189256 ACATATATACACATACATATATACGTATATATGTATATATGTATATATACATGTATATAT 189197

Qy 356 ACATATGTA-TACACATATATACACATATATACACACATATATACACATATATACACACA 414
Db 189196 GTATATGNNGTATACATATATACACGTATATATACGTATATGTATACATATATACACGTA 189137

Qy 415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
Db 189136 TATATACGTATATGTATACATATATACACGTATATATACGTGTATATACATGTATATATA 189077

Qy 475 TACACATATATATAATATACACACATATATATACACATATATACACACATATATACACAT 534
Db 189076 CGTATATGT-----ATACATATACATATATACACGTGTATATACATATGTATATACACGT 189022

Qy 535 ATATACACATATATATACACATATATACACATATATACATATATACACATATATATACAT 594
Db 189021 GTATATATGTATGTATATACACGTGTATATATGTATACGTATATATACACGTGTATATAT 188962

Qy 595 ATATACACATATATACATATATACACATATATACATATATACACACATATATACACATAC 654
Db 188961 GTATACGTATATATACACGTGTATATATGTATACGTATATATACACGTGTATATATGTAT 188902

Qy 655 ATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATAT 714
Db 188901 ACGTATATATACACGTGTATATATATGTATACGTATATATACACGTGTATATATGTATAC 188842

Qy 715 ATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAG 774
Db 188841 GTATATATACACGTGTATATATGTATGTATATATACGTGTATATATATGTATACGTATAT 188782

Qy 775 AGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGG 834
Db 188781 ATACGTATATATAAATGTAT--ATATGTATATGTGTATATATACGTATATATACATGTAT 188724

Qy 835 AAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATA 894
Db 188723 ATATACGTATATATACACATATACATATATGGTACGTATATATACATATGTGTATATATG 188664

Qy 895 TATATATAGTGAAAGCAGTGTAAACCT 921
Db 188663 TATATATATACATACATATATATACAT 188637

RESULT 51
ADC87564
ID ADC87564 standard; DNA; 1114 BP.
XX
AC ADC87564;
XX
DT 01-JAN-2004 (first entry)
XX
DE Human GPCR gene SEQ ID NO:2017.
XX
KW ds; gene; human; GPCR;
KW guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1270724-A2.
XX
PD 02-JAN-2003.
XX
PF 18-JUN-2002; 2002EP-00013517.
XX
PR 18-JUN-2001; 2001JP-00246789.
XX
PA (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI Suwa M, Asai K, Akiyama Y, Aburatani H;
XX
DR WPI; 2003-315783/31.
DR P-PSDB; ADC87565.
XX
PT New polynucleotide, useful for preparing a composition for treating a
PT patient in need of increased or suppressed activity or expression of the
PT guanosine triphosphate-binding protein coupled receptor.
XX
PS Claim 1; SEQ ID NO 2017; 28pp; English.
XX
CC The invention relates to a novel polynucleotide encoding a guanosine
CC triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC the invention may have a use in gene therapy. The polynucleotide and
CC polypeptide are useful for preparing a composition for treating a patient
CC in need of increased or suppressed activity or expression of the
CC guanosine triphosphate-binding protein coupled receptor. The
CC polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC invention.
XX
SQ Sequence 1114 BP; 398 A; 107 C; 167 G; 442 T; 0 U; 0 Other;

Query Match 5.5%; Score 254.2; DB 10; Length 1114;
Best Local Similarity 64.0%; Pred. No. 2.5e-34;
Matches 399; Conservative 0; Mismatches 223; Indels 1; Gaps 1;

```
Qy        229 TATATATATATATATATATATATACACACACACACATATATATATACACATATATATA 288
              ||||| ||||| | ||| |||||||| | | | ||| |||| |||||     || ||||||
Db        212 TATATGTATATTTGTATGTATATACAAATATATACATATATGTATATGTGTATGTATATA 271

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              || ||||||||||||||    || |||| ||| |||||||| ||| | | |  | |||
Db        272 CAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATATA 331

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              | ||| | | || |||    || |||| ||| ||||| ||| |||| |  |  ||||||
Db        332 CAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATATA 391

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              || | ||||| | | || | ||   ||||| | | ||||||| |||| |  |  ||||||
Db        392 CAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATATA 451

Qy        469 CATATATACACATATATATAATATACACACATATATATACACATATATACACACATATAT 528
              || ||||| |||||||| |  |||  | | | | ||||| ||||||||    |  |||||
Db        452 CAAATATATACATATATGT-GTATGTATATACAAATATATACATATATGTGTATGTATAT 510

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACATATAT 588
              ||| |||||||| |||| | ||   ||||| | | ||||| | |||| |  |  |||||
Db        511 ACAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATAT 570

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
              | | |||||||||| |||| |  || |||| ||| ||||| | |||||     || ||||
Db        571 ACAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATAT 630

Qy        649 ACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATAT 708
              ||| | ||||||| | |  |  || || ||| | ||| | | ||||||  |||| |||||
Db        631 ACAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATAT 690

Qy        709 GTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAG 768
              | |||||||       | ||   ||      |  |  |   ||   ||    || |
Db        691 ACAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATAT 750

Qy        769 GAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAG 828
              ||| | |     | |   || | |    | ||  | ||       ||       |||
Db        751 ACAAATATATACATATATGTGTATGTATATACAAATATATACATATATGTGTATGTATAT 810

Qy        829 TAGGGGAAATACATAAATGTGTA 851
                  | ||||||| |||||||
Db        811 ACAAATATATACATATATGTGTA 833
```

```
RESULT 52
ACC84012
ID   ACC84012 standard; DNA; 2200 BP.
XX
AC   ACC84012;
XX
DT   22-SEP-2003  (first entry)
XX
DE   Pacific white shrimp skeletal muscle actin gene promoter.
XX
KW   Shrimp; actin; promoter; transgenic; disease resistance;
KW   stress tolerance; cold tolerance; ds.
XX
OS   Litopenaeus vannamei.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    878. .893
FT                   /*tag=  a
FT                   /note= "CATA-rich region"
FT   misc_feature    1071. .1078
FT                   /*tag=  a
FT                   /note= "CATA-rich region"
FT   misc_feature    1538. .1549
FT                   /*tag=  a
FT                   /note= "CACA-rich region"
FT   misc_feature    1554. .1567
FT                   /*tag=  a
FT                   /note= "CACA-rich region"
FT   misc_feature    1570. .1585
FT                   /*tag=  a
FT                   /note= "CACA-rich region"
XX
PN   WO2003048325-A2.
XX
PD   12-JUN-2003.
XX
PF   04-DEC-2002; 2002WO-US038523.
XX
PR   04-DEC-2001; 2001US-0336603P.
XX
PA   (UYHA-) UNIV HAWAII.
XX
PI   Sun PS,  Arakaki KL,  Sun SSM;
XX
DR   WPI; 2003-513755/48.
XX
PT   New isolated Beta-lactin nucleic acid promoter molecule from shrimp,
PT   useful for imparting resistance against a virus and increasing stress
PT   tolerance in an animal and for regulating growth of an animal.
XX
PS   Claim 52; Page 14-15; 97pp; English.
XX
CC   The present sequence is that of the promoter region of the Pacific white
CC   shrimp skeletal muscle actin gene. This promoter is suitable for inducing
CC   expression of a protein encoded by a nucleic acid molecular operably
CC   associated with the promoter molecule in an expression vector. It
CC   contains promoter-associated TATA and CAAT boxes approximately 500 bp
CC   upstream from the translation start site. A unique characteristic of the
CC   promoter are CACA-rich and CATA-rich regions located upstream of the TATA
CC   and CAAT boxes. The invention relates to actin and beta-actin promoter
CC   molecules from shrimp, expression cassettes including the promoter,
CC   expression vectors, host cells and transgenic animals transduced with the
CC   promoter. Methods for imparting resistance to a pathogen, for regulating
CC   growth, for improving stress tolerance, and for increasing cold tolerance
CC   in an animal, especially a shrimp, involve transforming the animal with a
CC   nucleic acid construct including the isolated actin and beta-actin
CC   promoter molecules of the invention
XX
SQ   Sequence 2200 BP; 837 A; 372 C; 263 G; 728 T; 0 U; 0 Other;
```

```
  Query Match             5.5%;  Score 254;  DB 9;  Length 2200;
  Best Local Similarity   71.7%;  Pred. No. 2.8e-34;
  Matches  360;  Conservative    0;  Mismatches  140;  Indels    2;  Gaps    2;

Qy        234 ATATATATATATATATATACACACACACACATATATATATACACATATATATACACAT 293
              | || ||||||||||||||   | |    | | |||| ||| || | | ||| ||| |||
Db        647 ACATTTATATATATATATATGTATATGTATATATATGTATTTATATACATAAATATACAC 706

Qy        294 ATATACATATATACACATATATACACATATATATACACATACATATGTACACATATATAT 353
              |||| ||| |||| | |||| || | ||| |||||  |  || |||| | || ||||||||
Db        707 ATATTCATGTATATATATATGTATATATACATATATATTTATATATATGCATATATATAT 766

Qy        354 ACACATATGTATACACATATATACACATATATACACACATATATACACATATATACACAC 413
                | |||| |||| | | |||| || ||  ||| | |  |||||| | | ||||||| |
Db        767 GTATATATATATATATACATATGCATATTCATATATATGTATATATATACATATACATAT 826

Qy        414 ATATATACACATATATACACATATATACACATATACACATATATACACATATATACATAT 473
              |||| || | ||||  | | ||||||  | ||  | | ||||||| | ||| |||||||
Db        827 ATATGTATATATATGCATAAATATATGTATATGAATATATATATATATATACATACATAC 886

Qy        474 ATACACATATATATAATATACACACATATATATACACATATATACACACATATATACACA 533
              ||||| |||||||| ||||| | | |||||||||  | |||||||  | |  ||| | |||||
Db        887 ATACATATATATAT-ATATATATATATATATATATATATATATATATATACACACACA 945

Qy        534 TATATACACATATATATACACATATATACACATATATACATATATACACATATATATACA 593
              | || |||  | |||||||||  | | |  ||| | | |||||||||    | ||  |||||
Db        946 CACATGCACGCACATATACACACACACAAACACACACACATATATATGTAAATGCATACA 1005

Qy        594 TATATACACATATATACATATATACACATATATACATATATACACACATATATACACATA 653
              ||| |  |||   | |||||||||||||||||||||||| | ||||| | | ||||| ||
```

```
Db       1006 TATGTGTACACGGACACATATATACACATATATACATACACACACA-ACAGATACATATG 1064

Qy        654 CATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATA 713

Db       1065 CATATACATACATACACATATATCTCTCTCTATATACATATATATGTATATATGTGTATA 1124

Qy        714 TATATATGCTCCAGAGTTCATA 735

Db       1125 TACTTATACATAAAAACACACA 1146


RESULT 53
AAV83939/c
ID   AAV83939 standard; DNA; 80595 BP.
XX
AC   AAV83939;
XX
DT   03-MAR-1999  (first entry)
XX
DE   HC-contig derived from normal human chromosome 10q25.2 region.
XX
KW   Yeast artificial chromosome; YAC; probe; eukaryotic chromosome;
KW   neocentromere; replication; extra-chromosomal element; segregation;
KW   cell division; artificial chromosome; gene therapy; mardel(10);
KW   human artificial chromosome; transgenic; chromosome 10; 10q25.2; ss.
XX
OS   Homo sapiens.
XX
PN   WO9851790-A1.
XX
PD   19-NOV-1998.
XX
PF   13-MAY-1998;   98WO-AU000352.
XX
PR   13-MAY-1997;   97AU-00006784.
PR   26-AUG-1997;   97AU-00008791.
XX
PA   (AMRA-) AMRAD OPERATIONS PTY LTD.
XX
PI   Choo K,  Du Sart D,  Cancilla MR;
XX
DR   WPI; 1999-009773/01.
XX
PT   New isolated nucleic acid comprising neocentromere sequences from
PT   eukaryotic chromosome - used to produce replicable, segregating
PT   artificial chromosomes that can carry large amounts of DNA for gene
PT   therapy.
XX
PS   Claim 8; Fig 6; 540pp; English.
XX
CC   The present sequence represents the HC-contig derived from normal human
CC   chromosome 10, 10q25.2 region. This region can be naturally mutated to
CC   produce an unusual chromosomal marker referred to as mardel(10). The
CC   mardel(10) marker is mitotically stable and contains a functional
CC   neocentromere at a location regarded as non-centromeric. This
CC   neocentromere maps to q25.2 on chromosome 10. The specification describes
CC   nucleic acid sequences derived from a eukaryotic chromosome, including a
CC   neocentromere or its functional derivative or hybrid, that are able, in a
CC   compatible cell, of replicating, acting as extra-chromosomal element and
CC   segregating during cell division. The sequences can be used to construct
CC   artificial chromosomes for use in gene therapy comprising a replicable,
CC   segregating nucleic acid that confers a specific phenotype on cells.
CC   Human artificial chromosomes can propagate in human cells and carry large
CC   amounts of DNA (e.g. therapeutic genes), and, being extra-chromosomal,
CC   they are not mutagenic. The artificial chromosomes are also useful for
CC   generation of transgenic plants and animals, in production of proteins
CC   and to make diagnostic reagents, e.g. for expression of cytokines;
CC   receptors and growth factors, or to increase the copy number of a gene in
CC   a cell. The constructs may also be used for functional and structural
CC   analysis of chromosomes
XX
SQ   Sequence 80595 BP; 23183 A; 16613 C; 16824 G; 23975 T; 0 U; 0 Other;

  Query Match             5.5%;  Score 253.6;  DB 2;  Length 80595;
  Best Local Similarity   71.9%;  Pred. No. 3.8e-34;
  Matches  361;  Conservative    0;  Mismatches  134;  Indels    7;  Gaps    2;

Qy        227 TTTATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286

Db      16026 TATATACATATTTATGTCTATATACACATATATACATATTTATGTCTATATACACATATA 15967

Qy        287 TACACATATATACATATATACACATATATACACA------TATATATACACATACATATG 340

Db      15966 TACATATTTATGTCTATATACACATATATACATATTTATGTCTATATACACATATATACA 15907

Qy        341 TACACATATATATACACATATGTATACACATATATACACATATATACACACATATATACA 400

Db      15906 TATTTATGTCTATATACACATATATACATATTTATGTCTATATACACATATATACATATT 15847

Qy        401 CATATATACACACATATATACACATATATACACATATATACACATATACACATATATACA 460

Db      15846 TATGTCTATATACACATATATACATATTTATGTCTATATACACATATATACATATTTATG 15787

Qy        461 CATATATACATATATACACATATATAT-AATATACACACATATATATACACATATATACA 519

Db      15786 TCTATATACACATATATACATATTTATGTCTATATACACATATATACATATTTATGTCTA 15727

Qy        520 CACATATATACACATATATACACATATATATACACATATATACACATATATACATATATA 579

Db      15726 TATACACATATATACATATTTATGTCTATATACACATATATACATATTTATGTCTATATA 15667

Qy        580 CACATATATATACATATATACACATATATACATATATACACATATATACATATATACACA 639

Db      15666 CACATATATACATATTTATGTATATATACACATATATACATATTTATGTCTATATACACA 15607

Qy        640 CATATATACACATACATATACACACACATAGATATACATATATATACACATATATATACG 699

Db      15606 TATATACATATTTATGTATATATTTATGTATATATGTGTGTATATGTACATACATATATA 15547

Qy        700 TATATATATGTATATATATATG 721

Db      15546 CATACATAAATACATATATATG 15525


RESULT 54
AAA10225/c
ID   AAA10225 standard; DNA; 106746 BP.
XX
AC   AAA10225;
XX
DT   03-JUL-2000  (first entry)
XX
DE   Human PCTA-1 genomic DNA, SEQ ID NO:1.
XX
KW   Prostate carcinoma antigen-1; PCTA-1; human; galectin-8 homologue;
KW   splice variant; prostate cancer; biallelic marker; polymorphism;
KW   detection; diagnosis; prognosis; drug screening; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   misc_signal     1. .68647
FT                   /*tag=  a
FT                   /note= "5' regulation region"
FT   promoter        66647. .68647
FT                   /*tag=  b
FT   exon            68648. .68471
FT                   /*tag=  c
FT                   /number
FT   intron          68742. .70646
FT                   /*tag=  d
FT                   /number
FT   exon            70647. .70794
FT                   /*tag=  e
FT                   /number= 1
FT   intron          70795. .82207
```

```
FT                   /*tag= f
FT                   /number= 1
FT   exon            82208. .82296
FT                   /*tag= g
FT                   /number= 2
FT   intron          82297. .83612
FT                   /*tag= h
FT                   /number= 2
FT   exon            83613. .83823
FT                   /*tag= i
FT                   /number= 3
FT   intron          83824. .85297
FT                   /*tag= j
FT                   /number= 3
FT   exon            85298. .85417
FT                   /*tag= k
FT                   /number= 4
FT   intron          85418. .86388
FT                   /*tag= l
FT                   /number= 4
FT   exon            86389. .86445
FT                   /*tag= m
FT                   /number= 5
FT   intron          86446. .87495
FT                   /*tag= n
FT                   /number= 5
FT   exon            87496. .87522
FT                   /*tag= o
FT                   /number= 6
FT   intron          87523. .87649
FT                   /*tag= p
FT                   /number= 6
FT   exon            87650. .87775
FT                   /*tag= q
FT                   /number= 6bis
FT   intron          87776. .88294
FT                   /*tag= r
FT                   /number= 6bis
FT   exon            88295. .88383
FT                   /*tag= s
FT                   /number= 7
FT   intron          88384. .89483
FT                   /*tag= t
FT                   /number= 7
FT   exon            89484. .89649
FT                   /*tag= u
FT                   /number= 8
FT   intron          89650. .92748
FT                   /*tag= v
FT                   /number= 8
FT   exon            92749. .97155
FT                   /*tag= w
FT                   /number= 9
FT   exon            92749. .92883
FT                   /*tag= x
FT                   /number= 9bis
FT   intron          92884. .95820
FT                   /*tag= y
FT                   /number= 9bis
FT   exon            95821. .97155
FT                   /*tag= z
FT                   /number= 9ter
FT   misc_signal     97155. .106746
FT                   /*tag= aa
FT                   /note= "3' regulation region"
XX
PN   WO9964590-A1.
XX
PD   16-DEC-1999.
XX
PF   04-JUN-1999;   99WO-IB001072.
XX
PR   05-JUN-1998;   98US-0088187P.
PR   28-SEP-1998;   98US-0102324P.
XX
PA   (GEST ) GENSET.
XX
PI   Blumenfeld M,  Bougueleret L,  Chumakov I;
XX
DR   WPI; 2000-338635/29.
DR   P-PSDB; AAY87403, AAY87404, AAY87405.
XX
PT   Novel polynucleotide used to develop products for detecting
PT   susceptibility to, diagnosis, prognosis and therapy of prostate cancer.
XX
PS   Claim 1; Page 277-306; 339pp; English.
XX
CC   This sequence represents the genomic DNA sequence encoding human prostate
CC   carcinoma tumour antigen-1 (PCTA-1). At least three PCTA-1 cDNAs
CC   (AAA10226, AAA10227, AAA10228) have been identified all of which comprise
CC   exons 1, 2, 3, 4, 5, 6, 7, 8, and additionally either contain exon 9,
CC   exons 6bis and 9, or exons 9bis and 9ter. PCTA-1 is a human homologue of
CC   rat galectin-8, which can mediate both cell-cell and cell-matrix
CC   interactions. 125 biallelic markers were identified in the genomic
CC   sequence, with 40 biallelic markers being located in exonic regions. Six
CC   of these alter the amino acid sequence of a PCTA-1 protein. 43 markers
CC   were found in the 5' regulatory region, including 16 in the promoter, 39
CC   in intronic sequences, and 3 in the 3' regulatory region. Alleles
CC   containing certain biallelic markers are associated with prostate cancer.
CC   These polymorphisms can be used as the basis of methods for determining
CC   genetic predisposition to prostate cancer, for diagnosis of this cancer
CC   and for the prognosis/detection of an eventual treatment response to
CC   therapeutic agents against prostate cancer. They can be used to design of
CC   drug screening protocols to provide accurate and efficient evaluation of
CC   therapeutic and side-effect potential of new or already existing
CC   medicaments, as well as for screening molecules which modulate or inhibit
CC   the expression of the PCTA-1 gene, preferably those active against
CC   prostate cancer
XX
SQ   Sequence 106746 BP; 29695 A; 23487 C; 23529 G; 30029 T; 0 U; 6 Other;

  Query Match             5.5%;  Score 253;  DB 3;  Length 106746;
  Best Local Similarity   69.4%;  Pred. No. 4.8e-34;
  Matches  416;  Conservative    0;  Mismatches  175;  Indels    8;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              |||||||||||||||||  ||||||||||||||||||| |||||||||||||| ||||||
Db      38717 AGTTCAAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAAAATT 38658

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   ||||||||| | ||||||||||||||||||||||   |||| | ||||||||||||
Db      38657 AGCTGGGCATGGTGGTGCACGCCTGTAATCCCAGCTACTCAGGAGACTGAGGCAGGAGAA 38598

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCCA 178
              ||||||||||||  || |||||| ||| || ||| ||| |   | ||| |||||||||
Db      38597 TCACTTGAACCTG--GGAGGTGGAGTTTGTGGTGAGCCAAAACTTTCCCACTGCACTCCA 38540

Qy        179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCAT-TTTATATATAT 237
              |||      |   | |   | || |   | |  | |      | |   ||    | ||||
Db      38539 GCCAGTATATATATATATATATACACACGTACACGTGTACACATATACACACATATGC 38480

Qy        238 ATATATATATATATACACACACACACATATATATATACACATATATATACACATATAT 297
              | |  | || ||| ||||||| |||||   | ||| || ||||||| || |  | || ||
Db      38479 ACACGTGTACATACACACACATACACATGCACATACATGCACATATGTACATGCGTACAT 38420

Qy        298 ACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACAC 357
              ||| | ||  |||||||||||| || || | | ||||     |||||| || || | |
Db      38419 ACACACATGTACATATATACACACATGTACATATGTACACCCATACACACATGTACATAT 38360

Qy        358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417
              ||      | ||| | ||  |||||| ||||| |||||| | |  || |||   |||
Db      38359 GTACACCCATACACACATGTACATATGTACACCCATATACATATGTACATATGTACACCC 38300

Qy        418 ATACACATATATACACATATATACACATATACACATATATACACATATATACATATATAC 477
              |||||||||| |||| || || || |||| |||  || |||   | |   | | | ||
Db      38299 ATACACATATGTACATATGTACACCCATACACATGTACATATGTACACCCATACACATGT 38240
```

```
Qy        478 ACATATATATAATATACACACATATATATACACATATATACACACATATATACACATATA 537
              |||||| |   | || |||| || || |||   ||||  | ||||||||| |||| |||||
Db      38239 ACATATGT---ACATCCACATATGTACATATGTATATCCATACACATATGTACATATATA 38183

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATACATAT 596
              |||||||| || | |  || ||| | || | |  |||| ||   |||||||||   ||
Db      38182 CACACATATGTACATATGTACATATATATTTGTGTATATCTATGTATATATATAAGAAT 38124


RESULT 55
AAV83940/c
ID   AAV83940 standard; DNA; 80240 BP.
XX
AC   AAV83940;
XX
DT   03-MAR-1999  (first entry)
XX
DE   NC-contig derived from mardel(10) on chromosome 10q25.2.
XX
KW   Yeast artificial chromosome; YAC; probe; eukaryotic chromosome;
KW   neocentromere; replication; extra-chromosomal element; segregation;
KW   cell division; artificial chromosome; gene therapy; mardel(10);
KW   human artificial chromosome; transgenic; chromosome 10; 10q25.2; ss.
XX
OS   Homo sapiens.
XX
PN   WO9851790-A1.
XX
PD   19-NOV-1998.
XX
PF   13-MAY-1998;   98WO-AU000352.
XX
PR   13-MAY-1997;   97AU-00006784.
PR   26-AUG-1997;   97AU-00008791.
XX
PA   (AMRA-) AMRAD OPERATIONS PTY LTD.
XX
PI   Choo K,  Du Sart D,  Cancilla MR;
XX
DR   WPI; 1999-009773/01.
XX
PT   New isolated nucleic acid comprising neocentromere sequences from
PT   eukaryotic chromosome - used to produce replicable, segregating
PT   artificial chromosomes that can carry large amounts of DNA for gene
PT   therapy.
XX
PS   Claim 9; Fig 16A; 540pp; English.
XX
CC   The present sequence represents the NC-contig derived from a mutated
CC   human chromosome 10, 10q25.2 region. The sequence contains an unusual
CC   chromosomal marker referred to as mardel(10). The mardel(10) marker is
CC   mitotically stable and contains a functional neocentromere at a location
CC   regarded as non-centromeric. This neocentromere maps to q25.2 on
CC   chromosome 10. The specification describes nucleic acid sequences derived
CC   from a eukaryotic chromosome, including a neocentromere or its functional
CC   derivative or hybrid, that are able, in a compatible cell, of
CC   replicating, acting as extra-chromosomal element and segregating during
CC   cell division. The sequences can be used to construct artificial
CC   chromosomes for use in gene therapy comprising a replicable, segregating
CC   nucleic acid that confers a specific phenotype on cells. Human artificial
CC   chromosomes can propagate in human cells and carry large amounts of DNA
CC   (e.g. therapeutic genes), and, being extra-chromosomal, they are not
CC   mutagenic. The artificial chromosomes are also useful for generation of
CC   transgenic plants and animals, in production of proteins and to make
CC   diagnostic reagents, e.g. for expression of cytokines, receptors and
CC   growth factors, or to increase the copy number of a gene in a cell. The
CC   constructs may also be used for functional and structural analysis of
CC   chromosomes
XX
SQ   Sequence 80240 BP; 23102 A; 16537 C; 16747 G; 23846 T; 0 U; 8 Other;

  Query Match             5.5%;  Score 252.6;  DB 2;  Length 80240;
  Best Local Similarity   71.9%;  Pred. No. 5.6e-34;
  Matches  360;  Conservative    0;  Mismatches  134;  Indels    7;  Gaps    2;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | || |||| ||| | |||||| | ||| | | ||| |  ||| | |||| ||| | |||
Db      15787 TATACATATTTATGTCTATATACACACATATATACATATTTATGTCTATATACACACATA 15728

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              |||| || |||  ||||||||||||||||||| || ||| |  | ||||| || || |||
Db      15727 TACATATTTATGTCTATATACACATATATACATATTTATGTCTATATACACATATATACA 15668

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              ||| |||    ||||  | | | ||| |||   || | || | ||| ||||| ||||||
Db      15667 TATTTATGTCTATATACACATATATACATATTTATGTCTATATACACATATATACATATT 15608

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              |  |     |||||||||||||||||| || |||    ||||||||||||||||| || ||
Db      15607 T--ATGTCTATATACACATATATACATATTTATGTCTATATACACATATATACATATTTA 15550

Qy        467 TACATATATACACATATATA-----TAATATACACACATATATATACACATATATACACA 521
              |   ||||||||||||||||     | || |  | | | | ||||| |||||| ||
Db      15549 TGTCTATATACACATATATACATATTTATGTCTATATACACATATATACATATTTATGTC 15490

Qy        522 CATATATACACATATATACACATATATATACACATATATACACATATATACATATATACA 581
              |||||| ||| ||| |||   || | |||| ||| ||||| |||||| ||  | |||| |
Db      15489 TATATACACATATACATATTTATGTCTATATACACATATATACATATTTATGTCTATATA 15430

Qy        582 CATATATATACATATATACACATATATACATATATACACATATATACATATATACACACA 641
              || |||||||||||| ||    |||||||| ||||| |||||| ||  |||||| |||||
Db      15429 CACATATATACATATTTATGTCTATATACACATATATACATATTTATGTATATATACACA 15370

Qy        642 TATATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTA 701
              |||||||| ||  || |  | | ||| | |||||||||| |||  | |||| |||   ||
Db      15369 TATATACATATTTATGTCTATATACACATATATACATATTTATGTATATATTTATGTATA 15310

Qy        702 TATATATGTATATATATATGC 722
              ||| | ||||||| || || |
Db      15309 TATGTGTGTATATGTACATAC 15289


RESULT 56
ABV72623/c
ID   ABV72623 standard; DNA; 82938 BP.
XX
AC   ABV72623;
XX
DT   27-NOV-2002  (first entry)
XX
DE   Human transporter protein encoding gene.
XX
KW   Human; transporter protein; hTP; gamma-aminobutyric-acid; GABA;
KW   gene therapy; protein therapy; vaccine; gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1120. .81932
FT                   /*tag=  a
FT                   /product= "Transporter protein"
FT                   /note= "Contains 8 introns"
FT   exon            1120. .1223
FT                   /*tag=  b
FT                   /number= 1
FT   intron          1224. .27733
FT                   /*tag=  c
FT                   /number= 1
FT   variation       replace(2419,C)
FT                   /*tag=  d
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(3191,A)
FT                   /*tag=  e
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(4195,T)
```

```
FT                   /*tag= f
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(5910,C)
FT                   /*tag= g
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(7119,T)
FT                   /*tag= h
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(7529,G)
FT                   /*tag= h
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(8855,C)
FT                   /*tag= i
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(8956,A)
FT                   /*tag= j
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(10220,G)
FT                   /*tag= k
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(11628,C)
FT                   /*tag= l
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(11670,A)
FT                   /*tag= m
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(11671,G)
FT                   /*tag= n
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(11931,C)
FT                   /*tag= o
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(15984,G)
FT                   /*tag= p
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(17735,A)
FT                   /*tag= q
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(17910,A)
FT                   /*tag= r
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(18130,C)
FT                   /*tag= s
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19293. .19294,TWA)
FT                   /*tag= t
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19294. .19295,ATT)
FT                   /*tag= u
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19296. .19297,CYT)
FT                   /*tag= v
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19303. .19304,TYC)
FT                   /*tag= w
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19304. .19305,CYC)
FT                   /*tag= x
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19307. .19308,TMC)
FT                   /*tag= y
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19308. .19309,CCC)
FT                   /*tag= z
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19313. .19314,TMA)
FT                   /*tag= aa
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19314. .19315,AWT)
FT                   /*tag= ab
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19315. .19316,TCC)
FT                   /*tag= ac
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19316. .19317,CYT)
FT                   /*tag= ad
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19317. .19318,TWA)
FT                   /*tag= ae
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19318. .19319,ATT)
FT                   /*tag= af
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(19860,G)
FT                   /*tag= ag
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(20843,G)
FT                   /*tag= ah
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(21144,T)
FT                   /*tag= ai
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(21371,G)
FT                   /*tag= aj
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(22061. .22062,A)
FT                   /*tag= ak
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(22321,T)
FT                   /*tag= al
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(22470. .22471,A)
FT                   /*tag= am
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(22472. .22473,A)
FT                   /*tag= an
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(24990,C)
FT                   /*tag= ao
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(25284,A)
FT                   /*tag= ap
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(25934,G)
FT                   /*tag= aq
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(27509. .27510,TAT)
FT                   /*tag= ar
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(27513. .27514,TTT)
FT                   /*tag= as
FT                   /standard_name= "Single nucleotide polymorphism"
FT   exon            27734. .27882
FT                   /*tag= at
FT                   /number= 2
FT   intron          27883. .41045
FT                   /*tag= au
FT                   /number= 2
FT   variation       replace(28608,G)
FT                   /*tag= av
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(28627,C)
FT                   /*tag= aw
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(29709,T)
FT                   /*tag= ax
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(29788,T)
FT                   /*tag= ay
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(29878,T)
FT                   /*tag= az
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(30294,T)
FT                   /*tag= ba
FT                   /standard_name= "Single nucleotide polymorphism"
```

```
FT   variation       replace(30896. .30897,T)
FT                   /*tag= bb
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(31815,G)
FT                   /*tag= bc
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(31999,TTT)
FT                   /*tag= bd
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(32026,T)
FT                   /*tag= be
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(32146,A)
FT                   /*tag= bf
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(32184,S)
FT                   /*tag= bg
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(32214. .32215,T)
FT                   /*tag= bh
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(32215,C)
FT                   /*tag= bi
```

Query Match 5.5%; Score 252.6; DB 6; Length 82938;
Best Local Similarity 70.7%; Pred. No. 5.6e-34;
Matches 350; Conservative 0; Mismatches 144; Indels 1; Gaps 1;

```
Qy       227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
Db      7454 TATATACACATATGTATACATATATACATATATATGTATATACGTATATAGCTGTATATA 7395

Qy       287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
Db      7394 TATACACATGTATATAGCTGTATATATACATACATGTATATACATATACATATATACATA 7335

Qy       347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
Db      7334 TACACATATACATACATGTATATACATATACATATATACATACATGTATATACATATATA 7275

Qy       407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
Db      7274 CATATGTATACATGTGTATCTATATACATATATACATATGTATACATGTGTATCTATATA 7215

Qy       467 TACATATATACACATATATAT-AATATACACACATATATATACACATATATACACACATA 525
Db      7214 CATATATACATATGTATACATGTGTATCTATATACATATATACATATGTATACATGTGTA 7155

Qy       526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
Db      7154 TCTATATACATATATACATATGTATACATGTGTATCTATATACATATATACATATGTATA 7095

Qy       586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
Db      7094 CATGTGTATCTATATACATATATACATATGTATACATGTGTATCTATATACATATATACA 7035

Qy       646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
Db      7034 TATGTATACATGTGTATCTATATACATATATACATATGTATACATGTGTGTATGTATACA 6975

Qy       706 TATGTATATATATAT 720
Db      6974 TATGTGTGTGTATAT 6960
```

```
RESULT 57
ADC87564/c
ID   ADC87564 standard; DNA; 1114 BP.
XX
AC   ADC87564;
XX
DT   01-JAN-2004  (first entry)
XX
DE   Human GPCR gene SEQ ID NO:2017.
XX
KW   ds; gene; human; GPCR;
KW   guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1270724-A2.
XX
PD   02-JAN-2003.
XX
PF   18-JUN-2002; 2002EP-00013517.
XX
PR   18-JUN-2001; 2001JP-00246789.
XX
PA   (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA   (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI   Suwa M,  Asai K,  Akiyama Y,  Aburatani H;
XX
DR   WPI; 2003-315783/31.
DR   P-PSDB; ADC87565.
XX
PT   New polynucleotide, useful for preparing a composition for treating a
PT   patient in need of increased or suppressed activity or expression of the
PT   guanosine triphosphate-binding protein coupled receptor.
XX
PS   Claim 1; SEQ ID NO 2017; 28pp; English.
XX
CC   The invention relates to a novel polynucleotide encoding a guanosine
CC   triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC   the invention may have a use in gene therapy. The polynucleotide and
CC   polypeptide are useful for preparing a composition for treating a patient
CC   in need of increased or suppressed activity or expression of the
CC   guanosine triphosphate-binding protein coupled receptor. The
CC   polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC   invention.
XX
SQ   Sequence 1114 BP; 398 A; 107 C; 167 G; 442 T; 0 U; 0 Other;
```

Query Match 5.4%; Score 251; DB 10; Length 1114;
Best Local Similarity 63.7%; Pred. No. 8.7e-34;
Matches 397; Conservative 0; Mismatches 225; Indels 1; Gaps 1;

```
Qy       230 ATATATATATATATATATATATACACACACACACATATATATATACACATATATATAC 289
Db       864 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 805

Qy       290 ACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATAT 349
Db       804 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 745

Qy       350 ATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATAC 409
Db       744 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 685

Qy       410 ACACATATATACACATATATACACATATATACACATATACACATATATACACATATATAC 469
Db       684 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 625

Qy       470 ATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATATAT 528
Db       624 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 565

Qy       529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACATATAT 588
Db       564 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 505

Qy       589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
Db       504 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 445

Qy       649 ACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATAT 708
```

```
                | | |||||||   | | ||   |||| | ||| | | | || ||||||  | ||||||
Db        444 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 385

Qy        709 GTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAG 768
              || | |||||||        | |||         || |   ||    | | ||| |
Db        384 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 325

Qy        769 GAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAG 828
              | | | |  | | |    |||||    | |  | || |     |  | | | |||
Db        324 ATACACATATATGTATATATTTGTATATACATACACATATATGTATATATTTGTATATAC 265

Qy        829 TAGGGGAAATACATAAATGTGTA 851
                  | ||||||| |||| ||
Db        264 ATACACATATACATATATGTATA 242


RESULT 58
ADE95974
ID   ADE95974 standard; DNA; 96594 BP.
XX
AC   ADE95974;
XX
DT   12-FEB-2004  (first entry)
XX
DE   Human SYK gene genomic DNA sequence.
XX
KW   cancer diagnosis; cancer treatment; carcinoma; cytostatic; gene therapy;
KW   lymphoma; breast cancer; prostate cancer; leukaemia; ds; human; SYK.
XX
OS   Homo sapiens.
XX
PN   WO2003039484-A2.
XX
PD   15-MAY-2003.
XX
PF   08-NOV-2002; 2002WO-US036071.
XX
PR   08-NOV-2001; 2001US-00052482.
XX
PA   (SAGR-) SAGRES DISCOVERY.
XX
PI   Morris DW,  Engelhard EK;
XX
DR   WPI; 2003-441462/41.
XX
PT   New carcinoma associated nucleic acids and proteins, useful for screening
PT   drug candidates, or for diagnosing and treating carcinomas, e.g.
PT   lymphoma, breast cancer, prostate cancer or leukemia.
XX
PS   Claim 1; SEQ ID NO 232; 793pp; English.
XX
CC   This invention relates to novel recombinant nucleic acids for use in
CC   diagnosis and treatment of cancer, especially carcinomas, as well as the
CC   use of compositions in screening methods. The compositions of the
CC   invention may have cytostatic activity whilst the disclosed sequences may
CC   be useful for gene therapy. The carcinoma associated nucleic acids and
CC   proteins are useful for diagnosing and treating carcinomas, for example
CC   lymphoma, breast cancer, prostate cancer or leukaemia, or for screening
CC   drug candidates or bioactive agents capable of binding to, or modulating
CC   the activity of, a carcinoma associated protein. The present sequence is
CC   the genomic DNA sequence of the human SYK gene which is a carcinoma
CC   associated gene of the invention.
XX
SQ   Sequence 96594 BP; 27524 A; 20558 C; 21159 G; 26914 T; 0 U; 439 Other;

  Query Match             5.4%;  Score 251;  DB 10;  Length 96594;
  Best Local Similarity   69.5%;  Pred. No. 1.1e-33;
  Matches  354;  Conservative    0;  Mismatches  152;  Indels    3;  Gaps    1;

Qy        229 TATATATATATATATATATATATACACACACACACATATATATATACACATATATATA 288
              || ||| | |||| | |||||||||| ||| | ||||||||| | | |||| ||| ||| |
Db      90981 TACATACACATATGTGTATATATACATACATATACACATATGTGTGTACATATACATACA 91040

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              || | | |||  | | |||| ||| ||||||| | ||| ||| ||   ||  |  |||||
Db      91041 CATACACATATGTGTGTACATATACATACACACACATACACATATGTGTACATGTACATA 91100

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              ||| ||| | |  |||    | || |||| | ||| ||| ||| ||||    | ||| ||
Db      91101 CATACACATACACATATGTGTACATGTACATACATACACATACACATATGTGTACATGTA 91160

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              || ||||| | | ||| ||||    | || |||| ||||||||||||| ||| || | ||
Db      91161 CATACATACACATACACATATGTGTACATGTACATATATACACATATACACATATGTGTA 91220

Qy        469 CATATATACACATATATATAATATACACACATATATATACACATATATACACACATATAT 528
              ||| || | | ||| |    |||||||||| || | || |    | ||||| ||||||||||
Db      91221 CATGTACATATATACAC---ATATACACATATGTGTACATGTACATATATACACATATAC 91277

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACATATAT 588
              ||| || | ||||  || ||||||||||||||| ||| || | ||||| || | | | | ||
Db      91278 ACATATGTGTACATGTACATATACACATATACACATATGTGTACATGTACATATACACAT 91337

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
              ||||| ||||    | ||  ||||||| |||  | || |   | ||| |||||||| |
Db      91338 ATACACATATGTGTACATGCACATATACACATGTGTACATGCACATATACACATATGTGT 91397

Qy        649 ACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATAT 708
              ||||||| ||| ||||| ||    | || ||| ||||||||| |||| |  |||||||||
Db      91398 ACATACACATATACACATATGTGTGTATATACATATACACACATATGTGTGTATATACAT 91457

Qy        709 GTATATATATATGCTCCAGAGTTCATAAG 737
              |||| |||||| |       | |||| |
Db      91458 ATATACATATATACATATATATACATATG 91486


RESULT 59
ADA02726
ID   ADA02726 standard; DNA; 96595 BP.
XX
AC   ADA02726;
XX
DT   06-NOV-2003  (first entry)
XX
DE   Human SYK carcinoma associated gene, SEQ ID NO:1244.
XX
KW   Human; carcinoma associated; oncogene; carcinoma; cancer; breast;
KW   prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO2003057146-A2.
XX
PD   17-JUL-2003.
XX
PF   26-DEC-2002; 2002WO-US041414.
XX
PR   26-DEC-2001; 2001US-00035832.
XX
PA   (SAGR-) SAGRES DISCOVERY.
XX
PI   Morris DW;
XX
DR   WPI; 2003-587068/55.
XX
PT   New recombinant nucleic acid encoding carcinoma associated protein,
PT   useful for preparing compositions for treating carcinomas.
XX
PS   Claim 1; SEQ ID NO 1244; 245pp; English.
XX
CC   The invention relates to recombinant carcinoma associated (CA) nucleic
CC   acid sequences from mouse and human (ADA01482-ADA03094), and to
CC   recombinant carcinoma associated proteins (CAP) encoded by them. The
CC   invention also encompasses expression vectors and host cells comprising a
CC   CA nucleic acid, a polypeptide (especially an antibody) that specifically
```

```
CC   binds to the protein, and a biochip comprising CA nucleic acid or
CC   fragments thereof. The sequences of the invention were identified using
CC   oncogenic retroviruses, which insert into the genome of the host organism
CC   at random. Many of these do not carry transduced host oncogenes or
CC   pathogenic trans-acting viral genes, meaning that cancer incidence is a
CC   direct consequence of the effects of proviral integration into host
CC   protooncogenes. The CA nucleic acid sequences can be used to diagnose
CC   carcinoma (especially breast cancer, prostate cancer, lymphoma or
CC   leukaemia) or a propensity to carcinoma by determination of the sequence
CC   of a CA gene, or by determination of CA gene expression in particular
CC   tissues. CA nucleic acids, proteins and antibodies are also useful as
CC   therapeutic agents and in screening and evaluating drug candidates. The
CC   present sequence represents a specifically claimed human CA nucleic acid
CC   sequence of the invention. Note: The complete sequence data for this
CC   patent did not form part of the printed specification, but was obtained
CC   in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 96595 BP; 27524 A; 20559 C; 21158 G; 26915 T; 0 U; 439 Other;

  Query Match             5.4%;  Score 251;  DB 9;  Length 96595;
  Best Local Similarity   69.5%;  Pred. No. 1.1e-33;
  Matches  354;  Conservative    0;  Mismatches  152;  Indels    3;  Gaps    1;

Qy        229 TATATATATATATATATATATATACACACACACACACATATATATATACACATATATATA 288
              || ||| | |||| | |||||||||| ||| | |||||||| | | |||| ||| ||| |
Db      90982 TACATACACATATGTGTATATATACATACATATACACATATGTGTGTACATATACATACA 91041

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              || | | |||  | | |||| ||| ||||||| | ||| ||| ||   ||  |  |||||
Db      91042 CATACACATATGTGTGTACATATACATACACACACATACACATATGTGTACATGTACATA 91101

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              ||| ||| | |  |||     | || |||| | ||| ||| ||| ||||     | || ||
Db      91102 CATACACATACACATATGTGTACATGTACATACATACACATACACATATGTGTACATGTA 91161

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              || |||||  | | ||| ||||    | || |||| |||||||||||||| ||| || | ||
Db      91162 CATACATACACATACACATATGTGTACATGTACATATATACACATATACACATATGTGTA 91221

Qy        469 CATATATACACATATATATAATATACACACATATATATACACATATATACACACATATAT 528
              ||| || | | ||| |    |||||||||| || | || |    | ||||| |||||||||
Db      91222 CATGTACATATATACAC---ATATACACATATGTGTACATGTACATATATACACATATAC 91278

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACATATAT 588
              ||| || | ||||  || ||||||||||||||| ||| || | ||||| || | | | ||
Db      91279 ACATATGTGTACATGTACATATACACATATACACATATGTGTACATGTACATATACACAT 91338

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
              ||||| ||||    | ||  ||||||| |||  | || |   | ||| |||||||| | |
Db      91339 ATACACATATGTGTACATGCACATATACACATGTGTACATGCACATATACACATATGTGT 91398

Qy        649 ACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATAT 708
              ||||||| ||| ||||| ||    | || ||| |||||||| |||| |  ||||||| ||
Db      91399 ACATACACATATACACATATGTGTGTATATACATATACACACATATGTGTGTATATACAT 91458

Qy        709 GTATATATATATGCTCCAGAGTTCATAAG 737
              |||| |||||| |       | |||| |
Db      91459 ATATACATATATACATATATATACATATG 91487


RESULT 60
ADB72464
ID   ADB72464 standard; DNA; 96595 BP.
XX
AC   ADB72464;
XX
DT   04-DEC-2003  (first entry)
XX
DE   Human SYK gene.
XX
KW   human; ds; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;
KW   cancer; neoplasm; adenocarcinoma; sarcoma; gene.
XX
OS   Homo sapiens.
XX
PN   WO2003008583-A2.
XX
PD   30-JAN-2003.
XX
PF   26-DEC-2001; 2001WO-US051291.
XX
PR   02-MAR-2001; 2001US-00798586.
PR   23-OCT-2001; 2001US-00004113.
PR   08-NOV-2001; 2001US-00052482.
PR   30-NOV-2001; 2001US-00997722.
PR   20-DEC-2001; 2001US-00034650.
XX
PA   (SAGR-) SAGRES DISCOVERY.
XX
PI   Morris DW,  Engelhard EK;
XX
DR   WPI; 2003-239337/23.
XX
PT   New recombinant nucleic acid, useful for treating carcinomas, lymphomas,
PT   cancers, neoplasm, adenocarcinoma, or sarcomas.
XX
PS   Claim 1; SEQ ID NO 292; 2304pp; English.
XX
CC   The invention relates to a novel recombinant nucleic acid comprising a
CC   nucleotide sequence selected from any of the 660 sequences fully defined
CC   in the specification. A polynucleotide of the invention has cytostatic
CC   activity, and may have a use in gene therapy, or in a vaccine. The
CC   recombinant nucleic acids and polypeptides are useful for treating
CC   carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and
CC   sarcomas. The present sequence represents a human gene of the invention.
XX
SQ   Sequence 96595 BP; 27524 A; 20559 C; 21158 G; 26915 T; 0 U; 439 Other;

  Query Match             5.4%;  Score 251;  DB 10;  Length 96595;
  Best Local Similarity   69.5%;  Pred. No. 1.1e-33;
  Matches  354;  Conservative    0;  Mismatches  152;  Indels    3;  Gaps    1;

Qy        229 TATATATATATATATATATATATACACACACACACACATATATATATACACATATATATA 288
              || ||| | |||| | |||||||||| ||| | |||||||| | | |||| ||| ||| |
Db      90982 TACATACACATATGTGTATATATACATACATATACACATATGTGTGTACATATACATACA 91041

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              || | | |||  | | |||| ||| ||||||| | ||| ||| ||   ||  |  |||||
Db      91042 CATACACATATGTGTGTACATATACATACACACACATACACATATGTGTACATGTACATA 91101

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              ||| ||| | |  |||     | || |||| |'||| ||| ||| ||||     | || ||
Db      91102 CATACACATACACATATGTGTACATGTACATACATACACATACACATATGTGTACATGTA 91161

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              || |||||  | | ||| ||||    | || |||| |||||||||||||| ||| || | ||
Db      91162 CATACATACACATACACATATGTGTACATGTACATATATACACATATACACATATGTGTA 91221

Qy        469 CATATATACACATATATATAATATACACACATATATATACACATATATACACACATATAT 528
              ||| || | | ||| |    |||||||||| || | || |    | ||||| |||||||||
Db      91222 CATGTACATATATACAC---ATATACACATATGTGTACATGTACATATATACACATATAC 91278

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACATATAT 588
              ||| || | ||||  || ||||||||||||||| ||| || | ||||| || | | | ||
Db      91279 ACATATGTGTACATGTACATATACACATATACACATATGTGTACATGTACATATACACAT 91338

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
              ||||| ||||    | ||  ||||||| |||  | || |   | ||| |||||||| | |
Db      91339 ATACACATATGTGTACATGCACATATACACATGTGTACATGCACATATACACATATGTGT 91398

Qy        649 ACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATAT 708
              ||||||| ||| ||||| ||    | || ||| |||||||| |||| |  ||||||| ||
Db      91399 ACATACACATATACACATATGTGTGTATATACATATACACACATATGTGTGTATATACAT 91458

Qy        709 GTATATATATATGCTCCAGAGTTCATAAG 737
```

```
Db      91459 ATATACATATATACATATATATACATATG 91487


RESULT 61
ADQ97712/c
ID   ADQ97712 standard; DNA; 103661 BP.
XX
AC   ADQ97712;
XX
DT   07-OCT-2004  (first entry)
XX
DE   Mouse cancer associated sequence MD10-032, SEQ ID 689.
XX
KW   Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Mouse; ds.
XX
OS   Mus musculus.
XX
PN   WO2004060304-A2.
XX
PD   22-JUL-2004.
XX
PF   22-DEC-2003; 2003WO-US041389.
XX
PR   27-DEC-2002; 2002US-00330773.
XX
PA   (SAGR-) SAGRES DISCOVERY INC.
XX
PI   Morris DW,  Malandro MS;
XX
DR   WPI; 2004-543781/52.
XX
PT   New isolated cancer associated nucleic acids comprising at least 10
PT   contiguous nucleotides, useful for diagnosing, preventing and/or treating
PT   cancers such as leukemia and lymphoma.
XX
PS   Claim 1; SEQ ID NO 689; 199pp; English.
XX
CC   The present invention relates to cancer associated sequences (ADQ97025-
CC   ADQ98004). The sequences are useful for the diagnosis, prevention and/or
CC   treatment of cancer, such as leukemia and lymphoma. Note: The sequence
CC   data for this patent did not form part of the printed specification, but
CC   was obtained in electronic formate directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 103661 BP; 25663 A; 25358 C; 25585 G; 26717 T; 0 U; 338 Other;

  Query Match             5.4%;  Score 251;  DB 12;  Length 103661;
  Best Local Similarity   72.6%;  Pred. No. 1.1e-33;
  Matches  339;  Conservative    0;  Mismatches  125;  Indels    3;  Gaps    1;

Qy        230 ATATATATATATATATATATATACACACACACACATATATATATACACATATATATAC 289
              | | | | | | | | | | ||| | |||||||||| | | | | ||||| | | ||||
Db      73981 ACACACACACACACACACACACACATATACACACACACACACACACACACACACATAC 73922

Qy        290 ACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATAT 349
              ||| | ||| | | | ||||| | | ||||||| | | ||||| ||| |    ||||| ||
Db      73921 ACACACATATACACACACACACACACACATACACACACACACACACACACACACACAT 73862

Qy        350 ATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATAC 409
              ||| ||||| |   | ||||| | | ||| ||| | ||||||| | | ||||| ||||||
Db      73861 ATACACACACACACACACACACACACACACATATACACACACACACACACACACATATAC 73802

Qy        410 ACACATATATACACATATATACACATATATACACATATACACATATATACACATATATAC 469
              ||||| | | ||||| | | ||||| |||||||||| | ||||| | | ||||||| | ||
Db      73801 ACACACACACACACACACACACACACACATATACACACACACACACACACACATACACAC 73742

Qy        470 ATATATACACATATATATA---ATATACACACATATATATACACATATATACACACATAT 526
              | | | ||||| | |||||    | | ||||||| | | | ||||||||| ||||||| |
Db      73741 ACACACACACACACATATACACACACACACACACACACACACACATATACACACACAC 73682

Qy        527 ATACACATATATACACATATATATACACATATATACACATATATACATATATACACATAT 586
              | ||||| ||| ||||| ||| | ||||| | | ||||||| | ||| | | ||||| |
Db      73681 ACACACACATACACACACATACACACACACACACACACATACACACACACACACACAC 73622

Qy        587 ATATACATATATACACATATATACATATATACACATATATACATATATACACACATATAT 646
              ||| ||| | | ||||||| | ||| ||||||||| | | ||| | | ||||||| ||||
Db      73621 ATACACACACACACACATACACACACATATACACACACACACACACACACACACATAT 73562

Qy        647 ACACATACATATACACACACATAGATATACATATATATACACATATA 693
              ||||| ||| | ||| ||||| | ||||||| | | | ||||| | |
Db      73561 ACACACACACACACATACACACACATATACACACACACACACACA 73515


RESULT 62
ABD32998/c
ID   ABD32998 standard; DNA; 113079 BP.
XX
AC   ABD32998;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Human cancer-associated genomic DNA HD21-021.
XX
KW   Human; ds; cancer-associated protein; gene; cytostatic; cancer;
KW   leukaemia; lymphoma; CAP.
XX
OS   Homo sapiens.
XX
PN   WO2004074320-A2.
XX
PD   02-SEP-2004.
XX
PF   17-FEB-2004; 2004WO-US004730.
XX
PR   14-FEB-2003; 2003US-00367094.
PR   14-MAR-2003; 2003US-00388838.
PR   15-APR-2003; 2003US-00417375.
PR   13-JUN-2003; 2003US-00461862.
PR   15-SEP-2003; 2003US-00663431.
PR   15-DEC-2003; 2003US-00737318.
XX
PA   (SAGR-) SAGRES DISCOVERY INC.
XX
PI   Morris DW,  Morris DW,  Malandro MS;
XX
DR   WPI; 2004-652914/63.
XX
PT   New isolated cancer-associated polynucleotides and polypeptides useful
PT   for diagnosing, preventing or treating cancers, especially lymphoma and
PT   leukemia, or in screening for agents that modulate cancer.
XX
PS   claim 16; seqid 764; 310pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising at least 10
CC   contiguous nucleotides of any of the 233 polynucleotide sequences given
CC   in the specification, or its complement. The nucleic acids encode cancer-
CC   associated proteins. Also included are an expression vector comprising
CC   the isolated nucleic acid cited above, a host cell comprising the above
CC   recombinant nucleic acid or expression vector, a microarray for detecting
CC   a cancer-associated (CA) nucleic acid comprising at least one probe
CC   comprising at least 10 contiguous nucleotides of any of the above-
CC   mentioned nucleotide sequences, an isolated polypeptide (encoded within
CC   an open reading frame of a CA sequence selected from any of the 95
CC   polynucleotide sequences as mentioned in the specification, or its
CC   complement), an isolated antibody, (or its antigen binding fragment) that
CC   binds to the above polypeptide, a hybridoma that produces the above
CC   monoclonal antibody, a pharmaceutical composition comprising the above
CC   antibody and a pharmaceutical excipient, a kit for detecting cancer
CC   cells(comprising the antibody cited above, methods for diagnosing cancer
CC   or for detecting the presence or absence of cancer cells in an
CC   individual, a method for inhibiting growth of cancer cells in an
CC   individual, a method for delivering a therapeutic agent to cancer cells
CC   in an individual, an electronic library comprising the above
CC   polynucleotide or polypeptide (or their fragments), methods of screening
CC   for anticancer activity or for a bioactive agent capable of modulating
```

```
CC   the activity of a CA protein (CAP), methods for detecting cancer
CC   associated with expression of a polypeptide in a test cell sample, a
CC   method for treating cancers and a method for inhibiting the expression of
CC   CA gene in a cell. The composition and methods are useful for detecting,
CC   diagnosing, preventing and treating cancers, especially lymphoma and
CC   leukaemia. These may also be used in screening for agents that modulate
CC   cancer. The present sequence is a human CAP genomic sequence. Note: The
CC   sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 113079 BP; 27712 A; 27684 C; 28156 G; 29527 T; 0 U; 0 Other;

  Query Match             5.4%;  Score 250.6;  DB 13;  Length 113079;
  Best Local Similarity   73.3%;  Pred. No. 1.2e-33;
  Matches  335;  Conservative    0;  Mismatches  119;  Indels    3;  Gaps    1;

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              ||||||||||||||| ||| | | | ||||| | ||| | ||||||||| |||||||||
Db      31278 ATATATATATATACAGACATATATATATATAGACATATATATATATATAGACATATATAT 31219

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACAT 359
              ||||| ||| |||||||  | | | ||||| | ||| | |   || | ||| | | ||| ||
Db      31218 ATATAGACATATATATATAGAGACATATATATATATAGACATATATATAGAGAGACATAT 31159

Qy        360 ATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATAT 419
              || |||| ||||||||| | | | | | | | ||| | ||| ||||||| | |||||||||
Db      31158 ATATATAGACATATATATAGAGAGACATATATATAGAGACATATATATAGAGACATATAT 31099

Qy        420 ACACATATATACACATATATACACATATACACATATATACACATATATACATATATACAC 479
              | | | | ||| | ||| | | | ||||| ||||||||| | | ||||| ||||| |||
Db      31098 ATAGAGACATATATATAGACATATATATAGACATATATATAGACATATATATATAGACAT 31039

Qy        480 ATATATATA---ATATACACACATATATATACACATATATACACACATATATACACATAT 536
              ||||| | |   ||||| ||| ||||||| ||| |||||| || ||||||||| | | ||
Db      31038 ATATAGACATATATATAGACATATATATAGACATATATATGCAGACATATATATAGACAT 30979

Qy        537 ATACACATATATATACACATATATACACATATATACATATATACACATATATATACATAT 596
              ||| | ||| | ||| | ||  | ||| ||||||| | ||||| | | ||||||| | |
Db      30978 ATATATATAGACATATATATGCAGACATATATATAGACATATATAGACATATATATAGAC 30919

Qy        597 ATACACATATATACATATATACACATATATACATATATACACACATATATACACATACAT 656
              ||| | ||||| ||||||||| | | ||||| ||||| ||| | ||||||| | ||| ||
Db      30918 ATATATATATAGACATATATATAGACATATATATATAAACATATATATATATATATAT 30859

Qy        657 ATACACACACATAGATATACATATATATACACATATA 693
              ||| | | | | ||| | | | | | | | | | | |
Db      30858 ATATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA 30822


RESULT 63
ABL33999/c
ID   ABL33999 standard; DNA; 21537 BP.
XX
AC   ABL33999;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 1972.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP007537.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful for
PT   diagnosis and treatment of diseases associated with abnormal cytosine
PT   methylation.
XX
PS   Claim 1; SEQ ID NO 1972; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention
XX
SQ   Sequence 21537 BP; 6538 A; 333 C; 4426 G; 10240 T; 0 U; 0 Other;

  Query Match             5.4%;  Score 250;  DB 6;  Length 21537;
  Best Local Similarity   68.1%;  Pred. No. 1.5e-33;
  Matches  391;  Conservative    0;  Mismatches  180;  Indels    3;  Gaps    3;

Qy        181 CCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATA 240
              |||    | || |    || |  |||||      |  | |        | ||||||||||||
Db       6005 CCAAATTAAAATACAATAACGCGATCTCGATTCACTACACTATATATATATATATATATA 5946

Qy        241 TATATATATATACACACACACACATATATATATACACATATATATACACATATATACA 300
              |||||||||||| ||||| |||| |||||||||||| | |||||||||| | |||||||
Db       5945 TATATATATATATACACATACAC-CATATATATATATATATATATACATACATATACA 5887

Qy        301 TATATACACATATATACACATATATATACACATACATATGTACACATATATATACACATA 360
              |  |    |  || | | | || | | |   |  | |||| | | ||| ||
Db       5886 AAACAATTACTTACTAAATACACAAATCAATAAAATAAAAAATTCATAAACAAACATATC 5827

Qy        361 TGTATACACATATATACACATATATACACACATATATACACATATATACACACATATATA 420
              | | |  ||||||||| |  |||||  |      |||| | ||||||| | | |||||||
Db       5826 TCTCTTTACATATATATATCTATATCGATCACACTATATATATATATATATATATATA 5767

Qy        421 CACATATATACACATATATACACATATACACATATATACACATATATACATATATACACA 480
              |||| ||||| | ||||||| | ||| |||||||||||| | ||||||| | | | | | |
Db       5766 CACACATATATATATATATATATACACACATATATATATATATATATACACACATATA 5707

Qy        481 TATATATAATATACACACATATATATACACATATATACACACATATATACACATATATAC 540
              ||||||| | ||||||||||||||||| | ||||||| ||||| ||||| | |||||||
Db       5706 TATATAT-ACATACACACATATATATATATATATATACACACATATATATATATATAT 5648

Qy        541 ACATATATATACACATATATACACATATATACATATATACACATATATATACATATATAC 600
              | ||||||||| | ||||||| | ||||||| ||||||| | ||||| |  ||||||||
Db       5647 ATATATATATATATATATATATATATATATATATATATATATATATACACGCATATATAT 5588

Qy        601 ACATATATACATATATACACATATATACATATATACACACATATATACACATACATATAC 660
              || |||||||||||||| | |    ||| |||||||| | |||||||||| | ||| |||||
Db       5587 ACGTATATACATATATATAAACTCATATATATATATATACATATATATATATATATAT 5528

Qy        661 ACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATAT 720
              | | | ||| |   ||| |||||| |  | | | ||      || | | ||||  |  ||
Db       5527 ATATATATAAACTCACA-ATATATCCTTAAACAAATTATATAATCTCTATATACCTCAAT 5469

Qy        721 GCTCCAGAGTTCATAAGAGGTAGCAGTTGATTAC 754
                ||   | |  | || |   | | |  | |||
Db       5468 TTTCTCAAATATAAAACAAAAATAATTATAATAC 5435
```

```
RESULT 64
ADA02930
ID   ADA02930 standard; DNA; 96596 BP.
XX
AC   ADA02930;
XX
DT   06-NOV-2003  (first entry)
XX
DE   Human PIK3R1 carcinoma associated gene, SEQ ID NO:1448.
XX
KW   Human; carcinoma associated; oncogene; carcinoma; cancer; breast;
KW   prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO2003057146-A2.
XX
PD   17-JUL-2003.
XX
PF   26-DEC-2002; 2002WO-US041414.
XX
PR   26-DEC-2001; 2001US-00035832.
XX
PA   (SAGR-) SAGRES DISCOVERY.
XX
PI   Morris DW;
XX
DR   WPI; 2003-587068/55.
XX
PT   New recombinant nucleic acid encoding carcinoma associated protein,
PT   useful for preparing compositions for treating carcinomas.
XX
PS   Claim 1; SEQ ID NO 1448; 245pp; English.
XX
CC   The invention relates to recombinant carcinoma associated (CA) nucleic
CC   acid sequences from mouse and human (ADA01482-ADA03094), and to
CC   recombinant carcinoma associated proteins (CAP) encoded by them. The
CC   invention also encompasses expression vectors and host cells comprising a
CC   CA nucleic acid, a polypeptide (especially an antibody) that specifically
CC   binds to the protein, and a biochip comprising CA nucleic acid or
CC   fragments thereof. The sequences of the invention were identified using
CC   oncogenic retroviruses, which insert into the genome of the host organism
CC   at random. Many of these do not carry transduced host oncogenes or
CC   pathogenic trans-acting viral genes, meaning that cancer incidence is a
CC   direct consequence of the effects of proviral integration into host
CC   protooncogenes. The CA nucleic acid sequences can be used to diagnose
CC   carcinoma (especially breast cancer, prostate cancer, lymphoma or
CC   leukaemia) or a propensity to carcinoma by determination of the sequence
CC   of a CA gene, or by determination of CA gene expression in particular
CC   tissues. CA nucleic acids, proteins and antibodies are also useful as
CC   therapeutic agents and in screening and evaluating drug candidates. The
CC   present sequence represents a specifically claimed human CA nucleic acid
CC   sequence of the invention. Note: The complete sequence data for this
CC   patent did not form part of the printed specification, but was obtained
CC   in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 96596 BP; 27639 A; 18627 C; 19687 G; 30643 T; 0 U; 0 Other;

  Query Match             5.4%;  Score 249.6;  DB 9;  Length 96596;
  Best Local Similarity   70.5%;  Pred. No. 1.8e-33;
  Matches  347;  Conservative    0;  Mismatches  144;  Indels    1;  Gaps    1;

Qy        229 TATATATATATATATATATATATACACACACACACATATATATATACACATATATATA 288
              | | || ||||||| || || || ||   | | |||  | || ||||| |||| || ||
Db      57110 TGTGTACATATATACATGTAGATGCATGTAGATACATGTGTACATATATACATGTAGATG 57169

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              ||  || |||||| ||||||  || |||||  || ||| |   ||||||  |  | ||||
Db      57170 CATGTAGATACATGTATACATGTAGATACATGTAGATACATGTATACATGTAGATACATG 57229

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              || |||||  |||  ||| | | || || | | || || | |||| |||||| ||||| |
Db      57230 TAGATACATGTATACATATATACATGTAGATACATGTAGATACATGTATACATATATACA 57289

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              | ||| ||  | | ||||| |||| || | |||| || ||||||||||  || |||
Db      57290 TGTAGATACATGTATACATATATACATGTAGATACATGTATACATATATACATGTAGATA 57349

Qy        469 CATATATACACATATATATAATATACACACATATATATACACATATATACACACATATAT 528
              ||| |||||| ||||| ||  || | |||  || ||| |   ||| ||| | |||| ||
Db      57350 CATGTATACATATATACAT-GTAGATACATGTAGATACATGTATACATATATACATGTAG 57408

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACATATAT 588
              | |||| || | |||| |||| | | ||| ||  | ||| ||  ||||| | ||||| ||
Db      57409 ATACATGTAGATACATGTATACATATATACATGTAGATACATGTATATAGATACATACAT 57468

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
              | | ||||| | ||||| ||||||||||  | |||| |||||||||  ||| ||||||
Db      57469 AGATATATAGATACATAGATACATATATTTATATATGTACATATATCTACATATATATCT 57528

Qy        649 ACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATAT 708
              | ||| | ||| |   ||||| |||| |||||||||  |  || |||||  | |||||||
Db      57529 ATATATACATATATCTACATATATATCTATATATATATATCTACATATATATCTATATAT 57588

Qy        709 GTATATATATAT 720
              ||| | ||||||
Db      57589 GTAGAAATATAT 57600


RESULT 65
ADB72668
ID   ADB72668 standard; DNA; 96596 BP.
XX
AC   ADB72668;
XX
DT   04-DEC-2003  (first entry)
XX
DE   Human PIK3R1 gene.
XX
KW   human; ds; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;
KW   cancer; neoplasm; adenocarcinoma; sarcoma; gene.
XX
OS   Homo sapiens.
XX
PN   WO2003008583-A2.
XX
PD   30-JAN-2003.
XX
PF   26-DEC-2001; 2001WO-US051291.
XX
PR   02-MAR-2001; 2001US-00798586.
PR   23-OCT-2001; 2001US-00004113.
PR   08-NOV-2001; 2001US-00052482.
PR   30-NOV-2001; 2001US-00997722.
PR   20-DEC-2001; 2001US-00034650.
XX
PA   (SAGR-) SAGRES DISCOVERY.
XX
PI   Morris DW,  Engelhard EK;
XX
DR   WPI; 2003-239337/23.
XX
PT   New recombinant nucleic acid, useful for treating carcinomas, lymphomas,
PT   cancers, neoplasm, adenocarcinoma, or sarcomas.
XX
PS   Claim 1; SEQ ID NO 496; 2304pp; English.
XX
CC   The invention relates to a novel recombinant nucleic acid comprising a
CC   nucleotide sequence selected from any of the 660 sequences fully defined
CC   in the specification. A polynucleotide of the invention has cytostatic
CC   activity, and may have a use in gene therapy, or in a vaccine. The
CC   recombinant nucleic acids and polypeptides are useful for treating
```

```
CC   carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and
CC   sarcomas. The present sequence represents a human gene of the invention.
XX
SQ   Sequence 96596 BP; 27639 A; 18627 C; 19687 G; 30643 T; 0 U; 0 Other;

  Query Match             5.4%;  Score 249.6;  DB 10;  Length 96596;
  Best Local Similarity   70.5%;  Pred. No. 1.8e-33;
  Matches  347;  Conservative    0;  Mismatches  144;  Indels    1;  Gaps    1;

Qy        229 TATATATATATATATATATATATACACACACACACATATATATATACACATATATATA 288
              | | || ||||||| || || || ||    | | |||  | || ||||| |||| || ||
Db      57110 TGTGTACATATATACATGTAGATGCATGTAGATACATGTGTACATATATACATGTAGATG 57169

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              ||  || ||||||  ||||||  || |||||  || |||  |   ||||||  |  | ||||
Db      57170 CATGTAGATACATGTATACATGTAGATACATGTAGATACATGTATACATGTAGATACATG 57229

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              || |||||  |||  ||| | | || || | | || || | |||| ||||||  ||||| |
Db      57230 TAGATACATGTATACATATATACATGTAGATACATGTAGATACATGTATACATATATACA 57289

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
                | ||| ||  | | ||||| |||| || | |||| || ||||||||||  || |||
Db      57290 TGTAGATACATGTATACATATATACATGTAGATACATGTATACATATATACATGTAGATA 57349

Qy        469 CATATATACACATATATATAATATACACACATATATATACACATATATACACACATATAT 528
              ||| |||||| ||||| ||  || | |||  || |||  |   ||| ||| | |||| ||
Db      57350 CATGTATACATATATACAT-GTAGATACATGTAGATACATGTATACATATATACATGTAG 57408

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACACATATAT 588
              | |||| || | |||| |||| | | ||| ||  | ||| ||  ||||| | ||||| ||
Db      57409 ATACATGTAGATACATGTATACATATATACATGTAGATACATGTATATAGATACATACAT 57468

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
              | | ||||| | ||||| ||||||||||  | |||| ||||||||  ||| ||||||
Db      57469 AGATATATAGATACATAGATACATATATTTATATATGTACATATATCTACATATATATCT 57528

Qy        649 ACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATAT 708
              | ||| | ||| |   ||||| ||||  |||||||||  |  || |||||  | |||||||
Db      57529 ATATATACATATATCTACATATATATCTATATATATATATCTACATATATATCTATATAT 57588

Qy        709 GTATATATATAT 720
              ||| | ||||||
Db      57589 GTAGAAATATAT 57600


RESULT 66
ADC85410
ID   ADC85410 standard; DNA; 96596 BP.
XX
AC   ADC85410;
XX
DT   01-JAN-2004  (first entry)
XX
DE   Human Pik3r1 genomic sequence.
XX
KW   Cytostatic; gene therapy; vaccine; cancer; carcinoma-associated gene; CA;
KW   secreted; transmembrane; intracellular; ds.
XX
OS   Homo sapiens.
XX
PN   WO2003045230-A2.
XX
PD   05-JUN-2003.
XX
PF   02-DEC-2002; 2002WO-US038582.
XX
PR   30-NOV-2001; 2001US-00997722.
XX
PA   (SAGR-) SAGRES DISCOVERY.
XX
PI   Morris DW,  Engelhard EK;
XX
DR   WPI; 2003-513603/48.
XX
PT   New recombinant nucleic acid comprising a nucleotide sequence of any of
PT   the carcinoma-associated (CA) genes, useful for screening for drug
PT   candidates for diagnosing or treating carcinomas.
XX
PS   Claim 1; SEQ ID NO 196; 983pp; English.
XX
CC   The invention relates to a recombinant nucleic acid comprising a
CC   nucleotide sequence selected from any of the fully defined carcinoma-
CC   associated (CA) genes from the 50 tables given in the specification. The
CC   CA proteins are secreted, transmembrane or intracellular proteins. The
CC   recombinant nucleic acids are useful for screening for drug candidates
CC   for diagnosing or treating carcinomas. Sequences given in ADC85215-
CC   ADC85514 represent CA genes of the invention.
XX
SQ   Sequence 96596 BP; 27639 A; 18627 C; 19687 G; 30643 T; 0 U; 0 Other;

  Query Match             5.4%;  Score 249.6;  DB 10;  Length 96596;
  Best Local Similarity   70.5%;  Pred. No. 1.8e-33;
  Matches  347;  Conservative    0;  Mismatches  144;  Indels    1;  Gaps    1;

Qy        229 TATATATATATATATATATATATACACACACACACATATATATATACACATATATATA 288
              | | || ||||||| || || || ||    | | |||  | || ||||| |||| || ||
Db      57110 TGTGTACATATATACATGTAGATGCATGTAGATACATGTGTACATATATACATGTAGATG 57169

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              ||  || ||||||  ||||||  || |||||  || |||  |   ||||||  |  | ||||
Db      57170 CATGTAGATACATGTATACATGTAGATACATGTAGATACATGTATACATGTAGATACATG 57229

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              || |||||  |||  ||| | | || || | | || || | |||| ||||||  ||||| |
Db      57230 TAGATACATGTATACATATATACATGTAGATACATGTAGATACATGTATACATATATACA 57289

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
                | ||| ||  | | ||||| |||| || | |||| || ||||||||||  || |||
Db      57290 TGTAGATACATGTATACATATATACATGTAGATACATGTATACATATATACATGTAGATA 57349

Qy        469 CATATATACACATATATATAATATACACACATATATATACACATATATACACACATATAT 528
              ||| |||||| ||||| ||  || | |||  || |||  |   ||| ||| | |||| ||
Db      57350 CATGTATACATATATACAT-GTAGATACATGTAGATACATGTATACATATATACATGTAG 57408

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACACATATAT 588
              | |||| || | |||| |||| | | ||| ||  | ||| ||  ||||| | ||||| ||
Db      57409 ATACATGTAGATACATGTATACATATATACATGTAGATACATGTATATAGATACATACAT 57468

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
              | | ||||| | ||||| ||||||||||  | |||| ||||||||  ||| ||||||
Db      57469 AGATATATAGATACATAGATACATATATTTATATATGTACATATATCTACATATATATCT 57528

Qy        649 ACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATAT 708
              | ||| | ||| |   ||||| ||||  |||||||||  |  || |||||  | |||||||
Db      57529 ATATATACATATATCTACATATATATCTATATATATATATCTACATATATATCTATATAT 57588

Qy        709 GTATATATATAT 720
              ||| | ||||||
Db      57589 GTAGAAATATAT 57600


RESULT 67
ADM74525
ID   ADM74525 standard; DNA; 96596 BP.
XX
AC   ADM74525;
XX
DT   01-JUL-2004  (first entry)
XX
DE   Human carcinoma associated (CA) nucleic acid #97.
XX
KW   Human; carcinoma associated nucleic acid; CA nucleic acid; gene; ds;
KW   carcinoma associated protein; CAP; carcinoma; leukaemia; lymphoma;
KW   cytostatic.
XX
```

OS Homo sapiens.
XX
PN US2004072154-A1.
XX
PD 15-APR-2004.
XX
PF 30-NOV-2001; 2001US-00997722.
XX
PR 22-DEC-2000; 2000US-00747377.
PR 02-MAR-2001; 2001US-00798586.
XX
PA (MORR/) MORRIS D W.
PA (ENGE/) ENGELHARD E K.
XX
PI Morris DW, Engelhard EK;
XX
DR WPI; 2004-328562/30.
XX
PT New carcinoma associated gene or protein, useful for preparing a
PT composition for diagnosing or treating carcinoma e.g., leukemia or
PT lymphoma.
XX
PS Claim 1; SEQ ID NO 196; 29pp; English.
XX
CC The invention relates to new recombinant nucleic acids. The invention
CC also relates to a host cell comprising a recombinant nucleic acid or
CC expression vector, an expression vector comprising a recombinant nucleic
CC acid, a recombinant protein, a method of screening for drug candidates, a
CC method of screening for a bioactive agent capable of binding to a
CC carcinoma associated protein (CAP) encoded by a nucleotide sequence, a
CC method of screening for a bioactive agent capable of modulating the
CC activity of a CAP, a method of evaluating the effect of a candidate
CC carcinoma drug, a method of diagnosing carcinoma, a method for inhibiting
CC the activity of a CAP, a method of treating carcinomas, a method of
CC neutralising the effect of a CAP and a method of diagnosing carcinoma or
CC propensity to carcinoma. A method of evaluating the effect of a candidate
CC carcinoma drug comprises administering the drug to a patient, removing a
CC cell sample from the patient and determining alterations in the
CC expression or activation of a gene comprising the nucleotide sequence. A
CC method of diagnosing carcinoma comprises determining the expression of
CC one or more genes comprising the nucleic acid sequence in a first tissue
CC type of a first individual and comparing the expression of the gene from
CC a second normal tissue type from the first individual or a second
CC unaffected individual, where a difference in the expression indicates
CC that the first individual has carcinoma. A method of inhibiting the
CC activity of a CAP comprises binding an inhibitor to the CAP. Treating
CC carcinomas comprises administering to a patient an inhibitor of CAP.
CC Neutralising the effect of a CAP comprises contacting an agent specific
CC for the CAP. The polypeptide specifically binds to the protein encoded by
CC the nucleic acid. It comprises an antibody that specifically binds to the
CC protein encoded by the nucleic acid. The nucleic acids are useful for
CC preparing a composition for diagnosing or treating carcinoma e.g.,
CC leukaemia or lymphoma. This sequence represents a human carcinoma
CC associated (CA) nucleic acid of the invention. Note: The sequence data
CC for this patent did not form part of the printed specification but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 96596 BP; 27639 A; 18627 C; 19687 G; 30643 T; 0 U; 0 Other;

Query Match 5.4%; Score 249.6; DB 12; Length 96596;
Best Local Similarity 70.5%; Pred. No. 1.8e-33;
Matches 347; Conservative 0; Mismatches 144; Indels 1; Gaps 1;

```
Qy        229 TATATATATATATATATATATATACACACACACACATATATATATACACATATATATA 288
              | | || ||||||| || || || ||   | | |||  | || ||||| |||| || ||
Db      57110 TGTGTACATATATACATGTAGATGCATGTAGATACATGTGTACATATATACATGTAGATG 57169

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              ||  || ||||||  ||||||  || |||||  || ||| |   ||||||  |  | ||||
Db      57170 CATGTAGATACATGTATACATGTAGATACATGTAGATACATGTATACATGTAGATACATG 57229

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              || |||||  |||  ||| | | || || | | || || | |||| |||||| ||||| |
Db      57230 TAGATACATGTATACATATATACATGTAGATACATGTAGATACATGTATACATATATACA 57289

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              | ||| ||  | | ||||| |||| || | |||| || ||||||||||  || |||
Db      57290 TGTAGATACATGTATACATATATACATGTAGATACATGTATACATATATACATGTAGATA 57349

Qy        469 CATATATACACATATATATAATATACACACATATATATACACATATATACACACATATAT 528
              ||| |||||| ||||| ||  || | |||  || ||| |   ||| ||| | |||| ||
Db      57350 CATGTATACATATATACAT-GTAGATACATGTAGATACATGTATACATATATACATGTAG 57408

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACACATATAT 588
              | |||| || | |||| |||| | | ||| ||  | ||| ||  ||||| | ||||| ||
Db      57409 ATACATGTAGATACATGTATACATATATACATGTAGATACATGTATATAGATACATACAT 57468

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
              | | ||||| | ||||| ||||||||||  | |||| |||||||||  ||| ||||||
Db      57469 AGATATATAGATACATAGATACATATATTTATATATGTACATATATCTACATATATATCT 57528

Qy        649 ACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATAT 708
              | ||| | ||| |   ||||| |||| |||||||||  |  || |||||  | |||||||
Db      57529 ATATATACATATATCTACATATATATCTATATATATATATCTACATATATATCTATATAT 57588

Qy        709 GTATATATATAT 720
              ||| | ||||||
Db      57589 GTAGAAATATAT 57600
```

RESULT 68
ADC87556/c
ID ADC87556 standard; DNA; 1342 BP.
XX
AC ADC87556;
XX
DT 01-JAN-2004 (first entry)
XX
DE Human GPCR gene SEQ ID NO:2009.
XX
KW ds; gene; human; GPCR;
KW guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1270724-A2.
XX
PD 02-JAN-2003.
XX
PF 18-JUN-2002; 2002EP-00013517.
XX
PR 18-JUN-2001; 2001JP-00246789.
XX
PA (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI Suwa M, Asai K, Akiyama Y, Aburatani H;
XX
DR WPI; 2003-315783/31.
DR P-PSDB; ADC87557.
XX
PT New polynucleotide, useful for preparing a composition for treating a
PT patient in need of increased or suppressed activity or expression of the
PT guanosine triphosphate-binding protein coupled receptor.
XX
PS Claim 1; SEQ ID NO 2009; 28pp; English.
XX
CC The invention relates to a novel polynucleotide encoding a guanosine
CC triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC the invention may have a use in gene therapy. The polynucleotide and
CC polypeptide are useful for preparing a composition for treating a patient
CC in need of increased or suppressed activity or expression of the
CC guanosine triphosphate-binding protein coupled receptor. The
CC polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC invention.

```
XX
SQ   Sequence 1342 BP; 397 A; 191 C; 193 G; 436 T; 0 U; 125 Other;

  Query Match             5.4%;  Score 248.4;  DB 10;  Length 1342;
  Best Local Similarity   69.9%;  Pred. No. 2.5e-33;
  Matches  348;  Conservative    0;  Mismatches  143;  Indels    7;  Gaps    2;

Qy        229 TATATATATATATATATATATATACACACACACACACATATATATATACACATATATATA 288
              || | | | |||| | |||||||||| | ||||||  | | | ||||| |  | ||||||
Db        891 TACACACACATATGTGTATATATACATATACACACCTACACACATATATATGTGTATATA 832

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              ||||||| ||| | ||| |  | |||| | | ||||| | |||||  | | |||||||
Db        831 TACATATACACACACATATATGTGTATATATACATATACATACATATGTGTATACACATG 772

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              |||||   || || | || || ||||| |||||||| | ||| |||| || |||||||||
Db        771 TATATGTGCACATATGTATACGTATATGCACATATACATACATATATGTATACATATATA 712

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              |||| |||| | ||  ||| || || ||| | |||||||  |||| | ||  |||| |||
Db        711 CACATATATGTGCATGTATGTATACGTATGTGCACATATGTACATGTGTATGCATAAATA 652

Qy        469 CATATATACACATATATATA-ATATACACACATATATATACA------CATATATACACA 521
              |||||||  |||| | |||| ||| | |||||||||| ||||       | ||| | |||
Db        651 CATATATGTACATGTGTATACATAAATACACATATATGTACATGCGTGTACATAAATACA 592

Qy        522 CATATATACACATATATACACATATATATACACATATATACACATATATACATATATACA 581
              |||||||  ||||   || ||||| ||| ||| |||| ||||  | ||||||||||||||
Db        591 CATATATGTACATGCGTATACATAAATACACATATATGTACATGTGTATACATATATACA 532

Qy        582 CATATATATACATATATACACATATATACATATATACACATATATACATATATACACACA 641
              ||||||| |||||   || |||||||||||| ||||  |||| | |  | ||| | |||||
Db        531 CATATATGTACATGCGTATACATATATACACATATGTACATGTGTGTACATAAATACACA 472

Qy        642 TATATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTA 701
              ||| |||| ||   || | | | ||| | || |||||
Db        471 TATGTACATATGTGTACATAAATACACATATGTACATNNNNNNNNNNNNNNNNNNNNNNG 412

Qy        702 TATATATGTATATATATA 719
              || || | ||||||||||
Db        411 TACATGTATATATATATA 394


RESULT 69
ADC87320/c
ID   ADC87320 standard; DNA; 1363 BP.
XX
AC   ADC87320;
XX
DT   01-JAN-2004  (first entry)
XX
DE   Human GPCR gene SEQ ID NO:1773.
XX
KW   ds; gene; human; GPCR;
KW   guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1270724-A2.
XX
PD   02-JAN-2003.
XX
PF   18-JUN-2002; 2002EP-00013517.
XX
PR   18-JUN-2001; 2001JP-00246789.
XX
PA   (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA   (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI   Suwa M,  Asai K,  Akiyama Y,  Aburatani H;
XX
DR   WPI; 2003-315783/31.
DR   P-PSDB; ADC87321.
XX
PT   New polynucleotide, useful for preparing a composition for treating a
PT   patient in need of increased or suppressed activity or expression of the
PT   guanosine triphosphate-binding protein coupled receptor.
XX
PS   Claim 1; SEQ ID NO 1773; 28pp; English.
XX
CC   The invention relates to a novel polynucleotide encoding a guanosine
CC   triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC   the invention may have a use in gene therapy. The polynucleotide and
CC   polypeptide are useful for preparing a composition for treating a patient
CC   in need of increased or suppressed activity or expression of the
CC   guanosine triphosphate-binding protein coupled receptor. The
CC   polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC   invention.
XX
SQ   Sequence 1363 BP; 424 A; 175 C; 269 G; 495 T; 0 U; 0 Other;

  Query Match             5.4%;  Score 248.2;  DB 10;  Length 1363;
  Best Local Similarity   71.7%;  Pred. No. 2.7e-33;
  Matches  367;  Conservative    0;  Mismatches  139;  Indels    6;  Gaps    3;

Qy        227 TTTATATATATATATATATATATATACACACACACACATATATATATACACATATATA 286
              | || | ||||||| |||||| || | | | || |||| |   |||||||   | |||||
Db       1136 TGTACACATATATACATATATGTACATATATACGCACATACGCATATATATGTACATATA 1077

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              |||||| |||  |||||||  |||||||||||||||  ||| |  ||||||| ||||||
Db       1076 TACACACATACGCATATATGTACATATATACACATATGCATATATGTACATATATACACA 1017

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||  ||| |  ||  |||| ||| |||| || |||| |||| | ||| | || ||||||
Db       1016 TATGCATATATGTACATATATACACATATGCATATATGTACATATATACACACGCATATA 957

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              |  ||| |||| |||||||||    || |||||     |||||||| ||| | |  | ||
Db        956 TGTACATATATACACACATAT----ACGTATATGTGTGTATACACACATACATATGTGTA 901

Qy        467 TACATATATACACATATATATAATATACACACATATATATACACATATATACACACATAT 526
              |||| | ||||  |||||  | || |||||| ||| ||||| || |||||| ||||||||
Db        900 TACACACATACGTATATACGT-ATGTACACATATACATATATACGTATATATACACATAT 842

Qy        527 ATACACATATATACACATATATATACACATATATACACATATATACATATATACACATAT 586
              | ||  ||||| |||||||||| ||| |  || ||| ||  | ||| | |||||    ||
Db        841 ACACCTATATACACACATATACATATATGTACATATACGCACATATACATATATGTGTAC 782

Qy        587 ATATACATATATACACATATATACATATATACACATATATACA-TATATACACACATATA 645
              ||||| | | ||||| ||||  |||||||||||||| || |||| |||||||||||||| ||
Db        781 ATATATACACATACATATATGCACATATATACACACATGTACACTATATACACACATGTA 722

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              ||| || || ||| | |||||  |   | |||  || | ||||  | | ||| ||| | |
Db        721 CACATATGCACATATATACACACATGCACACACGTACACACACGCACACATATGTACACA 662

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAG 737
              |||||  | || | | | |     |  |||| |
Db        661 CATGTACACATGTGTACACATATGCCCATATG 630


RESULT 70
ABT10748
ID   ABT10748 standard; cDNA; 121600 BP.
XX
AC   ABT10748;
XX
DT   04-DEC-2002  (first entry)
XX
DE   Human breast cancer associated coding sequence SEQ ID NO: 882.
XX
KW   Human; breast specific gene; breast cancer; differential expression;
KW   cytostatic; gene therapy; gene; ss.
XX
```

```
OS   Homo sapiens.
XX
PN   WO200259271-A2.
XX
PD   01-AUG-2002.
XX
PF   25-JAN-2002; 2002WO-US002176.
XX
PR   25-JAN-2001; 2001US-0263757P.
PR   25-APR-2001; 2001US-0286090P.
PR   23-MAY-2001; 2001US-0292517P.
XX
PA   (GENE-) GENE LOGIC INC.
XX
PI   Orr MS,  Nation M,  Diggans JC,  Zeng W;
XX
DR   WPI; 2002-674803/72.
XX
PT   Diagnosing breast cancer in a patient comprises detecting the level of
PT   gene expression in cell or tissue samples, where a differential gene
PT   expression is indicative of breast cancer.
XX
PS   Claim 1; SEQ ID NO 882; 260pp + Sequence Listing; English.
XX
CC   The present invention relates to methods of diagnosing breast cancer in a
CC   patient, which comprise detecting the level of expression in a tissue
CC   sample of two or more genes selected from those shown in ABT09867-
CC   ABT11112, where a differential expression of the genes indicates breast
CC   cancer. The methods are useful in diagnosing, treating, detecting the
CC   progression, and in monitoring treatment of breast cancer in patients.
CC   The methods are also useful as a screening tool for agents that modulate
CC   the onset or progression of breast cancer. The breast cancer genes may be
CC   used as diagnostic markers for the prediction or identification of the
CC   malignant state of breast tissue, for confirming the type and progression
CC   of cancer, and for drug screening and assays. The present sequence is a
CC   coding sequence of the invention. Note: The sequence data for this patent
CC   did not form part of the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub.published_pct_sequences
XX
SQ   Sequence 121600 BP; 38372 A; 24652 C; 24048 G; 34528 T; 0 U; 0 Other;

  Query Match             5.4%;  Score 247.4;  DB 6;  Length 121600;
  Best Local Similarity   68.6%;  Pred. No. 4.4e-33;
  Matches  341;  Conservative    0;  Mismatches  156;  Indels    0;  Gaps    0;

Qy        224 CATTTTATATATATATATATATATATATACACACACACACATATATATATACACATAT 283
              ||   ||||||||| |||||||  ||||||  |    | | |   ||||||  |||| | || |
Db      43894 CAGCCTATATATATCTATATATCTATATATATCTATATATCTATATATCTATATATATCT 43953

Qy        284 ATATACACATATATACATATATACACATATATACACATATATATACACATACATATGTAC 343
              |||||   |||||||  ||||||   ||||||  | || | |||| |  || |||| ||
Db      43954 ATATATCTATATATATCTATATATCTATATATCTATATCTCTATATATCTATATATCTAT 44013

Qy        344 ACATATATATACACATATGTATACACATATATACACATATATACACACATATATACACAT 403
              |  | ||||||   || | ||||    |||||    ||||||  | | || |||| | ||
Db      44014 ATCTCTATATATCTATCTATATATCTCTATATCTCTATATATCTATATATCTATATATAT 44073

Qy        404 ATATACACACATATATACACATATATACACATATATACACATATACACATATATACACAT 463
              ||||  |   ||||||  | |||| || | |||| || | || || | || |||| | | |
Db      44074 CTATATATCTATATATCTATATATCTATATATATCTATATATCTATATATCTATATATCT 44133

Qy        464 ATATACATATATACACATATATATAATATACACACATATATATACACATATATACACACA 523
              |||||  |||||| |  ||||||| || || | |  |||||||  | |||| || | |
Db      44134 ATATATCTATATATATCTATATATCATCTATATATCTATATATCTATATATCTATATATC 44193

Qy        524 TATATACACATATATACACATATATATACACATATATACACATATATACATATATACACA 583
              ||||||   ||||||  | ||||||||  | ||||||  | |||| || ||||||  | |
Db      44194 TATATATCTATATATCTATATATATATCTATATATATCTATATATCTATATATATCTATA 44253

Qy        584 TATATATACATATATACACATATATACATATATACACATATATACATATATACACACATA 643
              ||| |||| || |||| | || |||| || |||| | || |||||||| |||| |   |||
Db      44254 TATCTATATATCTATATATATCTATATATCTATATATATCTATACATCTATATATCTATA 44313

Qy        644 TATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATA 703
              ||||   ||| || || | |   ||| ||||  |||||| || | |||| ||||  | ||
Db      44314 TATATCTATATATCTATATATCTATATATATCTATATATCTATATATATCTATATATCTA 44373

Qy        704 TATATGTATATATATAT 720
              ||||| | ||||||| |
Db      44374 TATATATCTATATATCT 44390


RESULT 71
ADQ18307
ID   ADQ18307 standard; DNA; 121600 BP.
XX
AC   ADQ18307;
XX
DT   26-AUG-2004  (first entry)
XX
DE   Human soft tissue sarcoma-upregulated DNA - SEQ ID 1125.
XX
KW   soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO2004048938-A2.
XX
PD   10-JUN-2004.
XX
PF   26-NOV-2003; 2003WO-US038193.
XX
PR   26-NOV-2002; 2002US-0429739P.
XX
PA   (PROT-) PROTEIN DESIGN LABS INC.
XX
PI   Aziz N,  Ginsburg WM,  Zlotnik A;
XX
DR   WPI; 2004-441208/41.
XX
PT   Early detection of soft tissue sarcoma comprises determining expression
PT   of a gene in a first soft tissue sample and a normal soft tissue sample
PT   and comparing the gene expression, also useful in treating soft tissue
PT   sarcoma.
XX
PS   Example 2; SEQ ID NO 1125; 210pp; English.
XX
CC   The invention relates to a novel method for detecting soft tissue sarcoma
CC   which comprises obtaining a first soft tissue sample from an individual
CC   and a normal soft tissue sample from the same or different individual,
CC   determining the expression of a gene in both samples and comparing the
CC   expression of the gene in both soft tissue samples, where a higher level
CC   of protein expression in the first soft tissue sample indicates the
CC   presence of soft tissue sarcoma. The method of the invention has
CC   cytostatic applications and may be useful for detecting soft tissue
CC   sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC   acid sequences may be useful in diagnostic and screening applications.
CC   The current sequence is that of a human soft tissue sarcoma-upregulated
CC   DNA of the invention. The current sequence is not shown within the
CC   specification per se but was submitted in CD format by the inventor.
XX
SQ   Sequence 121600 BP; 38372 A; 24652 C; 24048 G; 34528 T; 0 U; 0 Other;

  Query Match             5.4%;  Score 247.4;  DB 12;  Length 121600;
  Best Local Similarity   68.6%;  Pred. No. 4.4e-33;
  Matches  341;  Conservative    0;  Mismatches  156;  Indels    0;  Gaps    0;

Qy        224 CATTTTATATATATATATATATATATATACACACACACACATATATATATACACATAT 283
              ||   ||||||||| |||||||  ||||||  |    | | |   ||||||  |||| | || |
Db      43894 CAGCCTATATATATCTATATATCTATATATATCTATATATCTATATATCTATATATATCT 43953

Qy        284 ATATACACATATATACATATATACACATATATACACATATATATACACATACATATGTAC 343
              |||||   |||||||  ||||||   ||||||  | || | |||| |  || |||| ||
Db      43954 ATATATCTATATATATCTATATATCTATATATCTATATCTCTATATATCTATATATCTAT 44013
```

```
Qy        344 ACATATATATACACATATGTATACACATATATACACATATATACACACATATATACACAT 403
              |  | ||||||   || | ||||    |||||    ||||||  | | || |||| | ||
Db      44014 ATCTCTATATATCTATCTATATATCTCTATATCTCTATATATCTATATATCTATATATAT 44073

Qy        404 ATATACACACATATATACACATATATACACATATATACACATATACACATATATACACAT 463
              ||||  |   ||||||  | ||||  || | ||||  || | ||  || | ||  |||| |  |
Db      44074 CTATATATCTATATATCTATATATCTATATATATCTATATATCTATATATCTATATATCT 44133

Qy        464 ATATACATATATACACATATATATAATATACACACATATATATACACATATATACACACA 523
              |||||  ||||||  |  ||||||| || || |  |  |||||||  | ||||  || | |
Db      44134 ATATATCTATATATATCTATATATCATCTATATATCTATATATCTATATATCTATATATC 44193

Qy        524 TATATACACATATATACACATATATATACACATATATACACATATATACATATATACACA 583
              ||||||   ||||||  | ||||||||  | ||||||  | ||||  || ||||||  | |
Db      44194 TATATATCTATATATCTATATATATATCTATATATATCTATATATCTATATATATCTATA 44253

Qy        584 TATATATACATATATACACATATATACATATATACACATATATACATATATACACACATA 643
              ||| ||||  || ||||  | ||  |||| ||  |||| | ||  |||||||  ||||  |   |||
Db      44254 TATCTATATATCTATATATATCTATATATCTATATATATCTATACATCTATATATCTATA 44313

Qy        644 TATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATA 703
              ||||   ||| || || | |   ||| ||||  |||||| || | |||| ||||  | ||
Db      44314 TATATCTATATATCTATATATCTATATATATCTATATATCTATATATATCTATATATCTA 44373

Qy        704 TATATGTATATATATAT 720
              ||||| | ||||||| |
Db      44374 TATATATCTATATATCT 44390


RESULT 72
ACC79619
ID   ACC79619 standard; DNA; 71843 BP.
XX
AC   ACC79619;
XX
DT   27-OCT-2003  (revised)
DT   06-AUG-2003  (first entry)
XX
DE   Zebrafish protein kinase C lambda related genomic DNA SEQ ID NO:3.
XX
KW   Zebrafish; Danio rerio; protein kinase C lambda; PKC-lambda; enzyme;
KW   cardiant; cytostatic; nephrotropic; heart disease; cancer; organ failure;
KW   epithelial-epithelial cell interaction; epithelial cell polarity;
KW   heart failure; kidney failure; gene; ds.
XX
OS   Danio rerio.
XX
PN   WO2003023048-A2.
XX
PD   20-MAR-2003.
XX
PF   06-SEP-2002; 2002WO-US028410.
XX
PR   06-SEP-2001; 2001US-0317653P.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Peterson R,  Fishman MC;
XX
DR   WPI; 2003-313257/30.
XX
PT   Diagnosing or prognosing a disease or condition related to Protein Kinase
PT   C lambda, by analyzing sample nucleic acid obtained from the test subject
PT   to determine mutation in gene encoding Protein kinase C lambda.
XX
PS   Disclosure; Page 63-83; 88pp; English.
XX
CC   The present invention describes a method (M1) for determining whether a
CC   test subject has, or is at risk of developing, a disease or condition
CC   related to protein kinase C lambda (PKC-lambda) (I). M1 comprises
CC   analysing a nucleic acid molecule of a sample from the subject to
CC   determine whether the subject has a mutation in a gene encoding (I),
CC   where the presence of a mutation indicates that the subject has, or is at
CC   risk of developing, a disease or condition related to (I). (I) has
CC   cardiant, cytostatic and nephrotropic activities. M1 is useful for
CC   determining whether a test subject has, or is at risk of developing, a
CC   disease or condition related to (I), e.g. heart disease, cancer, or
CC   condition associated with epithelial-epithelial cell interactions or
CC   epithelial cell polarity. (I) is useful for treating organ failure such
CC   as heart or kidney failure. (I) is also useful for raising antibodies
CC   specific for various regions of (I). M1 is suitable for detecting an
CC   increased likelihood of diseases or conditions associated with (I), so
CC   that appropriate intervention can be instituted before any symptoms
CC   occur. This is useful with patients in high risk groups for such diseases
CC   or conditions. Diagnostic method facilitates determination of the
CC   etiology of such an existing disease or condition in a patient, so that
CC   an appropriate approach to treatment can be selected. The present
CC   sequence represents zebrafish genomic DNA for PKC-lambda, which is used
CC   in the exemplification of the present invention. (Updated on 27-OCT-2003
CC   to standardise OS field)
XX
SQ   Sequence 71843 BP; 24017 A; 12903 C; 12592 G; 21955 T; 0 U; 376 Other;

  Query Match             5.3%;  Score 246.6;  DB 8;  Length 71843;
  Best Local Similarity   63.1%;  Pred. No. 5.9e-33;
  Matches  313;  Conservative    0;  Mismatches  181;  Indels    2;  Gaps    2;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | |||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db      68244 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 68303

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| |||||||||| | | ||||| | | | ||||| ||| |
Db      68304 TATATATATATATATATATACACACACACACACACACACACACACACACACANNNNNNNN 68363

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406

Db      68364 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 68423

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
                                                  ||||| | ||||||| | |||||
Db      68424 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNATATATATATATATATATATA 68483

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | |||||||||| ||||| | | |||||||||| | ||||||| | | |||
Db      68484 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 68543

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | |||| |||| | ||||||| | ||||||| ||||||| | |||
Db      68544 TATATATATATATATATATATNTATATATATATATATATATATATATATATATATATA 68603

Qy        586 TATATACATATATACACATATATACATATATA-CACATATATACATATATACACACATAT 644
              |||||| ||||||| | ||||||| |||||||    ||||||  ||||||| | | ||||
Db      68604 TATATATATATATATATATATATATATATATAGTGTATATATGTATATATATATATATAT 68663

Qy        645 ATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATAT 704
              ||| | ||| ||||| |    | ||| ||||| |||||||||| | |||||||||  |||||
Db      68664 ATATATATATATATATATNTATATATATATATATATATATATATATATATATATATAT 68723

Qy        705 ATATGTATATATATAT 720
              |||| ||||||||||
Db      68724 ATATATATATATATAT 68739


RESULT 73
AAS46591/c
ID   AAS46591 standard; DNA; 6286 BP.
XX
AC   AAS46591;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Tumour suppressor gene derived chemically modified sequence #313.
XX
KW   Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer;
```

KW tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW cytosine methylation; ds.
XX
OS Homo sapiens.
XX
PN WO200168912-A2.
XX
PD 20-SEP-2001.
XX
PF 15-MAR-2001; 2001WO-EP002955.
XX
PR 15-MAR-2000; 2000DE-01013847.
PR 06-APR-2000; 2000DE-01019058.
PR 07-APR-2000; 2000DE-01019173.
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2001-602752/68.
XX
PT Fragments of chemically modified genes associated with tumor suppressor
PT genes and oncogenes, useful in designing primers and probes for analyzing
PT diseases associated with cytosine methylation state e.g. cancer.
XX
PS Claim 1; SEQ ID NO 313; 27pp; English.
XX
CC The invention relates to a nucleic acid comprising a sequence of 18
CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC bisulphite, of genes associated with tumour suppression and oncogenes
CC having a sequence taken from 536 (actually 533 since numbers 408, 458 and
CC 500 are missing from the sequence listing) sequences (Ss) and sequences
CC complementary to (Ss). The nucleic acid may be a peptide nucleic acid-
CC oligomer (PNA) of at least 9 nucleotides and may form part of a set of
CC probes for detecting the cytosine methylation state and/or single
CC nucleotide polymorphisms and also to be used in an array for analysing
CC diseases associated with CpG dinucleotides e.g. cancers and tumours. The
CC probes can also be used in a method for ascertaining genetic and/or
CC epigenetic parameters for the diagnosis and/or therapy of existing
CC diseases or the predisposition to specific diseases, by analysing
CC cytosine methylations. The parameters may be compared to another set of
CC genetic and/or epigenetic parameters, the differences serving as basis
CC for diagnosis and/or prognosis events which are disadvantageous to
CC patients. The present sequence is one of the 533 genomic sequences
CC derived from tumour suppressor genes and oncogenes. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 6286 BP; 1508 A; 124 C; 1538 G; 3116 T; 0 U; 0 Other;

Query Match 5.3%; Score 245; DB 4; Length 6286;
Best Local Similarity 57.3%; Pred. No. 1e-32;
Matches 443; Conservative 0; Mismatches 330; Indels 0; Gaps 0;

```
Qy       174 CTCCAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATAT 233
Db      4589 CACCCACCCATCCCCATTCATAAACATTTACTATAAAATAAATCAATAAATAAATAATAA 4530

Qy       234 ATATATATATATATATATACACACACACACACATATATATATACACATATATATACACAT 293
Db      4529 ATTAATAAATAAATAAATAAATAAATAAATACATAAATAAATAAATAAATAAATAAATAA 4470

Qy       294 ATATACATATATACACATATATACACATATATATACACATACATATGTACACATATATAT 353
Db      4469 ATAAATAAATAAATAAATAAATACATAAATAAATAAATAAATAAATAAATAAATAAATAA 4410

Qy       354 ACACATATGTATACACATATATACACATATATACACACATATATACACATATATACACAC 413
Db      4409 ATAAATAAATAAATACATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 4350

Qy       414 ATATATACACATATATACACATATATACACATATACACATATATACACATATATACATAT 473
Db      4349 ATAAATACATAAATAAATAAATAAATAAATAAATAAATACATAAATAAATAAATAA 4290

Qy       474 ATACACATATATATAATATACACACATATATATACACATATATACACACATATATACACA 533
Db      4289 ATAAATAAATAAATAAATAAACAACAAATAAATAAATAAATAAATAAATAAATAAATAA 4230

Qy       534 TATATACACATATATATACACATATATACACATATATACATATATACACATATATATACA 593
Db      4229 TAAATAAATACATAAATAAATAAATAAATAAATACATAAATAAATAAATAAATAAATAAA 4170

Qy       594 TATATACACATATATACATATATACACATATATACATATATACACACATATATACACATA 653
Db      4169 TAAATAAATAAATACATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAA 4110

Qy       654 CATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATA 713
Db      4109 TAAATAAATACATAAATAAATAAATAAATAAATAAATAAATAAATAAATACATAAATAAA 4050

Qy       714 TATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAA 773
Db      4049 TAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATACATAAATAAATAAA 3990

Qy       774 GAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGG 833
Db      3989 TAAATAAATAAATAAATAAATAAATAAATAAATACATAAATAAATAAATAAATACATAAA 3930

Qy       834 GAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATAT 893
Db      3929 TAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAACAACAAATAAAT 3870

Qy       894 ATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATG 946
Db      3869 AAATAAATAAATAAATAAATAAATAAATAAATACATAAATAAATAAACG 3817
```

RESULT 74
ADC87574/c
ID ADC87574 standard; DNA; 1309 BP.
XX
AC ADC87574;
XX
DT 01-JAN-2004 (first entry)
XX
DE Human GPCR gene SEQ ID NO:2027.
XX
KW ds; gene; human; GPCR;
KW guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1270724-A2.
XX
PD 02-JAN-2003.
XX
PF 18-JUN-2002; 2002EP-00013517.
XX
PR 18-JUN-2001; 2001JP-00246789.
XX
PA (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI Suwa M, Asai K, Akiyama Y, Aburatani H;
XX
DR WPI; 2003-315783/31.
DR P-PSDB; ADC87575.
XX
PT New polynucleotide, useful for preparing a composition for treating a
PT patient in need of increased or suppressed activity or expression of the
PT guanosine triphosphate-binding protein coupled receptor.
XX
PS Claim 1; SEQ ID NO 2027; 28pp; English.

```
XX
CC   The invention relates to a novel polynucleotide encoding a guanosine
CC   triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC   the invention may have a use in gene therapy. The polynucleotide and
CC   polypeptide are useful for preparing a composition for treating a patient
CC   in need of increased or suppressed activity or expression of the
CC   guanosine triphosphate-binding protein coupled receptor. The
CC   polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC   invention.
XX
SQ   Sequence 1309 BP; 463 A; 205 C; 200 G; 441 T; 0 U; 0 Other;

  Query Match             5.3%;  Score 243.2;  DB 10;  Length 1309;
  Best Local Similarity   70.7%;  Pred. No. 1.9e-32;
  Matches  355;  Conservative    0;  Mismatches  138;  Indels    9;  Gaps    2;

Qy        227 TTTATATATATATATATATATATATACACACACACACATATATATATACACATATATA 286
              | || |||| || || ||| ||||||| || |    || |||| ||||||| |||| ||
Db        553 TATACATATGTACATCTATGTATATACGCATATGTGCATATATGTATATACGCATACGTA 494

Qy        287 TACACATATATACATATATACACATATATACACATATATATACAC------ATACATATG 340
              ||     |||||||  ||||| |||||| | || |||| |||| ||       |||||||||
Db        493 TATGTGTATATACGCATATATACATATGTGCATATATGTATATACGCATATATACATATG 434

Qy        341 TACACATATATATACACATATGTATACACATATATACACATATATACACACATATATACA 400
              | || |||| |||| ||   || |||||| || |  | | || |||| || ||||||||||
Db        433 TGCATATATGTATATACGCATATATACATATGTGCATATATGTATATACGCATATATACA 374

Qy        401 CATATATACACACATATATACACATATATACACATATATACACATATACACATATATACA 460
              || |  | | |  |||||||  |||||||||| | || | || || | || ||||  | |
Db        373 TATGTGCATATATGTATATACGCATATATACATACATGTGCATATGTGCATATATTCATA 314

Qy        461 CATATATACATATATACACATATATATAATATACACACATATATATACACATATATACAC 520
              || ||  || ||||   | ||||    || |||| ||| |||| |||| |||| ||||
Db        313 AGTACATCTATGTATATGTACATAT---ATGTACATACACATATGTACATATATGTACAT 257

Qy        521 ACATATATACACATATATACACATATATATACACATATATACACATATATACATATATAC 580
              ||| ||||  ||||||||  ||||| | ||| ||||||||  ||||||| | | |||||
Db        256 ACACATATGTACATATATGTACATACACATATACATATATGTACATATACATACATATAT 197

Qy        581 ACATATATATACATATATACACATATATACATATATACACATATATACATATATACACAC 640
              ||| ||| || ||||  ||| |||  || ||||| | | || |||| ||| | | | ||
Db        196 GCATGTATGTATATATGCACATATACCTATATATACATATATGTATATATACACATATAC 137

Qy        641 ATATATACACATACATATACACACACATAGATATACATATATATACACATATATATACGT 700
              ||||||  |||||  | || | ||||| | ||  | |||| ||||| |||||||  |  |
Db        136 ATATATGTACATATGTGTATATACACACATATGCATATATGTATACGCATATATGCATAT 77

Qy        701 ATATATATGTATATATATATGC 722
              || |||||  | | ||||||||
Db         76 ATGTATATACACACATATATGC 55


RESULT 75
ABX08336_08
Continuation (9 of 17) of ABX08336 from base 800001 (Human phosphodiesterase 4D (PDE4D)
WP Sequence split into 17 fragments  LOCUS ABX08336 Accession Abx08336
WP      Fragment Name       Begin     End
WP      ABX08336_00             1    110000
WP      ABX08336_01        100001    210000
WP      ABX08336_02        200001    310000
WP      ABX08336_03        300001    410000
WP      ABX08336_04        400001    510000
WP      ABX08336_05        500001    610000
WP      ABX08336_06        600001    710000
WP      ABX08336_07        700001    810000
WP      ABX08336_08        800001    910000
WP      ABX08336_09        900001   1010000
WP      ABX08336_10       1000001   1110000
WP      ABX08336_11       1100001   1210000
WP      ABX08336_12       1200001   1310000
WP      ABX08336_13       1300001   1410000
WP      ABX08336_14       1400001   1510000
WP      ABX08336_15       1500001   1610000
WP      ABX08336_16       1600001   1691080

  Query Match             5.2%;  Score 242;  DB 6;  Length 110000;
  Best Local Similarity   70.0%;  Pred. No. 3.7e-32;
  Matches  368;  Conservative    1;  Mismatches  151;  Indels    6;  Gaps    3;

Qy        196 GAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACAC 255
              ||  ||| |    | | | |||| | |    | |||||||| ||||||||  |||||
Db      67577 GAGTTGTGAATATATATATAAAATATGTGTGTGTATATATACATATATATGCATATATGT 67636

Qy        256 ACACACACACATATATATATACACATATATATACACATATATACATATATACACATATAT 315
              ||  | | ||||||||| |||| | |  |||||  || ||||||||||||||||||||||
Db      67637 ACCTATATACATATAT-TATATATACCTATATGAACCTATATACATATATACACATATAT 67695

Qy        316 ACACATATATATACACATACATATGTACACATATAT----ATACACATATGTATACACAT 371
              || |||||| ||| ||||| || :  || |||||     ||| ||| || | |  ||
Db      67696 GAACCTATATACACATATACACATAYGTACCTATATACACATATACACATATGTGTACCC 67755

Qy        372 ATATACACATATATACACACATATATACACATATATACACACATATATACACATATATAC 431
              |||||||||||||| ||| |  ||  || | | | ||| |||||||| |  ||  ||||||
Db      67756 ATATACACATATACACATATGTACCTATATACACATATACACATATGTGTACCCATATAC 67815

Qy        432 ACATATATACACATATACACATATATACACATATATACATATATACACATATATATAATA 491
              ||||||| ||| || |  ||  |||||||||||| ||||||  |  ||  |||||  | |
Db      67816 ACATATACACATATGTGTACCCATATACACATATACACATATGTGTACCCATATA-CACA 67874

Qy        492 TACACACATATATATACACATATATACACACATATATACACATATATACACATATATATA 551
              || |||||||| | ||| |||||| ||| | | | |||    ||   | | | | |||||
Db      67875 TATACACATATGTGTACCCATATACACATATACACATATGTGTACCCATATACACATATA 67934

Qy        552 CACATATATACACATATATACATATATACACATATATATACATATATACACATATATACA 611
              ||||||| |  ||  ||||||| |||||||||||| | |||  |||||||||||||| |||
Db      67935 CACATATGTGTACCCATATACACATATACACATATGTGTACCCATATACACATATACACA 67994

Qy        612 TATATACACATATATACATATATACACACATATATACACATACATATACACACACATAGA 671
              | | ||| ||||||| ||||||| |||        ||    | | | ||||||||||    ||
Db      67995 TGTGTACCCATATACACATATACACATGTGTACCCATATACACATATACACATGTGTACC 68054

Qy        672 TATACATATATATACACATATATATACGTATATATATGTATATATA 717
              ||| | | |||||  |||||| | |||  |||||     ||| | ||
Db      68055 CATATACACATATATGCATATGTGTACCCATATACGCATATGTGTA 68100


RESULT 76
ADJ25985_08
Continuation (9 of 17) of ADJ25985 from base 800001 (Human phosphodiesterase 4D (PDE4D) c
WP Sequence split into 17 fragments  LOCUS ADJ25985 Accession Adj25985
WP      Fragment Name       Begin     End
WP      ADJ25985_00             1    110000
WP      ADJ25985_01        100001    210000
WP      ADJ25985_02        200001    310000
WP      ADJ25985_03        300001    410000
WP      ADJ25985_04        400001    510000
WP      ADJ25985_05        500001    610000
WP      ADJ25985_06        600001    710000
WP      ADJ25985_07        700001    810000
WP      ADJ25985_08        800001    910000
WP      ADJ25985_09        900001   1010000
WP      ADJ25985_10       1000001   1110000
WP      ADJ25985_11       1100001   1210000
WP      ADJ25985_12       1200001   1310000
WP      ADJ25985_13       1300001   1410000
WP      ADJ25985_14       1400001   1510000
WP      ADJ25985_15       1500001   1610000
WP      ADJ25985_16       1600001   1691139

  Query Match             5.2%;  Score 242;  DB 12;  Length 110000;
  Best Local Similarity   70.0%;  Pred. No. 3.7e-32;
  Matches  368;  Conservative    1;  Mismatches  151;  Indels    6;  Gaps    3;

Qy        196 GAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACAC 255
              ||  ||| |    | | | |||| | |    | |||||||| ||||||||  |||||
```

```
Db      67636 GAGTTGTGAATATATATATAAAATATGTGTGTGTATATATACATATATATGCATATATGT 67695

Qy        256 ACACACACACATATATATATACACATATATATACACATATATACATATATACACATATAT 315
              ||  | | ||||||||| |||| | |  |||||  || ||||||||||||||||||||||
Db      67696 ACCTATATACATATAT-TATATATACCTATATGAACCTATATACATATATACACATATAT 67754

Qy        316 ACACATATATATACACATACATATGTACACATATAT----ATACACATATGTATACACAT 371
              || |||||| ||| ||||| || :  || |||||    ||| ||| || | |  ||
Db      67755 GAACCTATATACACATATACACATAYGTACCTATATACACATATACACATATGTGTACCC 67814

Qy        372 ATATACACATATATACACACATATATACACATATATACACACATATATACACATATATAC 431
              |||||||||||||| ||| |  ||  || | | ||| |||||||| |  ||  ||||||
Db      67815 ATATACACATATACACATATGTACCTATATACACATATACACATATGTGTACCCATATAC 67874

Qy        432 ACATATATACACATATACACATATATACACATATATACATATATACACATATATATAATA 491
              ||||||| ||| || |  ||  |||||||||||||| |||||| |  ||  |||||  | |
Db      67875 ACATATACACATATGTGTACCCATATACACATATACACATATGTGTACCCATATA-CACA 67933

Qy        492 TACACACATATATATACACATATATACACACATATATACACATATATACACATATATATA 551
              || |||||||| | ||| |||||| ||| | | | |||    ||   | | | | |||||
Db      67934 TATACACATATGTGTACCCATATACACATATACACATATGTGTACCCATATACACATATA 67993

Qy        552 CACATATATACACATATATACATATATACACATATATATACATATATACACATATATACA 611
              |||||||  |  ||  ||||||| |||||||||||| | |||  ||||||||||||| |||
Db      67994 CACATATGTGTACCCATATACACATATACACATATGTGTACCCATATACACATATACACA 68053

Qy        612 TATATACACATATATACATATATACACACATATATACACATACATATACACACACATAGA 671
              | | ||| |||||| ||||||| |||    ||   | | | ||||||||||    ||
Db      68054 TGTGTACCCATATACACATATACACATGTGTACCCATATACACATATACACATGTGTACC 68113

Qy        672 TATACATATATATACACATATATATACGTATATATATGTATATATA 717
              ||| | | |||||  ||||| | |||  |||||    ||| | ||
Db      68114 CATATACACATATATGCATATGTGTACCCATATACGCATATGTGTA 68159


RESULT 77
ADN97989_08
Continuation (9 of 17) of ADN97989 from base 800001 (Human phosphodiesterase 4D genomic
WP Sequence split into 17 fragments  LOCUS ADN97989 Accession Adn97989
WP      Fragment Name       Begin      End
WP      ADN97989_00             1    110000
WP      ADN97989_01        100001    210000
WP      ADN97989_02        200001    310000
WP      ADN97989_03        300001    410000
WP      ADN97989_04        400001    510000
WP      ADN97989_05        500001    610000
WP      ADN97989_06        600001    710000
WP      ADN97989_07        700001    810000
WP      ADN97989_08        800001    910000
WP      ADN97989_09        900001   1010000
WP      ADN97989_10       1000001   1110000
WP      ADN97989_11       1100001   1210000
WP      ADN97989_12       1200001   1310000
WP      ADN97989_13       1300001   1410000
WP      ADN97989_14       1400001   1510000
WP      ADN97989_15       1500001   1610000
WP      ADN97989_16       1600001   1691138

  Query Match             5.2%;  Score 242;  DB 12;  Length 110000;
  Best Local Similarity   70.0%;  Pred. No. 3.7e-32;
  Matches  368;  Conservative    1;  Mismatches  151;  Indels    6;  Gaps    3;

Qy        196 GAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACAC 255
              ||  ||| |    | | | |||| | |    | ||||||| ||||||||  |||||
Db      67635 GAGTTGTGAATATATATATAAAATATGTGTGTGTATATATACATATATATGCATATATGT 67694

Qy        256 ACACACACACATATATATATACACATATATATACACATATATACATATATACACATATAT 315
              ||  | | ||||||||| |||| | |  |||||  || ||||||||||||||||||||||
Db      67695 ACCTATATACATATAT-TATATATACCTATATGAACCTATATACATATATACACATATAT 67753

Qy        316 ACACATATATATACACATACATATGTACACATATAT----ATACACATATGTATACACAT 371
              || |||||| ||| ||||| || :  || |||||    ||| ||| || | |  ||
Db      67754 GAACCTATATACACATATACACATAYGTACCTATATACACATATACACATATGTGTACCC 67813

Qy        372 ATATACACATATATACACACATATATACACATATATACACACATATATACACATATATAC 431
              |||||||||||||| ||| |  ||  || | | ||| |||||||| |  ||  ||||||
Db      67814 ATATACACATATACACATATGTACCTATATACACATATACACATATGTGTACCCATATAC 67873

Qy        432 ACATATATACACATATACACATATATACACATATATACATATATACACATATATATAATA 491
              ||||||| ||| || |  ||  |||||||||||||| |||||| |  ||  |||||  | |
Db      67874 ACATATACACATATGTGTACCCATATACACATATACACATATGTGTACCCATATA-CACA 67932

Qy        492 TACACACATATATATACACATATATACACACATATATACACATATATACACATATATATA 551
              || |||||||| | ||| |||||| ||| | | | |||    ||   | | | | |||||
Db      67933 TATACACATATGTGTACCCATATACACATATACACATATGTGTACCCATATACACATATA 67992

Qy        552 CACATATATACACATATATACATATATACACATATATATACATATATACACATATATACA 611
              |||||||  |  ||  ||||||| |||||||||||| | |||  ||||||||||||| |||
Db      67993 CACATATGTGTACCCATATACACATATACACATATGTGTACCCATATACACATATACACA 68052

Qy        612 TATATACACATATATACATATATACACACATATATACACATACATATACACACACATAGA 671
              | | ||| |||||| ||||||| |||    ||   | | | ||||||||||    ||
Db      68053 TGTGTACCCATATACACATATACACATGTGTACCCATATACACATATACACATGTGTACC 68112

Qy        672 TATACATATATATACACATATATATACGTATATATATGTATATATA 717
              ||| | | |||||  ||||| | |||  |||||    ||| | ||
Db      68113 CATATACACATATATGCATATGTGTACCCATATACGCATATGTGTA 68158


RESULT 78
ADO50281_08
Continuation (9 of 17) of ADO50281 from base 800001 (Human phosphodiesterase 4D (PDE4D) g
WP Sequence split into 17 fragments  LOCUS ADO50281 Accession Ado50281
WP      Fragment Name       Begin      End
WP      ADO50281_00             1    110000
WP      ADO50281_01        100001    210000
WP      ADO50281_02        200001    310000
WP      ADO50281_03        300001    410000
WP      ADO50281_04        400001    510000
WP      ADO50281_05        500001    610000
WP      ADO50281_06        600001    710000
WP      ADO50281_07        700001    810000
WP      ADO50281_08        800001    910000
WP      ADO50281_09        900001   1010000
WP      ADO50281_10       1000001   1110000
WP      ADO50281_11       1100001   1210000
WP      ADO50281_12       1200001   1310000
WP      ADO50281_13       1300001   1410000
WP      ADO50281_14       1400001   1510000
WP      ADO50281_15       1500001   1610000
WP      ADO50281_16       1600001   1691134

  Query Match             5.2%;  Score 242;  DB 12;  Length 110000;
  Best Local Similarity   70.0%;  Pred. No. 3.7e-32;
  Matches  368;  Conservative    1;  Mismatches  151;  Indels    6;  Gaps    3;

Qy        196 GAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACAC 255
              ||  ||| |    | | | |||| | |    | ||||||| ||||||||  |||||
Db      67636 GAGTTGTGAATATATATATAAAATATGTGTGTGTATATATACATATATATGCATATATGT 67695

Qy        256 ACACACACACATATATATATACACATATATATACACATATATACATATATACACATATAT 315
              ||  | | ||||||||| |||| | |  |||||  || ||||||||||||||||||||||
Db      67696 ACCTATATACATATAT-TATATATACCTATATGAACCTATATACATATATACACATATAT 67754

Qy        316 ACACATATATATACACATACATATGTACACATATAT----ATACACATATGTATACACAT 371
              || |||||| ||| ||||| || :  || |||||    ||| ||| || | |  ||
Db      67755 GAACCTATATACACATATACACATAYGTACCTATATACACATATACACATATGTGTACCC 67814

Qy        372 ATATACACATATATACACACATATATACACATATATACACACATATATACACATATATAC 431
              |||||||||||||| ||| |  ||  || | | ||| |||||||| |  ||  ||||||
Db      67815 ATATACACATATACACATATGTACCTATATACACATATACACATATGTGTACCCATATAC 67874

Qy        432 ACATATATACACATATACACATATATACACATATATACATATATACACATATATATAATA 491
              ||||||| ||| || |  ||  |||||||||||||| |||||| |  ||  |||||  | |
Db      67875 ACATATACACATATGTGTACCCATATACACATATACACATATGTGTACCCATATA-CACA 67933
```

```
Qy        492 TACACACATATATATACACATATATACACACATATATACACATATATACACATATATATA 551
              || |||||||| | ||| |||||| ||| | | | |||     ||    | | | | |||||
Db      67934 TATACACATATGTGTACCCATATACACATATACACATATGTGTACCCATATACACATATA 67993

Qy        552 CACATATATACACATATATACATATATACACATATATATACATATATACACATATATACA 611
              ||||||| |  ||  ||||||| |||||||||||| | |||  |||||||||||| |||
Db      67994 CACATATGTGTACCCATATACACATATACACATATGTGTACCCATATACACATATACACA 68053

Qy        612 TATATACACATATATACATATATACACACATATATACACATACATATACACACACATAGA 671
              | | ||| |||||| ||||||| |||     ||  | | | ||||||||||    ||
Db      68054 TGTGTACCCATATACACATATACACATGTGTACCCATATACACATATACACATGTGTACC 68113

Qy        672 TATACATATATATACACATATATATACGTATATATATGTATATATA 717
              ||| | | |||||  |||||  | |||  |||||     ||| | ||
Db      68114 CATATACACATATATGCATATGTGTACCCATATACGCATATGTGTA 68159


RESULT 79
ADQ97965
ID   ADQ97965 standard; DNA; 152096 BP.
XX
AC   ADQ97965;
XX
DT   07-OCT-2004  (first entry)
XX
DE   Mouse cancer associated sequence MD11-032, SEQ ID 942.
XX
KW   Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Mouse; ds.
XX
OS   Mus musculus.
XX
PN   WO2004060304-A2.
XX
PD   22-JUL-2004.
XX
PF   22-DEC-2003; 2003WO-US041389.
XX
PR   27-DEC-2002; 2002US-00330773.
XX
PA   (SAGR-) SAGRES DISCOVERY INC.
XX
PI   Morris DW,  Malandro MS;
XX
DR   WPI; 2004-543781/52.
XX
PT   New isolated cancer associated nucleic acids comprising at least 10
PT   contiguous nucleotides, useful for diagnosing, preventing and/or treating
PT   cancers such as leukemia and lymphoma.
XX
PS   Claim 1; SEQ ID NO 942; 199pp; English.
XX
CC   The present invention relates to cancer associated sequences (ADQ97025-
CC   ADQ98004). The sequences are useful for the diagnosis, prevention and/or
CC   treatment of cancer, such as leukemia and lymphoma. Note: The sequence
CC   data for this patent did not form part of the printed specification, but
CC   was obtained in electronic formate directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 152096 BP; 42690 A; 31533 C; 31445 G; 45238 T; 0 U; 1190 Other;

  Query Match             5.2%;  Score 241.4;  DB 12;  Length 152096;
  Best Local Similarity   68.5%;  Pred. No. 4.8e-32;
  Matches  346;  Conservative    0;  Mismatches  158;  Indels    1;  Gaps    1;

Qy        230 ATATATATATATATATATATATACACACACACACACATATATATATACACATATATATAC 289
              | | | | | | | | | | | ||| |||||||||| | | | ||| | ||| | | | ||
Db      78632 ACACACACACACACACACACACACACATACACACACATACACACATACATACACACACAC 78691

Qy        290 ACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATAT 349
              ||| | |||||||  | |||||  | |  |||||||  | | |||    |||  |    ||| | |
Db      78692 ACACACATACATACACACACACACACACATACACACACATGCACACACACACATACAC 78751

Qy        350 ATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATAC 409
              | |||||||  |    |||||||  | |||||||  |||||||||||  | | |||  | | | ||
Db      78752 ACATACACACACACATACACACACATACACACATATACACACACACACACATACACAC 78811

Qy        410 ACACATATATACACATATATACACATATATACACATATACACATATATACACATATATAC 469
              | |||  | |||||||  | |  |||  | |  |||||  |||||||  |||  |||||  |||  ||
Db      78812 ATACACACATACACACACACACATACACACACACACATACACACATACACACACATACAC 78871

Qy        470 ATATATACACATATATATAATATACACACATATATATACACATATATACACACATATATA 529
              | | | |||||  | |    | | |||||||  | |||  |||||  | |  |||||||  | | |
Db      78872 ACACACACACACACACCACACACACACACACACACATACACACACACACACACACACACA 78931

Qy        530 CACATATATACACATATATATACACATATATACACATATATACATATATACACATATATA 589
              ||||||  | |||||  |||  | |||||  | |  |||  | |  |||  | |  |||  | | |
Db      78932 CACATACACACACACATACACACACACACACACATACACACACACACACATACACACACA 78991

Qy        590 TACATATATATACACATATATATACATATATACACATATATACATATATACACACATATATACA 649
              |||||  | |||||  | |  |||  |||  |||||  | |  |||  | |||||||||  | |  |||
Db      78992 CACATACACACACACACACACACATACACACACACACACACACACATACACACACACACACA 79051

Qy        650 CATACATATATACACACACATAGATATACA-TATATATATACACATATATATACGTATATATAT 708
              ||  |||||  |||||||||  | | | |||  | | |  |||||  | | | ||  | | ||  |
Db      79052 CACACATACACACACACACACACACACACACCACACACACACACACACACACACACACATGT 79111

Qy        709 GTATATATATATGCTCCAGAGTTCA 733
              |  | |  |||  |||  |  | ||
Db      79112 GCGTGTGCATACACTCGTGCATGCA 79136


RESULT 80
ADC87404/c
ID   ADC87404 standard; DNA; 1441 BP.
XX
AC   ADC87404;
XX
DT   01-JAN-2004  (first entry)
XX
DE   Human GPCR gene SEQ ID NO:1857.
XX
KW   ds; gene; human; GPCR;
KW   guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1270724-A2.
XX
PD   02-JAN-2003.
XX
PF   18-JUN-2002; 2002EP-00013517.
XX
PR   18-JUN-2001; 2001JP-00246789.
XX
PA   (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA   (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI   Suwa M,  Asai K,  Akiyama Y,  Aburatani H;
XX
DR   WPI; 2003-315783/31.
DR   P-PSDB; ADC87405.
XX
PT   New polynucleotide, useful for preparing a composition for treating a
PT   patient in need of increased or suppressed activity or expression of the
PT   guanosine triphosphate-binding protein coupled receptor.
XX
PS   Claim 1; SEQ ID NO 1857; 28pp; English.
XX
CC   The invention relates to a novel polynucleotide encoding a guanosine
CC   triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC   the invention may have a use in gene therapy. The polynucleotide and
CC   polypeptide are useful for preparing a composition for treating a patient
CC   in need of increased or suppressed activity or expression of the
CC   guanosine triphosphate-binding protein coupled receptor. The
CC   polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC   invention.
```

```
XX
SQ   Sequence 1441 BP; 518 A; 214 C; 219 G; 490 T; 0 U; 0 Other;

  Query Match             5.2%;  Score 241.2;  DB 10;  Length 1441;
  Best Local Similarity   69.5%;  Pred. No. 4.2e-32;
  Matches  357;  Conservative    0;  Mismatches  153;  Indels    4;  Gaps    2;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | || ||||||||| ||| | |||| |   | | | |||||||||| |||||||||| |||
Db        702 TATACATATATATAGATACACATATGCGTATATATACACATATATGTATACACATACATA 643

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              | |  ||||| ||||||||  | | | ||||| || | |||||| |||  |||  |||
Db        642 TGCGTATATACACATATATGTATACACATACATATGTGTATACATATATGTATACACATG 583

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              ||||  |  | |||| |||  |||||||||   ||||| ||| |.   |||| |||||||
Db        582 CATATGTGTATATATTTATGAACATATATATGTATATACACATATGCGTATATACATATA 523

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              |  | ||| ||||  | |||| ||   | | ||| |  |||||||||| |||    ||||
Db        522 TGTATACACATATGTATATATTTATGTACACATATATGTATACACATACATATGTGTATA 463

Qy        467 TACATATATACACATATATATAATATACACACATATATATACACATATATACACACATAT 526
              ||  ||| ||||||||||     |||||||| | |||| |  | |||| |||||| || |
Db        462 TATTTATGTACACATATAT---GTATACACATACATATGTGTATATATTTACACATATGT 406

Qy        527 ATACACATATATACACATATATATACACATATATACACATATATACATATATACACATAT 586
              ||||||||| ||| |  | ||||||    | | | ||| || |||||| ||| | |  | |
Db        405 GTACACATACATATATGTGTATATATTTACACACACATATGTATACACATACATATGTGT 346

Qy        587 ATATACATATATACACATATATACATATATACACATATATACATA-TATACACACATATA 645
              |||||  |||  |||||||| || | | ||||  || | || ||| |  |||| || ||
Db        345 ATATATTTATGCACACATATGTATACACATACGTATGTGTATATATTGCACACATATGTA 286

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              ||||||||||||    | | ||  ||  ||| |||| || ||| ||| ||| || ||||
Db        285 TACACATACATATGTGTATATATTTATGCACACATATGTATACACATACATATGTGTATA 226

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAG 739
              | | |||||| | |||       | |||| | |
Db        225 TGTTTATATACACATGTATACACATGCATATGTG 192


RESULT 81
ADC87576
ID   ADC87576 standard; DNA; 1276 BP.
XX
AC   ADC87576;
XX
DT   01-JAN-2004  (first entry)
XX
DE   Human GPCR gene SEQ ID NO:2029.
XX
KW   ds; gene; human; GPCR;
KW   guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1270724-A2.
XX
PD   02-JAN-2003.
XX
PF   18-JUN-2002; 2002EP-00013517.
XX
PR   18-JUN-2001; 2001JP-00246789.
XX
PA   (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA   (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI   Suwa M,  Asai K,  Akiyama Y,  Aburatani H;
XX
DR   WPI; 2003-315783/31.
DR   P-PSDB; ADC87577.
XX
PT   New polynucleotide, useful for preparing a composition for treating a
PT   patient in need of increased or suppressed activity or expression of the
PT   guanosine triphosphate-binding protein coupled receptor.
XX
PS   Claim 1; SEQ ID NO 2029; 28pp; English.
XX
CC   The invention relates to a novel polynucleotide encoding a guanosine
CC   triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC   the invention may have a use in gene therapy. The polynucleotide and
CC   polypeptide are useful for preparing a composition for treating a patient
CC   in need of increased or suppressed activity or expression of the
CC   guanosine triphosphate-binding protein coupled receptor. The
CC   polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC   invention.
XX
SQ   Sequence 1276 BP; 416 A; 198 C; 221 G; 441 T; 0 U; 0 Other;

  Query Match             5.2%;  Score 240.8;  DB 10;  Length 1276;
  Best Local Similarity   69.9%;  Pred. No. 4.9e-32;
  Matches  341;  Conservative    0;  Mismatches  142;  Indels    5;  Gaps    1;

Qy        230 ATATATATATATATATATATATACACACACACACATATATATATACACATATATATAC 289
              | ||||| |||| || ||||||||| |   ||| | |||  |||||||| || || ||
Db        793 ACATATACATATGTACATATATACATATGTACATATATACGTATATACATATGTACATCT 852

Qy        290 ACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATAT 349
              |  |||||||  ||||  |||||||  | |||   ||||  |||  | | || ||  ||
Db        853 ATGTATATACGCATATGTGCATATATGTATATACGCATACGTATATGTGTATATACGCAT 912

Qy        350 ATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATAC 409
              |||||||  || || ||| |  ||||||| |||||||||| |   |  ||| |  |||||||
Db        913 ATATACATATGTGCATATATGTATATACGCATATATACATATGTGCATATATGTATATAC 972

Qy        410 ACACATATATACACATATATACACATATATACACATATACACATATATACACATATATAC 469
              || ||||| | |  |  ||| |  ||||||| |||||| |||||| | || |||| ||
Db        973 GCATATATACATATGTGCATATATGTATATACGCATATATACATATGTGCATATATGTATAT 1032

Qy        470 ATATATACACATATATATAATATACACACATATATATACACATATATACACACATATATA 529
              |||   | | |||       ||| |    ||||| |  | | ||  |||   |||| |||
Db       1033 ATACGCATATATA-----CATACATGTGCATATGTGCATATATTCATAAGTACATCTATG 1087

Qy        530 CACATATATACACATATATATACACATATATACACATATATACATATATACACACATATATA 589
              | || || | | || || |||||||||| |||| |||| |||||| | | |  || |||
Db       1088 TATATGTACATATATGTACATACACATATGTACATATATGTACATACACATATGTACATA 1147

Qy        590 TACATATATACACATATATACATATATACACATATATACATATATACACACATATATACA 649
              ||  || |||||||||| | |||| |||| ||| ||||||||||  ||   || |||| |
Db       1148 TATGTACATACACATATACATATGTACATATACATACATATATGCATGTATGTATATA 1207

Qy        650 CATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATATG 709
                 |||||||  | | ||| |||||  |||||||||||||||| |||  | || |||||
Db       1208 TGCACATATACCTATATATACATATATGTATATATACACATATACATATATGTACATATG 1267

Qy        710 TATATATA 717
              | ||||||
Db       1268 TGTATATA 1275


RESULT 82
ABL33869/c
ID   ABL33869 standard; DNA; 13573 BP.
XX
AC   ABL33869;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 1842.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
```

```
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP007537.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful for
PT   diagnosis and treatment of diseases associated with abnormal cytosine
PT   methylation.
XX
PS   Claim 1; SEQ ID NO 1842; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention
XX
SQ   Sequence 13573 BP; 4074 A; 204 C; 2749 G; 6546 T; 0 U; 0 Other;

  Query Match             5.2%;  Score 240.4;  DB 6;  Length 13573;
  Best Local Similarity   69.2%;  Pred. No. 6.4e-32;
  Matches  537;  Conservative    0;  Mismatches  206;  Indels   33;  Gaps   14;

Qy          4 TCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAA--TACAAAATTAG 61
              | || |||| ||| | ||| ||    ||| ||| ||||||| |||||     |||||||
Db       6406 TTAAAACCAACCTAAACAATATAACAAAATCCCGTCTCTACAAAAAAATACAAAAATTAA 6347

Qy         62 ACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAATC 121
               ||  | |  | | |||| ||| |||||||| ||||||   |  |  |   |  | ||||
Db       6346 CCAAACTTAATAACACACACCTATAATCCCAACTACTTAAAAAACTAAAATAAAAAAATC 6287

Qy        122 ACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAGCC 181
              |||| ||||           | |       | ||  ||| | ||| | ||||||| ||
Db       6286 ACTTAAACCCAAAAAATAAAAAC-TACAATAAACCATAATCGTACCACTACACTCCAACC 6228

Qy        182 CAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATAT---A 238
              |  ||| || |   ||   | |    | |||||| | |       | |||||||||   |
Db       6227 TAAACAATAAAACAAAACCCTAACAAAAAAAAAAATATATATATATATATATATATAA 6168

Qy        239 TATATATATATATACACACACACACATATATATAT-ACACATATATATACACATATAT 297
              || |||||||||||  | | |||   |||||||||| | ||||||||||| || |||||
Db       6167 TACATATATATATACGTATATACATTTATATATATATAATACATATATATATACGTATAT 6108

Qy        298 ACATATATACACATAT-ATACACATATATATACACATACATATGTACACATATATATACA 356
              |||| |||| | |||| ||||| ||||||||||  ||| | || || | ||||||| | |
Db       6107 ACATTTATATATATATAATACATATATATATACGTATATACATTTATATATATATA-ATA 6049

Qy        357 CATATGTATACACATATATACACATATAT-------ACACACATATATACACATATATAC 409
              ||||| |||| || ||||||||  |||||       | ||| ||||||| || |||||||
Db       6048 CATATATATATACGTATATACATTTATATATATATAATACATATATATATACGTATATAC 5989

Qy        410 ACACATAT-------ATACACATATATACACATATATACACATATACACATATATACACA 462
              |    ||||       ||||| ||||||| || ||||||||  |||| | ||||| | |||
Db       5988 ATTTATATATATATAATACATATATATATACGTATATACATTTATATATATATA-ATACA 5930

Qy        463 TATATACATATATACACATATATATAATATACACACATATATATACACATATATACACAC 522
              |||||| |||||||||  ||||||||     | ||| |||||||||| ||||| |   |
Db       5929 TATATATATATATACATTTATATATATATAATACATATATATATACATATATACATTTAT 5870

Qy        523 ATATAT---ACACATATATACACATATATATACACATATATACACATATATACATATATA 579
              ||||||   | |||||||||| | | |||||||||  ||||||  | ||| ||||||||||
Db       5869 ATATATATAATACATATATATATACATATATACATTTATATATATATA-ATACATATATA 5811

Qy        580 CACATATATATACATATATACACAT---ATATACATATATACACATATATACATATATAC 636
              | | ||||||||||  |||| | ||   ||| | ||||||| ||||||||||| ||||
Db       5810 TATACATATATACATTTATATATATATAATACATATATATATACATATATACATTTATAT 5751

Qy        637 ACACATATATACACATACATATACACACACATAGATATACATATAT-ATACACATATATA 695
              | | ||| ||||| ||| ||||||| | | | |  |||| |||||| ||||| |||||||
Db       5750 ATATATA-ATACATATATATATACATATATACATTTATATATATATAATACATATATATA 5692

Qy        696 TACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGAT 751
              ||| |||||| || ||||||||||| | || |  |    | |  ||    || ||
Db       5691 TACATATATACATTTATATATATATAATACATATATATATACATATATACATTTAT 5636


RESULT 83
ABK83460/c
ID   ABK83460 standard; cDNA; 165199 BP.
XX
AC   ABK83460;
XX
DT   14-AUG-2002  (first entry)
XX
DE   Human cDNA differentially expressed in granulocytic cells #31.
XX
KW   Human; ss; granulocytic cell; DNA chip; bacterial infection;
KW   viral infection; parasitic infection; protozoal infection;
KW   fungal infection; sterile inflammatory disease; psoriasis;
KW   rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KW   cardiac reperfusion injury; renal reperfusion injury; ARDS;
KW   adult respiratory distress syndrome; inflammatory bowel disease;
KW   Crohn's disease; ulcerative colitis; periodontal disease;
KW   granulocyte activation; chronic inflammation; allergy.
XX
OS   Homo sapiens.
XX
PN   WO200228999-A2.
XX
PD   11-APR-2002.
XX
PF   03-OCT-2001; 2001WO-US030821.
XX
PR   03-OCT-2000; 2000US-0237189P.
XX
PA   (GENE-) GENE LOGIC INC.
XX
PI   Beazer-Barclay Y,  Weissman SM,  Yamaga S,  Vockley J;
XX
DR   WPI; 2002-435328/46.
XX
PT   Detecting granulocyte activation by detecting differential expression of
PT   genes associated with granulocyte activation, which serves as diagnostic
PT   markers that is useful for monitoring disease states and drug toxicity.
XX
PS   Claim 1; SEQ ID NO 31; 114pp; English.
XX
CC   The invention relates to detecting (M1) granulocyte (GC) activation
CC   (GCA), by detecting the level of expression of gene(s) (Gs) identified by
CC   DNA chip analysis as given in the specification, and comparing the
CC   expression level to an expression level in an unactivated GC, where
CC   differential expression of Gs is indicative of GCA. Also included are
CC   modulating (M2) GA by contacting GC with an agent that alters the
CC   expression of at least one gene in Gs; (2) screening (M3) for an agent
CC   capable of modulating GCA or an inflammation (especially chronic) in a
```

CC tissue, an allergic response in a subject, exposure of a subject to a
CC pathogen or sterile inflammatory disease using the gene expression
CC profile; (3) detecting (M4) an inflammation (especially chronic) in a
CC tissue, an allergic response in a subject, exposure of a subject to a
CC pathogen or sterile inflammatory disease, by detecting the level of
CC expression in a sample of the tissue of gene(s) from Gs, where the level
CC of expression of the gene is indicative of inflammation; (4) treating
CC (M5) an inflammation (especially chronic) or in a tissue, an allergic
CC response in a subject, exposure of a subject to a pathogen or sterile
CC inflammatory disease, by contacting a tissue having inflammation with an
CC agent that modulates the expression of gene(s) from Gs in the tissue. M1
CC is useful for detecting GCA; M2 is useful for modulating GA; M3 is useful
CC for screening an agent capable of modulating GCA preferably in an
CC inflammation in a tissue; M4 is useful for detecting an inflammation
CC (especially chronic) in a tissue, an allergic response in a subject,
CC exposure of a subject to a pathogen or sterile inflammatory disease (e.g.
CC psoriasis, rheumatoid arthritis, glomerulonephritis, asthma, thrombosis,
CC cardiac reperfusion injury, renal reperfusion injury, ARDS, adult
CC respiratory distress syndrome, inflammatory bowel disease, Crohn's
CC disease, ulcerative colitis, periodontal disease; also bacterial
CC infection, viral infection, parasitic infection, protozoal infection,
CC fungal infection and M5 is useful for treating one of the above
CC conditions. The present sequence represents a gene differentially
CC expressed in granulocytes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 165199 BP; 48510 A; 33223 C; 34406 G; 49060 T; 0 U; 0 Other;

Query Match 5.2%; Score 240.2; DB 6; Length 165199;
Best Local Similarity 66.3%; Pred. No. 7.7e-32;
Matches 377; Conservative 0; Mismatches 186; Indels 6; Gaps 2;

```
Qy      234 ATATATATATATATATATACACACACACACATATATATATACACATATATATACACAT 293
            ||  |||||||||||||| ||||||  | | | |||||||||||| | |||||||||  |||
Db   139452 ATGGATATATATATATACACACGGATATATATATATATATATAGATATATATATCCAT 139393

Qy      294 ATATACATATATACACATATATACACATATATATACACATACATATGTACACATATATAT 353
            ||||  ||||||| |   ||||| | || ||||||   ||| |||| |  | ||||  ||
Db   139392 ATATGGATATATATATCGATATATATATCTATATATGGATATATATATCTATATATGGAT 139333

Qy      354 ACACATATGTATACACATATATACACATATATACACACATATATACACATATATACACAC 413
            | | ||||| ||| | |||| || | ||  ||| | |   ||||| | ||  ||| | |
Db   139332 ATATATATGGATATATATATCTATATATGGATATATATGGATATATATATGGATATATAT 139273

Qy      414 ATATATACACATATATACACATATATACACATATACACATATATACACATATATACATAT 473
            ||||||   |||||||  |    ||||| | || || | ||  ||| | ||||  | ||||
Db   139272 CTATATATGGATATATATATGGATATATATATCTATATATGGATATATATATGGATATAT 139213

Qy      474 ATACACATATATATA-ATATACACACATATATATACACATATATACACACATATATACAC 532
            |||   ||||||  | ||||| |   |||||||||   ||||||  | | ||||||  |
Db   139212 ATATCTATATATGGATATATATATGGATATATATATCTATATATGGATATATATATGGAT 139153

Qy      533 ATATATACACATATATATACACATATATACACATATATACATATATACACATATATATAC 592
            |||||||   ||||||  | | ||||||  | |||||||  ||||||   |||||||  |
Db   139152 ATATATATCTATATATGGATATATATATGGATATATATAACTATATATGGATATATACAT 139093

Qy      593 ATATATACACATATATACATATATACACATATATACATATATACACACATATATACACAT 652
            ||||||   ||||||| ||  ||      |||||| | ||||| | | |||||||   ||
Db   139092 CTATATATGGATATATATATTGAT-----TATATATAGATATATAGATATATATATGGAT 139038

Qy      653 ACATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTAT 712
            | |||||   | | || |||||| ||  ||||| | || |||||| | |||||||||  ||
Db   139037 ATATATATCTATATATGGATATATATGGATATATATATCTATATATGGATATATATGGAT 138978

Qy      713 ATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAA 772
            |||||||||    | |  |     | |  |||    |  ||    | |    |||||   |
Db   138977 ATATATATGGATTATATATAGATATATATAGATATATATATGGATTATATATAGATATAT 138918

Qy      773 AGAGAGTTTGACAGCAGTGTATTGTGAGA 801
            | |||  |  | | |  | |||  |  | |
Db   138917 ATAGATATATATATCTATATATGGATATA 138889
```

RESULT 84
ABD32673
ID ABD32673 standard; DNA; 91359 BP.
XX
AC ABD32673;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human cancer-associated genomic DNA HD13-075.
XX
KW Human; ds; cancer-associated protein; gene; cytostatic; cancer;
KW leukaemia; lymphoma; CAP.
XX
OS Homo sapiens.
XX
PN WO2004074320-A2.
XX
PD 02-SEP-2004.
XX
PF 17-FEB-2004; 2004WO-US004730.
XX
PR 14-FEB-2003; 2003US-00367094.
PR 14-MAR-2003; 2003US-00388838.
PR 15-APR-2003; 2003US-00417375.
PR 13-JUN-2003; 2003US-00461862.
PR 15-SEP-2003; 2003US-00663431.
PR 15-DEC-2003; 2003US-00737318.
XX
PA (SAGR-) SAGRES DISCOVERY INC.
XX
PI Morris DW, Morris DW, Malandro MS;
XX
DR WPI; 2004-652914/63.
XX
PT New isolated cancer-associated polynucleotides and polypeptides useful
PT for diagnosing, preventing or treating cancers, especially lymphoma and
PT leukemia, or in screening for agents that modulate cancer.
XX
PS claim 16; seqid 221; 310pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising at least 10
CC contiguous nucleotides of any of the 233 polynucleotide sequences given
CC in the specification, or its complement. The nucleic acids encode cancer-
CC associated proteins. Also included are an expression vector comprising
CC the isolated nucleic acid cited above, a host cell comprising the above
CC recombinant nucleic acid or expression vector, a microarray for detecting
CC a cancer-associated (CA) nucleic acid comprising at least one probe
CC comprising at least 10 contiguous nucleotides of any of the above-
CC mentioned nucleotide sequences, an isolated polypeptide (encoded within
CC an open reading frame of a CA sequence selected from any of the 95
CC polynucleotide sequences as mentioned in the specification, or its
CC complement), an isolated antibody, (or its antigen binding fragment) that
CC binds to the above polypeptide, a hybridoma that produces the above
CC monoclonal antibody, a pharmaceutical composition comprising the above
CC antibody and a pharmaceutical excipient, a kit for detecting cancer
CC cells(comprising the antibody cited above, methods for diagnosing cancer
CC or for detecting the presence or absence of cancer cells in an
CC individual, a method for inhibiting growth of cancer cells in an
CC individual, a method for delivering a therapeutic agent to cancer cells
CC in an individual, an electronic library comprising the above
CC polynucleotide or polypeptide (or their fragments), methods of screening
CC for anticancer activity or for a bioactive agent capable of modulating
CC the activity of a CA protein (CAP), methods for detecting cancer
CC associated with expression of a polypeptide in a test cell sample, a
CC method for treating cancers and a method for inhibiting the expression of
CC CA gene in a cell. The composition and methods are useful for detecting,
CC diagnosing, preventing and treating cancers, especially lymphoma and
CC leukaemia. These may also be used in screening for agents that modulate
CC cancer. The present sequence is a human CAP genomic sequence. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO

```
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 91359 BP; 24503 A; 18193 C; 19029 G; 29634 T; 0 U; 0 Other;

  Query Match             5.2%;  Score 238.8;  DB 13;  Length 91359;
  Best Local Similarity   68.8%;  Pred. No. 1.3e-31;
  Matches  373;  Conservative    0;  Mismatches  162;  Indels    7;  Gaps    3;

Qy        204 ATCTCAGAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACAC 263
              | || || |   |    || ||   | |      || |||||| || |||| |    | | |
Db      19526 AGCTGAGTAGCATTCAATGGTGTATATATAATCTATATACATCTATACATCTATATATCT 19585

Qy        264 ACATATATATATACACATATATATACACATATATACATATATACACATATATACACATAT 323
              | ||| |||||||||| || ||| || | || | || ||| ||| | | ||||  | ||||
Db      19586 ATATACATATATACATATGTATGTATATATTTCTATATACATATATACATATGTATATAT 19645

Qy        324 ATATACACATACATATGTACACATATATATACACATATGTATACACATATATACACATAT 383
              | ||  | |||||| | || |||||| |||||| | |||||||||  ||| || |||||
Db      19646 AGATTGATATACATCTATATACATATGTATACATAGATGTATACATCTATGTATACATAG 19705

Qy        384 ATACACACATATATACACATATATACACACATATATACACATATATACACATATATACAC 443
              ||  | |||||       || |||| ||| |||| || ||||||||||| || ||||||
Db      19706 ATGTATACATA-----CTATGTATATACATATATGTATACATATATACATATGTATACAT 19760

Qy        444 ATATACACATATATACACATATATACATATATACACATATATATA-ATATACACACATAT 502
               |||  | | ||||| |||||| || ||||||  | ||||| ||| || || | |  |||
Db      19761 CTATGTATACATATATACATATGTATATATATCTATATATACATATATGTATATATGTAT 19820

Qy        503 ATATACACATATATACACACATATATACACATATATACACATATATATACACATATATAC 562
              ||||  | |||  || | |  |||||||| ||||| | | || |||||||| || ||||
Db      19821 ATATCTATATACGTATATAGGTATATACATATATAAATATATCTATATACATATGTATAT 19880

Qy        563 ACATATATACATATATACACATATATATACATATATACACATATATACATATATACACAT 622
              |  |||||| ||||||  | |||||| |||||||||  | ||| ||| || |||| |  |
Db      19881 ATCTATATAAATATATGGATATATATCTACATATATCTATATACATATATCTATATATCT 19940

Qy        623 ATATACATATATACACACATATATACACAT-ACATATACACACACATAGATATACATATA 681
              |||||||||||  |   | || |||  ||||  |||||  | | | ||| |||||||| ||
Db      19941 ATATACATATAGATGTATATGTATCTACATGGCATATGTATATAGATATATATACATCTA 20000

Qy        682 TATACACATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGT 741
              |||  | || ||||||| ||| |||||   | |||||||   ||| |  |    | || |
Db      20001 TATGTATATCTATATACATATCTATATAGCTGTATATATATGCCATATATATAGATAGAT 20060

Qy        742 AG 743
              ||
Db      20061 AG 20062


RESULT 85
ABL51954
ID   ABL51954 standard; DNA; 40023 BP.
XX
AC   ABL51954;
XX
DT   11-JUL-2002  (first entry)
XX
DE   Human solute carrier family 18 member 2 (SLC18A2) gene SEQ ID NO:1.
XX
KW   Human; solute carrier family 18 member 2; SLC18A2; vesicular monoamine;
KW   vesicular monoamine transporter; VMAT2; polymorphic site; SNP;
KW   single nucleotide polymorphism; antiinflammatory; neuroleptic;
KW   haplotyping; genotyping; respiratory inflammatory disease;
KW   neuropsychiatric disorder; monoaminergic brain system; gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   misc_feature    4024
FT                   /*tag=  a
FT                   /note= "PS1: polymorphic base C or T"
FT   misc_feature    6142
FT                   /*tag=  b
FT                   /note= "PS2: polymorphic base C or G"
FT   misc_feature    6211
FT                   /*tag=  c
FT                   /note= "PS3: polymorphic base G or A"
FT   misc_feature    6213
FT                   /*tag=  d
FT                   /note= "PS4: polymorphic base G or T"
FT   misc_feature    6397
FT                   /*tag=  e
FT                   /note= "PS5: polymorphic base C or T"
FT   misc_feature    6398
FT                   /*tag=  f
FT                   /note= "PS6: polymorphic base T or G"
FT   misc_feature    16326
FT                   /*tag=  g
FT                   /note= "PS7: polymorphic base T or C"
FT   misc_feature    16464
FT                   /*tag=  h
FT                   /note= "PS8: polymorphic base C or T"
FT   misc_feature    16855
FT                   /*tag=  i
FT                   /note= "PS9: polymorphic base T or C"
FT   misc_feature    16959
FT                   /*tag=  j
FT                   /note= "PS10: polymorphic base T or A"
FT   misc_feature    16997
FT                   /*tag=  k
FT                   /note= "PS11: polymorphic base T or C"
FT   misc_feature    17761
FT                   /*tag=  l
FT                   /note= "PS12: polymorphic base C or T"
FT   misc_feature    17763
FT                   /*tag=  m
FT                   /note= "PS13: polymorphic base T or C"
FT   misc_feature    17780
FT                   /*tag=  n
FT                   /note= "PS14: polymorphic base A or C"
FT   misc_feature    17897
FT                   /*tag=  o
FT                   /note= "PS15: polymorphic base G or C"
FT   misc_feature    20047
FT                   /*tag=  p
FT                   /note= "PS16: polymorphic base C or T"
FT   misc_feature    20070
FT                   /*tag=  q
FT                   /note= "PS17: polymorphic base T or A"
FT   misc_feature    29050
FT                   /*tag=  r
FT                   /note= "PS18: polymorphic base G or T"
FT   misc_feature    29230
FT                   /*tag=  s
FT                   /note= "PS19: polymorphic base C or T"
FT   misc_feature    29395
FT                   /*tag=  t
FT                   /note= "PS20: polymorphic base C or T"
FT   misc_feature    32343
FT                   /*tag=  u
FT                   /note= "PS21: polymorphic base C or T"
FT   misc_feature    32572
FT                   /*tag=  v
FT                   /note= "PS22: polymorphic base A or G"
FT   misc_feature    39464
FT                   /*tag=  w
FT                   /note= "PS23: polymorphic base T or C"
FT   misc_feature    39766
FT                   /*tag=  x
FT                   /note= "PS24: polymorphic base C or T"
FT   misc_feature    39848
FT                   /*tag=  y
FT                   /note= "PS25: polymorphic base A or C"
FT   misc_feature    39910
FT                   /*tag=  z
FT                   /note= "PS26: polymorphic base T or C"
```

```
FT   misc_feature    39984
FT                   /*tag=  aa
FT                   /note= "PS27: polymorphic base T or C"
XX
PN   WO200222652-A2.
XX
PD   21-MAR-2002.
XX
PF   17-SEP-2001; 2001WO-US042217.
XX
PR   15-SEP-2000; 2000US-0232895P.
XX
PA   (GENA-) GENAISSANCE PHARM INC.
XX
PI   Anastasio AE,  Han J,  Kliem SE,  Sausker EA;
XX
DR   WPI; 2002-393942/42.
XX
PT   Novel genetic variants of soluble carrier family 18 (vesicular
PT   monoamine), member 2 gene useful for screening drugs to treat diseases
PT   e.g. neuropsychiatric disorders involving monoaminergic brain systems.
XX
PS   Claim 15; Fig 1; 183pp; English.
XX
CC   The present invention describes an isolated polynucleotide (I) having a
CC   sequence (S1) comprising soluble carrier family 18 (vesicular monoamine),
CC   member 2 (SLC18A2) isogene selected from 49 isogenes with regions of a
CC   sequence (SS) of 40023 bp (see ABL51954), and defined by a corresponding
CC   set of polymorphisms whose locations and identities are given in the
CC   specification; or a sequence (S2) complementary to (S1). (I) has
CC   antiinflammatory and neuroleptic activities, and can be used in gene
CC   therapy. Methods from the present invention can be used for haplotyping
CC   and genotyping the SLC18A2 gene in an individual. SLC18A2 is also known
CC   as the vesicular monoamine transporter (VMAT2). (I) is useful in studying
CC   the expression and function of SLC18A2, and in expressing the SLC18A2
CC   protein for use in screening for candidate drugs to treat diseases
CC   related to SLC18A2 activity and in studying the effect of the variation
CC   on the biological activity of SLC18A2 as well as on the binding affinity
CC   of candidate drugs targeting SLC18A2 for the treatment of respiratory
CC   inflammatory diseases such as neuropsychiatric disorders involving
CC   monoaminergic brain systems. The present sequence represents the human
CC   SLC18A2 gene sequence, indicating the polymorphic sites in the sequence
CC   by the ambiguity code
XX
SQ   Sequence 40023 BP; 10020 A; 9128 C; 9557 G; 11291 T; 0 U; 27 Other;

  Query Match             5.1%;  Score 237.4;  DB 6;  Length 40023;
  Best Local Similarity   71.2%;  Pred. No. 2.2e-31;
  Matches  341;  Conservative    0;  Mismatches  136;  Indels    2;  Gaps    2;

Qy        230 ATATATATATATATATATATATACACACACACACACATATATATATACACATATATATAC 289
              | ||| | ||  ||| | | | |||||||| || ||| | ||| | ||||  | | |  |
Db      31298 ACATACACATTCATACACAAACACACACACTCATACACACATACACACACTGACACACGC 31357

Qy        290 ACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATAT 349
              | | | | ||| ||| ||||||| |  | |  | ||| |||||||| ||  ||| ||||
Db      31358 ATACACACACACATACACACATACACCCCCCCACATAAACACATACACATACACAAATAT 31417

Qy        350 ATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATAC 409
              | ||||||| || |  | ||||| |||||| || ||||||| | | |||||  | |   ||
Db      31418 ACATACACACATATGCATACATACATACACGTAGATACACATACACATACAAACACCCAC 31477

Qy        410 ACACATATATACACATATATACACATATATACACATATACACATATATACACATATATAC 469
               |||||| |||  || | ||||||| | | | |||||||||||| ||| ||||| |  |||
Db      31478 CCACATACATATGCACACATACACACACACATACATATACACACATACACACACCTAC 31537

Qy        470 ATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATATAT 528
              | | |||||||||| | | | | | | | |||  || ||||||| | ||| ||||| | |
Db      31538 ACACATACACATACACACACACACAAATACAAGTACATACACAAACATATACACAGACAC 31597

Qy        529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACATATAT 588
              | ||| ||  | ||| || || |||||||||||||||||| ||||| | ||||| |||||
Db      31598 ATACACATGCAAACACATGTACACACATATATACACATACATACACAGATACATATATAC 31657

Qy        589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
              ||| | ||| | ||| ||| | | | | |  |||| | ||||||| | | |||||  | |
Db      31658 ATATACATA-ATACACATAAATACACACATGCATACACACATATACATAGACATACACAG 31716

Qy        649 ACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATA 707
              | | | ||||||| | ||||| |||  |||| ||| |||||||| ||||  | ||| |
Db      31717 ATACATATATACATATACATACATACTCATAAATACACACATATACATACAAACATACA 31775


RESULT 86
ADR07428
ID   ADR07428 standard; cDNA; 1821 BP.
XX
AC   ADR07428;
XX
DT   04-NOV-2004  (first entry)
XX
DE   Full length human cDNA useful for treating neurological disease Seq 934.
XX
KW   gene; ss; human; oligo-capping method; diagnostic marker; gene therapy;
KW   osteoporosis; neurological disease; Alzheimer's disease;
KW   Parkinson's disease; dementia; short memory; cancer;
KW   sense or motor function; emotional reaction; fear response; panic;
KW   osteopathic; neuroprotective; nootropic; antiparkinsonian; cytostatic;
KW   tranquiliser.
XX
OS   Homo sapiens.
XX
PN   EP1447413-A2.
XX
PD   18-AUG-2004.
XX
PF   12-FEB-2004; 2004EP-00003145.
XX
PR   14-FEB-2003; 2003JP-00102207.
PR   09-MAY-2003; 2003JP-00131452.
XX
PA   (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI   Isogai T,  Yamamoto J,  Nishikawa T,  Isono Y,  Sugiyama T,  Otsuki T;
PI   Wakamatsu A,  Ishii S,  Nagai K,  Irie R;
XX
DR   WPI; 2004-583265/57.
DR   P-PSDB; ADR09384.
XX
PT   New 1995 cDNA, useful for treating osteoporosis, neurological diseases,
PT   Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.
XX
PS   Claim 1; SEQ ID NO 934; 2686pp; English.
XX
CC   This invention relates to novel, isolated full length human cDNA
CC   molecules and the encoded proteins thereof. Specifically, it refers to
CC   cDNA clones obtained by an oligo-capping method, where none of these
CC   clones are identical to any known human mRNAs. The present invention
CC   describes an immunoassay to identify agonists and antagonists, as well as
CC   antibodies, antisense molecules and siRNAs that can all be used to bind
CC   to and modulate expression of the cDNA molecules. As such, these
CC   molecules are useful for diagnostic markers or therapeutic targets for
CC   the various diseases or morbid states. In particular, they are useful in
CC   gene therapy for treating osteoporosis, neurological disease, Alzheimer's
CC   disease, Parkinson's disease, dementia, short memory and various cancers,
CC   as well as for maintaining equilibrium of sense or motor function, and
CC   for treating emotional reaction, fear response and panic. Accordingly,
CC   they exhibit osteopathic, neuroprotective, nootropic, antiparkinsonian,
CC   cytostatic and tranquiliser activities. This polynucleotide is a full
CC   length human cDNA sequence of the invention. NOTE: This sequence is not
CC   given in the sequence listing of the specification but can be obtained on
CC   CD-ROM from the European Patent Office, Vienna Sub-office.
XX
SQ   Sequence 1821 BP; 752 A; 616 C; 149 G; 304 T; 0 U; 0 Other;

  Query Match             5.1%;  Score 236.6;  DB 13;  Length 1821;
```

```
  Best Local Similarity   69.0%;  Pred. No. 2.6e-31;
  Matches  339;  Conservative    0;  Mismatches  149;  Indels    3;  Gaps    1;

Qy        230 ATATATATATATATATATATATACACACACACACACATATATATATACACATATATATAC 289
              | | ||  | ||| |    | || |||||||| ||||||  | ||||||||||  | | ||
Db         44 ACACATGCACATACACTCACATGCACACACATACACATGCACATATACACATGCACACAC 103

Qy        290 ACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATAT 349
              |||  ||||||| | ||  || ||||||||||  | |||||||  ||||   ||||| ||
Db        104 GCATGCATACATACACACGCACATATACACATGCACATACACACTCATACACACACACAT 163

Qy        350 ATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATAC 409
              ||| |||||||   || || | | || |||  ||  |||||||  | ||||||||  | |
Db        164 ATACACACATACACATGCATACACATGCACTCATGCACACACACACATACACATGCACAT 223

Qy        410 ACACATATATACACATATATACACATATATACACATATACACATATATACACATATATAC 469
              ||||| | || || | |||| |||||||| |||   | | |||||  |  |||||  | ||
Db        224 ACACAAACATGCAGAGATATGCACATATACACATGCACATACATACACTCACATGCACAC 283

Qy        470 ATATATACACATATATAT---AATATACACACATATATATACACATATATACACACATAT 526
              ||| | ||||||| | ||    ||||||| ||||  | |||||||  | || ||||||  |
Db        284 ATAGACACACATACACATGCACATATACATACATGCACATACACAAACATGCACACAGAC 343

Qy        527 ATACACATATATACACATATATATACACATATATACACATATATACATATATACACATAT 586
              | ||| | | ||||||| ||  | | ||| | ||||||  |||| |||| ||| ||| ||
Db        344 ACACATACACATACACACATGCACATACACAGATACACTCATATGCATACATATACACAT 403

Qy        587 ATATACATATATACACATATATACATATATACACATATATACATATATACACACATATAT 646
              |   ||||||| |||   | ||||| ||| ||||||  || || ||||||||||||  |
Db        404 ACGCACATATACACATGAACATACACATACACACATGCATGCACATATACACACATGCAC 463

Qy        647 ACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATAT 706
              ||||| |   | | | |||||  |   |||||  | | || |||| | | |  ||  | ||
Db        464 ACACACATGCACATATACACACATGCATACACACACATGCACACAGACACACACACATGC 523

Qy        707 ATGTATATATA 717
              |   | ||| |
Db        524 ACACACATACA 534


RESULT 87
ACC79619/c
ID   ACC79619 standard; DNA; 71843 BP.
XX
AC   ACC79619;
XX
DT   27-OCT-2003  (revised)
DT   06-AUG-2003  (first entry)
XX
DE   Zebrafish protein kinase C lambda related genomic DNA SEQ ID NO:3.
XX
KW   Zebrafish; Danio rerio; protein kinase C lambda; PKC-lambda; enzyme;
KW   cardiant; cytostatic; nephrotropic; heart disease; cancer; organ failure;
KW   epithelial-epithelial cell interaction; epithelial cell polarity;
KW   heart failure; kidney failure; gene; ds.
XX
OS   Danio rerio.
XX
PN   WO2003023048-A2.
XX
PD   20-MAR-2003.
XX
PF   06-SEP-2002; 2002WO-US028410.
XX
PR   06-SEP-2001; 2001US-0317653P.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Peterson R,  Fishman MC;
XX
DR   WPI; 2003-313257/30.
XX
PT   Diagnosing or prognosing a disease or condition related to Protein Kinase
PT   C lambda, by analyzing sample nucleic acid obtained from the test subject
PT   to determine mutation in gene encoding Protein kinase C lambda.
XX
PS   Disclosure; Page 63-83; 88pp; English.
XX
CC   The present invention describes a method (M1) for determining whether a
CC   test subject has, or is at risk of developing, a disease or condition
CC   related to protein kinase C lambda (PKC-lambda) (I). M1 comprises
CC   analysing a nucleic acid molecule of a sample from the subject to
CC   determine whether the subject has a mutation in a gene encoding (I),
CC   where the presence of a mutation indicates that the subject has, or is at
CC   risk of developing, a disease or condition related to (I). (I) has
CC   cardiant, cytostatic and nephrotropic activities. M1 is useful for
CC   determining whether a test subject has, or is at risk of developing, a
CC   disease or condition related to (I), e.g. heart disease, cancer, or
CC   condition associated with epithelial-epithelial cell interactions or
CC   epithelial cell polarity. (I) is useful for treating organ failure such
CC   as heart or kidney failure. (I) is also useful for raising antibodies
CC   specific for various regions of (I). M1 is suitable for detecting an
CC   increased likelihood of diseases or conditions associated with (I), so
CC   that appropriate intervention can be instituted before any symptoms
CC   occur. This is useful with patients in high risk groups for such diseases
CC   or conditions. Diagnostic method facilitates determination of the
CC   etiology of such an existing disease or condition in a patient, so that
CC   an appropriate approach to treatment can be selected. The present
CC   sequence represents zebrafish genomic DNA for PKC-lambda, which is used
CC   in the exemplification of the present invention. (Updated on 27-OCT-2003
CC   to standardise OS field)
XX
SQ   Sequence 71843 BP; 24017 A; 12903 C; 12592 G; 21955 T; 0 U; 376 Other;

  Query Match             5.1%;  Score 236;  DB 8;  Length 71843;
  Best Local Similarity   60.9%;  Pred. No. 3.9e-31;
  Matches  314;  Conservative    0;  Mismatches  198;  Indels    4;  Gaps    2;

Qy        209 AGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACACATA 268
              | ||||||||||    |  | |||||||||||||||||||||||| | | | | | | |||
Db      68772 AAAAAAAAAAAATNTCTATATATATATATATATATATATATATATATATATATATATA 68713

Qy        269 TATATATACACATATATATACACATATATACATATATACACATATATACACATATATATA 328
              |||||||| | ||||||||| | ||||||| ||||||| | ||||||| | |||||||||
Db      68712 TATATATATATATATATATATATATATATANATATATATATATATATATATATATATATA 68653

Qy        329 CACATACATATGTACAC---ATATATATACACATATGTATACACATATATACACATATAT 385
              | |||||||| |||||   |||||||| | |||| |||| | ||||||| | ||||||
Db      68652 TATATACATATATACACTATATATATATATATATATATATATATATATATATATATAT 68593

Qy        386 ACACACATATATACACATATATACACACATATATACACATATATACACATATATACACAT 445
              | | | ||||||| | ||||||| | | ||||||| | ||||||| | ||||||| | ||
Db      68592 ATATATATATATATATATATATATATANATATATATATATATATATATATATATATAT 68533

Qy        446 ATACACATATATACACATATATACATATATACACATATATATA-ATATACACACATATAT 504
              ||| | ||||||| | ||||||| ||||||| | ||||||||| ||||| | | ||||||
Db      68532 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 68473

Qy        505 ATACACATATATACACACATATATACACATATATACACATATATATACACATATATACAC 564
              ||| | ||||||
Db      68472 ATATATATATATNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 68413

Qy        565 ATATATACATATATACACATATATATACATATATACACATATATACATATATACACATAT 624
                                                                       | |
Db      68412 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTGT 68353

Qy        625 ATACATATATACACACATATATACACATACATATACACACACATAGATATACATATATAT 684
              |   | | |     | | |    |  ||||| | | | ||| ||||| ||||||||
Db      68352 GTGTGTGTGTGTGTGTGTGTGTGTGTGTATATATATATATATATATATATATATAT 68293

Qy        685 ACACATATATATACGTATATATATGTATATATATAT 720
              | | |||||||||  |||||||||| ||||||||||
Db      68292 ATATATATATATATATATATATATATATATATAT 68257


RESULT 88
```

```
ADC87374/c
ID   ADC87374 standard; DNA; 1192 BP.
XX
AC   ADC87374;
XX
DT   01-JAN-2004  (first entry)
XX
DE   Human GPCR gene SEQ ID NO:1827.
XX
KW   ds; gene; human; GPCR;
KW   guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1270724-A2.
XX
PD   02-JAN-2003.
XX
PF   18-JUN-2002; 2002EP-00013517.
XX
PR   18-JUN-2001; 2001JP-00246789.
XX
PA   (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA   (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI   Suwa M,  Asai K,  Akiyama Y,  Aburatani H;
XX
DR   WPI; 2003-315783/31.
DR   P-PSDB; ADC87375.
XX
PT   New polynucleotide, useful for preparing a composition for treating a
PT   patient in need of increased or suppressed activity or expression of the
PT   guanosine triphosphate-binding protein coupled receptor.
XX
PS   Claim 1; SEQ ID NO 1827; 28pp; English.
XX
CC   The invention relates to a novel polynucleotide encoding a guanosine
CC   triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC   the invention may have a use in gene therapy. The polynucleotide and
CC   polypeptide are useful for preparing a composition for treating a patient
CC   in need of increased or suppressed activity or expression of the
CC   guanosine triphosphate-binding protein coupled receptor. The
CC   polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC   invention.
XX
SQ   Sequence 1192 BP; 384 A; 185 C; 211 G; 412 T; 0 U; 0 Other;

  Query Match             5.1%;  Score 235.8;  DB 10;  Length 1192;
  Best Local Similarity   69.2%;  Pred. No. 3.5e-31;
  Matches  355;  Conservative    0;  Mismatches  147;  Indels   11;  Gaps    2;

Qy        214 AAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACATATATAT 273
              |||||| |      | |||| |||||||| | ||||| || | ||||||| |||| |||||
Db       1000 AAAAAATATTATATATATACATATATATGTTTATATGTATATACACACATACATGTATAT 941

Qy        274 ATACACATATATATACACATATATACATATATACACATATATACACATATATATACACAT 333
              ||   | |||||  || ||| ||||||  | |  ||| ||| || ||       ||||||||||
Db        940 ATGTATATATACATGCACGTATATATACACATATACACATGTATGTGCGTATATACACAT 881

Qy        334 ACATATGTACACATATATATACACATATGTATACACATATATACACATATATACACACAT 393
              |||| | || | | | |||||||||| | | | |  ||||||||    ||||| ||||||
Db        880 ACATGTATATATACACATATACACATGTATGTGCGTATATATAC----ATATATACACAT 825

Qy        394 ATATACACATATATACACACATATATACACATATATACACATATATACACATATACACAT 453
              |||   | | |||| ||| ||||| |||  | | | |   ||||||||  ||||||||||
Db        824 GTATGTGCGTGTATATACATATATACACATGTGTGTGCGTGTATATACATATATACACAT 765

Qy        454 ATATACACATATATACATATATACACATATATAT-------AATATACACACATATATAT 506
              | |   | | |||| | ||||| ||| | ||||          | || | |||||||||
Db        764 GTGTGTGCGTGTATATACATATATACACAGATATGTGTGTGCGTGTATATACATATATAC 705

Qy        507 ACACATATATACACACATATATACACATATATACACATATATATACACATATATACACAT 566
              ||| |||| |      |||||||||| ||| | ||| |||| | |    |||| |||| ||
Db        704 ACAGATATGTGTGTGTATATATACATATACACACAGATATGTGTGTGTATATGTACATAT 645

Qy        567 ATATACATATATACACATATATATACATATATACACATATATACATATATACACATATAT 626
              | | |||| | |  | ||||| ||| ||| ||| ||| | ||   | |||| | | ||||
Db        644 ACACACATGTGTGTATATATACATATATACATATACACACATGTGTGTATATATACATAT 585

Qy        627 ACATATATACACACATATATACACATACATATACACACACATAGATATACATATATATAC 686
              | | |||||||||| | | || | ||||||||| ||| | | | |   |||||||||
Db        584 ATACATATACACACGTGTGTATATATACATATATACATATACACACGTGTGTATATATAC 525

Qy        687 ACATATATATACGTATATATATGTATATATATA 719
              | ||||| | || || || || || ||||||||
Db        524 ATATATACACACATACATGTACGTGTATATATA 492


RESULT 89
ABL33998/c
ID   ABL33998 standard; DNA; 21537 BP.
XX
AC   ABL33998;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 1971.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP007537.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful for
PT   diagnosis and treatment of diseases associated with abnormal cytosine
PT   methylation.
XX
PS   Claim 1; SEQ ID NO 1971; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention
XX
SQ   Sequence 21537 BP; 5893 A; 333 C; 4680 G; 10631 T; 0 U; 0 Other;

  Query Match             5.1%;  Score 235.2;  DB 6;  Length 21537;
  Best Local Similarity   69.6%;  Pred. No. 5.1e-31;
  Matches  348;  Conservative    0;  Mismatches  148;  Indels    4;  Gaps    2;

Qy        197 AAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACA 256
```

```
Db      16071 AAATTAAAACACACAAAAACTATATAACTTATCCAAAAACACACTATAAATCCATATATA 16012

Qy        257 CACACACACATATATATATACACATATATATACACATATATACATATATACACATATATA 316
              | | | | |||||||||||| | ||||| | | | ||| || |||||||| | |||||
Db      16011 TATATATATATATATATATATATATACACATATATAAATCCATATATATATATATACG 15952

Qy        317 CACATATATATACACATACATATGTACACATATATATACACATATGTATACACATATATA 376
              || |||||||   | ||| ||||  ||||| |||||||||||||| || ||||||||||
Db      15951 CATATATATATACGTATATATATATACACACACATATATACACATATATACACACATATATA 15892

Qy        377 CACATATATACACACATATATACACATATATACACACATATATACACATATATACACATA 436
              |||||||| ||| | ||||||| | |||||||||||||| | ||| | ||||||||||| |
Db      15891 CACATATACACATATATATATATATATATACACACACACATATATATATATACACACA 15832

Qy        437 TATACACATATACACATATATACACATATATACATATATACACATATATATA---ATATA 493
              | | | ||||| | | | | |||||||||| |||||||||| | | |||   |||||
Db      15831 CACATATATATATATACACACACACATATATATATATATACACACACACATATATATATA 15772

Qy        494 CACACATATATATACACATATATACACACATATATACACATATATACACATATATATACA 553
              | | ||| | | |||||||||||| |     ||  | | | ||||| | ||||| | | |
Db      15771 TATATATACACACACACATATATATAATATAATCGATATAAATATATATATATAAAAAAA 15712

Qy        554 CATATATACACATATATACATATATACACATATATATACATATATACACATATATACATA 613
              |||||   |  |||| ||          |||  || || | || ||   |   ||  |
Db      15711 AATATACCTATTTATAAACTCTCTATTCCATTAATTTATACATCTAATAAACAATTATTT 15652

Qy        614 TATACACATATATACATATATACACACATATATACACATACATATACACACACATAGATA 673
              |||| | ||||||| |||||||||||||||||||| | ||| |||||  | | |||||  |||
Db      15651 TATATATATATATATATATATACACACATATATATA-ATATATATATATATACATACATA 15593

Qy        674 TACATATATATACACATATA 693
              | | ||| ||| |||||||
Db      15592 CATACATACATATACATATA 15573
```

RESULT 90
AAS46748
ID AAS46748 standard; DNA; 26997 BP.
XX
AC AAS46748;
XX
DT 18-DEC-2001 (first entry)
XX
DE Tumour suppressor gene derived chemically modified sequence #472.
XX
KW Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer;
KW tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW cytosine methylation; ds.
XX
OS Homo sapiens.
XX
PN WO200168912-A2.
XX
PD 20-SEP-2001.
XX
PF 15-MAR-2001; 2001WO-EP002955.
XX
PR 15-MAR-2000; 2000DE-01013847.
PR 06-APR-2000; 2000DE-01019058.
PR 07-APR-2000; 2000DE-01019173.
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2001-602752/68.
XX
PT Fragments of chemically modified genes associated with tumor suppressor
PT genes and oncogenes, useful in designing primers and probes for analyzing
PT diseases associated with cytosine methylation state e.g. cancer.
XX
PS Claim 1; SEQ ID NO 472; 27pp; English.
XX
CC The invention relates to a nucleic acid comprising a sequence of 18
CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC bisulphite, of genes associated with tumour suppression and oncogenes
CC having a sequence taken from 536 (actually 533 since numbers 408, 458 and
CC 500 are missing from the sequence listing) sequences (Ss) and sequences
CC complementary to (Ss). The nucleic acid may be a peptide nucleic acid-
CC oligomer (PNA) of at least 9 nucleotides and may form part of a set of
CC probes for detecting the cytosine methylation state and/or single
CC nucleotide polymorphisms and also to be used in an array for analysing
CC diseases associated with CpG dinucleotides e.g. cancers and tumours. The
CC probes can also be used in a method for ascertaining genetic and/or
CC epigenetic parameters for the diagnosis and/or therapy of existing
CC diseases or the predisposition to specific diseases, by analysing
CC cytosine methylations. The parameters may be compared to another set of
CC genetic and/or epigenetic parameters, the differences serving as basis
CC for diagnosis and/or prognosis events which are disadvantageous to
CC patients. The present sequence is one of the 533 genomic sequences
CC derived from tumour suppressor genes and oncogenes. Sequences with even
CC numbered Seq ID numbers are the complementary sequence of the
CC corresponding odd numbered sequence (e.g. ID 2 and ID1, ID 536 and ID
CC 535, except for those whose partner sequence is missing). Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 26997 BP; 7650 A; 354 C; 5924 G; 13069 T; 0 U; 0 Other;

Query Match 5.1%; Score 234.2; DB 4; Length 26997;
Best Local Similarity 69.7%; Pred. No. 7.6e-31;
Matches 345; Conservative 0; Mismatches 148; Indels 2; Gaps 2;

```
Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||  |||||||||||||| |   | | | | |||||| |||| |  ||||||
Db       5699 TATATATATACGTATATATATATATATTTATATATATATATATGTATATATGTATATA 5758

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||  |||| || | || |||| | |||||||||| | |||  ||| || | |
Db       5759 TATATATATATGTATATTTATATATGTATATATATATATATATA-ATATGTATATATATA 5817

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              ||||||||   |||| |||| | || |||| || |||||| | | || | |  | ||||
Db       5818 TATATATATGTATATATATATATATGTATATACGTATATATAAATATGTGTGTATATATG 5877

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | || | |||| |  |||||| | ||||||    ||||| |  ||| || |  ||||
Db       5878 TATATACGTGTATATATGTATATATATATATATGTGTATATATATGTATGTATATGTATA 5937

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || |||||||     |||||||| |||  | |  |||||||| | |||| |  | |  ||
Db       5938 TATATATATATGTGTATATATATGTATGTATATGTATATATATATATATGTGTATATGTA 5997

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||  | ||||||  | |||||||||  |  | ||||   |||| || ||||||  | |||
Db       5998 TATGTATATATATGTATATATATATATATGTGTATATGTATATGTATATATATGTATATA 6057

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              ||||||  || | || | |||||||  |||||||   |||||||  |||||||     |||||
Db       6058 TATATATATGTGTATATATATATGTATATATATGTATATATATATATATATGTGTATATA 6117

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              |  | |||  | |  | | |  || || |   | ||||||     |||||||| || | ||
Db       6118 TGTATATATGTGTGTATATATGTATATGTGTGTGTATATATGTTTATATATATGTGTGTA 6177

Qy        706 TATGTATATATATAT 720
              ||| ||||||||||||
Db       6178 TATATATATATATAT 6192
```

RESULT 91
AAV83939

```
ID   AAV83939 standard; DNA; 80595 BP.
XX
AC   AAV83939;
XX
DT   03-MAR-1999  (first entry)
XX
DE   HC-contig derived from normal human chromosome 10q25.2 region.
XX
KW   Yeast artificial chromosome; YAC; probe; eukaryotic chromosome;
KW   neocentromere; replication; extra-chromosomal element; segregation;
KW   cell division; artificial chromosome; gene therapy; mardel(10);
KW   human artificial chromosome; transgenic; chromosome 10; 10q25.2; ss.
XX
OS   Homo sapiens.
XX
PN   WO9851790-A1.
XX
PD   19-NOV-1998.
XX
PF   13-MAY-1998;   98WO-AU000352.
XX
PR   13-MAY-1997;   97AU-00006784.
PR   26-AUG-1997;   97AU-00008791.
XX
PA   (AMRA-) AMRAD OPERATIONS PTY LTD.
XX
PI   Choo K,  Du Sart D,  Cancilla MR;
XX
DR   WPI; 1999-009773/01.
XX
PT   New isolated nucleic acid comprising neocentromere sequences from
PT   eukaryotic chromosome - used to produce replicable, segregating
PT   artificial chromosomes that can carry large amounts of DNA for gene
PT   therapy.
XX
PS   Claim 8; Fig 6; 540pp; English.
XX
CC   The present sequence represents the HC-contig derived from normal human
CC   chromosome 10, 10q25.2 region. This region can be naturally mutated to
CC   produce an unusual chromosomal marker referred to as mardel(10). The
CC   mardel(10) marker is mitotically stable and contains a functional
CC   neocentromere at a location regarded as non-centromeric. This
CC   neocentromere maps to q25.2 on chromosome 10. The specification describes
CC   nucleic acid sequences derived from a eukaryotic chromosome, including a
CC   neocentromere or its functional derivative or hybrid, that are able, in a
CC   compatible cell, of replicating, acting as extra-chromosomal element and
CC   segregating during cell division. The sequences can be used to construct
CC   artificial chromosomes for use in gene therapy comprising a replicable,
CC   segregating nucleic acid that confers a specific phenotype on cells.
CC   Human artificial chromosomes can propagate in human cells and carry large
CC   amounts of DNA (e.g. therapeutic genes), and, being extra-chromosomal,
CC   they are not mutagenic. The artificial chromosomes are also useful for
CC   generation of transgenic plants and animals, in production of proteins
CC   and to make diagnostic reagents, e.g. for expression of cytokines,
CC   receptors and growth factors, or to increase the copy number of a gene in
CC   a cell. The constructs may also be used for functional and structural
CC   analysis of chromosomes
XX
SQ   Sequence 80595 BP; 23183 A; 16613 C; 16824 G; 23975 T; 0 U; 0 Other;

  Query Match             5.1%;  Score 234.2;  DB 2;  Length 80595;
  Best Local Similarity   66.3%;  Pred. No. 7.9e-31;
  Matches  352;  Conservative    0;  Mismatches  178;  Indels    1;  Gaps    1;

Qy        189 CAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATA 248
              ||  |   ||||| || | | |     |||||     ||   | ||| |||||||| |||
Db      15479 CATGTTTGAAATATCTTATAATTTTTAAAAATTAAAATGGGAGATACATATATATGTATT 15538

Qy        249 TATACACACACACACATATATATATACACATATATATACACATATATACATATATACA 308
              |||  |   | | |   || || ||||||||| |||||||| | ||||| | ||| | |
Db      15539 TATGTATGTATATATGTATGTACATATACACACATATATACATAAATATATACATAAATA 15598

Qy        309 CATATATACACATATATATACACATACATATGTACACATATATATACACATATGTATACA 368
              |||||    |||||| ||| | | || | || |  | |||||| |||||  |||  |
Db      15599 TGTATATATGTGTATATAGACATAAATATGTATATATGTGTATATATACATAAATATGTA 15658

Qy        369 CATATATACACATATATACACACATATATACACATATATACACACATATATACACATATA 428
              |||| |  | ||| | | | | || |||| |  | |||| | ||||| |||   |||||
Db      15659 TATATGTGTATATAGACATAAATATGTATATATGTGTATATAGACATAAATATGTATATA 15718

Qy        429 TACACATATATACACATATACACATATATACACATATATACATATATACACATATATATA 488
              |    ||||| ||| | |||   ||||||    ||||| ||||| |||   ||||||
Db      15719 TGTGTATATAGACATAAATATGTATATATGTGTATATAGACATAAATATGTATATATGT- 15777

Qy        489 ATATACACACATATATATACACATATATACACACATATATACACATATATACACATATAT 548
              ||||  | ||||| ||||  | ||||  |  | | |  ||| | || ||||  |  | |||
Db      15778 GTATATAGACATAAATATGTATATATGTGTATATAGACATAAATATGTATATATGTGTAT 15837

Qy        549 ATACACATATATACACATATATACATATATACACATATATATACATATATACACATATAT 608
              |||  ||||| |||   ||||||   ||||||  ||||| ||||  ||||||    |||||
Db      15838 ATAGACATAAATATGTATATATGTGTATATAGACATAAATATGTATATATGTGTATATAG 15897

Qy        609 ACATATATACACATATATACATATATACACACATATATACACATACATATACACACACAT 668
              |||||  |||   ||||||   ||||||  |||  | ||||   |  |||   | ||  | | ||||
Db      15898 ACATAAATATGTATATATGTGTATATAGACATAAATATGTATATATGTGTATATAGACAT 15957

Qy        669 AGATATACATATATATACACATATATATACGTATATATATGTATATATATA 719
              | ||||  ||||||  |   |  ||| | |||  |||  |||||  | |  ||||||
Db      15958 AAATATGTATATATGTGTATATAGACATAAATATGTATATATGTGTATATA 16008


RESULT 92
ABQ88198
ID   ABQ88198 standard; cDNA; 154902 BP.
XX
AC   ABQ88198;
XX
DT   18-SEP-2002  (first entry)
XX
DE   Human osteoblast differentiation related cDNA SEQ ID NO 105.
XX
KW   Human; osteoblast; stem cell differentiation; bone tissue deposition;
KW   osteoporosis; osteopathic; ss.
XX
OS   Homo sapiens.
XX
PN   WO200250301-A2.
XX
PD   27-JUN-2002.
XX
PF   18-DEC-2001; 2001WO-US048276.
XX
PR   18-DEC-2000; 2000US-0255882P.
PR   24-APR-2001; 2001US-0285691P.
XX
PA   (GENE-) GENE LOGIC INC.
PA   (PROC ) PROCTER & GAMBLE CO.
XX
PI   Ji D,  Axelrod DW,  Cook JS,  Jaiswal N,  Einstein R,  Houghton A;
PI   Mertz L;
XX
DR   WPI; 2002-557663/59.
XX
PT   Use of genes and their expression profiles associated with osteoblast
PT   differentiation for screening modulators bone formation, for diagnosing
PT   or treating e.g. osteoporosis, or as markers for the differentiation
PT   process.
XX
PS   Claim 1; SEQ ID NO 105; 78pp + Sequence Listing; English.
XX
CC   The invention relates to genes and their expression profiles are used
CC   for: (a) screening modulators of precursor stem cell differentiation into
CC   osteoblasts, or bone tissue deposition; (b) diagnosing abnormal
CC   deposition of bone tissue, abnormal rate of osteoblast formation or
CC   osteoporosis; or (c) treating or monitoring treatment of the conditions
CC   cited in (b), or monitoring the progression of bone tissue deposition.
```

CC Specific conditions include postmenopausal osteoporosis, glucocorticoid
CC osteoporosis or male osteoporosis, osteopenia, osteodystrophy, drug-
CC induced abnormalities in bone formation or bone loss, conditions that
CC involve altered bone metabolism (e.g. idiopathic juvenile osteoporosis),
CC skeletal disease linked to breast cancer, mastocytosis, Fanconi syndrome
CC or fibrous dysplasia. The present sequence is that of an osteoblast
CC differentiation associated cDNA marker of the invention. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 154902 BP; 43917 A; 31458 C; 32848 G; 46679 T; 0 U; 0 Other;

Query Match 5.1%; Score 234.2; DB 6; Length 154902;
Best Local Similarity 75.4%; Pred. No. 8.2e-31;
Matches 304; Conservative 0; Mismatches 98; Indels 1; Gaps 1;

```
Qy      286 ATACACATATATACATATATACACATATATACACATATATATACACATACATATGTACAC 345
            |||||| |||||||| ||||| | | ||||| | | ||||||||| ||| | ||||| ||
Db    31565 ATACATATATATACACATATATATACATATATATACATATATACATATATACATGTATAC 31624

Qy      346 ATATATATACACATATGTATACACATATATACACATATATACACACATATATACACATAT 405
            | ||| ||| | | || |||||||| || ||  |||||||||| | | | | |   ||
Db    31625 ACATACATATATACATATATACACACATTTATGCATATATACATAGACACACATTTATGC 31684

Qy      406 ATACACACATATATACACATATATACACATATATACACATATACACATATATACACATAT 465
            ||||| | ||| | |||   ||||| || ||||||||||||||  |||||||||  ||||
Db    31685 ATACATATATACACACATGCATATATACCTATATACACATATATGCATATATACCTATAT 31744

Qy      466 ATACATATATACACATATATATAATATACACACATATATATACACATATATACACACATA 525
            | |||||||| || ||||| || |||||||||||||  ||||| |||||||| ||||| |||
Db    31745 ACACATATATGCATATATACAT-ATATACACACATGCATATATACATATACACACATATA 31803

Qy      526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
            |  | | ||| ||| | | |||||||||||| |||| || ||||| | ||||| ||| | |
Db    31804 TGCATATATACATATATACATATATACACACATATGCATATATACATATATACACATACA 31863

Qy      586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
            |||||||| |||||||| ||||| | ||||| ||| | ||| ||||| ||| | || ||
Db    31864 TATATACACATATACACACATATATACATATATACACACATATATATACACATATATGTA 31923

Qy      646 TACACATACATATACACACACATAGATATACATATATATACAC 688
            |||||| ||||||| | | | ||| | || | |   |||
Db    31924 TACACACACATATATATATATATATTTTAGCTGGAGTCTCAC 31966
```

RESULT 93
ADC87318/c
ID ADC87318 standard; DNA; 1210 BP.
XX
AC ADC87318;
XX
DT 01-JAN-2004 (first entry)
XX
DE Human GPCR gene SEQ ID NO:1771.
XX
KW ds; gene; human; GPCR;
KW guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1270724-A2.
XX
PD 02-JAN-2003.
XX
PF 18-JUN-2002; 2002EP-00013517.
XX
PR 18-JUN-2001; 2001JP-00246789.
XX
PA (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI Suwa M, Asai K, Akiyama Y, Aburatani H;
XX
DR WPI; 2003-315783/31.
DR P-PSDB; ADC87319.
XX
PT New polynucleotide, useful for preparing a composition for treating a
PT patient in need of increased or suppressed activity or expression of the
PT guanosine triphosphate-binding protein coupled receptor.
XX
PS Claim 1; SEQ ID NO 1771; 28pp; English.
XX
CC The invention relates to a novel polynucleotide encoding a guanosine
CC triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC the invention may have a use in gene therapy. The polynucleotide and
CC polypeptide are useful for preparing a composition for treating a patient
CC in need of increased or suppressed activity or expression of the
CC guanosine triphosphate-binding protein coupled receptor. The
CC polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC invention.
XX
SQ Sequence 1210 BP; 391 A; 154 C; 200 G; 465 T; 0 U; 0 Other;

Query Match 5.1%; Score 233.6; DB 10; Length 1210;
Best Local Similarity 68.5%; Pred. No. 8.4e-31;
Matches 337; Conservative 0; Mismatches 154; Indels 1; Gaps 1;

```
Qy      230 ATATATATATATATATATATATACACACACACACATATATATATACACATATATATAC 289
            ||| || ||||   |||||||||||||| | |   | |   ||||||| |||| ||| ||
Db      885 ATACATGTATACGAATATATATACACATATATGTATACGAATATATATACATGTATGTAT 826

Qy      290 ACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATAT 349
            ||  ||||| ||| | | | || ||||  ||||||||||||||  || | |||  ||||
Db      825 ACGAATATATATACACATATATGTATACGAATATATATACACATGTATGTATACGAATAT 766

Qy      350 ATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATAC 409
            |||||||||| | | || ||  ||||| | | | ||  | |  ||||   | ||||||||
Db      765 ATATACACATGTATGTATACGAATATATATACACATGTATATGTATACGAATATATATAC 706

Qy      410 ACACATATATACACATATATACACATATATACACATATACACATATATACACATATATAC 469
            |||  |||||  | | | ||| | ||| | |||||| || |   ||| |||||  || ||
Db      705 ACATGTATATGTATACAAATATATATACACACACATGTATATGAATACACACACATGTAT 646

Qy      470 ATATATACACATATATAT-AATATACACACATATATATACACATATATACACACATATAT 528
            ||  ||||||| |||||| |||| |||||||||| ||||   ||| | ||||||||||  |
Db      645 ATGAATACACACATATATGAATACACACACATATGTATATGAATACACACACACATATGT 586

Qy      529 ACACATATATACACATATATATACACATATATACACATATATACATATATACACACATAT 588
            | |   ||| ||||| |  ||| | ||  ||||||| | |   || ||||   ||| |
Db      585 ATATGAATACACACACACGTATGTATATGAATACACACACACGTATGTATATGAATACAC 526

Qy      589 ATACATATATACACATATATACATATATACACATATATACATATATACACACATATATAC 648
            | |||  |||  | ||  ||||| | | ||  || ||||   ||| |||||||  |||
Db      525 ACACACGTATGTATATGAATACACACACACGTATGTATATGAATACACACACACGTATGT 466

Qy      649 ACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATAT 708
            | ||  ||| | |||||| ||  ||||   ||| |||||||  ||| ||  |  ||| ||
Db      465 ATATGAATACATACACACGTATGTATATGAATACATACACACGTATGTATATGAATACAT 406

Qy      709 GTATATATATAT 720
              | | ||||||
Db      405 ACACACATATAT 394
```

RESULT 94
AAF84800/c
ID AAF84800 standard; DNA; 110000 BP.
XX
AC AAF84800;
XX
DT 09-JUL-2001 (first entry)
XX
DE Nucleotide sequence of the human SPG4 gene.
XX
KW Human; SPG4 gene; spastin; PSF-AD; gene therapy;

```
KW   autosomal dominant familial spastic paraplegia; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             9932. .102009
FT                   /*tag= a
FT                   /note= "contains introns"
FT   exon            9932. .10471
FT                   /*tag= b
FT                   /number= 1
FT   intron          10472. .33718
FT                   /*tag= c
FT                   /number= 1
FT   exon            33719. .33805
FT                   /*tag= d
FT                   /number= 2
FT   intron          33806. .35748
FT                   /*tag= e
FT                   /number= 2
FT   exon            35749. .35832
FT                   /*tag= f
FT                   /number= 3
FT   intron          35833. .45022
FT                   /*tag= g
FT                   /number= 3
FT   exon            45023. .45118
FT                   /*tag= h
FT                   /number= 4
FT   intron          45119. .60863
FT                   /*tag= i
FT                   /number= 4
FT   exon            60864. .61051
FT                   /*tag= j
FT                   /number= 5
FT   intron          61052. .61927
FT                   /*tag= k
FT                   /number= 5
FT   exon            61928. .62061
FT                   /*tag= l
FT                   /number= 6
FT   intron          62062. .62344
FT                   /*tag= m
FT                   /number= 6
FT   exon            62345. .62438
FT                   /*tag= n
FT                   /number= 7
FT   intron          62439. .73173
FT                   /*tag= o
FT                   /number= 7
FT   exon            73174. .73248
FT                   /*tag= p
FT                   /number= 8
FT   intron          73249. .74633
FT                   /*tag= q
FT                   /number= 8
FT   exon            74634. .74705
FT                   /*tag= r
FT                   /number= 9
FT   intron          74706. .82788
FT                   /*tag= s
FT                   /number= 9
FT   exon            82789. .82864
FT                   /*tag= t
FT                   /number= 10
FT   intron          82865. .83102
FT                   /*tag= u
FT                   /number= 10
FT   exon            83103. .83194
FT                   /*tag= v
FT                   /number= 11
FT   intron          83195. .83334
FT                   /*tag= w
FT                   /number= 11
FT   exon            83335. .83414
FT                   /*tag= x
FT                   /number= 12
FT   intron          83415. .88129
FT                   /*tag= y
FT                   /number= 12
FT   exon            88130. .88172
FT                   /*tag= z
FT                   /number= 13
FT   intron          88173. .89561
FT                   /*tag= aa
FT                   /number= 13
FT   exon            89562. .89641
FT                   /*tag= ab
FT                   /number= 14
FT   intron          89642. .91162
FT                   /*tag= ac
FT                   /number= 14
FT   exon            91163. .91233
FT                   /*tag= ad
FT                   /number= 15
FT   intron          91234. .93443
FT                   /*tag= ae
FT                   /number= 15
FT   exon            93444. .93484
FT                   /*tag= af
FT                   /number= 16
FT   intron          93485. .100599
FT                   /*tag= ag
FT                   /number= 16
FT   exon            100600. .102009
FT                   /*tag= ah
FT                   /number= 17
XX
PN   FR2798138-A1.
XX
PD   09-MAR-2001.
XX
PF   03-SEP-1999;   99FR-00011097.
XX
PR   03-SEP-1999;   99FR-00011097.
XX
PA   (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI   Weissenbach J,  Hazan J;
XX
DR   WPI; 2001-283966/30.
XX
PT   New human nucleic acid from the SPG4 gene, useful e.g. for diagnosis of
PT   autosomal dominant familial spastic paraplegia and in drug screening.
XX
PS   Claim 2; Page 45-106; 145pp; French.
XX
CC   The present sequence represents a human SPG4 gene. The SPG4 gene encodes
CC   a spastin polypeptide. Mutations in the SPG4 gene are responsible for
CC   autosomal dominant familial spastic paraplegia. SPG4 polynucleotides, and
CC   their fragments, are used to screen DNA banks for sequences that encode
CC   spastin (particularly sequences in other mammals, specifically mice); to
CC   identify SPG4 mutations, or other genetic anomalies, particularly for
CC   diagnosis of autosomal dominant familial spastic paraplegia (PSF-AD); to
CC   identify promoters and other regulatory elements of the SPG4 gene; for
CC   detection and amplification; for recombinant production of spastin; and
CC   for diagnostic genotyping of PSF-AD
XX
SQ   Sequence 110000 BP; 30622 A; 21640 C; 22817 G; 34921 T; 0 U; 0 Other;

  Query Match             5.0%;  Score 232.8;  DB 5;  Length 110000;
  Best Local Similarity   67.2%;  Pred. No. 1.4e-30;
  Matches  449;  Conservative    0;  Mismatches  207;  Indels   12;  Gaps    8;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC--AAAAT 58
              ||||  ||| |||||||  |||||||||||||||||||| |||||||| ||||  |||||
```

```
Db      56570 AGTTTGAGAACAGCCTGGCCAACATGGTGAAACCCCATTTCTACTAAAGATACAAAAAAT 56511

Qy         59 TAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGA 118
              ||| | ||| ||||| || ||||||||||||| |||| |||| | || ||||||||
Db      56510 TAGCCTGGCGTGGTGCCACGTGCCTGTAATCCCAGGTACTCGGGAGACTGAAGCAGGAGA 56451

Qy        119 ATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCC 177
              ||  |||||||||| ||  |  |||||||||  || ||||||    ||| |||| |||
Db      56450 ATTGCTTGAACCTG--GGAAGCAGAGGTTGCGGTAAGCCAAGATAGCACCACTGCATTCC 56393

Qy        178 AGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATAT 237
              ||||  ||  || || || | |   ||||||  |||||||| ||    |    | | | |
Db      56392 AGCCTGGGGGAC-AGGGCAAGACTCCATCTCCAAAAAAAAACAAAAAAAACAAAAAAAAA 56334

Qy        238 ATATATATATATATACACACACACACATATATATATACACATATATATACACATATAT 297
              | | | | | |  || | |   | ||||| ||||| | | |||||||||     ||| ||
Db      56333 AAACAAAAAAAACTATATATGTATACACACATATACAAATACATATATAACTGTATACAT 56274

Qy        298 ACATATATACACATATATACACATATATATACACATACATATGTACACATATAT-ATACA 356
              | |  | || | ||| ||  |||  |||||| |||||  ||| |  | |||  | |||||
Db      56273 ATA-GTTTATATATACATTTACAGCTATATATACATAGTTATTTTGATATAACTAATACA 56215

Qy        357 CATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATA 416
              ||||   || | || |||| ||||||| | | |||||| ||   ||||||  | |   |
Db      56214 TATATAACTATATATGTATATACATATACATATACATATGTAACTATATATGTATATGCA 56155

Qy        417 TATACACATATATACACATATATACACATATACACATATATACACATATATACATATATA 476
              |||||| || ||   | |||| |  |   ||| |||||| ||   ||||||  ||||  |
Db      56154 TATACATATGTAACTATATATGTGTATGCATATACATATGTAACTATATATGTATATGCA 56095

Qy        477 CACATATATATAA-TATACACACATATATATACACATATATACACACATATATACACATA 535
              | | |||| ||| |||| |   ||||  ||| |||||| ||   | |||| ||    |
Db      56094 TATAGATATGTAACTATATATGTATATGCATATACATATGTAACTATATATGTATGTGTG 56035

Qy        536 TATACACATATATATACACATATATACACATATATACATATATACACATATATATACATA 595
              |||||| || ||  || | || |||   |||||| | |||||   | ||| | |||  |
Db      56034 TATACATATGTAACTATATATGTATGTGCATATACATATATAACTATATACAGATAGCT- 55976

Qy        596 TATACACATATATACATATATACACATATATACATATATACACACATATATACACATACA 655
              || |||||||||  ||||||  |||||||   || | ||   | ||  | ||  |   |
Db      55975 --TATACATATATAGTTATATATACATATAGTTATTTTTAGTTATATCAAAACTAAACAA 55918

Qy        656 TATACACA 663
              |||| | |
Db      55917 TATATATA 55910


RESULT 95
AAV83940
ID   AAV83940 standard; DNA; 80240 BP.
XX
AC   AAV83940;
XX
DT   03-MAR-1999  (first entry)
XX
DE   NC-contig derived from mardel(10) on chromosome 10q25.2.
XX
KW   Yeast artificial chromosome; YAC; probe; eukaryotic chromosome;
KW   neocentromere; replication; extra-chromosomal element; segregation;
KW   cell division; artificial chromosome; gene therapy; mardel(10);
KW   human artificial chromosome; transgenic; chromosome 10; 10q25.2; ss.
XX
OS   Homo sapiens.
XX
PN   WO9851790-A1.
XX
PD   19-NOV-1998.
XX
PF   13-MAY-1998;   98WO-AU000352.
XX
PR   13-MAY-1997;   97AU-00006784.
PR   26-AUG-1997;   97AU-00008791.
XX
PA   (AMRA-) AMRAD OPERATIONS PTY LTD.
XX
PI   Choo K,  Du Sart D,  Cancilla MR;
XX
DR   WPI; 1999-009773/01.
XX
PT   New isolated nucleic acid comprising neocentromere sequences from
PT   eukaryotic chromosome - used to produce replicable, segregating
PT   artificial chromosomes that can carry large amounts of DNA for gene
PT   therapy.
XX
PS   Claim 9; Fig 16A; 540pp; English.
XX
CC   The present sequence represents the NC-contig derived from a mutated
CC   human chromosome 10, 10q25.2 region. The sequence contains an unusual
CC   chromosomal marker referred to as mardel(10). The mardel(10) marker is
CC   mitotically stable and contains a functional neocentromere at a location
CC   regarded as non-centromeric. This neocentromere maps to q25.2 on
CC   chromosome 10. The specification describes nucleic acid sequences derived
CC   from a eukaryotic chromosome, including a neocentromere or its functional
CC   derivative or hybrid, that are able, in a compatible cell, of
CC   replicating, acting as extra-chromosomal element and segregating during
CC   cell division. The sequences can be used to construct artificial
CC   chromosomes for use in gene therapy comprising a replicable, segregating
CC   nucleic acid that confers a specific phenotype on cells. Human artificial
CC   chromosomes can propagate in human cells and carry large amounts of DNA
CC   (e.g. therapeutic genes), and, being extra-chromosomal, they are not
CC   mutagenic. The artificial chromosomes are also useful for generation of
CC   transgenic plants and animals, in production of proteins and to make
CC   diagnostic reagents, e.g. for expression of cytokines, receptors and
CC   growth factors, or to increase the copy number of a gene in a cell. The
CC   constructs may also be used for functional and structural analysis of
CC   chromosomes
XX
SQ   Sequence 80240 BP; 23102 A; 16537 C; 16747 G; 23846 T; 0 U; 8 Other;

  Query Match             5.0%;  Score 232.2;  DB 2;  Length 80240;
  Best Local Similarity   64.9%;  Pred. No. 1.7e-30;
  Matches  359;  Conservative    0;  Mismatches  193;  Indels    1;  Gaps    1;

Qy        217 AAAAAGGCATTTTATATATATATATATATATATACACACACACACATATATATATA 276
              |||||    |  |   | ||| |||||||| ||| ||   |   | | |  ||| || |||
Db      15240 AAAAATTAAAATGGGAGATACATATATATGTATTTATGTATGTATATATGTATGTACATA 15299

Qy        277 CACATATATATACACATATATACATATATACACATATATACACATATATATACACATACA 336
              ||| | ||||| ||||| ||| ||| ||| | || |||| |   | |||||| ||||| |
Db      15300 TACACACATATATACATAAATATATACATAAATATGTATATATGTGTATATAGACATAAA 15359

Qy        337 TATGTACACATATATATACACATATGTATACACATATATACACATATATACACACATATA 396
              |||||| | || | |||| | | ||  |||  |||||    ||||| ||| | ||||
Db      15360 TATGTATATATGTGTATATATACATAAATATGTATATATGTGTATATAGACATAAATATG 15419

Qy        397 TACACATATATACACACATATATACACATATATACACATATATACACATATACACATATA 456
              || | || | || | | | ||| | |  ||||||    |||||| ||||| | |  ||||
Db      15420 TATATATGTGTATATAGACATAAATATGTATATATGTGTATATAGACATAAATATGTATA 15479

Qy        457 TACACATATATACATATATACACATATATATAATATACACACATATATATACACATATAT 516
              |    ||||| ||||| |||   |||||| |  |||| | ||||| ||||  | |||| |
Db      15480 TGTGTATATAGACATAAATATGTATATATGT-GTATATAGACATAAATATGTATATATGT 15538

Qy        517 ACACACATATATACACATATATACACATATATATACACATATATACACATATATACATAT 576
                | | | ||| | || |||| |  | |||||| ||||| |||   ||||||   |||
Db      15539 GTATATAGACATAAATATGTATATATGTGTATATAGACATAAATATGTATATATGTGTAT 15598

Qy        577 ATACACATATATATACATATATACACATATATACATATATACACATATATACATATATAC 636
              ||| ||||| ||||  ||||||    ||||| ||||| |||   ||||||   ||||||
Db      15599 ATAGACATAAATATGTATATATGTGTATATAGACATAAATATGTATATATGTGTATATAG 15658

Qy        637 ACACATATATACACATACATATACACACACATAGATATACATATATATACACATATATAT 696
              ||| | ||||  | |||  | || | | ||||| ||||  ||||||  |  | ||| | ||
Db      15659 ACATAAATATGTATATATGTGTATATAGACATAAATATGTATATATGTGTATATAGACAT 15718

Qy        697 ACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCA 756
```

```
Db      15719 AAATATGTATATGTGTGTATATAGACATAAATATGTATATATGTGTGTATATAGACATAA 15778

Qy        757 CTGGGGATAGAGG 769

Db      15779 ATATGTATATATG 15791
```

RESULT 96
ACN44026/c
ID ACN44026 standard; DNA; 122923 BP.
XX
AC ACN44026;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human genomic sequence hCG24510.
XX
KW Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.
XX
OS Homo sapiens.
XX
PN WO2003073826-A2.
XX
PD 12-SEP-2003.
XX
PF 28-FEB-2003; 2003WO-US006235.
XX
PR 01-MAR-2002; 2002US-00087192.
XX
PA (SAGR-) SAGRES DISCOVERY.
XX
PI Morris DW;
XX
DR WPI; 2003-328604/31.
XX
PT Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
PT comprises a nucleotide sequence.
XX
PS Claim 1; SEQ ID NO 268; 0pp; English.
XX
CC The present invention relates to novel DNA and protein sequences which
CC are associated with carcinomas. The sequences are useful for: (i) for
CC screening drug candidates; (ii) for screening of bioactive agent capable
CC of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
CC a bioactive agent capable of modulating the activity of CAP; (iv) for
CC evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
CC carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
CC carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
CC (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
CC determining Carcinoma Associated (CA) gene copy number. In addition, the
CC CA genes are useful as DNA vaccines and the CAP are useful as markers of
CC carcinoma including lymphoma. The present sequence is one such CA coding
CC sequence. Note: This patent is an equivalent to basic patent
CC US2002182586A1, for which no sequence data was published
XX
SQ Sequence 122923 BP; 28628 A; 24391 C; 25218 G; 32559 T; 0 U; 12127 Other;

Query Match 5.0%; Score 231.8; DB 11; Length 122923;
Best Local Similarity 62.3%; Pred. No. 2.1e-30;
Matches 447; Conservative 0; Mismatches 257; Indels 13; Gaps 6;

```
Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAATTA 60

Db      43687 AGTTTGAGACCAGCCTGGGCAACATGGCAAAACCCTATCTCTACAAAAAATACAAAAATG 43628

Qy         61 GA-CAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119

Db      43627 CACCAGGTGTAGTGGCATGCACCTGTAGTCTCAACTACTTGGGGGGCCAAGGCAGAAGGA 43568

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCCA 178

Db      43567 TCGCTTGAACC--CGGGAGGTCAAGGTTGCAGTGAGCCAAGACCACACTACTGCACTCCA 43510

Qy        179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATA 238

Db      43509 GTCTAGGTGATAA-AGTGAGACCCTGTTTCAGAAAAAAAAA-------TTTATATATATA 43458

Qy        239 TATATATATATATACACACACACACATATATATATACACATATATATACACATATATA 298

Db      43457 TATGTGTGTGTGTGTGTGTGTGTGTGTGTGTATGTGTGTGTATATGTGTATATA 43398

Qy        299 CATATATACACATATATACACATATATATACACATACATATGTACACA-TATATATACAC 357

Db      43397 TGTGTGTGTCTATGTGTGTACATATATATAAAATTGTACATATAAAAATTATATAAGTAT 43338

Qy        358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417

Db      43337 ATATATGCACGTGTATATACGTATGTATGTGTGTGTATGTATGTGTGTGTATATATGTAT 43278

Qy        418 ATACACATATATACACATATATACACATATACACATATATACACATATATACATATATAC 477

Db      43277 GTGTATGTATATATGTATGTGTGTATATATATGTATGCATGTGTGTGTATGTATGTATGN 43218

Qy        478 ACATATATATAATATACACACATATATATACACATATATACACACATATATACACATATA 537

Db      43217 NNNNNNNNNNNNNNNNNNACACATATATATACACACATATATATACACACATATATATA 43158

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATACATATA 597

Db      43157 CACATATATATATACACATATATATACACATATATATACACACATATATATACACATATA 43098

Qy        598 TACACATATATACATATATACACATATATACATATATACACACATATATACACATACATA 657

Db      43097 TATACACACATACACACACACACACACACACACACACACACACACACACACATATATATATA 43038

Qy        658 TACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATAT 714

Db      43037 TATATGCCAAGAGTAAGTATTAGGGGAATAAAGCTTAAAATAAAAAAATAAGAAAAT 42981
```

RESULT 97
ADL13505/c
ID ADL13505 standard; DNA; 76146 BP.
XX
AC ADL13505;
XX
DT 06-MAY-2004 (first entry)
XX
DE Osteoarthritis-associated polymorphic nucleotide #37.
XX
KW ds; gene; osteopathic; antiinflammatory; antiarthritic; gene therapy;
KW joint space narrowing; osteophyte development; joint pain;
KW osteoarthritis; SNP; single nucleotide polymorphism.
XX
OS Homo sapiens.
XX
PN WO2003054166-A2.
XX
PD 03-JUL-2003.
XX
PF 19-DEC-2002; 2002WO-US041225.
XX
PR 20-DEC-2001; 2001US-0342603P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Jones KA, Schafer A;
XX
DR WPI; 2003-559141/52.
XX
PT Determining susceptibility of an individual to joint space narrowing,
PT osteophyte development and/or joint pain comprises identifying whether
PT the individual has at least one polymorphism in a polynucleotide encoding
PT a protein.
XX

```
PS   Disclosure; SEQ ID NO 37; 297pp; English.
XX
CC   The invention relates to a method of determining susceptibility of an
CC   individual to joint space narrowing and/or osteophyte development and/or
CC   joint pain comprising identifying whether the individual has at least one
CC   polymorphism in a polynucleotide encoding at least one of the protein
CC   listed in the specification. The methods, composition and agent are
CC   useful for modulating the susceptibility of an individual to joint space
CC   narrowing and/or osteophyte development and/or joint pain that is
CC   associated with a disease, preferably osteoarthritis. The cell line and
CC   the non-human animal are useful for screening for an agent for diagnosing
CC   an individual having susceptibility to joint space narrowing and/or
CC   osteophyte development and/or joint pain. This sequence corresponds to
CC   the polynucleotide encoding a protein listed in the specification. (Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification but was obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences).
XX
SQ   Sequence 76146 BP; 14629 A; 22827 C; 22905 G; 15785 T; 0 U; 0 Other;

  Query Match             5.0%;  Score 231;  DB 10;  Length 76146;
  Best Local Similarity   69.7%;  Pred. No. 2.8e-30;
  Matches  373;  Conservative    0;  Mismatches  150;  Indels   12;  Gaps    4;

Qy        209 AGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACACATA 268
              | |||| | | | |       ||  | ||  ||  | ||| || ||||||  ||||||||
Db      14496 ACAAAACACATATGTACACACATCCACATGCATCCACATACATGCACACATCCACACATA 14437

Qy        269 TATATATACACATATATATACACATATATACATATAT------ACACATATATACACATA 322
              ||  |  |||||  |||| |||||| || ||| ||||      |||||  || ||||  |
Db      14436 TACGTGCACACACCTATACACACATGTACACACATATACATGCACACACCTAGACACTCA 14377

Qy        323 TATATACACATACATATGTACACATATATATACACATATGTATACACATATATACACATA 382
              | || || |  ||| | |||||||  | ||||||||||| ||||||||||| |||||||
Db      14376 CACATGCATACCCATGTATACACATTCACATACACATATG-ATACACATATACACACATA 14318

Qy        383 TATACACACATATATACACATATATACACACATATATACACATATATACACATATATACA 442
              ||||||| ||  | |||||   || ||||||| |||||||||| || ||||| || |
Db      14317 TATACACAGATGCACACACACTCATGCACACATCTATACACATACATGCACATGTACATC 14258

Qy        443 CATATACACATATATACACATATATACATATATACACATATATATAATATACACACATAT 502
              || || | || ||  |  ||| |||||| | || |||| ||  |  | ||||||||||| |
Db      14257 CACATGCGCACATGCATGCATGTATACACACATGCACACATGCA-CACATACACACATGT 14199

Qy        503 ATATACACATATATACACACATATATACACATATATACACATATATATACACATATATAC 562
              | | ||||  | | |||| || | |||||||| ||||||||  || | |||||     ||
Db      14198 ACACACACGCACACACACGCACACATACACATTTATACACACGTACACACACAC----AC 14143

Qy        563 ACATATATACATATATACACATATATATACATATATACACATATATACATATATACACAT 622
              ||||  ||||||  | ||| | ||| |||||||  | ||||||||| || | |||||||
Db      14142 ACATGCATACATGCACACATACATACATACATACCTGCACATATACACCTGTATACACGC 14083

Qy        623 ATATACATATATACACACATATATACACATACATATACACACACATAGATATACATATAT 682
              ||| |||||| |||||||||| || |||| |  |||||||||   | | | ||||  |
Db      14082 ATACACATATGTACACACATACATGCACACATCTATACACACATGCATACACACATTCAC 14023

Qy        683 ATACACATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAG 737
              ||| |||||| ||||| |||| | ||  ||| | ||||    |       |  |||||
Db      14022 ATATACATATGTATACATATACACATACATACACATATATCCACACATAGATAAG 13968


RESULT 98
ADJ95414
ID   ADJ95414 standard; DNA; 9800 BP.
XX
AC   ADJ95414;
XX
DT   06-MAY-2004  (first entry)
XX
DE   Human nootropic protein coding sequence, SEQ ID 3.
XX
KW   Nootropic; Gene therapy; human; mental illness; gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             2. .250
FT                   /*tag=  a
FT                   /partial
FT                   /note= "No start codon given"
FT                   /product= "Human nootropic protein #3"
XX
PN   JP2003310264-A.
XX
PD   05-NOV-2003.
XX
PF   19-APR-2002; 2002JP-00117143.
XX
PR   19-APR-2002; 2002JP-00117143.
XX
PA   (KAZU-) ZH KAZUSA DNA KENKYUSHO.
XX
DR   WPI; 2004-027143/03.
DR   P-PSDB; ADJ95425.
XX
PT   Novel DNA encoding recombinant protein from cDNA library derived from
PT   human adult whole brain and human fetus whole brain, useful for diagnosis
PT   of disease such as mental illness.
XX
PS   Claim 1; SEQ ID NO 3; 91pp; Japanese.
XX
CC   The present invention relates to human proteins (ADJ95423-ADJ95433) and
CC   their coding sequences (ADJ95412-ADJ95422). The sequences are useful for
CC   diagnosis and treatment of disease such as mental illness.
XX
SQ   Sequence 9800 BP; 2648 A; 2439 C; 2333 G; 2380 T; 0 U; 0 Other;

  Query Match             5.0%;  Score 229.4;  DB 12;  Length 9800;
  Best Local Similarity   68.5%;  Pred. No. 4.8e-30;
  Matches  379;  Conservative    0;  Mismatches  161;  Indels   13;  Gaps    4;

Qy        185 GCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATA 244
              ||  | ||  ||| || |    || |     |||||       || | ||| ||| | | |||
Db       5611 GCCTCCAGTTCGAGATATACCCTAAATTTAAAAATTGTATATATAATACATACACACATA 5670

Qy        245 TATATATACA-----CACACACACACATATATATATACACATATATATACACATATATAC 299
              |||||||  |     ||||||  |||||  | |||  |  |  |  |  |  ||  ||  ||||| ||
Db       5671 CATATATAGACACACCACACATACACACACATACACATATACACACATGCATATATACAC 5730

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACAT 359
              |||||  ||||||  |||||||  | ||||||||||  ||||    |  |  |||  |  |  ||
Db       5731 ATATACACACATGCATACACACACATATACACATGCATACACATATATACACACACCACA 5790

Qy        360 ATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATAT 419
                      |||||||  |||||||  |||  |  |||  ||||  ||    |  |||||   ||||||
Db       5791 CATACACAACATATACACACATACATATACACATATGCATGCACATACATGCATATAC 5850

Qy        420 ACACATATATACACATATATACACATATACACATATATACACATATATACATATATACAC 479
              ||||||||||  ||||  |||  ||||||  ||  |||  |  ||||||||||||||   ||||||  |
Db       5851 ACACATATATGCACACATACACACATGTATACACACATACACATATATACTCATATACCC 5910

Qy        480 ATATATATA-ATATACACACATATATATACACATATATACACACATATATACACATATAT 538
                |  | |||  ||||  ||||||  |||||  |  ||||||    ||||||  |  |||  |  |  |  |
Db       5911 CCACACATATATATCCACACACATATAAATACATATCCCCACACAGACATATATACACAC 5970

Qy        539 AC---ACATATATATACACATATATACACATATATACATATATACACATATATATACATA 595
              ||   |||  |  |  |||||    |  |||  |  |||||  ||||  |||||  ||  |||  |  |||  |
Db       5971 ACTATACACACACATACATGCAGATATATATATACACATGTATACTCACATAAACACACA 6030

Qy        596 TATACACATATATACATATATACACATATATACATATATACACACATATATACACATACA 655
              |  |  |||||||  ||   |||  |  |||  |  |  ||||||||||  ||||  ||||
Db       6031 TGCATGCATATATGCAGGCATATATATACACATGTATATACACAAGTATACACACGTACA 6090

Qy        656 TATACACA----CACATAGATATACATATATATACACATATATATACGTATATATATGTA 711
              ||||||||      | ||  |||||||  |  ||||||||||| |||||  ||||||    |
Db       6091 TATACACATATTTAAATGTATATACACAGGTATACACATATGTATACACATATATGCACA 6150
```

```
Qy        712 TATATATATGCTC 724
              | || |||||  |
Db       6151 CACATGTATGCAC 6163


RESULT 99
ABD32843/c
ID   ABD32843 standard; DNA; 326002 BP.
XX
AC   ABD32843;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Human cancer-associated genomic DNA HD17-014.
XX
KW   Human; ds; cancer-associated protein; gene; cytostatic; cancer;
KW   leukaemia; lymphoma; CAP.
XX
OS   Homo sapiens.
XX
PN   WO2004074320-A2.
XX
PD   02-SEP-2004.
XX
PF   17-FEB-2004; 2004WO-US004730.
XX
PR   14-FEB-2003; 2003US-00367094.
PR   14-MAR-2003; 2003US-00388838.
PR   15-APR-2003; 2003US-00417375.
PR   13-JUN-2003; 2003US-00461862.
PR   15-SEP-2003; 2003US-00663431.
PR   15-DEC-2003; 2003US-00737318.
XX
PA   (SAGR-) SAGRES DISCOVERY INC.
XX
PI   Morris DW,  Morris DW,  Malandro MS;
XX
DR   WPI; 2004-652914/63.
XX
PT   New isolated cancer-associated polynucleotides and polypeptides useful
PT   for diagnosing, preventing or treating cancers, especially lymphoma and
PT   leukemia, or in screening for agents that modulate cancer.
XX
PS   claim 16; seqid 506; 310pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising at least 10
CC   contiguous nucleotides of any of the 233 polynucleotide sequences given
CC   in the specification, or its complement. The nucleic acids encode cancer-
CC   associated proteins. Also included are an expression vector comprising
CC   the isolated nucleic acid cited above, a host cell comprising the above
CC   recombinant nucleic acid or expression vector, a microarray for detecting
CC   a cancer-associated (CA) nucleic acid comprising at least one probe
CC   comprising at least 10 contiguous nucleotides of any of the above-
CC   mentioned nucleotide sequences, an isolated polypeptide (encoded within
CC   an open reading frame of a CA sequence selected from any of the 95
CC   polynucleotide sequences as mentioned in the specification, or its
CC   complement), an isolated antibody, (or its antigen binding fragment) that
CC   binds to the above polypeptide, a hybridoma that produces the above
CC   monoclonal antibody, a pharmaceutical composition comprising the above
CC   antibody and a pharmaceutical excipient, a kit for detecting cancer
CC   cells(comprising the antibody cited above, methods for diagnosing cancer
CC   or for detecting the presence or absence of cancer cells in an
CC   individual, a method for inhibiting growth of cancer cells in an
CC   individual, a method for delivering a therapeutic agent to cancer cells
CC   in an individual, an electronic library comprising the above
CC   polynucleotide or polypeptide (or their fragments), methods of screening
CC   for anticancer activity or for a bioactive agent capable of modulating
CC   the activity of a CA protein (CAP), methods for detecting cancer
CC   associated with expression of a polypeptide in a test cell sample, a
CC   method for treating cancers and a method for inhibiting the expression of
CC   CA gene in a cell. The composition and methods are useful for detecting,
CC   diagnosing, preventing and treating cancers, especially lymphoma and
CC   leukaemia. These may also be used in screening for agents that modulate
CC   cancer. The present sequence is a human CAP genomic sequence. Note: The
CC   sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 326002 BP; 84319 A; 72970 C; 76829 G; 91884 T; 0 U; 0 Other;

  Query Match             5.0%;  Score 229;  DB 13;  Length 326002;
  Best Local Similarity   70.1%;  Pred. No. 6.6e-30;
  Matches  323;  Conservative    0;  Mismatches  135;  Indels    3;  Gaps    1;

Qy        235 TATATATATATATATATACACACACACACATATATATATACACATATATATACACATA 294
              ||| | |||| | | | |||||||||||||||| | | | | || | || | ||||||| |
Db      46727 TATGTGTATAGACACACACACACACACACACACACACACACGCGTACACATACACACA 46668

Qy        295 TATACATATATACACATATATACACATATATATACACATACATATGTACACATATATATA 354
              || ||  | | | ||| | ||||||| | |||  || | ||| | | | ||| ||| |
Db      46667 TACACGCACACATACACACATACACACACATACGCATACACAGACGCATACACATACACG 46608

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              |||| ||  | ||||| | |||||||   | ||||||||| || |||| ||| | |||||
Db      46607 CACACATACACACACACACATACACACGCACACACACATACATGCACACATACATACACA 46548

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              |||  || | |||  || ||| | ||||||| ||||  ||| ||||| ||| ||  | |
Db      46547 CATACGCATACATACGCATATACACACACATACACACGCATACACACACATACACGCACA 46488

Qy        475 TACACATATATATAATATACACACATATATATACACATATATACACACATATATACACAT 534
              | ||| | ||   ||| ||||| | ||| | ||||||||| ||  ||| | |||||||
Db      46487 CATACACACAT---ATACACACAGAGATACACACACATATACACGGACACACATACACAC 46431

Qy        535 ATATACACATATATATACACATATATACACATATATACATATATACACATATATATACAT 594
              |   | ||| ||  |||||| ||| |  |||||| ||| ||| | ||||  ||| ||||
Db      46430 ACTCATACACATGCATACACACATACATGCATATACACACATACATACATGCATACACAT 46371

Qy        595 ATATACACATATATACATATATACACATATATACATATATACACACATATATACACATAC 654
              | ||||||| ||| |||||  |  ||| |   ||| ||| ||||||||| | || |||||
Db      46370 ACATACACACATACACATACCTGGACACACGCACACATACACACACATACACACGCATAC 46311

Qy        655 ATATACACACACATAGATATACATATATATACACATATATA 695
              | |||| ||||||  |   ||||| | | | | || | | ||
Db      46310 ACATACGCACACACATGCATACACACACACAGACCTCTCTA 46270


RESULT 100
ADC87570
ID   ADC87570 standard; DNA; 1672 BP.
XX
AC   ADC87570;
XX
DT   01-JAN-2004  (first entry)
XX
DE   Human GPCR gene SEQ ID NO:2023.
XX
KW   ds; gene; human; GPCR;
KW   guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1270724-A2.
XX
PD   02-JAN-2003.
XX
PF   18-JUN-2002; 2002EP-00013517.
XX
PR   18-JUN-2001; 2001JP-00246789.
XX
PA   (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
PA   (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
PI   Suwa M,  Asai K,  Akiyama Y,  Aburatani H;
XX
```

DR WPI; 2003-315783/31.
DR P-PSDB; ADC87571.
XX
PT New polynucleotide, useful for preparing a composition for treating a
PT patient in need of increased or suppressed activity or expression of the
PT guanosine triphosphate-binding protein coupled receptor.
XX
PS Claim 1; SEQ ID NO 2023; 28pp; English.
XX
CC The invention relates to a novel polynucleotide encoding a guanosine
CC triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC the invention may have a use in gene therapy. The polynucleotide and
CC polypeptide are useful for preparing a composition for treating a patient
CC in need of increased or suppressed activity or expression of the
CC guanosine triphosphate-binding protein coupled receptor. The
CC polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC invention.
XX
SQ Sequence 1672 BP; 499 A; 171 C; 300 G; 602 T; 0 U; 100 Other;

Query Match 5.0%; Score 228.8; DB 10; Length 1672;
Best Local Similarity 65.6%; Pred. No. 5.7e-30;
Matches 349; Conservative 0; Mismatches 182; Indels 1; Gaps 1;

```
Qy        204 ATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATATACACACACACAC 263
              || |  | | | | | |      | | |||||| | |||| | | |||| ||||| | ||
Db        832 ATGTGTGTATATACATATATATGTGTGTATATACACATATGTGTGTATATACACATATAT 891

Qy        264 ACATATATATATACACATATATATACACATATATACATATATACACATATATACACATAT 323
              |||||||| | || | | | ||||| |||||||   ||||||||||||||||||  | |
Db        892 ACATATATGTGTATATACACATATATACATATATGTGTATATACACATATATACATGTGT 951

Qy        324 ATATACACATACATATGTACACATATATATACACATATGTATACACATATATACACATAT 383
              | || | | |||||| |      | ||||||||||||| | |     ||||||||||||
Db        952 GTGTATATACACATATATGTGTGTGTATATACACATATATGTGTGTGTATATACACATAT 1011

Qy        384 ATACACACATATATACACATATATACACACATATATACACATATATACACATATATACAC 443
              ||       ||||||||||||||       |||||||||||||||     |||||||||
Db       1012 ATGTGTGTGTATATACACATATATGTGTGTGTATATACACATATATGTGTGTATATACAC 1071

Qy        444 ATATACACATATATACACATATATACATATATACACATATATATAATATACACACATATA 503
              |||||    | | || | | | ||| ||| ||| |  ||| | |  |||| | ||| |||
Db       1072 ATATATGTGTGTATATACACATATATACATATATCTATGTGT-GTATATATACACATA 1130

Qy        504 TATACACATATATACACACATATATACACATATATACACATATATATACACATATATACA 563
              ||||||  | | |       ||||||| ||||||||    ||||||| ||| | |||||
Db       1131 TATACATGTCTGTGTGTGTATATATATACATATATGTGTATATATACACATACATATATG 1190

Qy        564 CATATATACATATATACACATATATATACATATATACACATATATACATATATACACATA 623
                | | || |||| ||||||||||  || ||| ||     ||||||  | || | | |||
Db       1191 TGTGTGTATATATGTACACATATACGTATATACATGTGTGTATATATGTGTACATATATA 1250

Qy        624 TATACATATATACACACATATATACACATACATATACACACACATAGATATACATATATA 683
              | |  |||||||     |||||||    || ||||    | | |||  | |  ||||||
Db       1251 TGTGTATATATATGTGTATATATATGTGTATATATGTGTATATATATGTGTGTATATATG 1310

Qy        684 TACACATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATA 735
              |  | |||| | |  || |||||| | | ||||||||     | ||  |||
Db       1311 TGTATATATGTGTGTGTGTATATATATGTGTATATATGTGTGTGTGTGTATA 1362
```

Search completed: April 21, 2005, 00:47:59
Job time : 1530.38 secs

This Page Blank (uspto)

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 20, 2005, 22:49:10 ; Search time 12928.1 Seconds
(without alignments)
17319.707 Million cell updates/sec

Title: US-09-939-209A-3_COPY_1_4621
Perfect score: 4621
Sequence: 1 agttcaagaccagcctgagc.....ttgctttttagtcctaaaaa 4621

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	4621	100.0	20300	6	AX451337	AX451337 Sequence
	2	4605.8	99.7	165329	9	AL583850	AL583850 Human DNA
c	3	4214.2	91.2	191699	2	AC031977	AC031977 Homo sapi
	4	411.2	8.9	203216	2	AC097328	AC097328 Pan trogl
	5	405.4	8.8	104321	2	CR759943	CR759943 Homo sapi
c	6	405.2	8.8	181946	2	AC097330	AC097330 Pan trogl
	7	402.2	8.7	109081	9	AL358074	AL358074 Human DNA
	8	396.8	8.6	164844	9	AC009058	AC009058 Homo sapi
	9	395.2	8.6	130349	9	AC011593	AC011593 Homo sapi
c	10	394.8	8.5	196413	2	AC097269	AC097269 Pan trogl
	11	390.8	8.5	189662	9	AC015468	AC015468 Homo sapi
c	12	387.4	8.4	125685	9	AC073310	AC073310 Homo sapi
	13	385.8	8.3	139118	9	AL133477	AL133477 Human DNA
	14	382.6	8.3	216021	9	HUAC004787	AC004787 Homo sapi
	15	379.8	8.2	92564	9	AY007685	AY007685 Homo sapi
c	16	379.4	8.2	161252	2	AC114955	AC114955 Homo sapi
c	17	378	8.2	202305	9	AC114291	AC114291 Homo sapi
c	18	373.6	8.1	192139	9	AC021955	AC021955 Homo sapi
c	19	372	8.1	99596	9	AC106873	AC106873 Homo sapi
	20	371.4	8.0	38096	9	BX901949	BX901949 Human DNA
c	21	371.4	8.0	110000	2	BX324168_0	BX324168 Homo sapi
	22	370.4	8.0	252512	2	AC097326	AC097326 Pan trogl
c	23	368.2	8.0	170646	2	AC123545	AC123545 Pan trogl
	24	367	7.9	112762	9	HS323B6	Z83841 Human DNA s
c	25	365.4	7.9	163976	9	AC108483	AC108483 Homo sapi
	26	365.4	7.9	163998	9	AC022166	AC022166 Homo sapi
c	27	364.2	7.9	153201	2	AC092863	AC092863 Homo sapi
c	28	363.6	7.9	151549	9	AC092694	AC092694 Homo sapi
	29	357.4	7.7	217985	9	AC021193	AC021193 Homo sapi
c	30	357.2	7.7	151424	2	CR812471	CR812471 Danio rer
c	31	356.6	7.7	121600	9	HS141H5	AL049176 Human DNA
c	32	356.6	7.7	127172	2	CR788239	CR788239 Danio rer
	33	355.4	7.7	282240	2	CR753865	CR753865 Danio rer
c	34	355.2	7.7	149969	2	AC024180	AC024180 Homo sapi
c	35	355.2	7.7	151047	2	CR759776	CR759776 Danio rer
c	36	355.2	7.7	162996	9	AC006441	AC006441 Homo sapi
c	37	354.8	7.7	182430	9	AL590381	AL590381 Human DNA
c	38	354	7.7	252450	2	CR753861	CR753861 Danio rer
	39	353.6	7.7	163651	2	CR790381	CR790381 Danio rer
c	40	353.6	7.7	164994	2	AC141288	AC141288 Homo sapi
c	41	353	7.6	165979	2	CR376741	CR376741 Danio rer
c	42	353	7.6	248167	2	CR627491	CR627491 Danio rer
	43	352.4	7.6	178254	9	AL731541	AL731541 Human DNA
	44	352.4	7.6	204486	2	CR752655	CR752655 Danio rer
c	45	352.2	7.6	178427	2	CR392006	CR392006 Danio rer

ALIGNMENTS

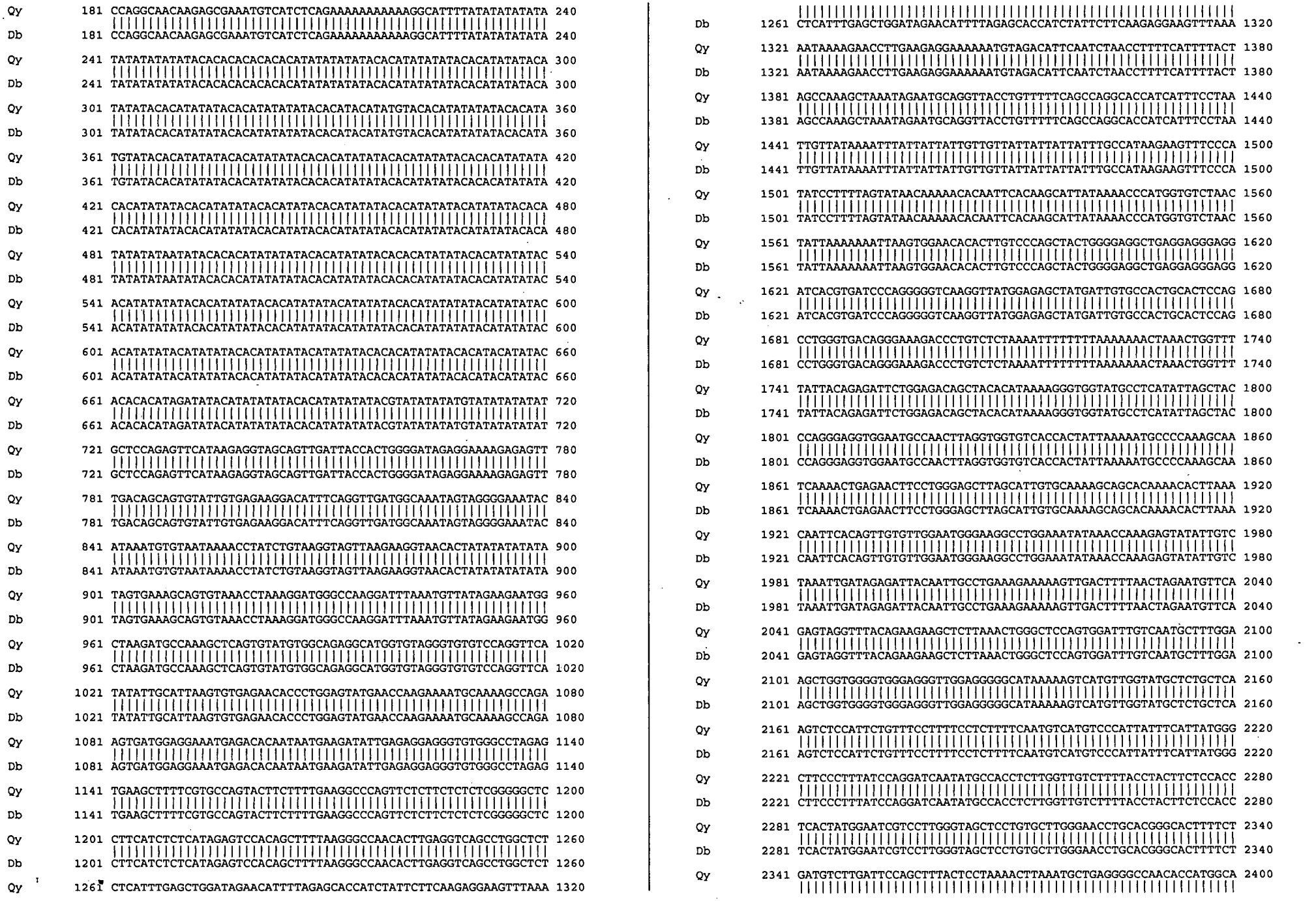
RESULT 1
AX451337
LOCUS AX451337 20300 bp DNA linear PAT 03-JUL-2002
DEFINITION Sequence 3 from Patent WO0216653.
ACCESSION AX451337
VERSION AX451337.1 GI:21698388
KEYWORDS .
SOURCE synthetic construct
ORGANISM synthetic construct
other sequences; artificial sequences.
REFERENCE 1
AUTHORS Levitt,P.R., Mirnics,K., Kodavali,V.C. and Nimgaonkar,V.L.
TITLE Methods and systems for facilitating the diagnosis and treatment of schizophrenia
JOURNAL Patent: WO 0216653-A 3 28-FEB-2002;
University of Pittsburgh (US)
FEATURES Location/Qualifiers
source 1. .20300
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="A genomic sequence containing RGS4 nucleic acid sequence and sequ ences upstream and downstream to the RGS4 nucleic acid sequence"
ORIGIN

Query Match 100.0%; Score 4621; DB 6; Length 20300;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4621; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAATTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAATTA 60

Qy         61 GACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAAT 120

Qy        121 CACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAGC 180
```



```
Db      2341 GATGTCTTGATTCCAGCTTTACTCCTAAAACTTAAATGCTGAGGGGCCAACACCATGGCA 2400

Qy      2401 GTGGTAGGGATGGGAATGGGGGTCTTGTAACACACTACATAAACTACACGAAATAAACTA 2460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2401 GTGGTAGGGATGGGAATGGGGGTCTTGTAACACACTACATAAACTACACGAAATAAACTA 2460

Qy      2461 CATGAAACTCAACATGTTTGCAAGACTCAGTTCACATCCATGAGGAGCTCATGCTTCTCC 2520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2461 CATGAAACTCAACATGTTTGCAAGACTCAGTTCACATCCATGAGGAGCTCATGCTTCTCC 2520

Qy      2521 CTCCTGCTCCCCTAGCACACATGATTATCTCTATTTGGAAATGTTTGGCATTTTTGGTGA 2580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2521 CTCCTGCTCCCCTAGCACACATGATTATCTCTATTTGGAAATGTTTGGCATTTTTGGTGA 2580

Qy      2581 AGTGAATGGTTCAATAACTTTCTCCACCATCAGAACAAAAGCTCTTTAAGGTTAGGGATG 2640
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2581 AGTGAATGGTTCAATAACTTTCTCCACCATCAGAACAAAAGCTCTTTAAGGTTAGGGATG 2640

Qy      2641 GGATCATACACACTTCCCTTGTCCAAGTCCCCATCACCCCTTATCTAGACAATTGCTACA 2700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2641 GGATCATACACACTTCCCTTGTCCAAGTCCCCATCACCCCTTATCTAGACAATTGCTACA 2700

Qy      2701 GTTTCCTACACACTCTTCTAACCTCTTGCAGTCTATTTTCATAAAACAGCTAGAGAACTT 2760
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2701 GTTTCCTACACACTCTTCTAACCTCTTGCAGTCTATTTTCATAAAACAGCTAGAGAACTT 2760

Qy      2761 TGAGATGTAAGTCAAAAAATAGAACATGTCGCTCTTTCCCATTGTTTTTGAAATAAAGTT 2820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2761 TGAGATGTAAGTCAAAAAATAGAACATGTCGCTCTTTCCCATTGTTTTTGAAATAAAGTT 2820

Qy      2821 CAACCCCCTTACCAGGGTCAACAAGGCCCTGCAATGATTTGGTCCTGTTAAAAATTCTTT 2880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2821 CAACCCCCTTACCAGGGTCAACAAGGCCCTGCAATGATTTGGTCCTGTTAAAAATTCTTT 2880

Qy      2881 AGCCTTAACTCATGCTGTTCTTCCTTACACTCACTGCATTCTAGCCATTGAGGTTTCTAT 2940
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2881 AGCCTTAACTCATGCTGTTCTTCCTTACACTCACTGCATTCTAGCCATTGAGGTTTCTAT 2940

Qy      2941 GCATCAAACTTTTTTTGGTCCCAGCACTGTGCACATCCTTCTGGGTAGAATGCCCCTTGA 3000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2941 GCATCAAACTTTTTTTGGTCCCAGCACTGTGCACATCCTTCTGGGTAGAATGCCCCTTGA 3000

Qy      3001 TTTGTATAATTAGCACCTCCTTCATCATTTAGGTCTTAGTATAACTACTACCTTCTTAGA 3060
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3001 TTTGTATAATTAGCACCTCCTTCATCATTTAGGTCTTAGTATAACTACTACCTTCTTAGA 3060

Qy      3061 GAAGCTCTGCTTCTTCATCCTATAAAAAAGTAAAATTCCTTACCCTGTTATTTTTTAAGT 3120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3061 GAAGCTCTGCTTCTTCATCCTATAAAAAAGTAAAATTCCTTACCCTGTTATTTTTTAAGT 3120

Qy      3121 CATCCGTGTTTCATTCTGTTAAAGTTCTTATCACAATTTATCATTATTTTATTTACAGTC 3180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3121 CATCCGTGTTTCATTCTGTTAAAGTTCTTATCACAATTTATCATTATTTTATTTACAGTC 3180

Qy      3181 ATGTGCCACATAACAATGTTTCAGTCAGGGATAGAACACAAATGTATCTGGCCCCATAAT 3240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3181 ATGTGCCACATAACAATGTTTCAGTCAGGGATAGAACACAAATGTATCTGGCCCCATAAT 3240

Qy      3241 ATTATAAGCTGAGAAATTTCTATTAACTAGTGATATCGCAGCCATCATAAGTGTAATGCA 3300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3241 ATTATAAGCTGAGAAATTTCTATTAACTAGTGATATCGCAGCCATCATAAGTGTAATGCA 3300

Qy      3301 GGACATTACCTTTTCTATGTTTAGATATGTTAGATACACAAATATATTTCATTGTGTTAT 3360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3301 GGACATTACCTTTTCTATGTTTAGATATGTTAGATACACAAATATATTTCATTGTGTTAT 3360

Qy      3361 AATTTCCTACAGTATTCAGTACAGTAACATGCTGTACAGGTTTGTAACCTAGGAGTAATA 3420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3361 AATTTCCTACAGTATTCAGTACAGTAACATGCTGTACAGGTTTGTAACCTAGGAGTAATA 3420

Qy      3421 GGCTATACCATACAGCTTAGGTGTGTAGTAGGCTATAACCATCTAGGTTTGTGTAAGTAC 3480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3421 GGCTATACCATACAGCTTAGGTGTGTAGTAGGCTATAACCATCTAGGTTTGTGTAAGTAC 3480

Qy      3481 ATTCTATGATATTCCCACAATGATGAAATCACCTAACTACACATTTCTCAGAATGTTTCA 3540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3481 ATTCTATGATATTCCCACAATGATGAAATCACCTAACTACACATTTCTCAGAATGTTTCA 3540

Qy      3541 CTGTTGTGAAGTGACCCATGACTATATTTTCCTATATACTTGATATTTTTGTGCATCTGC 3600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3541 CTGTTGTGAAGTGACCCATGACTATATTTTCCTATATACTTGATATTTTTGTGCATCTGC 3600

Qy      3601 CCATGAGAATGTAGTGTAAGATCAAAGGATGCAAGAATGGGTTCTATCCAGTATAGTACC 3660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3601 CCATGAGAATGTAGTGTAAGATCAAAGGATGCAAGAATGGGTTCTATCCAGTATAGTACC 3660

Qy      3661 CACTACACTGGTGGATGTCAATATGTATTTGTTAGATTAATATCTCAAGAATGAGCACCT 3720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3661 CACTACACTGGTGGATGTCAATATGTATTTGTTAGATTAATATCTCAAGAATGAGCACCT 3720

Qy      3721 TTCTCAGACACATAAAAGATGCTCAATATAAAAGTTTGTTGAACTGAACGTTATTGGCAA 3780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3721 TTCTCAGACACATAAAAGATGCTCAATATAAAAGTTTGTTGAACTGAACGTTATTGGCAA 3780

Qy      3781 ATGTAACATGATCGGATTTAAAGAGGAGCGAAACAGAGGTCTGGCTCAAACACCATACTT 3840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3781 ATGTAACATGATCGGATTTAAAGAGGAGCGAAACAGAGGTCTGGCTCAAACACCATACTT 3840

Qy      3841 CTAGAGTGCATAAGAGGTAGCAGTTGATTACCACTGGCGACAGGAGAAAAAAGAGCTTGA 3900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3841 CTAGAGTGCATAAGAGGTAGCAGTTGATTACCACTGGCGACAGGAGAAAAAAGAGCTTGA 3900

Qy      3901 CCGCAGGGTACTGTGAAGACATTTCAGGTTGATGGCACAGAACAGGGGAAATACATAAAT 3960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3901 CCGCAGGGTACTGTGAAGACATTTCAGGTTGATGGCACAGAACAGGGGAAATACATAAAT 3960

Qy      3961 GTGTGGGAATATTCAGTGGTCTGGGATGACTACATAGTAGAATATAATGAAGAAAAGAGT 4020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3961 GTGTGGGAATATTCAGTGGTCTGGGATGACTACATAGTAGAATATAATGAAGAAAAGAGT 4020

Qy      4021 GGAAGGGAAAGATGAAAAGTTGGAATGGGGATGAATTATGAAAGTACCAGAATGTTATGC 4080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4021 GGAAGGGAAAGATGAAAAGTTGGAATGGGGATGAATTATGAAAGTACCAGAATGTTATGC 4080

Qy      4081 TAAGGAATCTAGATTTTAAAATGTGAGGGCAAATTGAAGTCCTGGGCACGTTACAAAACT 4140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4081 TAAGGAATCTAGATTTTAAAATGTGAGGGCAAATTGAAGTCCTGGGCACGTTACAAAACT 4140

Qy      4141 AGAGGTCATAAAGTTTACCCTAATTTACCAAGATTTCCTAGAGGATCTATAATTGGAATC 4200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4141 AGAGGTCATAAAGTTTACCCTAATTTACCAAGATTTCCTAGAGGATCTATAATTGGAATC 4200

Qy      4201 CAGATCTGCCTCTCTGTAAAGTTCAAGCACTTTCCATGACACCATACTGTTTCTTTCCAC 4260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4201 CAGATCTGCCTCTCTGTAAAGTTCAAGCACTTTCCATGACACCATACTGTTTCTTTCCAC 4260

Qy      4261 CTGCACAATGCAAATGAACTCTTATGAAACTGCTGTTTCTATCCTGGGCTAAATGTTGCA 4320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4261 CTGCACAATGCAAATGAACTCTTATGAAACTGCTGTTTCTATCCTGGGCTAAATGTTGCA 4320

Qy      4321 GAAAAAAGATTTAATCTTTGGGATAAGGCTATTTTGGGTTTTCTCCTACTTCTTGGGAAA 4380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4321 GAAAAAAGATTTAATCTTTGGGATAAGGCTATTTTGGGTTTTCTCCTACTTCTTGGGAAA 4380

Qy      4381 CAAGGTTTTCTTCCCCTGGCTAATTAAGTGTGGTATTGTTCTTCCAGGGAAATCAGTGAT 4440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4381 CAAGGTTTTCTTCCCCTGGCTAATTAAGTGTGGTATTGTTCTTCCAGGGAAATCAGTGAT 4440

Qy      4441 GCATCACCTGCTGCTATCAAATGTCAGGGTTGGAGTTCCTGATTTATTGCATGTGCCCAC 4500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4441 GCATCACCTGCTGCTATCAAATGTCAGGGTTGGAGTTCCTGATTTATTGCATGTGCCCAC 4500

Qy      4501 AAAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCT 4560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4501 AAAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCT 4560
```

```
Qy       4561 GTTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAA 4620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4561 GTTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAA 4620

Qy       4621 A 4621
              |
Db       4621 A 4621


RESULT 2
AL583850
LOCUS       AL583850              165329 bp    DNA     linear   PRI 15-NOV-2001
DEFINITION  Human DNA sequence from clone RP11-430G6 on chromosome 1, complete
            sequence.
ACCESSION   AL583850
VERSION     AL583850.5  GI:16973044
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 165329)
  AUTHORS   Tracey,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Nov 16, 2001 this sequence version replaced gi:15020514.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            RP11-430G6 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-430G6 It may be shorter because we sequence overlapping
            sections only once, except for a short overlap.
            The true right end of clone RP11-430G6 is at 165329 in this
            sequence. The true right end of clone RP11-331H2 is at 2000 in this
            sequence.
FEATURES             Location/Qualifiers
     source          1. .165329
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-430G6"
                     /clone_lib="RPCI-11.2"
ORIGIN

  Query Match             99.7%;  Score 4605.8;  DB 9;  Length 165329;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 4618;  Conservative    0;  Mismatches    2;  Indels    1;  Gaps    1;
```

```
Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAATTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71683 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAATTA 71742

Qy         61 GACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71743 GACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAAT 71802

Qy        121 CACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71803 CACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAGC 71862

Qy        181 CCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71863 CCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATA 71922

Qy        241 TATATATATATACACACACACACATATATATATACACATATATATACACATATATACA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71923 TATATATATATACACACACACACATATATATATACACATATATATACACATATATACA 71982

Qy        301 TATATACACATATATACACATATATATACACATACATATGTACACATATATATACACATA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71983 TATATACACATATATACACATATATATACACATACATATGTACACATATATATACACATA 72042

Qy        361 TGTATACACATATATACACATATATACACACATATATACACATATATACACACATATATA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72043 TGTATACACATATATACACATATATACACACATATATACACATATATACACACATATATA 72102

Qy        421 CACATATATACACATATATACACATATACACATATATACACATATATACATATATACACA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72103 CACATATATACACATATATACACATATACACATATATACACATATATACATATATACACA 72162

Qy        481 TATATATAATATACACACATATATATACACATATATACACACATATATACACATATATAC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72163 TATATATAATATACACACATATATATACACATATATACACACATATATACACATATATAC 72222

Qy        541 ACATATATATACACATATATACACATATATACATATATACACATATATATACATATATAC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72223 ACATATATATACACATATATACACATATATACATATATACACATATATATACATATATAC 72282

Qy        601 ACATATATACATATATACACATATATACATATATACACACATATATACACATACATATAC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72283 ACATATATACATATATACACATATATACATATATACACACATATATACACATACATATAC 72342

Qy        661 ACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72343 ACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATAT 72402

Qy        721 GCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72403 GCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTT 72462

Qy        781 TGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72463 TGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATAC 72522

Qy        841 ATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATATATATA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72523 ATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATATATATA 72582

Qy        901 TAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTATAGAAGAATGG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72583 TAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTATAGAAGAATGG 72642

Qy        961 CTAAGATGCCAAAGCTCAGTGTATGTGGCAGAGGCATGGTGTAGGGTGTGTCCAGGTTCA 1020
              ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db      72643 CTAAGATGCCAAAGCTCAGTGTATGTGGCAGAGGCATGGTGGAGGGTGTGTCCAGGTTCA 72702

Qy       1021 TATATTGCATTAAGTGTGAGAACACCCTGGAGTATGAACCAAGAAAATGCAAAAGCCAGA 1080
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db      72703 TATATTGCATTAAGTGTGAGAA-ACCCTGGAGTATGAACCAAGAAAATGCAAAAGCCAGA 72761
```

```
Qy       1081 AGTGATGGAGGAAATGAGACACAATAATGAAGATATTGAGAGGAGGGTGTGGGCCTAGAG 1140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72762 AGTGATGGAGGAAATGAGACACAATAATGAAGATATTGAGAGGAGGGTGTGGGCCTAGAG 72821

Qy       1141 TGAAGCTTTTCGTGCCAGTACTTCTTTTGAAGGCCCAGTTCTCTTCTCTCTCGGGGGCTC 1200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72822 TGAAGCTTTTCGTGCCAGTACTTCTTTTGAAGGCCCAGTTCTCTTCTCTCTCGGGGGCTC 72881

Qy       1201 CTTCATCTCTCATAGAGTCCACAGCTTTTAAGGGCCAACACTTGAGGTCAGCCTGGCTCT 1260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72882 CTTCATCTCTCATAGAGTCCACAGCTTTTAAGGGCCAACACTTGAGGTCAGCCTGGCTCT 72941

Qy       1261 CTCATTTGAGCTGGATAGAACATTTTAGAGCACCATCTATTCTTCAAGAGGAAGTTTAAA 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72942 CTCATTTGAGCTGGATAGAACATTTTAGAGCACCATCTATTCTTCAAGAGGAAGTTTAAA 73001

Qy       1321 AATAAAAGAACCTTGAAGAGGAAAAAATGTAGACATTCAATCTAACCTTTTCATTTTACT 1380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73002 AATAAAAGAACCTTGAAGAGGAAAAAATGTAGACATTCAATCTAACCTTTTCATTTTACT 73061

Qy       1381 AGCCAAAGCTAAATAGAATGCAGGTTACCTGTTTTTCAGCCAGGCACCATCATTTCCTAA 1440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73062 AGCCAAAGCTAAATAGAATGCAGGTTACCTGTTTTTCAGCCAGGCACCATCATTTCCTAA 73121

Qy       1441 TTGTTATAAAATTTATTATTATTGTTGTTATTATTATTATTTGCCATAAGAAGTTTCCCA 1500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73122 TTGTTATAAAATTTATTATTATTGTTGTTATTATTATTATTTGCCATAAGAAGTTTCCCA 73181

Qy       1501 TATCCTTTTAGTATAACAAAAACACAATTCACAAGCATTATAAAACCCATGGTGTCTAAC 1560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73182 TATCCTTTTAGTATAACAAAAACACAATTCACAAGCATTATAAAACCCATGGTGTCTAAC 73241

Qy       1561 TATTAAAAAAAATTAAGTGGAACACACTTGTCCCAGCTACTGGGGAGGCTGAGGAGGGAGG 1620
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73242 TATTAAAAAAAATTAAGTGGAACACACTTGTCCCAGCTACTGGGGAGGCTGAGGAGGGAGG 73301

Qy       1621 ATCACGTGATCCCAGGGGGTCAAGGTTATGGAGAGCTATGATTGTGCCACTGCACTCCAG 1680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73302 ATCACGTGATCCCAGGGGGTCAAGGTTATGGAGAGCTATGATTGTGCCACTGCACTCCAG 73361

Qy       1681 CCTGGGTGACAGGGAAAGACCCTGTCTCTAAAATTTTTTTTAAAAAAACTAAACTGGTTT 1740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73362 CCTGGGTGACAGGGAAAGACCCTGTCTCTAAAATTTTTTTTAAAAAAACTAAACTGGTTT 73421

Qy       1741 TATTACAGAGATTCTGGAGACAGCTACACATAAAAGGGTGGTATGCCTCATATTAGCTAC 1800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73422 TATTACAGAGATTCTGGAGACAGCTACACATAAAAGGGTGGTATGCCTCATATTAGCTAC 73481

Qy       1801 CCAGGGAGGTGGAATGCCAACTTAGGTGGTGTCACCACTATTAAAAATGCCCCAAAGCAA 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73482 CCAGGGAGGTGGAATGCCAACTTAGGTGGTGTCACCACTATTAAAAATGCCCCAAAGCAA 73541

Qy       1861 TCAAAACTGAGAACTTCCTGGGAGCTTAGCATTGTGCAAAAGCAGCACAAAACACTTAAA 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73542 TCAAAACTGAGAACTTCCTGGGAGCTTAGCATTGTGCAAAAGCAGCACAAAACACTTAAA 73601

Qy       1921 CAATTCACAGTTGTGTTGGAATGGGAAGGCCTGGAAATATAAACCAAAGAGTATATTGTC 1980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73602 CAATTCACAGTTGTGTTGGAATGGGAAGGCCTGGAAATATAAACCAAAGAGTATATTGTC 73661

Qy       1981 TAAATTGATAGAGATTACAATTGCCTGAAAGAAAAAGTTGACTTTTAACTAGAATGTTCA 2040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73662 TAAATTGATAGAGATTACAATTGCCTGAAAGAAAAAGTTGACTTTTAACTAGAATGTTCA 73721

Qy       2041 GAGTAGGTTTACAGAAGAAGCTCTTAAACTGGGCTCCAGTGGATTTGTCAATGCTTTGGA 2100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73722 GAGTAGGTTTACAGAAGAAGCTCTTAAACTGGGCTCCAGTGGATTTGTCAATGCTTTGGA 73781

Qy       2101 AGCTGGTGGGGTGGGAGGGTTGGAGGGGGCATAAAAAGTCATGTTGGTATGCTCTGCTCA 2160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73782 AGCTGGTGGGGTGGGAGGGTTGGAGGGGGCATAAAAAGTCATGTTGGTATGCTCTGCTCA 73841

Qy       2161 AGTCTCCATTCTGTTTCCTTTTCCTCTTTTCAATGTCATGTCCCATTATTTCATTATGGG 2220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73842 AGTCTCCATTCTGTTTCCTTTTCCTCTTTTCAATGTCATGTCCCATTATTTCATTATGGG 73901

Qy       2221 CTTCCCTTTATCCAGGATCAATATGCCACCTCTTGGTTGTCTTTTACCTACTTCTCCACC 2280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73902 CTTCCCTTTATCCAGGATCAATATGCCACCTCTTGGTTGTCTTTTACCTACTTCTCCACC 73961

Qy       2281 TCACTATGGAATCGTCCTTGGGTAGCTCCTGTGCTTGGGAACCTGCACGGGCACTTTTCT 2340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73962 TCACTATGGAATCGTCCTTGGGTAGCTCCTGTGCTTGGGAACCTGCACGGGCACTTTTCT 74021

Qy       2341 GATGTCTTGATTCCAGCTTTACTCCTAAAACTTAAATGCTGAGGGGCCAACACCATGGCA 2400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74022 GATGTCTTGATTCCAGCTTTACTCCTAAAACTTAAATGCTGAGGGGCCAACACCATGGCA 74081

Qy       2401 GTGGTAGGGATGGGAATGGGGGTCTTGTAACACACTACATAAACTACACGAAATAAACTA 2460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74082 GTGGTAGGGATGGGAATGGGGGTCTTGTAACACACTACATAAACTACACGAAATAAACTA 74141

Qy       2461 CATGAAACTCAACATGTTTGCAAGACTCAGTTCACATCCATGAGGAGCTCATGCTTCTCC 2520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74142 CATGAAACTCAACATGTTTGCAAGACTCAGTTCACATCCATGAGGAGCTCATGCTTCTCC 74201

Qy       2521 CTCCTGCTCCCCTAGCACACATGATTATCTCTATTTGGAAATGTTTGGCATTTTTGGTGA 2580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74202 CTCCTGCTCCCCTAGCACACATGATTATCTCTATTTGGAAATGTTTGGCATTTTTGGTGA 74261

Qy       2581 AGTGAATGGTTCAATAACTTTCTCCACCATCAGAACAAAAGCTCTTTAAGGTTAGGGATG 2640
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74262 AGTGAATGGTTCAATAACTTTCTCCACCATCAGAACAAAAGCTCTTTAAGGTTAGGGATG 74321

Qy       2641 GGATCATACACACTTCCCTTGTCCAAGTCCCCATCACCCCTTATCTAGACAATTGCTACA 2700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74322 GGATCATACACACTTCCCTTGTCCAAGTCCCCATCACCCCTTATCTAGACAATTGCTACA 74381

Qy       2701 GTTTCCTACACACTCTTCTAACCTCTTGCAGTCTATTTTCATAAAACAGCTAGAGAACTT 2760
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74382 GTTTCCTACACACTCTTCTAACCTCTTGCAGTCTATTTTCATAAAACAGCTAGAGAACTT 74441

Qy       2761 TGAGATGTAAGTCAAAAAATAGAACATGTCGCTCTTTCCCATTGTTTTTGAAATAAAGTT 2820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74442 TGAGATGTAAGTCAAAAAATAGAACATGTCGCTCTTTCCCATTGTTTTTGAAATAAAGTT 74501

Qy       2821 CAACCCCCTTACCAGGGTCAACAAGGCCCTGCAATGATTTGGTCCTGTTAAAAATTCTTT 2880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74502 CAACCCCCTTACCAGGGTCAACAAGGCCCTGCAATGATTTGGTCCTGTTAAAAATTCTTT 74561

Qy       2881 AGCCTTAACTCATGCTGTTCTTCCTTACACTCACTGCATTCTAGCCATTGAGGTTTCTAT 2940
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74562 AGCCTTAACTCATGCTGTTCTTCCTTACACTCACTGCATTCTAGCCATTGAGGTTTCTAT 74621

Qy       2941 GCATCAAACTTTTTTTGGTCCCAGCACTGTGCACATCCTTCTGGGTAGAATGCCCCTTGA 3000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74622 GCATCAAACTTTTTTTGGTCCCAGCACTGTGCACATCCTTCTGGGTAGAATGCCCCTTGA 74681

Qy       3001 TTTGTATAATTAGCACCTCCTTCATCATTTAGGTCTTAGTATAACTACTACCTTCTTAGA 3060
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74682 TTTGTATAATTAGCACCTCCTTCATCATTTAGGTCTTAGTATAACTACTACCTTCTTAGA 74741

Qy       3061 GAAGCTCTGCTTCTTCATCCTATAAAAAAGTAAAATTCCTTACCCTGTTATTTTTTAAGT 3120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74742 GAAGCTCTGCTTCTTCATCCTATAAAAAAGTAAAATTCCTTACCCTGTTATTTTTTAAGT 74801

Qy       3121 CATCCGTGTTTCATTCTGTTAAAGTTCTTATCACAATTTATCATTATTTTATTTACAGTC 3180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74802 CATCCGTGTTTCATTCTGTTAAAGTTCTTATCACAATTTATCATTATTTTATTTACAGTC 74861

Qy       3181 ATGTGCCACATAACAATGTTTCAGTCAGGGATAGAACACAAATGTATCTGGCCCCATAAT 3240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74862 ATGTGCCACATAACAATGTTTCAGTCAGGGATAGAACACAAATGTATCTGGCCCCATAAT 74921

Qy       3241 ATTATAAGCTGAGAAATTTCTATTAACTAGTGATATCGCAGCCATCATAAGTGTAATGCA 3300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      74922 ATTATAAGCTGAGAAATTTCTATTAACTAGTGATATCGCAGCCATCATAAGTGTAATGCA 74981

Qy       3301 GGACATTACCTTTTCTATGTTTAGATATGTTAGATACACAAATATATTTCATTGTGTTAT 3360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74982 GGACATTACCTTTTCTATGTTTAGATATGTTAGATACACAAATATATTTCATTGTGTTAT 75041

Qy       3361 AATTTCCTACAGTATTCAGTACAGTAACATGCTGTACAGGTTTGTAACCTAGGAGTAATA 3420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75042 AATTTCCTACAGTATTCAGTACAGTAACATGCTGTACAGGTTTGTAACCTAGGAGTAATA 75101

Qy       3421 GGCTATACCATACAGCTTAGGTGTGTAGTAGGCTATAACCATCTAGGTTTGTGTAAGTAC 3480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75102 GGCTATACCATACAGCTTAGGTGTGTAGTAGGCTATAACCATCTAGGTTTGTGTAAGTAC 75161

Qy       3481 ATTCTATGATATTCCCACAATGATGAAATCACCTAACTACACATTTCTCAGAATGTTTCA 3540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75162 ATTCTATGATATTCCCACAATGATGAAATCACCTAACTACACATTTCTCAGAATGTTTCA 75221

Qy       3541 CTGTTGTGAAGTGACCCATGACTATATTTTCCTATATACTTGATATTTTTGTGCATCTGC 3600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75222 CTGTTGTGAAGTGACCCATGACTATATTTTCCTATATACTTGATATTTTTGTGCATCTGC 75281

Qy       3601 CCATGAGAATGTAGTGTAAGATCAAAGGATGCAAGAATGGGTTCTATCCAGTATAGTACC 3660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75282 CCATGAGAATGTAGTGTAAGATCAAAGGATGCAAGAATGGGTTCTATCCAGTATAGTACC 75341

Qy       3661 CACTACACTGGTGGATGTCAATATGTATTTGTTAGATTAATATCTCAAGAATGAGCACCT 3720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75342 CACTACACTGGTGGATGTCAATATGTATTTGTTAGATTAATATCTCAAGAATGAGCACCT 75401

Qy       3721 TTCTCAGACACATAAAAGATGCTCAATATAAAAGTTTGTTGAACTGAACGTTATTGGCAA 3780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75402 TTCTCAGACACATAAAAGATGCTCAATATAAAAGTTTGTTGAACTGAACGTTATTGGCAA 75461

Qy       3781 ATGTAACATGATCGGATTTAAAGAGGAGCGAAACAGAGGTCTGGCTCAAACACCATACTT 3840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75462 ATGTAACATGATCGGATTTAAAGAGGAGCGAAACAGAGGTCTGGCTCAAACACCATACTT 75521

Qy       3841 CTAGAGTGCATAAGAGGTAGCAGTTGATTACCACTGGCGACAGGAGAAAAAAGAGCTTGA 3900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75522 CTAGAGTGCATAAGAGGTAGCAGTTGATTACCACTGGCGACAGGAGAAAAAAGAGCTTGA 75581

Qy       3901 CCGCAGGGTACTGTGAAGACATTTCAGGTTGATGGCACAGAACAGGGGAAATACATAAAT 3960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75582 CCGCAGGGTACTGTGAAGACATTTCAGGTTGATGGCACAGAACAGGGGAAATACATAAAT 75641

Qy       3961 GTGTGGGAATATTCAGTGGTCTGGGATGACTACATAGTAGAATATAATGAAGAAAAGAGT 4020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75642 GTGTGGGAATATTCAGTGGTCTGGGATGACTACATAGTAGAATATAATGAAGAAAAGAGT 75701

Qy       4021 GGAAGGGAAAGATGAAAAGTTGGAATGGGGATGAATTATGAAAGTACCAGAATGTTATGC 4080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75702 GGAAGGGAAAGATGAAAAGTTGGAATGGGGATGAATTATGAAAGTACCAGAATGTTATGC 75761

Qy       4081 TAAGGAATCTAGATTTTAAAATGTGAGGGCAAATTGAAGTCCTGGGCACGTTACAAAACT 4140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75762 TAAGGAATCTAGATTTTAAAATGTGAGGGCAAATTGAAGTCCTGGGCACGTTACAAAACT 75821

Qy       4141 AGAGGTCATAAAGTTTACCCTAATTTACCAAGATTTCCTAGAGGATCTATAATTGGAATC 4200
              ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db      75822 AGAGGTCATAAAGTTTACCCTAATTTACCAAGATTTCCTAGAGGATCTAGAATTGGAATC 75881

Qy       4201 CAGATCTGCCTCTCTGTAAAGTTCAAGCACTTTCCATGACACCATACTGTTTCTTTCCAC 4260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75882 CAGATCTGCCTCTCTGTAAAGTTCAAGCACTTTCCATGACACCATACTGTTTCTTTCCAC 75941

Qy       4261 CTGCACAATGCAAATGAACTCTTATGAAACTGCTGTTTCTATCCTGGGCTAAATGTTGCA 4320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75942 CTGCACAATGCAAATGAACTCTTATGAAACTGCTGTTTCTATCCTGGGCTAAATGTTGCA 76001

Qy       4321 GAAAAAAGATTTAATCTTTGGGATAAGGCTATTTTGGGTTTTCTCCTACTTCTTGGGAAA 4380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76002 GAAAAAAGATTTAATCTTTGGGATAAGGCTATTTTGGGTTTTCTCCTACTTCTTGGGAAA 76061

Qy       4381 CAAGGTTTTCTTCCCCTGGCTAATTAAGTGTGGTATTGTTCTTCCAGGGAAATCAGTGAT 4440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76062 CAAGGTTTTCTTCCCCTGGCTAATTAAGTGTGGTATTGTTCTTCCAGGGAAATCAGTGAT 76121

Qy       4441 GCATCACCTGCTGCTATCAAATGTCAGGGTTGGAGTTCCTGATTTATTGCATGTGCCCAC 4500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76122 GCATCACCTGCTGCTATCAAATGTCAGGGTTGGAGTTCCTGATTTATTGCATGTGCCCAC 76181

Qy       4501 AAAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCT 4560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76182 AAAGCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCT 76241

Qy       4561 GTTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAA 4620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      76242 GTTTACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAA 76301

Qy       4621 A 4621
              |
Db      76302 A 76302
```

```
RESULT 3
AC031977/c
LOCUS       AC031977              191699 bp    DNA     linear   HTG 12-APR-2001
DEFINITION  Homo sapiens chromosome 1 clone RP11-288O18, WORKING DRAFT
            SEQUENCE, 3 unordered pieces.
ACCESSION   AC031977
VERSION     AC031977.7  GI:13194952
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 191699)
  AUTHORS   Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
            Federspiel,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,
            Mao,J., Komp,C., Kottler,S., Lam,B., Marathe,R., Miranda,M.,
            Morehouse,A.J., Nguyen,M., Oefner,P., Palm,C.J., Ramirez,D.,
            Southwick,A.M., Webb,C., Wilhelmy,J., Yu,S. and Davis,R.W.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 191699)
  AUTHORS   Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
            Federspiel,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,
            Mao,J., Marathe,R., Morehouse,A.J., Oefner,P., Palm,C.J.,
            Ramirez,D., Wilhelmy,J., Yu,S. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-APR-2000) DNA Sequencing and Technology Center,
            Stanford University, 855 California Avenue, Palo Alto, CA 94304,
            USA
COMMENT     On Mar 4, 2001 this sequence version replaced gi:9665085.
            -------------- Genome Center
                 Center: Stanford DNA Sequencing and Technology Development
            Center
                 Center code: SDSTDC
                 Web site: http://sequence-www.stanford.edu/group/human/
                 Contact: hum-info@sequence.stanford.edu
            -------------- Project Information
                 Center project name: 880
                 Center clone name: RP11-288O18
            -------------- Summary Statistics
                 Sequencing Vector: M13mp18; X02513; 100% of reads
                 Sequencing Vector: plasmid; plasmid_accession; 0% of reads
                 Chemistry: Dye-primer; 1% of reads
                 Chemistry: Dye-terminator Big Dye; 99% of reads
                 Assembly program: Phrap; version 0.990319
                 Consensus quality: 190680 bases at least Q40
                 Consensus quality: 191287 bases at least Q30
                 Consensus quality: 191336 bases at least Q20
                 Insert size: 195548; agarose-fp
                 Insert size: 191499; sum-of-contigs
                 Quality coverage:  7.9x in Q20 bases; agarose-fp
                 Quality coverage:  8.1x in Q20 bases; sum-of-contigs.
```

```
           * NOTE: This is a 'working draft' sequence. It currently
           * consists of 3 contigs. The true order of the pieces
           * is not known and their order in this sequence record is
           * arbitrary. Gaps between the contigs are represented as
           * runs of N, but the exact sizes of the gaps are unknown.
           * This record will be updated with the finished sequence
           * as soon as it is available and the accession number will
           * be preserved.
           *       1    12646: contig of 12646 bp in length
           *   12647    12746: gap of unknown length
           *   12747    94961: contig of 82215 bp in length
           *   94962    95061: gap of unknown length
           *   95062   191699: contig of 96638 bp in length.
FEATURES             Location/Qualifiers
     source          1. .191699
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-288O18"
                     /clone_lib="RPCI human BAC library 11"
     misc_feature    1. .12646
                     /note="assembly_name:Contig5"
     misc_feature    12747. .94961
                     /note="assembly_name:Contig6
                     clone_end:T7"
     misc_feature    95062. .191699
                     /note="assembly_name:Contig7
                     clone_end:SP6"
ORIGIN

  Query Match             91.2%;  Score 4214.2;  DB 2;  Length 191699;
  Best Local Similarity   96.7%;  Pred. No. 0;
  Matches 4289;  Conservative    0;  Mismatches  133;  Indels   15;  Gaps    4;

Qy        194 GCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATATAC 253
              |||.
Db      12749 GCGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 12690

Qy        254 ACACACACACACATATATATATACACATATATATACACATATATACATATATACACATAT 313
                                                  |||| ||||| | | |
Db      12689 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTACACATATATATACAC 12630

Qy        314 ATACACATATATATACACATACATATGTACACATATATATACACATATGTATACACATAT 373
              |||   ||| | ||| | | ||||||| |||||| |||||||||||||| || ||| ||||
Db      12629 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACATATAT 12570

Qy        374 ATACACATATATACACACATATATACACATATATACACACATATATACACATATATACAC 433
              | |||||||||||||||||||||||||||||||    |||||||||||||||||||||||
Db      12569 ACACACATATATACACACATATATACACATAT----ACACATATATACACATATATACAC 12514

Qy        434 ATATATACACATATACACATATATACACATATATACATATATACACATATATATAATATA 493
              |||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db      12513 ATATATACACATATATACATATATACACATATATACATATATACACATATATATAATATA 12454

Qy        494 CACACATATATATACACATATATACACACATATATACACATATATACACATAT------A 547
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||      |
Db      12453 CACACATATATATACACATATATACACACATATATACACATATATACACATATATACACA 12394

Qy        548 TATACACATATATACACATATATACATATATACACATATATATACATATATACACATATA 607
              |||| |||||||||||||||| | ||||| ||| ||||||||||||||||||||||||||
Db      12393 TATATACATATATACACATATATATACATATATACACATATATACATATATACACATATA 12334

Qy        608 TACATATATACACATATATACATATATACACACATATATACACATACATATACACACACA 667
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12333 TACATATATACACATATATACATATATACACACATATATACACATACATATACACACACA 12274

Qy        668 TAGATATACATATATATACACATATATATACGTATATATATGTATATATATATGCTCCAG 727
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db      12273 TAGATATACATATATATACACATATATATGTATATATATGTATATATATATGCTCCAG 12214

Qy        728 AGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGC 787
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db      12213 AGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGC 12154

Qy        788 AGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATG 847
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12153 AGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATG 12094

Qy        848 TGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACAC----TATATATATATATAG 903
              |||||||||||||||||||||||||||||||||||||||||||    |||||||||||||||
Db      12093 TGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATATATATATATATAG 12034

Qy        904 TGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTATAGAAGAATGGCTA 963
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12033 TGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTATAGAAGAATGGCTA 11974

Qy        964 AGATGCCAAAGCTCAGTGTATGTGGCAGAGGCATGGTGTAGGGTGTGTCCAGGTTCATAT 1023
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11973 AGATGCCAAAGCTCAGTGTATGTGGCAGAGGCATGGTGGAGGGTGTGTCCAGGTTCATAT 11914

Qy       1024 ATTGCATTAAGTGTGAGAACACCCTGGAGTATGAACCAAGAAAATGCAAAAGCCAGAAGT 1083
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db      11913 ATTGCATTAAGTGTGAGAA-ACCCTGGAGTATGAACCAAGAAAATGCAAAAGCCAGAAGT 11855

Qy       1084 GATGGAGGAAATGAGACACAATAATGAAGATATTGAGAGGAGGGTGTGGGCCTAGAGTGA 1143
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11854 GATGGAGGAAATGAGACACAATAATGAAGATATTGAGAGGAGGGTGTGGGCCTAGAGTGA 11795

Qy       1144 AGCTTTTCGTGCCAGTACTTCTTTTGAAGGCCCAGTTCTCTTCTCTCTCGGGGGCTCCTT 1203
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11794 AGCTTTTCGTGCCAGTACTTCTTTTGAAGGCCCAGTTCTCTTCTCTCTCGGGGGCTCCTT 11735

Qy       1204 CATCTCTCATAGAGTCCACAGCTTTTAAGGGCCAACACTTGAGGTCAGCCTGGCTCTCTC 1263
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11734 CATCTCTCATAGAGTCCACAGCTTTTAAGGGCCAACACTTGAGGTCAGCCTGGCTCTCTC 11675

Qy       1264 ATTTGAGCTGGATAGAACATTTTAGAGCACCATCTATTCTTCAAGAGGAAGTTTAAAAAT 1323
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11674 ATTTGAGCTGGATAGAACATTTTAGAGCACCATCTATTCTTCAAGAGGAAGTTTAAAAAT 11615

Qy       1324 AAAAGAACCTTGAAGAGGAAAAAATGTAGACATTCAATCTAACCTTTTCATTTTACTAGC 1383
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11614 AAAAGAACCTTGAAGAGGAAAAAATGTAGACATTCAATCTAACCTTTTCATTTTACTAGC 11555

Qy       1384 CAAAGCTAAATAGAATGCAGGTTACCTGTTTTTCAGCCAGGCACCATCATTTCCTAATTG 1443
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11554 CAAAGCTAAATAGAATGCAGGTTACCTGTTTTTCAGCCAGGCACCATCATTTCCTAATTG 11495

Qy       1444 TTATAAAATTTATTATTATTGTTGTTATTATTATTATTTGCCATAAGAAGTTTCCCATAT 1503
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11494 TTATAAAATTTATTATTATTGTTGTTATTATTATTATTTGCCATAAGAAGTTTCCCATAT 11435

Qy       1504 CCTTTTAGTATAACAAAAACACAATTCACAAGCATTATAAAACCCATGGTGTCTAACTAT 1563
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11434 CCTTTTAGTATAACAAAAACACAATTCACAAGCATTATAAAACCCATGGTGTCTAACTAT 11375

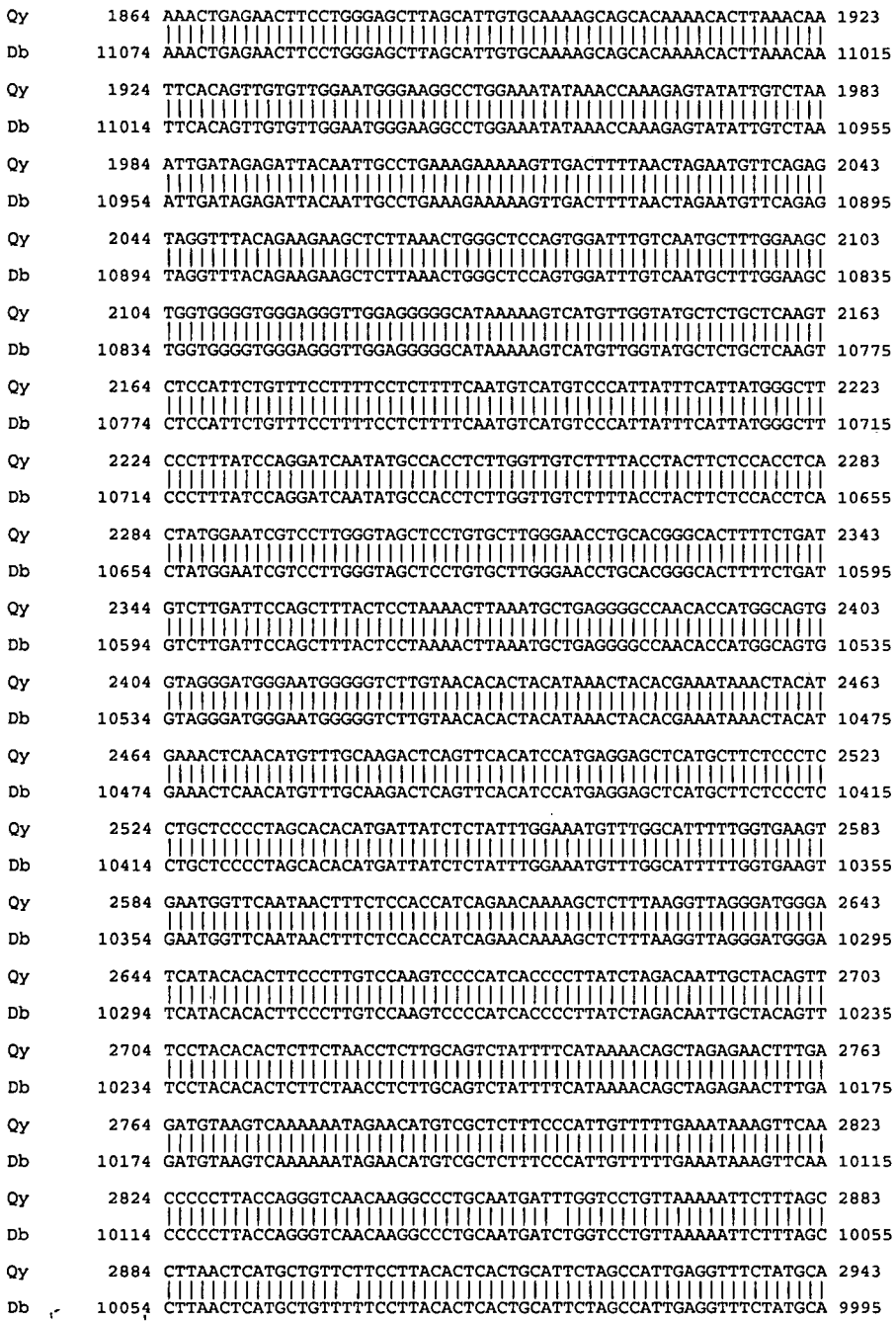
Qy       1564 TAAAAAAATTAAGTGGAACACACTTGTCCCAGCTACTGGGGAGGCTGAGGAGGGAGGATC 1623
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db      11374 TAAAAAAATTAAGTGGAACACACTTGTCCCAGCTACTGGGGAGGCCGAGGAGGGAGGATC 11315

Qy       1624 ACGTGATCCCAGGGGGTCAAGGTTATGGAGAGCTATGATTGTGCCACTGCACTCCAGCCT 1683
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11314 ACGTGATCCCAGGGGGTCAAGGTTATGGAGAGCTATGATTGTGCCACTGCACTCCAGCCT 11255

Qy       1684 GGGTGACAGGGAAAGACCCTGTCTCTAAAATTTTTTTTAAAAAAACTAAACTGGTTTTAT 1743
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11254 GGGTGACAGGGAAAGACCCTGTCTCTAAAATTTTTTTTAAAAAAACTAAACTGGTTTTAT 11195

Qy       1744 TACAGAGATTCTGGAGACAGCTACACATAAAAGGGTGGTATGCCTCATATTAGCTACCCA 1803
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11194 TACAGAGATTCTGGAGACAGCTACACATAAAAGGGTGGTATGCCTCATATTAGCTACCCA 11135

Qy       1804 GGGAGGTGGAATGCCAACTTAGGTGGTGTCACCACTATTAAAAATGCCCCAAAGCAATCA 1863
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11134 GGGAGGTGGAATGCCAACTTAGGTGGTGTCACCACTATTAAAAATGCCCCAAAGCAATCA 11075
```


Qy 1864 AAACTGAGAACTTCCTGGGAGCTTAGCATTGTGCAAAAGCAGCACAAAACACTTAAACAA 1923
Db 11074 AAACTGAGAACTTCCTGGGAGCTTAGCATTGTGCAAAAGCAGCACAAAACACTTAAACAA 11015
Qy 1924 TTCACAGTTGTGTTGGAATGGGAAGGCCTGGAAATATAAACCAAAGAGTATATTGTCTAA 1983
Db 11014 TTCACAGTTGTGTTGGAATGGGAAGGCCTGGAAATATAAACCAAAGAGTATATTGTCTAA 10955
Qy 1984 ATTGATAGAGATTACAATTGCCTGAAAGAAAAAGTTGACTTTTAACTAGAATGTTCAGAG 2043
Db 10954 ATTGATAGAGATTACAATTGCCTGAAAGAAAAAGTTGACTTTTAACTAGAATGTTCAGAG 10895
Qy 2044 TAGGTTTACAGAAGAAGCTCTTAAACTGGGCTCCAGTGGATTTGTCAATGCTTTGGAAGC 2103
Db 10894 TAGGTTTACAGAAGAAGCTCTTAAACTGGGCTCCAGTGGATTTGTCAATGCTTTGGAAGC 10835
Qy 2104 TGGTGGGGTGGGAGGGTTGGAGGGGGCATAAAAAGTCATGTTGGTATGCTCTGCTCAAGT 2163
Db 10834 TGGTGGGGTGGGAGGGTTGGAGGGGGCATAAAAAGTCATGTTGGTATGCTCTGCTCAAGT 10775
Qy 2164 CTCCATTCTGTTTCCTTTTCCTCTTTTCAATGTCATGTCCCATTATTTCATTATGGGCTT 2223
Db 10774 CTCCATTCTGTTTCCTTTTCCTCTTTTCAATGTCATGTCCCATTATTTCATTATGGGCTT 10715
Qy 2224 CCCTTTATCCAGGATCAATATGCCACCTCTTGGTTGTCTTTTACCTACTTCTCCACCTCA 2283
Db 10714 CCCTTTATCCAGGATCAATATGCCACCTCTTGGTTGTCTTTTACCTACTTCTCCACCTCA 10655
Qy 2284 CTATGGAATCGTCCTTGGGTAGCTCCTGTGCTTGGGAACCTGCACGGGCACTTTTCTGAT 2343
Db 10654 CTATGGAATCGTCCTTGGGTAGCTCCTGTGCTTGGGAACCTGCACGGGCACTTTTCTGAT 10595
Qy 2344 GTCTTGATTCCAGCTTTACTCCTAAAACTTAAATGCTGAGGGGCCAACACCATGGCAGTG 2403
Db 10594 GTCTTGATTCCAGCTTTACTCCTAAAACTTAAATGCTGAGGGGCCAACACCATGGCAGTG 10535
Qy 2404 GTAGGGATGGGAATGGGGGTCTTGTAACACACTACATAAACTACACGAAATAAACTACAT 2463
Db 10534 GTAGGGATGGGAATGGGGGTCTTGTAACACACTACATAAACTACACGAAATAAACTACAT 10475
Qy 2464 GAAACTCAACATGTTTGCAAGACTCAGTTCACATCCATGAGGAGCTCATGCTTCTCCCTC 2523
Db 10474 GAAACTCAACATGTTTGCAAGACTCAGTTCACATCCATGAGGAGCTCATGCTTCTCCCTC 10415
Qy 2524 CTGCTCCCCTAGCACACATGATTATCTCTATTTGGAAATGTTTGGCATTTTTGGTGAAGT 2583
Db 10414 CTGCTCCCCTAGCACACATGATTATCTCTATTTGGAAATGTTTGGCATTTTTGGTGAAGT 10355
Qy 2584 GAATGGTTCAATAACTTTCTCCACCATCAGAACAAAAGCTCTTTAAGGTTAGGGATGGGA 2643
Db 10354 GAATGGTTCAATAACTTTCTCCACCATCAGAACAAAAGCTCTTTAAGGTTAGGGATGGGA 10295
Qy 2644 TCATACACACTTCCCTTGTCCAAGTCCCCATCACCCCTTATCTAGACAATTGCTACAGTT 2703
Db 10294 TCATACACACTTCCCTTGTCCAAGTCCCCATCACCCCTTATCTAGACAATTGCTACAGTT 10235
Qy 2704 TCCTACACACTCTTCTAACCTCTTGCAGTCTATTTTCATAAAACAGCTAGAGAACTTTGA 2763
Db 10234 TCCTACACACTCTTCTAACCTCTTGCAGTCTATTTTCATAAAACAGCTAGAGAACTTTGA 10175
Qy 2764 GATGTAAGTCAAAAAATAGAACATGTCGCTCTTTCCCATTGTTTTTGAAATAAAGTTCAA 2823
Db 10174 GATGTAAGTCAAAAAATAGAACATGTCGCTCTTTCCCATTGTTTTTGAAATAAAGTTCAA 10115
Qy 2824 CCCCCTTACCAGGGTCAACAAGGCCCTGCAATGATTTGGTCCTGTTAAAAATTCTTTAGC 2883
Db 10114 CCCCCTTACCAGGGTCAACAAGGCCCTGCAATGATCTGGTCCTGTTAAAAATTCTTTAGC 10055
Qy 2884 CTTAACTCATGCTGTTCTTCCTTACACTCACTGCATTCTAGCCATTGAGGTTTCTATGCA 2943
Db 10054 CTTAACTCATGCTGTTTTTCCTTACACTCACTGCATTCTAGCCATTGAGGTTTCTATGCA 9995
Qy 2944 TCAAACTTTTTTTGGTCCCAGCACTGTGCACATCCTTCTGGGTAGAATGCCCCTTGATTT 3003
Db 9994 TCAAACTTTTTTTGGTCCCAGCACTGTGCACATCCTTCTGGGTAGAATGCCCCTTGATTT 9935
Qy 3004 GTATAATTAGCACCTCCTTCATCATTTAGGTCTTAGTATAACTACTACCTTCTTAGAGAA 3063
Db 9934 GTATAATTAGCACCTCCTTCATCATTTAGGTCTTAGTATAACTACTACCTTCTTAGAGAA 9875
Qy 3064 GCTCTGCTTCTTCATCCTATAAAAAAGTAAAATTCCTTACCCTGTTATTTTTTAAGTCAT 3123
Db 9874 GCTCTGCTTCTTCATCCTATAAAAAAGTAAAATTCCTTACCCTGTTATTTTTTAAGTCAT 9815
Qy 3124 CCGTGTTTCATTCTGTTAAAGTTCTTATCACAATTTATCATTATTTTATTTACAGTCATG 3183
Db 9814 CCGTGTTTCATTCTGTTAAAGTTCTTATCACAATTTATCATTATTTTATTTACAGTCATG 9755
Qy 3184 TGCCACATAACAATGTTTCAGTCAGGGATAGAACACAAATGTATCTGGCCCCATAATATT 3243
Db 9754 TGCCACATAACAATGTTTCAGTCAGGGATAGAACACAAATGTATCTGGCCCCATAATATT 9695
Qy 3244 ATAAGCTGAGAAATTTCTATTAACTAGTGATATCGCAGCCATCATAAGTGTAATGCAGGA 3303
Db 9694 ATAAGCTGAGAAATTTCTATTAACTAGTGATATCGCAGCCATCATAAGTGTAATGCAGGA 9635
Qy 3304 CATTACCTTTTCTATGTTTAGATATGTTAGATACACAAATATATTTCATTGTGTTATAAT 3363
Db 9634 CATTACCTTTTCTATGTTTAGATATGTTAGATACACAAATATATTTCATTGTGTTATAAT 9575
Qy 3364 TTCCTACAGTATTCAGTACAGTAACATGCTGTACAGGTTTGTAACCTAGGAGTAATAGGC 3423
Db 9574 TTCCTACAGTATTCAGTACAGTAACATGCTGTACAGGTTTGTAACCTAGGAGTAATAGGC 9515
Qy 3424 TATACCATACAGCTTAGGTGTGTAGTAGGCTATAACCATCTAGGTTTGTGTAAGTACATT 3483
Db 9514 TATACCATACAGCTTAGGTGTGTAGTAGGCTATAACCATCTAGGTTTGTGTAAGTACATT 9455
Qy 3484 CTATGATATTCCCACAATGATGAAATCACCTAACTACACATTTCTCAGAATGTTTCACTG 3543
Db 9454 CTATGATATTCCCACAATGATGAAATCACCTAACTACACATTTCTCAGAATGTTTCACTG 9395
Qy 3544 TTGTGAAGTGACCCATGACTATATTTTCCTATATACTTGATATTTTTGTGCATCTGCCCA 3603
Db 9394 TTGTGAAGTGACCCATGACTATATTTTCCTATATACTTGATATTTTTGTGCATCTGCCCA 9335
Qy 3604 TGAGAATGTAGTGTAAGATCAAAGGATGCAAGAATGGGTTCTATCCAGTATAGTACCCAC 3663
Db 9334 TGAGAATGTAGTGTAAGATCAAAGGATGCAAGAATGGGTTCTATCCAGTATAGTACCCAC 9275
Qy 3664 TACACTGGTGGATGTCAATATGTATTTGTTAGATTAATATCTCAAGAATGAGCACCTTTC 3723
Db 9274 TACACTGGTGGATGTCAATATGTATTTGTTAGATTAATATCTCAAGAATGAGCACCTTTC 9215
Qy 3724 TCAGACACATAAAAGATGCTCAATATAAAAGTTTGTTGAACTGAACGTTATTGGCAAATG 3783
Db 9214 TCAGACACATAAAAGATGCTCAATATAAAAGTTTGTTGAACTGAACGTTATTGGCAAATG 9155
Qy 3784 TAACATGATCGGATTTAAAGAGGAGCGAAACAGAGGTCTGGCTCAAACACCATACTTCTA 3843
Db 9154 TAACATGATCGGATTTAAAGAGGAGCGAAACAGAGGTCTGGCTCAAACACCATACTTCTA 9095
Qy 3844 GAGTGCATAAGAGGTAGCAGTTGATTACCACTGGCGACAGGAGAAAAAAGAGCTTGACCG 3903
Db 9094 GAGTGCATAAGAGGTAGCAGTTGATTACCACTGGCGACAGGAGAAAAAAGAGCTTGACCG 9035
Qy 3904 CAGGGTACTGTGAAGACATTTCAGGTTGATGGCACAGAACAGGGGAAATACATAAATGTG 3963
Db 9034 CAGGGTACTGTGAAGACATTTCAGGTTGATGGCACAGAACAGGGGAAATACATAAATGTG 8975
Qy 3964 TGGGAATATTCAGTGGTCTGGGATGACTACATAGTAGAATATAATGAAGAAAAGAGTGGA 4023
Db 8974 TGGGAATATTCAGTGGTCTGGGATGACTACATAGTAGAATATAATGAAGAAAAGAGTGGA 8915
Qy 4024 AGGGAAAGATGAAAAGTTGGAATGGGGATGAATTATGAAAGTACCAGAATGTTATGCTAA 4083

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8914 AGGGAAAGATGAAAAGTTGGAATGGGGATGAATTATGAAAGTACCAGAATGTTATGCTAA 8855

Qy       4084 GGAATCTAGATTTTAAAATGTGAGGGCAAATTGAAGTCCTGGGCACGTTACAAAACTAGA 4143
                |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db       8854 GGAATCTAGATTTTAAAATGTGAGGGCAAATTGAAGTCCAGGGCACGTTACAAAACTAGA 8795

Qy       4144 GGTCATAAAGTTTACCCTAATTTACCAAGATTTCCTAGAGGATCTATAATTGGAATCCAG 4203
                |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db       8794 GGTCATAAAGTTTACCCTAATTTACCAAGATTTCCTAGAGGATCTAGAATTGGAATCCAG 8735

Qy       4204 ATCTGCCTCTCTGTAAAGTTCAAGCACTTTCCATGACACCATACTGTTTCTTTCCACCTG 4263
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8734 ATCTGCCTCTCTGTAAAGTTCAAGCACTTTCCATGACACCATACTGTTTCTTTCCACCTG 8675

Qy       4264 CACAATGCAAATGAACTCTTATGAAACTGCTGTTTCTATCCTGGGCTAAATGTTGCAGAA 4323
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8674 CACAATGCAAATGAACTCTTATGAAACTGCTGTTTCTATCCTGGGCTAAATGTTGCAGAA 8615

Qy       4324 AAAAGATTTAATCTTTGGGATAAGGCTATTTTGGGTTTTCTCCTACTTCTTGGGAAACAA 4383
                |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db       8614 AAAAGATTTAATCTTTGGGATAAGGCTATTTTGGGTTTTCTCCTCCTTCTTGGGAAACAA 8555

Qy       4384 GGTTTTCTTCCCCTGGCTAATTAAGTGTGGTATTGTTCTTCCAGGGAAATCAGTGATGCA 4443
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8554 GGTTTTCTTCCCCTGGCTAATTAAGTGTGGTATTGTTCTTCCAGGGAAATCAGTGATGCA 8495

Qy       4444 TCACCTGCTGCTATCAAATGTCAGGGTTGGAGTTCCTGATTTATTGCATGTGCCCACAAA 4503
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8494 TCACCTGCTGCTATCAAATGTCAGGGTTGGAGTTCCTGATTTATTGCATGTGCCCACAAA 8435

Qy       4504 GCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTGTT 4563
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8434 GCTTGGTGCAAAGAATTGGACACATTTCCCAAAAGTAAGACATACTGGGAAGTCCCTGTT 8375

Qy       4564 TACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAA 4620
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8374 TACCTTCCTGGTATACAGCATCCTCCAGCCCCATATCTTTGCTTTTTAGTCCTAAAA 8318


RESULT 4
AC097328
LOCUS       AC097328               203216 bp    DNA     linear   HTG 20-OCT-2001
DEFINITION  Pan troglodytes clone RP43-174K14, WORKING DRAFT SEQUENCE, 21
            unordered pieces.
ACCESSION   AC097328
VERSION     AC097328.1  GI:16118151
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 203216)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 203216)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-OCT-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: ZUAM
                Center clone name: RP43-174K14
            -------------- Summary Statistics
                Sequencing vector: Plasmid; M77789
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 196812 bases at least Q40
                Consensus quality: 208422 bases at least Q30
                Consensus quality: 218080 bases at least Q20
                Estimated insert size: 228347; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 7.1x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 21 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    30770: contig of 30770 bp in length
            *    30771    30870: gap of unknown length
            *    30871    52356: contig of 21486 bp in length
            *    52357    52456: gap of unknown length
            *    52457    72697: contig of 20241 bp in length
            *    72698    72797: gap of unknown length
            *    72798    90396: contig of 17599 bp in length
            *    90397    90496: gap of unknown length
            *    90497    96134: contig of 5638 bp in length
            *    96135    96234: gap of unknown length
            *    96235   114433: contig of 18199 bp in length
            *   114434   114533: gap of unknown length
            *   114534   125184: contig of 10651 bp in length
            *   125185   125284: gap of unknown length
            *   125285   135194: contig of 9910 bp in length
            *   135195   135294: gap of unknown length
            *   135295   140273: contig of 4979 bp in length
```

```
            *   140274   140373: gap of unknown length
            *   140374   148490: contig of 8117 bp in length
            *   148491   148590: gap of unknown length
            *   148591   153435: contig of 4845 bp in length
            *   153436   153535: gap of unknown length
            *   153536   162229: contig of 8694 bp in length
            *   162230   162329: gap of unknown length
            *   162330   167843: contig of 5514 bp in length
            *   167844   167943: gap of unknown length
            *   167944   175297: contig of 7354 bp in length
            *   175298   175397: gap of unknown length
            *   175398   179409: contig of 4012 bp in length
            *   179410   179509: gap of unknown length
            *   179510   185931: contig of 6422 bp in length
            *   185932   186031: gap of unknown length
            *   186032   190690: contig of 4659 bp in length
            *   190691   190790: gap of unknown length
            *   190791   194808: contig of 4018 bp in length
            *   194809   194908: gap of unknown length
            *   194909   198788: contig of 3880 bp in length
            *   198789   198888: gap of unknown length
            *   198889   201055: contig of 2167 bp in length
            *   201056   201155: gap of unknown length
            *   201156   203216: contig of 2061 bp in length.
FEATURES             Location/Qualifiers
     source          1. .203216
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="RP43-174K14"
ORIGIN

  Query Match             8.9%;  Score 411.2;  DB 2;  Length 203216;
  Best Local Similarity   76.7%;  Pred. No. 1e-71;
  Matches  554;  Conservative    0;  Mismatches  163;  Indels    5;  Gaps    4;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCA-TCTCTACTAAAAATACAAAATT 59
              ||||||| |||||||||| |||||||||||||||||  |||||||||||| |||||||||
Db     114617 AGTTCAAAACCAGCCTGACCAACATGGTGAAACCCCCGTCTCTACTAAAAACACAAAATT 114676

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   |||||||||  | ||  ||||||| ||||||||||   ||| ||   ||||||||||
Db     114677 AGCTGGGCATGGTGGTGCATGCCTGTAGTCCCAGCTACTCAGGAAGCTAAGGCAGGAGAA 114736

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGC-AAGATCATGCCATTGCACTCCA 178
              ||||||||||  | || || |||||||| | || |  |||||||||| |||||||||
Db     114737 TCACTTGAACC--CAGGAGGCGGAGGTTGCAGTGAACTGAGATCATGCCACTGCACTCCA 114794

Qy        179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAGGCATTTTATATATATA 238
              |||   | ||||| |||        |  |    | | |    |       | | ||||||||
Db     114795 GCCTGAGTAACAAAAGCATCTCAAAAAATATATATATATGTATATATGTGTGTATATATA 114854

Qy        239 TATATATATATATACACACACACACATATATATATACACATATATATACACATATATA 298
              | | |||||||||       | | | |  | ||||||||     | | |||| |   | ||||
Db     114855 TGTGTATATATATGTGTGTATATATATGTGTATATATATGTGTGTGTATATATGTGTATA 114914

Qy        299 CATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACA 358
              |||||||  | |||||||  | |||||||||  | |||  |||| ||  | |||||||||  | |
Db     114915 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 114974

Qy        359 TATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATA 418
              |||  ||||  |  |||||||  | |||||||  | |  |||||||  | |||||||  | |  |||||
Db     114975 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 115034

Qy        419 TACACATATATACACATATATACACATATACACATATATACACATATATACATATATACA 478
              ||  | |||||||  | |||||||  | |||||  | |||||||  | |||||||  |||||||  |
Db     115035 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 115094

Qy        479 CATATATATAATATACACACATATATATACACATATATACACACATATATACACACATATAT 538
              ||||||||  |||||  | |  |||||||||  | |||||||  | |  |||||||  | ||||||
Db     115095 TATATATAT-ATATATATATATATATATATATATATATATATATATATATATATATATAT 115153

Qy        539 ACACATATATATACACATATATACACATATATACATATATACACATATATATACATATAT 598
              | | |||||||||  | |||||||  | |||||||  |||||||  | |||||||||  ||||||
Db     115154 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 115213

Qy        599 ACACATATATACATATATACACATATATACATATATACACACATATATACACATACATAT 658
              | | |||||||  |||||||  | |||||||  |||||||  | |  |||||||  | |||  ||||
Db     115214 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 115273

Qy        659 ACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATAT 718
              | | |   |||  |||||  |||||||||  | |||||||||   |||||||||  |||||||||
Db     115274 ATATATCTATATATATATATATATATATATATATATATATATATATATATATATAT 115333

Qy        719 AT 720
              ||
Db     115334 AT 115335


RESULT 5
CR759943
LOCUS       CR759943              104321 bp    DNA     linear   HTG 01-OCT-2004
DEFINITION  Homo sapiens chromosome 6 clone DAAP-21A6, WORKING DRAFT SEQUENCE.
ACCESSION   CR759943
VERSION     CR759943.2  GI:53748680
KEYWORDS    HTG; HTGS_PHASE2; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 104321)
  AUTHORS   Phillimore,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 1, 2004 this sequence version replaced gi:51965768.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bAP21A6
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 103081 bases at least Q40
            Consensus quality: 103313 bases at least Q30
            Consensus quality: 103455 bases at least Q20
            Insert size: 104321; sum-of-contigs
            Insert size: 110232; 0.4% error; agarose-fp
            Quality coverage: 7.35x in Q20 bases; sum-of-contigs Quality
            coverage: 7.39x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   104321: contig of 104321 bp in length.
FEATURES             Location/Qualifiers
     source          1. .104321
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="DAAP-21A6"
                     /clone_lib="APD"
     misc_feature    1. .104321
                     /note="assembly_fragment:01274"
ORIGIN
```

```
  Query Match             8.8%;  Score 405.4;  DB 2;  Length 104321;
  Best Local Similarity   75.3%;  Pred. No. 1.6e-70;
  Matches  558;  Conservative    0;  Mismatches  176;  Indels    7;  Gaps    4;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC---AAAA 57
              ||||| ||||||||||   || ||||||||||||||||||||||||||||||    ||||
Db      96311 AGTTCGAGACCAGCCTGGCCAGCATGGTGAAACCCCATCTCTACTAAAAATACAAAAAAA 96370

Qy         58 TTAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAG 117
              |||| ||||  ||||    |  | || || ||||| |||| | |||||| | |||||| |
Db      96371 TTAGCCAGGTGTGGTAGCGCGTGTCTCTAGTCCCACCTACCTGGGAGGCTGTGGCAGGCG 96430

Qy        118 AATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGC-AAGATCATGCCATTGCACTC 176
              |||  ||| ||||  |  | || ||||||||| | ||||  ||||||  |||| |||||||
Db      96431 AATTGCTTCAACC--CGTGAGGCGGAGGTTGCAGTGAGCTGAGATCACGCCACTGCACTC 96488

Qy        177 CAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATA 236
              |||||  |||||| ||||| | |      |||||| |||||||| |       | ||||||||
Db      96489 CAGCCTGGGCAAC-AGAGCAAGAATCAGTCTCAAAAAAAAAATATATATATATATATATA 96547

Qy        237 TATATATATATATATACACACACACACATATATATATACACATATATATACACATATA 296
              |||||||||||||||| | | | | | | |||||||||||| | ||||||||| | |||||
Db      96548 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 96607

Qy        297 TACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACA 356
              || ||||||| | ||||||| | ||||||||| | ||| |||| || | |||||||||  |
Db      96608 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 96667

Qy        357 CATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATA 416
              |||| |||| | ||||||| | ||||||| | | ||||||| | ||||||| | | |||
Db      96668 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 96727

Qy        417 TATACACATATATACACATATATACACATATACACATATATACACATATATACATATATA 476
              |||| | ||||||| | ||||||| | ||||| | ||||||| | ||||||| ||| |||
Db      96728 TATATATATATATATATATATATATATATATATATATATATATATATATATATATAAATA 96787

Qy        477 CACATATATATAATATACACACATATATATACACATATATACACACATATATACACATAT 536
              | ||||| |||  |           ||| | ||||||| | | ||||||| | ||||
Db      96788 TAAATATAAATATAAATATATATATATAATATATATATATATATATATATATATATAT 96847

Qy        537 ATACACATATATATACACATATATACACATATATACATATATACACATATATATACATAT 596
              ||| | ||||||||| | ||||||| | ||||||| ||||||| | |||||||||  ||||
Db      96848 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 96907

Qy        597 ATACACATATATACATATATACACATATATACATATATACACACATATATACACATACAT 656
              ||| | ||||||| ||||||| | ||||||| ||||||| | | ||||||| | ||| ||
Db      96908 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 96967

Qy        657 ATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATAT 716
              ||| | | | ||| ||||| |||| |||| | |||| | ||  ||  | ||||| |||||
Db      96968 ATATATATATATATATATATATATGTATATATATATGTGTATATACGTGTATGTGTATAT 97027

Qy        717 ATATGCTCCAGAGTTCATAAG 737
              |||       | ||  ||| |
Db      97028 ATACATGTATGTGTATATATG 97048


RESULT 6
AC097330/c
LOCUS       AC097330              181946 bp    DNA     linear   HTG 20-OCT-2001
DEFINITION  Pan troglodytes clone RP43-54O15, WORKING DRAFT SEQUENCE, 16
            unordered pieces.
ACCESSION   AC097330
VERSION     AC097330.1  GI:16118153
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 181946)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 181946)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-OCT-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: ZUAP
                Center clone name: RP43-54O15
            -------------- Summary Statistics
                Sequencing vector: Plasmid; M77789
                Sequencing vector: M13; L08821
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 169928 bases at least Q40
                Consensus quality: 176100 bases at least Q30
                Consensus quality: 180968 bases at least Q20
                Estimated insert size: 183838; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 6.9x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 16 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
```

```
           * as soon as it is available and the accession number will
           * be preserved.
           *        1    17715: contig of 17715 bp in length
           *    17716    17815: gap of unknown length
           *    17816    44117: contig of 26302 bp in length
           *    44118    44217: gap of unknown length
           *    44218    67185: contig of 22968 bp in length
           *    67186    67285: gap of unknown length
           *    67286    84740: contig of 17455 bp in length
           *    84741    84840: gap of unknown length
           *    84841   100199: contig of 15359 bp in length
           *   100200   100299: gap of unknown length
           *   100300   112190: contig of 11891 bp in length
           *   112191   112290: gap of unknown length
           *   112291   124548: contig of 12258 bp in length
           *   124549   124648: gap of unknown length
           *   124649   134134: contig of 9486 bp in length
           *   134135   134234: gap of unknown length
           *   134235   143821: contig of 9587 bp in length
           *   143822   143921: gap of unknown length
           *   143922   152987: contig of 9066 bp in length
           *   152988   153087: gap of unknown length
           *   153088   159063: contig of 5976 bp in length
           *   159064   159163: gap of unknown length
           *   159164   166547: contig of 7384 bp in length
           *   166548   166647: gap of unknown length
           *   166648   171663: contig of 5016 bp in length
           *   171664   171763: gap of unknown length
           *   171764   177471: contig of 5708 bp in length
           *   177472   177571: gap of unknown length
           *   177572   179822: contig of 2251 bp in length
           *   179823   179922: gap of unknown length
           *   179923   181946: contig of 2024 bp in length.
FEATURES             Location/Qualifiers
     source          1. .181946
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="RP43-54O15"
ORIGIN

  Query Match             8.8%;  Score 405.2;  DB 2;  Length 181946;
  Best Local Similarity   68.3%;  Pred. No. 1.6e-70;
  Matches  634;  Conservative    0;  Mismatches  288;  Indels    6;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||| |||||| |||| | | ||||||||||||| ||||||||| |||||| ||||||
Db     155131 AGTTCGAGACCAACCTGGACTATATGGTGAAACCCCGTCTCTACTAGAAATACAAAAATT 155072

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   |||||||||  |||||||||||||||||| |||||| ||||||  ||||||||||
Db     155071 AGCTGGGCATGGTGGCACACGCCTGTAATCCCAACTACTTGGGAGGCTAAGGCAGGAGAA 155012

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCCA 178
              |||||||||||   | || ||  |||||||| | ||| | |||||   ||| |||||||||
Db     155011 TCACTTGAAC--TCGGGAGGCAGAGGTTGCAGTGAGCCCAGATCGCACCACTGCACTCCA 154954

Qy        179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAGGCATTTTATATATATA 238
              ||| |||| |       |   || |||     | |||| |||| |     |       | | |
Db     154953 GCCTAGGCGAGTGAGACTCCATCTCAAAAAAAAAAAGAAAAGAAAATTGAGGGCTGGACA 154894

Qy        239 TATATATATATACACACACACACACATATATATATACACATATATATACACATATATA 298
              |      || |     |  | ||    | |  ||||| | |||||||||  | |||||||
Db     154893 CAGTGGCTTACACCTGTAATCCCAGCACTTTGGATATATATATATATATATATATATA 154834

Qy        299 CATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACA 358
              ||||||| | ||||||| | ||||||||| | ||| |||| || | |||||||||  | |
Db     154833 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 154774

Qy        359 TATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATA 418
              ||| |||| | ||||||| | ||||||| | | ||||||| | ||||||| | | |||||
Db     154773 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 154714

Qy        419 TACACATATATACACATATATACACATATACACATATATACACATATATACATATATACA 478
              || | ||||||| | ||||||| | ||||| | ||||||| | ||||||| ||||||| |
Db     154713 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 154654

Qy        479 CATATATATA-ATATACACACATATATATACACATATATACACACATATATACACATATA 537
              |||||||||  |||||  | | |||||||||  | |||||||  | |  |||||||  | |||||
Db     154653 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 154594

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATACATATA 597
              || | |||||||||  | |||||||  | |||||||  |||||||  | |||||||||  |||||
Db     154593 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 154534

Qy        598 TACACATATATACATATATACACATATATACATATATACACACATATATACACATACATA 657
              || | |||||||  |||||||  | |||||||  |||||||  | |  |||||||  | |||  |||
Db     154533 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 154474

Qy        658 TACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATA 717
              || | | | |||  |||||  |||||||||  | |||||||||  |||||||||  ||||||||
Db     154473 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 154414

Qy        718 TATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGA 777
              |||      |  |||       |  |  ||    |    |||  |   | | | |
Db     154413 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 154354

Qy        778 GTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAA 837
                | | |    | |||    | |   | || |      |       | |||       | |
Db     154353 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 154294

Qy        838 TACATAAATGTGTAATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTAACACTATATATA 896
              || ||| || | ||   | |  ||| | || |  ||  ||    |  ||    ||||||||
Db     154293 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 154234

Qy        897 TATATAGTGAAAGCAGTGTAAACCTAAA 924
              ||||||   | |    | || |  || |
Db     154233 TATATATATATATATATATATATATA 154206


RESULT 7
AL358074
LOCUS       AL358074               109081 bp    DNA     linear   PRI 21-NOV-2001
DEFINITION  Human DNA sequence from clone RP11-423C15 on chromosome 9, complete
            sequence.
ACCESSION   AL358074
VERSION     AL358074.22  GI:17048358
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 109081)
  AUTHORS   Hammond,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Nov 22, 2001 this sequence version replaced gi:17043779.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
```

http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr9
RP11-423C15 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
IMPORTANT: This sequence is not the entire insert of clone RP11-423C15 It may be shorter because we sequence overlapping sections only once, except for a short overlap.
The true right end of clone RP11-423C15 is at 109081 in this sequence. The true left end of clone RP11-770B5 is at 78953 in this sequence. The true right end of clone RP11-12A16 is at 2000 in this sequence.

```
FEATURES             Location/Qualifiers
     source          1. .109081
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /clone="RP11-423C15"
                     /clone_lib="RPCI-11.2"
     misc_feature    complement(65197. .65223)
                     /note="Single clone region. Sequence from reads from a
                     short insert library derived from a single pUC clone.
                     Restriction digest data confirm the assembly."
     misc_feature    65533. .65757
                     /note="Single clone region. Sequence from reads from a
                     short insert library derived from a single pUC clone.
                     Restriction digest data confirm the assembly."
     misc_feature    98714. .98750
                     /note="Single clone region. Sequence from reads from a
                     short insert library derived from a single pUC clone.
                     Restriction digest data confirm the assembly."
     misc_feature    98834. .98842
                     /note="Single clone region. Sequence from reads from a
                     short insert library derived from a single pUC clone.
                     Restriction digest data confirm the assembly."
     misc_feature    99111. .99161
                     /note="Single clone region. Sequence from reads from a
                     short insert library derived from a single pUC clone.
                     Restriction digest data confirm the assembly"
     misc_feature    99163. .99191
                     /note="Single clone region. Sequence from reads from a
                     short insert library derived from a single pUC clone.
                     Restriction digest data confirm the assembly"
ORIGIN
```

Query Match 8.7%; Score 402.2; DB 9; Length 109081;
Best Local Similarity 78.2%; Pred. No. 6.7e-70;
Matches 564; Conservative 0; Mismatches 138; Indels 19; Gaps 6;

```
Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC--AAAAT 58
              ||||  ||||||||||  ||||||||||||||||| |||||||||||||||  ||||
Db      11620 AGTTTGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAAAAAC 11679

Qy         59 TAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGA 118
              ||| || || |||||  || | ||||||||||||||||||  ||  || ||||||||||
Db      11680 TAGCCAAGCGTGGTGGCACGCACCTGTAATCCCAGCTACTCGGGCAGCTGAGGCAGGAGA 11739

Qy        119 ATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCC 177
              ||| ||||||||  |  | ||  ||||||||| | ||| ||||||||||||||||||||
Db      11740 ATCGCTTGAACC--CAAGAGGCAGAGGTTGCAGTGAGCCAAGATCATGCCATTGCACTCC 11797

Qy        178 AGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATAT 237
              ||| || | ||||||||||||||  |  |||| ||||||||||         ||||||||
Db      11798 AGCTCAAGTAACAAGAGCGAAATTCCGCCTCAAAAAAAAAAAA---------ATATATAT 11848

Qy        238 ATATATATATATATACACACACACACATATATATATACACATATATATACACATATAT 297
              ||||||||| | | |||||||||| |||  |||||||| ||||| ||||| ||| |||||
Db      11849 ATATATATACACACACACACACATACATGTATATATACACACACATATACACATATATAC 11908

Qy        298 ACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACAC 357
              ||| ||||| ||| ||| ||||| |||||||||||||| | |   ||| ||||| ||| ||
Db      11909 ACACATATATACACATACACACACATATATACACATATACACACACATATATACATATAC 11968

Qy        358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417
              | |   ||| | | ||| ||||| ||||| |||| |||| ||||| | ||||||| ||||
Db      11969 ACACACATATATACATACACACACATATATACACGTATACACACACACATACACATATAT 12028

Qy        418 ATACACATATATACACATATATACACATATACACATATATACACATATATACATATATAC 477
              | |||||||||||||||||| ||||| ||| |||  |||||| | | | ||||||| ||
Db      12029 ACACACATATATACACATATACACACACATATACACGTATACATACACACACATATACAC 12088

Qy        478 ACATATATATA-ATATACACACATATATATACACATATATACACACATATATACACATAT 536
                |||||||  |  |||| | |||  |||||| ||| ||||| |||| ||||||||||||||
Db      12089 GTATATATACACGTATATATACACGTATATATACACATATATACACGTATATACACATAT 12148

Qy        537 ATACACATATATATACACATATATACACATATATACATATATACAC----ATATATATAC 592
              |||||| |||||||||||||||||||| |||||||| ||||| ||    ||||||||||
Db      12149 ATACACGTATATATACACATATATACACGTATATACACATATATACACGTATATATATAC 12208

Qy        593 ATATATACACATATATACATATATACACATATATACATATATACACACATATATACACAT 652
              | | ||| | ||| | ||| ||||||||| | | |||||||| ||||| ||| |||||
Db      12209 ACACATATATATACACACACATATACACACACACACATATATATACACACATACACACAC 12268

Qy        653 ACATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTAT 712
              ||| | ||||||||| | ||||| |||||  |   |    |||   |   | |  ||||
Db      12269 ACACACACACACACACACATATATATATAGTTGGCCGGGCATAGGGGCTCACACCTGTAA 12328

Qy        713 A 713
              |
Db      12329 A 12329
```

RESULT 8
AC009058
LOCUS AC009058 164844 bp DNA linear PRI 13-MAR-2003
DEFINITION Homo sapiens chromosome 16 clone RP11-292B23, complete sequence.
ACCESSION AC009058
VERSION AC009058.10 GI:28933531
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 164844)
AUTHORS DOE Joint Genome Institute, Stanford Human Genome Center and Los Alamos National Laboratory.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 164844)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 164844)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (04-DEC-2001) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 4 (bases 1 to 164844)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (09-DEC-2001) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 5 (bases 1 to 164844)
AUTHORS DOE Joint Genome Institute, Stanford Human Genome Center and Los Alamos National Laboratory.
TITLE Direct Submission
JOURNAL Submitted (13-MAR-2003) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT On Mar 13, 2003 this sequence version replaced gi:17432456.
Draft Sequence Produced by DOE Joint Genome Institute

```
           www.jgi.doe.gov
           Finishing Completed at Stanford Human Genome Center and Los Alamos
           National Laboratory
           www-shgc.stanford.edu
           Quality: Phrap Quality >=40 100% of Sequence;
           Estimated Total Number of Errors is 0.
FEATURES             Location/Qualifiers
     source          1. .164844
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone="RP11-292B23"
ORIGIN

  Query Match             8.6%;  Score 396.8;  DB 9;  Length 164844;
  Best Local Similarity   75.4%;  Pred. No. 7.3e-69;
  Matches  559;  Conservative    0;  Mismatches  173;  Indels    9;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAATTA 60
              ||||| | |||| ||||  |||||| ||||||||||||||||||||||||    |||||||
Db     140343 AGTTCGAAACCACCCTGGCCAACATAGTGAAACCCCATCTCTACTAAAATACAAAATTA 140402

Qy         61 GACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAAT 120
              | ||||| |||||  | ||||||||||||||||||||  |||||| |||||||||| ||
Db     140403 GCCAGGCGTGGTGGCAGACGCCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGAGGAT 140462

Qy        121 CACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCCAG 179
              ||||||||  | || ||  |||||||| | ||| | ||||  ||||| || |||||||
Db     140463 TGCTTGAACC--CAGGAGGCAGAGGTTGCAGTGAGCCCAGATTGTGCCACTGGACTCCAG 140520

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  ||  |||       | |         | | | | | |        |||||||||
Db     140521 CCTGGGAGACAGAGTAAGACTCCGTCTCAAATACATATATACATACACACATATATATAC 140580

Qy        240 ATATATATATATACACACACACACACATATATATATACACATATATATACAC----ATAT 295
              | ||||| ||| ||||| | | | ||| ||||| ||||||| ||| ||||||    ||||
Db     140581 ACATATACATACACACATATATATACACATATACATACACACATACATACACACATATAT 140640

Qy        296 ATACATATATACACATATATACACATATATATACACATACATATGTACACATATATAT-A 354
              | ||| ||||||| |||||||||||||||||||||| ||||| || || ||| |||||| |
Db     140641 ACACACATATACATATATATACACATATATATACATATACACATATATACACATATATAA 140700

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              | | |   ||| | ||| ||| | | | ||| |||||||||||||| ||||| ||||| |
Db     140701 TATACACACATATATATACATATATACACATATACACATATATACATATATACACACATA 140760

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              |||| |||||||||||||||||||||| ||| |||||||||||| ||| | |||||||
Db     140761 TATACACACATATATATACACATATATACATATACACACATATATACATATACACACATATA 140820

Qy        475 TACACATATATATAATATACACACATATATATACACATATATACACACATATATACACAT 534
              ||||||| | |  | ||||||||| | ||| ||||||| | ||||| | | | ||||||
Db     140821 CACACATACACA-CACATACACACACACATACACACATACACACACACACACACACAT 140879

Qy        535 ATATACACATATATATACACATATATACACATATATACATATATACACATATATATACAT 594
              | | | ||| | ||| ||||| ||||||||| ||| |  || | | ||| | |||||||
Db     140880 ACACATACACACATACACACACATATACACACATACATGTACACATACACACATATACAC 140939

Qy        595 ATATACACATATATACATATATACACATATATACATATATACACACATATATACACATAC 654
              ||||||||| |||||||||||||||| |||||||||||||||||||||||| |||||
Db     140940 ATATACACACATATACATATATACACACATATACATATATACACACATATATATACATAT 140999

Qy        655 ATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATAT 714
              ||| ||| | | | | ||||||| |||| |||| ||||| |||  || | | |  |||||
Db     141000 ATACACATATATACATATATACACATATGTACATATATACATATATACACACACATATAT 141059

Qy        715 ATATATGCTCCAGAGTTCATA 735
              | |||   |     | ||||
Db     141060 GTGTATATACACACATACATA 141080


RESULT 9
```

```
AC011593
LOCUS       AC011593              130349 bp    DNA     linear   PRI 07-NOV-2001
DEFINITION  Homo sapiens BAC clone RP11-437E9 from 2, complete sequence.
ACCESSION   AC011593
VERSION     AC011593.8  GI:14190772
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 130349)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
   PUBMED   9847074
REFERENCE   2  (bases 1 to 130349)
  AUTHORS   Pearman,C. and Cotton,M.
  TITLE     The sequence of Homo sapiens BAC clone RP11-437E9
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 130349)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 130349)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-MAY-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 130349)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-NOV-2001) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On May 23, 2001 this sequence version replaced gi:13518261.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0437E09
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. John D.
            McPherson, Department of Genetics, Washington University, St. Louis
            MO.  For additional information about the map position of this
            sequence, see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries.  Genomics 51:1-8.  The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
```

and coworkers at the Roswell Park Cancer Institute
(http://bacpac.med.buffalo.edu)
VECTOR: pBACe3.6
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-289J14, 200 bp overlap; the clone sequenced to the right is AC061984. Actual start of this clone is at base position 150577 of RP11-289J14; actual end is at base position 130349 of RP11-437E9.

FEATURES	Location/Qualifiers
source	1. .130349
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="2"
	/map="2"
	/clone="RP11-437E9"
	/clone_lib="RPCI-11"
repeat_region	1813. .2297
	/rpt_family="MaLR"
repeat_region	2618. .3136
	/rpt_family="MaLR"
repeat_region	3200. .3370
	/rpt_family="MaLR"
repeat_region	3522. .3856
	/rpt_family="Alu"
repeat_region	3866. .4290
	/rpt_family="MaLR"
repeat_region	4325. .4805
	/rpt_family="ERV1"
repeat_region	4910. .5014
	/rpt_family="MER1_type"
repeat_region	5258. .5600
	/rpt_family="L1"
repeat_region	5601. .5772
	/rpt_family="Alu"
repeat_region	5773. .6644
	/rpt_family="L1"
repeat_region	6773. .7356
	/rpt_family="L1"
repeat_region	7389. .7516
	/rpt_family="L1"
repeat_region	7529. .8188
	/rpt_family="L1"
repeat_region	8189. .8490
	/rpt_family="Alu"
repeat_region	8491. .8590
	/rpt_family="L1"
repeat_region	8605. .8943
	/rpt_family="L1"
repeat_region	9017. .9042
	/rpt_family="(CA)n"
repeat_region	9062. .9104
	/rpt_family="MER53"
repeat_region	9105. .9394
	/rpt_family="Alu"
repeat_region	9395. .9508
	/rpt_family="MER53"
repeat_region	9786. .9966
	/rpt_family="C-rich"
repeat_region	9946. .10036
	/rpt_family="(TCC)n"
repeat_region	10080. .10374
	/rpt_family="Alu"
repeat_region	12491. .12521
	/rpt_family="(CA)n"
repeat_region	13138. .13321
	/rpt_family="MIR"
repeat_region	13414. .13442
	/rpt_family="(TAA)n"
repeat_region	13615. .13925
	/rpt_family="Alu"
repeat_region	16183. .16266
	/rpt_family="MIR"
repeat_region	16629. .16997
	/rpt_family="MaLR"
repeat_region	17643. .17821
	/rpt_family="Tc2"
repeat_region	18157. .18200
	/rpt_family="(TG)n"
repeat_region	19591. .19619
	/rpt_family="AT_rich"
repeat_region	21204. .21375
	/rpt_family="MaLR"
repeat_region	21376. .21940
	/rpt_family="ERVL"
repeat_region	21962. .22901
	/rpt_family="L1"
repeat_region	22909. .23443
	/rpt_family="L1"
repeat_region	23492. .23987
	/rpt_family="L1"
repeat_region	23988. .24016
	/rpt_family="(TTTTG)n"
repeat_region	24017. .24634
	/rpt_family="L1"
repeat_region	24636. .25726
	/rpt_family="L1"
repeat_region	25727. .26040
	/rpt_family="Alu"
repeat_region	26041. .27537
	/rpt_family="L1"
repeat_region	27607. .28123
	/rpt_family="ERVL"
repeat_region	28282. .28808
	/rpt_family="ERVL"
repeat_region	28809. .29192
	/rpt_family="MaLR"
repeat_region	29193. .29577
	/rpt_family="ERVL"
repeat_region	29628. .29895
	/rpt_family="Alu"
repeat_region	29960. .30264
	/rpt_family="ERVL"
repeat_region	30265. .30573
	/rpt_family="Alu"
repeat_region	30574. .30753
	/rpt_family="ERVL"
repeat_region	30774. .32056
	/rpt_family="ERVL"
repeat_region	32100. .32242
	/rpt_family="ERVL"
repeat_region	32477. .32750
	/rpt_family="MaLR"
repeat_region	32751. .32847
	/rpt_family="L1"
repeat_region	32848. .33164
	/rpt_family="Alu"
repeat_region	33165. .33302

Query Match 8.6%; Score 395.2; DB 9; Length 130349;
Best Local Similarity 67.8%; Pred. No. 1.6e-68;
Matches 645; Conservative 0; Mismatches 293; Indels 14; Gaps 6;

```
Qy          4 TCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC--AAAATTAG 61
              ||||||||| ||||   |||||||||||| ||||||||||||||||||||  ||||||||
Db      39934 TCAAGACCATCCTGGCTAACATGGTGAATCCCCATCTCTACTAAAAATACAAAAAATTAG 39993

Qy         62 ACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAATC 121
               ||||| |||||      ||||||| |||||||||||  |||||| ||||||||||||||
Db      39994 CCAGGCGTGGTGGCGGGTGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATG 40053

Qy        122 ACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAGCC 181
              | ||||||  | || |||||||| ||||   |||| |||||   ||||||||||||||||
Db      40054 GCGTGAACC--CGGGAGGTGGAGCTTGCAATGAGCCAGATCGCACCATTGCACTCCAGCC 40111

Qy        182 CAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTT---TATATATATA 238
```

```
Db      40112 TGGACAAC-GGAGCAAGACTCCATCTAAAAAATATATATAAATATATACATAAATATATA 40170

Qy        239 TATATATATATATACACACACACACATATATATATACACATATATATACACATATATA 298
Db      40171 TATATAAATATAGACATAAATATATATATATAAATATAGACATAAATATATATGCATAAA 40230

Qy        299 CATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACA 358
Db      40231 TATATATATGCATAAATATATATAAAAATATATATAAATATATACATAAATATATATAAA 40290

Qy        359 TATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATA 418
Db      40291 TATATACAAAAATATATATAAATATATAAAAAAATATATAAATATATATACACATATATA 40350

Qy        419 TACACATATATACACATATATACACATATACACATATATACACATATATACATATATACA 478
Db      40351 AATATATATACATACATATATAAACATATATACATAAATATATATGTATAAATATATATA 40410

Qy        479 CATATATATA---ATATACACACATATATATACACATATATACACACATATATACACATA 535
Db      40411 CACATAAATATATGTATGAATATATATACATAAATATATATGTATAAATATATATACATA 40470

Qy        536 TATACACATATATATACACATATATACACATATATACATATATACACATATATATACATA 595
Db      40471 AATATATAAAGATATATACATAAATATATATAAATATATATACATAAATATATATAAATA 40530

Qy        596 TATACACATATATACATATATACACATATATACATATATACACACATATATACACATACA 655
Db      40531 TATATAAATAGATATATAAATATATATATAAATATATAAATATATATATAAATATATAAA 40590

Qy        656 TATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATA 715
Db      40591 TATATAAAAATAGATATATAAATATATATATAAATATATAAATATATATATAAATATATA 40650

Qy        716 TATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGA 775
Db      40651 TAAATATATAAATATATATATAAATATATATAAATATATAAATATATATAAATATATAAA 40710

Qy        776 GAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGA 835
Db      40711 TATATATATAAATATATATAAATATATAAATATATATAAATATATATAAATATATAAATA 40770

Qy        836 AATACATAAATGTGTAATAAAACCTATCTGTA---AGGTAGTTAAGAAGGTAACACTATA 892
Db      40771 TATATATAAATATATATAAATATATAAATATATATATAAATATATAAATATATAAATATA 40830

Qy        893 TATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAA 944
Db      40831 TATATAAATATATAAATATATATAAATATATATAAATATATAAATATATATA 40882


RESULT 10
AC097269/c
LOCUS       AC097269               196413 bp    DNA     linear   HTG 20-OCT-2001
DEFINITION  Pan troglodytes clone RP43-11J15, WORKING DRAFT SEQUENCE, 18
            unordered pieces.
ACCESSION   AC097269
VERSION     AC097269.1  GI:16117532
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 196413)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 196413)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-OCT-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: ZUAW
                Center clone name: RP43-11J15
            -------------- Summary Statistics
                Sequencing vector: Plasmid; M77789
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 190674 bases at least Q40
                Consensus quality: 194102 bases at least Q30
                Consensus quality: 196834 bases at least Q20
                Estimated insert size: 194298; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 8.1x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 18 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    17120: contig of 17120 bp in length
            *    17121    17220: gap of unknown length
            *    17221    36306: contig of 19086 bp in length
            *    36307    36406: gap of unknown length
            *    36407    58958: contig of 22552 bp in length
```

```
            *    58959    59058: gap of unknown length
            *    59059    82373: contig of 23315 bp in length
            *    82374    82473: gap of unknown length
            *    82474    96867: contig of 14394 bp in length
            *    96868    96967: gap of unknown length
            *    96968   109418: contig of 12451 bp in length
            *   109419   109518: gap of unknown length
            *   109519   120755: contig of 11237 bp in length
            *   120756   120855: gap of unknown length
            *   120856   134377: contig of 13522 bp in length
            *   134378   134477: gap of unknown length
            *   134478   144432: contig of 9955 bp in length
            *   144433   144532: gap of unknown length
            *   144533   155013: contig of 10481 bp in length
            *   155014   155113: gap of unknown length
            *   155114   161933: contig of 6820 bp in length
            *   161934   162033: gap of unknown length
            *   162034   168622: contig of 6589 bp in length
            *   168623   168722: gap of unknown length
            *   168723   175724: contig of 7002 bp in length
            *   175725   175824: gap of unknown length
            *   175825   180213: contig of 4389 bp in length
            *   180214   180313: gap of unknown length
            *   180314   186640: contig of 6327 bp in length
            *   186641   186740: gap of unknown length
            *   186741   190642: contig of 3902 bp in length
            *   190643   190742: gap of unknown length
            *   190743   194248: contig of 3506 bp in length
            *   194249   194348: gap of unknown length
            *   194349   196413: contig of 2065 bp in length.
FEATURES             Location/Qualifiers
     source          1. .196413
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="RP43-11J15"
ORIGIN

  Query Match             8.5%;  Score 394.8;  DB 2;  Length 196413;
  Best Local Similarity   75.8%;  Pred. No. 1.8e-68;
  Matches  547;  Conservative    0;  Mismatches  157;  Indels   18;  Gaps    4;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC------- 53
              |||||||||||||||||| |||||||||| |||||||||||||| |||||| |
Db     196321 AGTTCAAGACCAGCCTGACCAACATGGTGGAACCCCATCTCTAGTAAAAAAAAAAAAAAA 196262

Qy         54 -------AAAATTAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGC 106
                     |||||||| |||| ||||| || ||||| ||||| | |||||| ||||||
Db     196261 AAAAAAAAAAATTAGCCAGGTGTGGTGGGGCATGCCTGAAATCCTAACTACTTGGGAGGC 196202

Qy        107 CGAGGCAGGAGAATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGC-AAGATCATG 165
              ||||||||| |||  |||||| |   |  | ||||||||||||    ||||  | |||| |
Db     196201 TGAGGCAGGAAAATTGCTTGAAGC--CAAGAGGTGGAGGTTGCAATGAGCTGATATCACG 196144

Qy        166 CCATTGCACTCCAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAGGCA 225
              || |||||||||||||  |||||||| |  ||||   ||||||| |||||| |   | |
Db     196143 CCGTTGCACTCCAGCCTGGGCAACAAAAATGAAACTCCATCTCAAAAAAAAGAGGGGCCG 196084

Qy        226 TTTTATATATATATATATATATATACACACACACACACATATATATATACACATATAT 285
                 |  |  ||  |  || | || || | | | | | | ||||||||||| | ||||||
Db     196083 GGAGAGGTGGATCAAGCTAGAAATCTATATATATATATATATATATATATATATATAT 196024

Qy        286 ATACACATATATACATATATACACATATATACACATATATATACACATACATATGTACAC 345
              ||| | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db     196023 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 195964

Qy        346 ATATATATACACATATGTATACACATATATACACATATATACACACATATATACACATAT 405
              ||||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | ||||
Db     195963 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 195904

Qy        406 ATACACACATATATACACATATATACACATATATACACATATACACATATATACACATAT 465
              ||| | | ||||||| | |||||||| | ||||||| | ||||| | ||||||| | ||||
Db     195903 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 195844

Qy        466 ATACATATATACACATATATATAATATACACACATATATATACACATATATACACACATA 525
              ||| ||||||| | |||||||| ||||| | | ||||||||| | ||||||| | | |||
Db     195843 ATATATATATATATATATATAT-ATATATATATATATATATATATATATATATATATA 195785

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | |||||||| | ||||||| | ||||||| ||||||| | |||
Db     195784 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 195725

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db     195724 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 195665

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | ||||||||   |||  |
Db     195664 TATATATATATATATATATATATATATATATATATATATATATATATATATATCTATATCCA 195605

Qy        706 TA 707
              |
Db     195604 CA 195603


RESULT 11
AC015468
LOCUS       AC015468              189662 bp    DNA     linear   PRI 01-MAY-2001
DEFINITION  Homo sapiens chromosome 8, clone RP11-369E15, complete sequence.
ACCESSION   AC015468
VERSION     AC015468.5  GI:13899433
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 189662)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 8, clone RP11-369E15
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 189662)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 189662)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M.,
            Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K.,
            Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W.,
            Iliev,I., Johnson,R., Jones,C., Karatas,A., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Sougnez,C., Spencer,B., Stange-Thomann,N.,
```

```
            Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
            Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAY-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 1, 2001 this sequence version replaced gi:12313808.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L2466
                Center clone name: 369_E_15
            --------------.
FEATURES             Location/Qualifiers
     source          1. .189662
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="RP11-369E15"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   1. .100
                     /rpt_family="MLT1A1"
     repeat_region   1594. .1741
                     /rpt_family="GA-rich"
     repeat_region   complement(2782. .2957)
                     /rpt_family="MIR"
     repeat_region   3311. .3471
                     /rpt_family="MIR"
     repeat_region   3964. .4067
                     /rpt_family="MIR"
     repeat_region   complement(4528. .4771)
                     /rpt_family="L2"
     repeat_region   4776. .4871
                     /rpt_family="MER4A"
     repeat_region   4869. .5223
                     /rpt_family="MER4A"
     repeat_region   complement(5255. .5765)
                     /rpt_family="L2"
     repeat_region   complement(5805. .6003)
                     /rpt_family="THE1A"
     repeat_region   6004. .6035
                     /rpt_family="(TTA)n"
     repeat_region   complement(6151. .6319)
                     /rpt_family="SVA"
     repeat_region   complement(6320. .6468)
                     /rpt_family="THE1A"
     repeat_region   8722. .8751
                     /rpt_family="(TTTTA)n"
     repeat_region   9155. .9553
                     /rpt_family="MSTB"
     repeat_region   9599. .9982
                     /rpt_family="THE1C"
     repeat_region   complement(10113. .10320)
                     /rpt_family="AluY"
     repeat_region   10822. .10970
                     /rpt_family="MIR3"
     unsure          11210. .11215
                     /note="<30 qual SNGL region"
     repeat_region   11794. .12080
                     /rpt_family="AluSx"
     repeat_region   12974. .13010
                     /rpt_family="(TG)n"
     repeat_region   complement(13091. .13304)
                     /rpt_family="MER106A"
     repeat_region   13484. .13858
                     /rpt_family="MLT1C"
     repeat_region   14025. .14260
                     /rpt_family="L2"
     repeat_region   14584. .14679
                     /rpt_family="MIR"
     repeat_region   14722. .14825
                     /rpt_family="L2"
     repeat_region   15167. .15332
                     /rpt_family="MER33"
     repeat_region   15436. .15788
                     /rpt_family="L2"
     repeat_region   complement(16078. .16165)
                     /rpt_family="AluJ/FRAM"
     repeat_region   16532. .17094
                     /rpt_family="MLT1F1"
     repeat_region   17653. .17674
                     /rpt_family="(TAA)n"
     repeat_region   complement(17675. .17892)
                     /rpt_family="MER7C"
     repeat_region   complement(17897. .18650)
                     /rpt_family="LTR17"
     repeat_region   complement(18651. .18996)
                     /rpt_family="MER7C"
     repeat_region   18999. .19103
                     /rpt_family="(TAA)n"
     repeat_region   complement(19203. .19283)
                     /rpt_family="MIR"
     repeat_region   20464. .20523
                     /rpt_family="A-rich"
     repeat_region   complement(21279. .21902)
                     /rpt_family="MER41B"
     repeat_region   21922. .22092
                     /rpt_family="MIR"
     repeat_region   22249. .22373
                     /rpt_family="MIR"
     repeat_region   23956. .24058
                     /rpt_family="L2"
     repeat_region   complement(24069. .24198)
                     /rpt_family="MIR"
     repeat_region   25149. .25213
                     /rpt_family="L2"
     repeat_region   complement(25783. .26130)
                     /rpt_family="MLT1I"
     repeat_region   complement(26174. .26539)
                     /rpt_family="MLT1A1"
     repeat_region   complement(28187. .28528)
                     /rpt_family="AluY"
     repeat_region   complement(28792. .29263)
                     /rpt_family="Charlie8"
     repeat_region   complement(30652. .30730)
                     /rpt_family="MER53"
     repeat_region   complement(30986. .31117)
                     /rpt_family="L1ME1"
     repeat_region   31113. .31556
                     /rpt_family="L1ME2"
     repeat_region   31590. .32174
                     /rpt_family="L1ME2"
     repeat_region   32162. .32588
                     /rpt_family="L1ME3"
     repeat_region   33242. .33490
                     /rpt_family="MIR"
     repeat_region   complement(34416. .34803)
                     /rpt_family="MSTB"
     repeat_region   complement(35044. .35350)
                     /rpt_family="AluSg"
     repeat_region   36407. .36695
                     /rpt_family="AluSx"
     repeat_region   complement(37670. .37790)
                     /rpt_family="FLAM_A"
     repeat_region   complement(38065. .38132)
                     /rpt_family="L2"
     repeat_region   complement(38504. .38590)
```

```
                  /rpt_family="MIR"
     repeat_region   39044. .39228
                  /rpt_family="MIR3"
     repeat_region   complement(40317. .40473)
                  /rpt_family="MIR"

  Query Match             8.5%;  Score 390.8;  DB 9;  Length 189662;
  Best Local Similarity   65.9%;  Pred. No. 1.1e-67;
  Matches  643;  Conservative    0;  Mismatches  327;  Indels    6;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAAT-ACAAAATT 59
              ||||| |||||  ||||  ||||||| ||||||||||||||||||||||| | ||||||
Db      65793 AGTTCGAGACCCACCTGGCCAACATGATGAAACCCCATCTCTACTAAAAATGAAAAAATT 65852

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGA-GGCCGAGGCAGGAGA 118
              ||   ||  |||||  |   || | ||||||||||||||  ||| ||| ||||||| |||
Db      65853 AGCTGGGTGTGGTGGCAGGTGCATCTAATCCCAGCTACTCTGGAGGGCTGAGGCAGAAGA 65912

Qy        119 ATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCC 177
              ||  ||||||||  | || |||||| ||||| | ||| | |||||  ||||||||||||
Db      65913 ATTGCTTGAACC--CAGGAGGTGGACGTTGCAGTGAGCCGAGATCGCGCCATTGCACTCC 65970

Qy        178 AGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATAT 237
              ||||  |||||||||||| |||     ||||| |||||||||         |||||||||
Db      65971 AGCCTGGGCAACAAGAGCAAAACTCTGTCTCAAAAAAAAAAATTTTATAAATATATATAT 66030

Qy        238 ATATATATATATATACACACACACACATATATATATACACATATATATACACATATAT 297
              | |||||||||| | | | | | | | ||||||| ||| | ||| ||||| | | ||||
Db      66031 AAATATATATATAAATATATATAAATATATATATAAATATATATAAATATATATAAATAT 66090

Qy        298 ACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACAC 357
              | ||||| | | ||||| | | ||||| ||| | ||| |||| || | |||||||| |
Db      66091 ATATATAAATATATATAAATATATATAAATATATATAAATATATATAAATATATATATAA 66150

Qy        358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417
              |||| |||| | ||||| | | ||||||| | | ||||||| | ||||||| | | ||||
Db      66151 ATATATATAAATATATAAAAATATATATATAAATATATATAAATATATATAAATATATAT 66210

Qy        418 ATACACATATATACACATATATACACATATACACATATATACACATATATACATATATAC 477
              | | | ||||||| | ||||||| | ||||| | ||||||| | | | ||| |||||||
Db      66211 AAATATATATATAAATATATATAAATATATATAAATATATATATAAAAATATATATATAA 66270

Qy        478 ACATATATATAATATACACACATATATATACACATATATACACACATATATACACATATA 537
              | ||||||| |||||| | | | ||||||| | ||||||| | | ||||||| | || ||
Db      66271 ATATATATA-AATATATATATAAATATATATAAATATATATATAAATATATAAATATGTA 66329

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATACATATA 597
              || | ||||||| | |  |||||| | || |||| | |||||  | |||| |||| | |||
Db      66330 TATAAATATATAAATATGTATATAAATATGTATATAAATATATAAATATGTATATAAATA 66389

Qy        598 TACACATATATACATATATACACATATATACATATATACACACATATATACACATACATA 657
              || | ||| ||| ||| |||    ||||| | ||||||| | |  ||||||| | ||   |||
Db      66390 TATATATAAATATATAAATATGTATATAAATATATATATAAATATATATAAATATGTATA 66449

Qy        658 TACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATA 717
              || | | | ||| ||||  |||||  ||| | |||||||  |  ||||||  |  ||||||||
Db      66450 TAAATATATATAAATATGTATATAAATATATATATATAAATATATATAAAAATATATATA 66509

Qy        718 TATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGA 777
              ||        | |  |||         |  | | ||    |   ||| |   ||| | |
Db      66510 TAAATATATAAATATATATATAAATATATATAAATATATATAAATATATATATAAATATA 66569

Qy        778 GTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAA 837
                | |    | ||     | |  | || |         |    |||||         | |
Db      66570 TAAATATATATATATAAATATATATATAAATATATATAAAAATATAAATATATATAAATA 66629

Qy        838 TACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATATAT 897
              || ||| |  | ||  | |  |       |  | ||  |   |    |||| ||||
Db      66630 TATATATAAATATATATATAAAAAATATATAAAAATATATATATAAATATATATAAATAT 66689

Qy        898 ATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTATAGAAGAA 957
              |||||   | |       || |  || || |  |   ||  | | |||| || |      ||
Db      66690 ATATAAATATATATAAATATATATATAGTAGTGAAGAATCACTAAAATATTTTTATGAAA 66749

Qy        958 TGGCTAAGATGCCAAA 973
                ||  ||  | | |
Db      66750 AAAGTATGAATCTAGA 66765


RESULT 12
AC073310/c
LOCUS       AC073310               125685 bp    DNA     linear   PRI 08-OCT-2003
DEFINITION  Homo sapiens BAC clone RP11-49G5 from 7, complete sequence.
ACCESSION   AC073310
VERSION     AC073310.7  GI:13176604
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 125685)
  AUTHORS   Sulston,J.E. and Wilson,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
   PUBMED   9847074
REFERENCE   2  (bases 1 to 125685)
  AUTHORS   Haakenson,W., Hannah,C. and Kang,K.
  TITLE     The sequence of Homo sapiens BAC clone RP11-49G5
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 125685)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUN-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 125685)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 125685)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2001) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   6  (bases 1 to 125685)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2001) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   7  (bases 1 to 125685)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-APR-2003) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   8  (bases 1 to 125685)
  AUTHORS   Wilson,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-OCT-2003) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Mar 1, 2001 this sequence version replaced gi:11024926.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0049G05
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.
```

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see http://www.nhgri.nih.gov/DIR/GTB/CHR7 , send mailto:egreen@nhgri.nih.gov , or see http://genome.wustl.edu

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at http://www.chori.org
VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-409M7, 200 bp overlap the clone sequenced to the right is RP11-374N8. Actual start of this clone is at base position 14572 of RP11-409M7 actual end is at base position 125685 of RP11-49G5.

FEATURES Location/Qualifiers

```
     source          1. .125685
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7"
                     /clone="RP11-49G5"
                     /clone_lib="RPCI-11"
     repeat_region   1013. .1151
                     /rpt_family="CR1"
     repeat_region   2238. .2394
                     /rpt_family="MIR"
     repeat_region   2548. .2668
                     /rpt_family="L2"
     repeat_region   3206. .5231
                     /rpt_family="L1"
     repeat_region   5694. .5715
                     /rpt_family="AT_rich"
     repeat_region   7142. .7353
                     /rpt_family="MER1_type"
     repeat_region   7490. .7666
                     /rpt_family="ERVL"
     repeat_region   7671. .7771
                     /rpt_family="L1"
     repeat_region   7951. .7985
                     /rpt_family="AT_rich"
     repeat_region   9572. .9809
                     /rpt_family="L1"
     repeat_region   9811. .9831
                     /rpt_family="AT_rich"
     repeat_region   9895. .10017
                     /rpt_family="MaLR"
     repeat_region   10320. .10375
                     /rpt_family="MIR"
     repeat_region   10395. .10732
                     /rpt_family="L1"
     repeat_region   10911. .11118
                     /rpt_family="(TA)n"
     unsure          10914. .11684
                     /note="unresolved simple sequence repeat"
     repeat_region   11130. .11196
                     /rpt_family="L1"
     repeat_region   11217. .11663
                     /rpt_family="(TA)n"
     repeat_region   11667. .11918
                     /rpt_family="Alu"
     repeat_region   11919. .11970
                     /rpt_family="L1"
     repeat_region   12356. .12508
                     /rpt_family="L1"
     repeat_region   12646. .12813
                     /rpt_family="L1"
     repeat_region   12863. .13111
                     /rpt_family="L1"
     repeat_region   13280. .13592
                     /rpt_family="L1"
     repeat_region   13768. .14032
                     /rpt_family="L1"
     repeat_region   14048. .14489
                     /rpt_family="L1"
     repeat_region   15108. .15378
                     /rpt_family="Alu"
     repeat_region   15563. .15584
                     /rpt_family="AT_rich"
     repeat_region   17280. .17442
                     /rpt_family="MIR"
     repeat_region   18313. .18467
                     /rpt_family="L1"
     repeat_region   18468. .18676
                     /rpt_family="L1"
     repeat_region   18677. .18701
                     /rpt_family="AT_rich"
     repeat_region   18797. .19155
                     /rpt_family="MaLR"
     repeat_region   19594. .19911
                     /rpt_family="Alu"
     repeat_region   21451. .21639
                     /rpt_family="MER53"
     repeat_region   22809. .22986
                     /rpt_family="L2"
     repeat_region   23094. .23371
                     /rpt_family="L1"
     repeat_region   23372. .23678
                     /rpt_family="Alu"
     repeat_region   23679. .23806
                     /rpt_family="L1"
     repeat_region   23820. .23999
                     /rpt_family="L1"
     repeat_region   24445. .24470
                     /rpt_family="AT_rich"
     repeat_region   24596. .24649
                     /rpt_family="(CA)n"
     repeat_region   26292. .26765
                     /rpt_family="L1"
     repeat_region   26766. .27123
                     /rpt_family="MaLR"
     repeat_region   27124. .27202
                     /rpt_family="L1"
     repeat_region   27203. .27556
                     /rpt_family="MaLR"
```

```
  Query Match             8.4%;  Score 387.4;  DB 9;  Length 125685;
  Best Local Similarity   67.7%;  Pred. No. 5.5e-67;
  Matches  647;  Conservative    0;  Mismatches  296;  Indels   12;  Gaps    7;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              |||||||||||||||   |||||||||||||||||| | | ||||||||||| ||||||
Db      11886 AGTTCAAGACCAGCCTAGCCAACATGGTGAAACCCCGTTTTTACTAAAAATACAAAAATT 11827

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              || ||||| ||||| |    ||||||||||||||| ||   ||||||  ||||||||||
```

```
Db      11826 AGCCAGGCGTGGTGGT---TGCCTGTAATCCCAGCCACACGGGAGGCTAAGGCAGGAGAA 11770

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGC-AAGATCATGCCATTGCACTCCA 178

Db      11769 TTGCTTGAATCTG--GGAGACGGAGGTTGCAGTGAGCTGAGATCACGCCACTGCACTCCA 11712

Qy        179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATA 238

Db      11711 GCCTGGGTGAC-AGAGTGAGACTCCATCTCAAAAAAAAAAAAAAAAAGTCATATATATATA 11653

Qy        239 TATATATATATATACACACACACACATATATATATACACATATATATACACATATATA 298

Db      11652 TATATATATATATATATATATATATATAAAAATATATATATAAATATATATATAAATATA 11593

Qy        299 CATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACA 358

Db      11592 TATATATAAATATATTTATAAATATATATATAAATATATTTATAAATATATATAAATATA 11533

Qy        359 TATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATA 418

Db      11532 TATATATAAATATATATATAAATATATGTATAAAAATATATATATAAATATAAATATATA 11473

Qy        419 TACACATATATACACATATATACACATATACACATATATACACATATATACATATATACA 478

Db      11472 TATAAATATATATATAAAAATATATATATAAATATAAATATATATATAAATATATATATA 11413

Qy        479 CATATATATA---ATATACACACATATATATACACATATATACACACATATATACACATA 535

Db      11412 AATATATATATATATATAAATATATATAAATATATATATATATATATATATAAATATCTA 11353

Qy        536 TATACACATATATATACACATATATACACATATATACATATATACACATATATATACATA 595

Db      11352 TATAAACATATATATATAAATATATATATAAACATATATATATAAATATATATATAAACA 11293

Qy        596 TATACACATATATACATATATACACATATATACATATATACACACATATATACACATACA 655

Db      11292 TATATATATAAATATATATATAAACATATATATATAAATATATATATATATATATATA 11233

Qy        656 TATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATA 715

Db      11232 TATATATATATATATAAACATATATATATTCAAATATATATATAAACATATATATATAAA 11173

Qy        716 TATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGA 775

Db      11172 TATATATATAAACATATATATATAAATATATATATAAACATATATAATATAACATATATA 11113

Qy        776 GAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGA 835

Db      11112 TATATATATATATATATAAATATATATATAAATATATATATAAATATATATATATAAATA 11053

Qy        836 AATACATAAATGTGTAATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTAACACTATATA 894

Db      11052 TATATATAAATATATATATATATATATATATATAAATATATATATATAAATATATATATA 10993

Qy        895 TATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTA 949

Db      10992 TAAATATATATATATATATATATATATAAATATATATATAAATATATATATATAA 10938


RESULT 13
AL133477
LOCUS       AL133477               139118 bp    DNA     linear   PRI 02-FEB-2001
DEFINITION  Human DNA sequence from clone RP11-172F4 on chromosome 9q22.2-31.1,
            complete sequence.
ACCESSION   AL133477
VERSION     AL133477.16  GI:12666198
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 139118)
  AUTHORS   Lloyd,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Feb 5, 2001 this sequence version replaced gi:12584335.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence has been finished according to sequence map criteria
            as follows.  An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated repeat sequence elements.  Where the sequence is
            ambiguous, there is an annotation using the 'unsure' feature key.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr9
            RP11-172F4 is from the library RPCI-11.1 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-172F4 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true left end of clone RP11-172F4 is at 1 in this sequence. The
            true left end of clone RP11-240L7 is at 139019 in this sequence.
            The true right end of clone RP11-392G7 is at 43511 in this
            sequence.
FEATURES             Location/Qualifiers
     source          1. .139118
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /map="q22.2-31.1"
                     /clone="RP11-172F4"
                     /clone_lib="RPCI-11.1"
     repeat_region   45. .78
                     /note="L1MB2 repeat: matches 6125. .6158 of consensus"
     repeat_region   195. .361
                     /note="MER20 repeat: matches 39. .218 of consensus"
     repeat_region   488. .688
                     /note="MLT1B repeat: matches 205. .390 of consensus"
     repeat_region   689. .986
                     /note="AluY repeat: matches 1. .297 of consensus"
     repeat_region   987. .1144
                     /note="MLT1B repeat: matches 5. .205 of consensus"
     repeat_region   1149. .1500
                     /note="MST-INTERNAL repeat: matches 1289. .1651 of
                     consensus"
     repeat_region   1812. .2130
                     /note="MLT1-INTERNAL repeat: matches 404. .728 of
                     consensus"
     repeat_region   2175. .2472
                     /note="MLT2FB repeat: matches 101. .414 of consensus"
     repeat_region   2729. .2832
                     /note="MER5B repeat: matches 2. .80 of consensus"
     repeat_region   2833. .3146
                     /note="AluSx repeat: matches 1. .308 of consensus"
     repeat_region   3147. .3236
                     /note="MER5B repeat: matches 80. .177 of consensus"
     repeat_region   5694. .6008
                     /note="AluSx repeat: matches 1. .306 of consensus"
     repeat_region   6183. .6318
                     /note="MER5A repeat: matches 10. .189 of consensus"
```

```
repeat_region   6556. .6717
                /note="FRAM repeat: matches -5. .161 of consensus"
repeat_region   7313. .7619
                /note="AluSx repeat: matches 1. .311 of consensus"
repeat_region   7907. .8217
                /note="AluY repeat: matches 1. .311 of consensus"
repeat_region   8545. .8837
                /note="AluSg repeat: matches 1. .293 of consensus"
repeat_region   9101. .9389
                /note="AluJo repeat: matches 1. .282 of consensus"
repeat_region   9399. .9701
                /note="AluY repeat: matches 1. .311 of consensus"
repeat_region   9702. .10014
                /note="AluSx repeat: matches 1. .311 of consensus"
repeat_region   10050. .10292
                /note="AluSp repeat: matches 66. .310 of consensus"
repeat_region   10300. .10615
                /note="AluSx repeat: matches 4. .312 of consensus"
repeat_region   14460. .14754
                /note="AluJo repeat: matches 13. .308 of consensus"
repeat_region   14846. .15042
                /note="AluJb repeat: matches 96. .296 of consensus"
repeat_region   15166. .15281
                /note="AluJb repeat: matches 9. .96 of consensus"
repeat_region   15506. .15781
                /note="AluSg repeat: matches 1. .276 of consensus"
repeat_region   16165. .16466
                /note="AluSx repeat: matches 1. .300 of consensus"
repeat_region   18757. .19057
                /note="AluSx repeat: matches 1. .301 of consensus"
repeat_region   19097. .19144
                /note="12 copies 4 mer tata 77% conserved"
repeat_region   19359. .19662
                /note="AluSg repeat: matches 1. .304 of consensus"
repeat_region   19743. .20060
                /note="AluSx repeat: matches 1. .305 of consensus"
repeat_region   21872. .22179
                /note="AluSx repeat: matches 1. .299 of consensus"
repeat_region   27049. .27427
                /note="MLT1B repeat: matches 7. .390 of consensus"
repeat_region   29141. .29384
                /note="MIR repeat: matches 2. .260 of consensus"
misc_feature    29833. .30559
                /note="CpG island"
                /evidence=not_experimental
repeat_region   31569. .31653
                /note="MER91A repeat: matches 3. .63 of consensus"
repeat_region   31939. .32227
                /note="AluY repeat: matches 1. .291 of consensus"
repeat_region   32228. .32323
                /note="48 copies 2 mer gg 67% conserved"
repeat_region   32248. .32297
                /note="10 copies 5 mer gacgg 88% conserved"
repeat_region   32324. .32378
                /note="11 copies 5 mer aagaa 78% conserved"
repeat_region   33239. .33518
                /note="AluJo repeat: matches 1. .289 of consensus"
repeat_region   34313. .34610
                /note="AluSq repeat: matches 1. .298 of consensus"
repeat_region   35793. .35899
                /note="MIR repeat: matches 1. .116 of consensus"
repeat_region   37385. .37604
                /note="L2 repeat: matches 2436. .2659 of consensus"
repeat_region   38070. .38181
                /note="L2 repeat: matches 2563. .2729 of consensus"
repeat_region   40862. .41123
                /note="AluSq repeat: matches 39. .302 of consensus"
repeat_region   41722. .42016
                /note="AluSp repeat: matches 1. .297 of consensus"
repeat_region   43048. .43521
                /note="L1MB5 repeat: matches 4783. .5239 of consensus"
repeat_region   43522. .43695
                /note="FAM repeat: matches 1. .174 of consensus"
repeat_region   43696. .44552
                /note="L1MB5 repeat: matches 5239. .6138 of consensus"
repeat_region   44598. .44834
                /note="L1MB4 repeat: matches 5814. .6065 of consensus"
repeat_region   44835. .45133
                /note="AluSg repeat: matches 1. .295 of consensus"
repeat_region   45134. .45257
                /note="L1MB4 repeat: matches 6065. .6184 of consensus"
repeat_region   45312. .45611
                /note="AluJo repeat: matches 1. .293 of consensus"
repeat_region   45681. .45989
                /note="AluSg repeat: matches 1. .300 of consensus"
repeat_region   46204. .46347
                /note="L1ME2 repeat: matches 6007. .6153 of consensus"
repeat_region   46876. .47188
                /note="AluJo repeat: matches 1. .304 of consensus"
repeat_region   47262. .47560
                /note="AluSq repeat: matches 1. .301 of consensus"
repeat_region   47923. .48041
                /note="L1MC4 repeat: matches 6562. .6679 of consensus"
repeat_region   48060. .48148
                /note="L1ME1 repeat: matches 5347. .5428 of consensus"
repeat_region   48149. .48413
                /note="AluJo repeat: matches 2. .302 of consensus"
repeat_region   48414. .48641
                /note="L1ME1 repeat: matches 5428. .5637 of consensus"
repeat_region   48642. .48935
                /note="AluSg repeat: matches 1. .294 of consensus"
repeat_region   48936. .49372
                /note="L1ME1 repeat: matches 5637. .6160 of consensus"
repeat_region   50733. .50947
                /note="MIR repeat: matches 2. .257 of consensus"
repeat_region   51470. .51541
                /note="MIR repeat: matches 44. .124 of consensus"
repeat_region   51667. .51977
                /note="AluY repeat: matches 1. .311 of consensus"
repeat_region   52966. .53214
                /note="MIR repeat: matches 12. .262 of consensus"
repeat_region   53551. .54965
                /note="L2 repeat: matches 505. .2101 of consensus"
repeat_region   54984. .55272
                /note="AluSp repeat: matches 1. .291 of consensus"

 Query Match             8.3%;  Score 385.8;  DB 9;  Length 139118;
 Best Local Similarity   77.4%;  Pred. No. 1.1e-66;
 Matches  583;  Conservative    0;  Mismatches  132;  Indels   38;  Gaps    8;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC---AAAA 57
              ||||| ||||||||||  ||||||||||||||||||||||||||||||||    ||||
Db      97498 AGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAAAAA 97557

Qy         58 TTAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAG 117
              ||||   ||| |||||  ||| ||||||| || |||||||  | |||||| ||||||||||
Db      97558 TTAGCTGGGCGTGGTGGCACATGCCTGTAGTCTCAGCTACCTGGGAGGCTGAGGCAGGAG 97617

Qy        118 AATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTC 176
              |||| ||||||||  | || ||||||||||||  | ||| | |||||   ||| ||||||||
Db      97618 AATCGCTTGAACC--CAGGAGGTGGAGGTTGCAGTGAGCCGAGATCGCTCCACTGCACTC 97675

Qy        177 CAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCAT---------- 226
              | ||   ||| || |||||   |   | ||||| ||||||||||   ||
Db      97676 CGGCTTGGGCGAC-AGAGCATGACTCCGTCTCAAAAAAAAAAAAATATATATATATAT 97734

Qy        227 --------------TTTATATATATATATATATATATACACACACACACACATATAT 271
                           | |||||||||| | ||||||| ||| | ||| ||| | ||| ||
Db      97735 ATGTGTGTGTGTGTGTATATATATACACATATATACATATATACATACATATATACAT 97794

Qy        272 ATATACACATATATATACACATATATACATATATACACATATATACACATATATATACAC 331
              ||||||| | ||||||||||||||||||||| ||| ||||||||||||| ||||| | ||
Db      97795 ATATACATACATATATACACATATATACATACATATACATATATACACACATATACATAC 97854

Qy        332 ATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACACAC 391
              ||| | || || ||||||||| ||||| ||  ||| | | ||||| ||||||||||||||
```

```
Db      97855 ATATACATATATACATATATACACACACATACATATATACATATATACATATATACACAT 97914

Qy        392 ATATATACACATATATACACACATATATACACATATATACACATATATACACATATACAC 451
              ||||| | | ||| ||| | ||| ||||| ||| | ||| ||||| ||| ||||||| ||
Db      97915 ATATACATATATACATATATACACATATATACACACATATACATACATATACATATATAC 97974

Qy        452 ATATATACACATATATACATATATACACATATATATA-ATATACACACATATATATACAC 510
              |||||||||||||||||||||||||| |||||  || ||||||| | ||| ||||| ||
Db      97975 GTATATACACATATATACATATATACATATATACGTATATATACATATATACATATATAC 98034

Qy        511 ATATATACACACATATATACACATATATACACATATATATACACATATATACACATATAT 570
              |||||| ||| ||||| | | |||  || | ||| ||||| |||||||||| ||||||||
Db      98035 GTATATATACATATATACATATATACGTATATATACATATATACATATATATACATATAT 98094

Qy        571 ACATATATACACATAT----ATATACATATATACACATATATACATATATACACATATAT 626
              ||||||||| ||||||    ||||| | ||||| || ||||| |||||||| ||||||||
Db      98095 ACATATATATACATATATACATATATACATATATACGTATATGCATATATATACATATAT 98154

Qy        627 ACATATATACACACATATATACACATACATATACACACACATAGATATACATATATATAC 686
              ||||||||| ||| ||||||||| ||| | | | | | ||| | |||||||| ||||||
Db      98155 ACATATATATACATATATATACATATATACACATATATACACA-ATATACATACATATAC 98213

Qy        687 ACATATATATACGTATATATATGTATATATATA 719
              | ||||| | |  || |||||| || |||||||
Db      98214 ATATATACACATATACATATATATACATATATA 98246


RESULT 14
HUAC004787
LOCUS       HUAC004787            216021 bp    DNA     linear   PRI 30-OCT-2002
DEFINITION  Homo sapiens Chromosome 16 BAC clone CIT987SK-A-952F10, complete
            sequence.
ACCESSION   AC004787
VERSION     AC004787.1  GI:3337381
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 216021)
  AUTHORS   Loftus,B.J., Kim,U.J., Sneddon,V.P., Kalush,F., Brandon,R.,
            Fuhrmann,J., Mason,T., Crosby,M.L., Barnstead,M., Cronin,L.,
            Deslattes Mays,A., Cao,Y., Xu,R.X., Kang,H.L., Mitchell,S.,
            Eichler,E.E., Harris,P.C., Venter,J.C. and Adams,M.D.
  TITLE     Genome duplications and other features in 12 Mb of DNA sequence
            from human chromosome 16p and 16q
  JOURNAL   Genomics 60 (3), 295-308 (1999)
  MEDLINE   99425270
   PUBMED   10493829
REFERENCE   2  (bases 1 to 216021)
  AUTHORS   Adams,M.D., Loftus,B.J., Zhou,L., Crosby,M., Fuhrmann,J.,
            Mason,T.M., Brandon,R., Kim,U.J., Kerlavage,A.R. and Venter,J.C.
  TITLE     Homo sapiens Chromosome 16 BAC clone CIT987SK-A-952F10
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 216021)
  AUTHORS   Adams,M.D. and Loftus,B.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1998) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, Email:
            bjloftus@tigr.org
REFERENCE   4  (bases 1 to 216021)
  AUTHORS   Adams,M.D. and Loftus,B.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JUL-1998) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
COMMENT     On Jul 24, 1998 this sequence version replaced gi:3241936.
            Address all correspondence to: Mark Adams The Institute for Genomic
            Research 9712 Medical Center Dr, Rockville, MD 20850, USA e-mail
            address: humgen@tigr.org.   The orientation of the sequence is from
            SP6 end to T7 end. Genes were identified by a combination of five
            methods including: XGRAIL (available by anonymous ftp from
            arthur.epm.ornl.gov), Genefinder (Phil Green, University of
            Washington), Genscan  (Chris Burge,
            http://gnomic.stanford.edu/~chris/GENSCANW.html)searches of the
            complete sequence against a peptide database,and the Human gene
            Index database at TIGR (http://www.tigr.org/tdb/hgi/hgi.html).
            Genes without pepetide homolgy having spliced EST hits are termed
            'Unknown gene product'. Genes encoding tRNAs are predicted by
            tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
FEATURES             Location/Qualifiers
     source          1. .216021
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /map="#16q21-22"
                     /clone="A-952F10"
     STS             27765. .27872
                     /note="7766, STS1-cSRL-27g3-uA/cSRL-27g3-uZ, Chr. -, Homo
                     sapiens"
                     /db_xref="dbSTS:G02280"
     STS             73826. .73943
                     /note="7608, STS1-cSRL-24g1-uA/cSRL-24g1-uZ, Chr. -, Homo
                     sapiens"
                     /db_xref="dbSTS:G02122"
     STS             175801. .175945
                     /note="16084, CHLC.GCT10B02, Chr. -, Homo sapiens"
                     /db_xref="dbSTS:G09703"
     STS             175810. .175945
                     /note="16316, CHLC.GCT15C04, Chr. -, Homo sapiens"
                     /db_xref="dbSTS:G09935"
     STS             199463. .199572
                     /note="9824, WI-3555, Chr. 16, Homo sapiens"
                     /db_xref="dbSTS:G04338"
ORIGIN

  Query Match             8.3%;  Score 382.6;  DB 9;  Length 216021;
  Best Local Similarity   76.2%;  Pred. No. 4.4e-66;
  Matches  551;  Conservative    0;  Mismatches  159;  Indels   13;  Gaps    6;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAATTA 60
              ||||| | |||| ||||  |||||| |||||||||||||||||||||||     |||||||
Db      80976 AGTTCGAAACCACCCTGGCCAACATAGTGAAACCCCATCTCTACTAAAATACAAAAATTA 81035

Qy         61 GACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAAT 120
              | ||||| |||||   | ||||||||||||||||||||||   |||||| ||||||||| ||
Db      81036 GCCAGGCGTGGTGGCAGACGCCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGAGGAT 81095

Qy        121 CACTTGAACCTGCTGGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCCAG 179
              ||||||||  | || ||  |||||||| | ||| | ||||  ||||| || |||||||
Db      81096 TGCTTGAACC--CAGGAGGCAGAGGTTGCAGTGAGCCCAGATTGTGCCACTGGACTCCAG 81153

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  ||  || |||| | |   | ||||| | | | | | |      || | | ||||
Db      81154 CCTGGGAGAC-AGAGTAAGACTCCGTCTCAAATACATATATA-----CATACACACATAT 81207

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATACATAC 299
              ||||| | ||||||| ||||||| | ||||| | ||| ||||| | | | ||||| | ||
Db      81208 ATATACACATATACATACACACATATATATACACATATACATACACACATACATACACAC 81267

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACATATATATATACACAT 359
              |||||||||| ||||||| ||||||| ||| ||| |||  || ||| ||||||||||||
Db      81268 ATATATACACACATATACATATATATACACATATATATACATATACACATATATACACAT 81327

Qy        360 ATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATAT 419
              || || | | | | | | | ||||| | | | | | ||| ||| ||||| ||| |||||
Db      81328 ATATA-ATATACACACATATATATACATATATACACATATACACATATATACATATATAC 81386

Qy        420 ACACATATATACACATATATACACATATACACATATATACACATATATACATATATACAC 479
              |||||||||||||||||||||||||| ||| |||||||||||| ||| | |||||| ||||
Db      81387 ACACATATATACACATATATACATATACACACATATATACATATACACACATATACACAC 81446

Qy        480 ATATATATA---ATATACACACATATATATACACATATATACACACATATATACACATAT 536
              ||| | | |   | | ||||||||| | | | ||| | | | ||||| | | | ||| |
Db      81447 ATACACACACATACACACACACATACACACATACACACACATACACACACACATACACAC 81506
```

```
Qy        537 ATACACATATATATACACATATATACACATATATACATATATACACATATATATACATAT 596
              | | ||| | ||| ||||| |||||||||| ||| |  || | | ||| | ||||||| ||
Db      81507 ACATACACACATACACACACATATACACACATACATGTACACATACACACATATACACAT 81566

Qy        597 ATACACATATATACATATATACACATATATACATATATACACACATATATACACATACAT 656
              ||||||| |||||||||||||||| |||||||||||||||||||||||| ||||| ||
Db      81567 ATACACACATATACATATATACACACATATACATATATACACACATATATATACATATAT 81626

Qy        657 ATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATAT 716
              | ||| | | | ||||||| |||||||| ||||  | |  |||| |||| || | ||
Db      81627 ACACATATATACATATATACACATATATACATATATGCACATATATACATATATACACAT 81686

Qy        717 ATA 719
              |||
Db      81687 GTA 81689


RESULT 15
AY007685
LOCUS       AY007685               92564 bp    DNA     linear   PRI 22-FEB-2002
DEFINITION  Homo sapiens telomerase catalytic subunit (TERT) and sodium
            channel-like protein genes, complete cds.
ACCESSION   AY007685
VERSION     AY007685.2  GI:15991796
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 92564)
  AUTHORS   Leem,S.H., Londono-Vallejo,J.A., Kim,J.H., Bui,H., Tubacher,E.,
            Solomon,G., Park,J.E., Horikawa,I., Kouprina,N., Barrett,J.C. and
            Larionov,V.
  TITLE     The human telomerase gene: complete genomic sequence and analysis
            of tandem repeat polymorphisms in intronic regions
  JOURNAL   Oncogene 21 (5), 769-777 (2002)
  MEDLINE   21839826
   PUBMED   11850805
REFERENCE   2  (bases 1 to 92564)
  AUTHORS   Londono-Vallejo,J.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-SEP-2000) Centre d'Etudes du Polymorphisme Humain, 27
            rue Juliette Dodu, Paris 75010, France
REFERENCE   3  (bases 1 to 92564)
  AUTHORS   Londono-Vallejo,J.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-OCT-2001) Centre d'Etudes du Polymorphisme Humain, 27
            rue Juliette Dodu, Paris 75010, France
  REMARK    Sequence update by submitter
COMMENT     On Oct 10, 2001 this sequence version replaced gi:12642956.
FEATURES             Location/Qualifiers
     source          1. .92564
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /map="5p13"
     repeat_region   complement(2853. .3034)
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   complement(6174. .6429)
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   complement(7445. .7837)
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   complement(8199. .8341)
                     /rpt_family="MLT1"
                     /rpt_type=dispersed
     repeat_region   complement(10300. .10916)
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   12028. .12301
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   13088. .13597
                     /rpt_family="L1"
                     /rpt_type=dispersed
     repeat_region   13716. .13988
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   14271. .14556
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   14599. .17711
                     /rpt_family="L1"
                     /rpt_type=dispersed
     repeat_region   18109. .18398
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   complement(20818. .21123)
                     /rpt_family="LTR1"
                     /rpt_type=dispersed
     repeat_region   21650. .21912
                     /rpt_family="MLT1"
                     /rpt_type=dispersed
     repeat_region   complement(22236. .22484)
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   complement(22986. .23283)
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     gene            <24442. .64466
                     /gene="TERT"
     mRNA            join(<24442. .24660,24765. .26118,35473. .35668,37759. .37939,
                     38627. .38806,39301. .39456,45360. .45455,46436. .46521,
                     49007. .49120,51105. .51176,53048. .53236,57040. .57166,
                     58981. .59042,62229. .62353,63135. .63272,63809. .>63912)
                     /gene="TERT"
                     /product="telomerase catalytic subunit"
     CDS             join(24442. .24660,24765. .26118,35473. .35668,37759. .37939,
                     38627. .38806,39301. .39456,45360. .45455,46436. .46521,
                     49007. .49120,51105. .51176,53048. .53236,57040. .57166,
                     58981. .59042,62229. .62353,63135. .63272,63809. .63912)
                     /gene="TERT"
                     /codon_start=1
                     /product="telomerase catalytic subunit"
                     /protein_id="AAG23289.1"
                     /db_xref="GI:12642957"
                     /translation="MPRAPRCRAVRSLLRSHYREVLPLATFVRRLGPQGWRLVQRGDP
                     AAFRALVAQCLVCVPWDARPPPAAPSFRQVSCLKELVARVLQRLCERGAKNVLAFGFA
                     LLDGARGGPPEAFTTSVRSYLPNTVTDALRGSGAWGLLLRRVGDDVLVHLLARCALFV
                     LVAPSCAYQVCGPPLYQLGAATQARPPPHASGPRRRLGCERAWNHSVREAGVPLGLPA
                     PGARRRGGSASRSLPLPKRPRRGAAPEPERTPVGQGSWAHPGRTRGPSDRGFCVVSPA
                     RPAEEATSLEGALSGTRHSHPSVGRQHHAGPPSTSRPPRPWDTPCPPVYAETKHFLYS
                     SGDKEQLRPSFLLSSLRPSLTGARRLVETIFLGSRPWMPGTPRRLPRLPQRYWQMRPL
                     FLELLGNHAQCPYGVLLKTHCPLRAAVTPAAGVCAREKPQGSVAAPEEEDTDPRRLVQ
                     LLRQHSSPWQVYGFVRACLRRLVPPGLWGSRHNERRFLRNTKKFISLGKHAKLSLQEL
                     TWKMSVRDCAWLRRSPGVGCVPAAEHRLREEILAKFLHWLMSVYVVELLRSFFYVTET
                     TFQKNRLFFYRKSVWSKLQSIGIRQHLKRVQLRELSEAEVRQHREARPALLTSRLRFI
                     PKPDGLRPIVNMDYVVGARTFRREKRAERLTSRVKALFSVLNYERARRPGLLGASVLG
                     LDDIHRAWRTFVLRVRAQDPPPELYFVKVDVTGAYDTIPQDRLTEVIASIIKPQNTYC
                     VRRYAVVQKAAHGHVRKAFKSHVSTLTDLQPYMRQFVAHLQETSPLRDAVVIEQSSSL
                     NEASSGLFDVFLRFMCHHAVRIRGKSYVQCQGIPQGSILSTLLCSLCYGDMENKLFAG
                     IRRDGLLLRLVDDFLLVTPHLTHAKTFLRTLVRGVPEYGCVVNLRKTVVNFPVEDEAL
                     GGTAFVQMPAHGLFPWCGLLLDTRTLEVQSDYSSYARTSIRASLTFNRGFKAGRNMRR
                     KLFGVLRLKCHSLFLDLQVNSLQTVCTNIYKILLLQAYRFHACVLQLPFHQQVWKNPT
                     FFLRVISDTASLCYSILKAKNAGMSLGAKGAAGPLPSEAVQWLCHQAFLLKLTRHRVT
                     YVPLLGSLRTAQTQLSRKLPGTTLTALEAAANPALPSDFKTILD"
     repeat_region   27341. .27628
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   27650. .29650
                     /note="VNTR"
                     /rpt_type=tandem
                     /rpt_unit="gagtgaggcgccgtccccgggtgtccctgtcacgtgcagggt"
```

```
repeat_region   complement(31926. .32203)
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   complement(32555. .32835)
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   33569. .33861
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   complement(33873. .33967)
                /rpt_family="MER33"
                /rpt_type=dispersed
repeat_region   34111. .35453
                /note="VNTR"
                /rpt_type=tandem
repeat_region   complement(35993. .36281)
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   41485. .42324
                /note="VNTR"
                /rpt_type=tandem
                /rpt_unit="gtgggattggttttcatgtgtggggtaggtggggatct"
repeat_region   complement(42656. .42954)
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   43312. .43618
                /rpt_family="MER1"
                /rpt_type=dispersed
repeat_region   44225. .45115
                /note="VNTR"
                /rpt_type=tandem
                /rpt_unit="ggggtctgatgtgtggtgactgtggatggcggtcgt"
repeat_region   complement(48015. .48300)
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   complement(49647. .50276)
                /rpt_family="L1"
                /rpt_type=dispersed
repeat_region   53945. .54234
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   57630. .58735
                /note="VNTR"
                /rpt_type=tandem
                /rpt_unit="tgggcatctgcgtccactccccctctcctg"
repeat_region   61332. .61597
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   complement(66201. .66323)
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   66389. .66661
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   66661. .66736
                /note="microsatellite"
                /rpt_type=tandem
                /rpt_unit="aag"
repeat_region   67140. .67583
                /rpt_family="MLT1"
                /rpt_type=dispersed
mRNA            complement(join(<73264. .73377,73925. .74084,76991. .77080,
                78049. .78146,79551. .79661,81950. .82131,84724. .84861,
                85253. .85393,91909. .>92068))
                /product="sodium channel-like protein"
CDS             complement(join(73264. .73377,73925. .74084,76991. .77080,
                78049. .78146,79551. .79661,81950. .82131,84724. .84861,
                85253. .85393,91909. .92068))
                /codon_start=1
                /evidence=not_experimental
                /product="sodium channel-like protein"
                /protein_id="AAG23290.1"
                /db_xref="GI:12642958"
```

```
  Query Match             8.2%;  Score 379.8;  DB 9;  Length 92564;
  Best Local Similarity   75.0%;  Pred. No. 1.8e-65;
  Matches  542;  Conservative    0;  Mismatches  172;  Indels    9;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||| |||||||||||  ||| || | ||||||||||||| |||||||| || ||||||
Db      11380 AGTTCGAGACCAGCCTGGCCAATATAGCGAAACCCCATCTGTACTAAAAACACAAAAATT 11439

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   ||  |||||| ||| |||||||||||||| ||||  ||||||  ||||||||||
Db      11440 AGCTGGGTGTGGTGACACATGCCTGTAATCCCAGGTACTCAGGAGGCTAAGGCAGGAGAA 11499

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              |  |||||||  |  ||  | || |||     |   ||||||    ||| ||||||||||
Db      11500 TTGCTTGAACTGGGAGGCAGAGGTGGT-----GAGCCAAGATTGCACCAGTGCACTCCAG 11554

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  ||  || |||| ||||   ||||||| ||| ||| |||   | |  ||||| |||
Db      11555 CCTTGGTGAC-AGAGTGAAACTCCATCTCAAAAACAAACAAACAAAATACATATACATAA 11613

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              |||||||  | | | | | ||| | ||| ||||||| ||| ||||||| |  |||||||
Db      11614 ATATATATGCACATATATATACATATATAAATATATATACACATATATAAATCTATATAC 11673

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACATATAT-ATACACA 358
              |||||||| ||||| ||| ||||| ||  | |||||||| || |||||||| ||| |
Db      11674 ATATATACATATATACACATATATAAATCTATATACATATATATACATATATAATATATT 11733

Qy        359 TATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATA 418
              ||  |||| | |||||||| ||||| | | |||||||| | | ||||| | | ||| |
Db      11734 TACATATATAAATATATACATATATAAATATACATATATAAATACATATATAAATATACA 11793

Qy        419 TACACATATATACACATATATACACATATACACATATATACACATATATACATATATACA 478
              || | | ||||| |||||||| | ||| | | ||| ||||| ||||| | ||| ||| |
Db      11794 TATATAAATATATACATATATAAATATACATATATAAATACATATATAAATATACATATA 11853

Qy        479 CATATATATA-ATATACACACATATATATACACATATATACACACATATATACACATATA 537
               | ||||||| ||||| | | ||| ||||| | |||||||||| | ||| ||| |||||||
Db      11854 TAAATATATACATATATAAATATACATATATAAATATATACATATATAAATATACATATA 11913

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATACATATA 597
              |||| ||||| ||| ||||||| | | ||| ||| ||||| ||| ||||| ||| | |||
Db      11914 TACATATATAAATATACATATACATATATAAATATATATACACATATATAAATATACATA 11973

Qy        598 TACACATATATACATATATACACATATATACATATATACACACATATATACACATACATA 657
              || | ||||| | ||||| | | ||| ||| ||| ||| ||| ||||| | | ||| | |
Db      11974 TATAAATATACATATATAAATATATACATATATAAATATACATATATAAATATATATACA 12033

Qy        658 TACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATA 717
              || | | | ||| | ||| | |  |||| ||| |||||||| |||||| |||||||||
Db      12034 TATATAAATATATAAATATACAAGTATATACAAATATATACATATATAAATGTATATACG 12093

Qy        718 TAT 720
              |||
Db      12094 TAT 12096


RESULT 16
AC114955/c
LOCUS       AC114955              161252 bp    DNA     linear   HTG 14-MAR-2002
DEFINITION  Homo sapiens chromosome 5 clone RP11-117B23, WORKING DRAFT
            SEQUENCE, 9 unordered pieces.
ACCESSION   AC114955
VERSION     AC114955.1  GI:19424416
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 161252)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 5
```

```
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 161252)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-MAR-2002) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 451340 .
            Center clone name: RPCI-11_117B23
            -------
            Summary Statistics
            Consensus quality: 148939 bases at least Q40
            Consensus quality: 156608 bases at least Q30
            Consensus quality: 158826 bases at least Q20
            Estimated insert size: 175000; agarose-fp estimation
            Estimated insert size: 160452; sum-of-contigs estimation
            Quality coverage:  7.26 in Q20 bases; agarose-fp estimation
            Quality coverage:  7.92 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     3008: contig of 3008 bp in length
            *     3009     3108: gap of unknown length
            *     3109     9209: contig of 6101 bp in length
            *     9210     9309: gap of unknown length
            *     9310    15613: contig of 6304 bp in length
            *    15614    15713: gap of unknown length
            *    15714    23243: contig of 7530 bp in length
            *    23244    23343: gap of unknown length
            *    23344    39674: contig of 16331 bp in length
            *    39675    39774: gap of unknown length
            *    39775    60001: contig of 20227 bp in length
            *    60002    60101: gap of unknown length
            *    60102    85303: contig of 25202 bp in length
            *    85304    85403: gap of unknown length
            *    85404   117460: contig of 32057 bp in length
            *   117461   117560: gap of unknown length
            *   117561   161252: contig of 43692 bp in length.
FEATURES             Location/Qualifiers
     source          1. .161252
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="RP11-117B23"
                     /clone_lib="RPCI human BAC library 11"
ORIGIN

  Query Match             8.2%;  Score 379.4;  DB 2;  Length 161252;
  Best Local Similarity   68.7%;  Pred. No. 2e-65;
  Matches  600;  Conservative    0;  Mismatches  256;  Indels   17;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||| ||||||||||  ||| || | ||||||||||| ||||||||| || ||||||
Db     149753 AGTTCGAGACCAGCCTGGCCAATATAGCGAAACCCCATCTGTACTAAAAACACAAAAATT 149694

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   ||  |||||| ||| ||||||||||||| ||||  ||||||  ||||||||||
Db     149693 AGCTGGGTGTGGTGACACATGCCTGTAATCCCAGGTACTCAGGAGGCTAAGGCAGGAGAA 149634

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              | |||||||  |  ||  | || |||      |   ||||||    ||| ||||||||||
Db     149633 TTGCTTGAACTGGGAGGCAGAGGTGGT-----GAGCCAAGATTGCACCAGTGCACTCCAG 149579

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  ||  || |||| ||||    ||||||| ||| ||| |||    | |  ||||| |||
Db     149578 CCTTGGTGAC-AGAGTGAAACTCCATCTCAAAAACAAACAAACAAAATACATATACATAA 149520

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              |||||||  | | | | | ||| | ||| ||||||| ||| ||||||| |  |||||||
Db     149519 ATATATATGCACATATATATACATATATAAATATATATACACATATATAAATCTATATAC 149460

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACATATATATA----- 354
              |||||||| ||||| ||| ||||| ||  | |||||||| || | | |||||||
Db     149459 ATATATACATATATACACATATATAAATCTATATACATATATATATACATATATAATATA 149400

Qy        355 --CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACA 412
                |||||| || | | ||| ||| | | ||||||| | ||| ||||| ||||| | | |
Db     149399 TTTACATATATAAATATATACATATATAAATATACATATATAAATACATATATAAATATA 149340

Qy        413 CATATATACACATATATACACATATATACACATATACACATATATACACATATATACATA 472
              ||||||| | ||||| | | ||| ||| ||||||| | ||| ||| | ||| |||||||
Db     149339 CATATATAAATATATACATATATAAATATACATATATAAATACATATATAAATACATA 149280

Qy        473 TATACACATATATAT---AATATACACACATATATATACACATATATACACACATATATA 529
              |||| | ||| ||||    |||||| ||| ||||| ||| ||||||||| | | ||| |||
Db     149279 TATAAATATACATATATAAATATATACATATATAAATATACATATATAAATATATACATA 149220

Qy        530 CACATATATACACATATATATACACATATATACACATATATACATATATACACATATATA 589
              | | |||||| ||||| ||| | | ||||||| ||| ||| ||| ||| | ||| | |
Db     149219 TATAAATATACATATATACATATATAAATATACATATACATATATAAATATATATACACA 149160

Qy        590 TACATATATACACATATATACATATATACACATATATACATATATACACACATATATACA 649
              || ||| ||| ||||||||| ||||| | | ||| ||| ||| ||| | | ||||| | |
Db     149159 TATATAAATATACATATATAAATATACATATATAAATATATACATATATAAATATACATA 149100

Qy        650 CATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATATG 709
              ||| ||||| | ||| ||| | ||| ||| |||||||  |||||| |  |||||| ||
Db     149099 TATAAATATATATACATATATAAATATATAAATATACAAGTATATACAAATATATACATA 149040

Qy        710 TATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGG 769
              |||| | ||||    |     | | ||        | |  |   ||      |   ||| | |
Db     149039 TATAAACATATATACGTATATACGTATATACATGTATATACGTATATACGTATATACATG 148980

Qy        770 AAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGT 829
              | | |   |  ||     |  ||     | | | | || |     ||     |  ||
Db     148979 TATATACGTATATACGTATATACATGTATATACGTATATATATAAATGTATATACGTATA 148920

Qy        830 AGGGGAAATACATAAATGTGTAATAAAACCTAT 862
                | ||| ||||||  | |||  ||||| | |
Db     148919 TACATATATATATAAATATATAAAAAAACTTTT 148887


RESULT 17
AC114291/c
LOCUS       AC114291              202305 bp    DNA     linear   PRI 15-MAR-2003
DEFINITION  Homo sapiens chromosome 5 clone CTD-3080P12, complete sequence.
ACCESSION   AC114291
VERSION     AC114291.2  GI:28973811
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 202305)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 202305)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAR-2002) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 202305)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
```

```
  JOURNAL   Submitted (15-MAR-2003) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Mar 15, 2003 this sequence version replaced gi:19224940.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 99.9% of Sequence;
            Estimated Total Number of Errors is 0.2.
            NOTE: Transposon sequencing failed to verify number of repeat
            copies 173000-175780. Unsure number of repeat copies 173000-175780.
            BAC and subclones unstable 76080. Force join at 76080.
FEATURES             Location/Qualifiers
     source          1. .202305
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTD-3080P12"
     misc_feature    76080
                     /note="NOTE: BAC and subclones unstable 76080. Force join
                     at 76080."
     misc_feature    173000. .175780
                     /note="NOTE: Transposon sequencing failed to verify number
                     of repeat copies 173000-175780. Unsure number of repeat
                     copies 173000-175780"
ORIGIN

  Query Match             8.2%;  Score 378;  DB 9;  Length 202305;
  Best Local Similarity   66.3%;  Pred. No. 3.6e-65;
  Matches  640;  Conservative    0;  Mismatches  310;  Indels   16;  Gaps    6;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||| ||||||||||||  ||| || | ||||||||||||| |||||||| || ||||||
Db     192244 AGTTCGAGACCAGCCTGGCCAATATAGCGAAACCCCATCTGTACTAAAAACACAAAAATT 192185

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   ||  |||||| ||| ||||||||||||||| ||||  ||||||  ||||||||||
Db     192184 AGCTGGGTGTGGTGACACATGCCTGTAATCCCAGGTACTCAGGAGGCTAAGGCAGGAGAA 192125

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              |  |||||||  |  ||  | || |||     |   ||||||    ||| ||||||||||
Db     192124 TTGCTTGAACTGGGAGGCAGAGGTGGT-----GAGCCAAGATTGCACCAGTGCACTCCAG 192070

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  ||   || |||| ||||    ||||||| ||| ||| |||    | |   ||||| |||
Db     192069 CCTTGGTGAC-AGAGTGAAACTCCATCTCAAAAACAAACAAACAAAATACATATACATAA 192011

Qy        240 ATATATATATATACACACACACACACATATATATATACACATATATATACACATATATAC 299
              ||||||||  | | | | | ||| | ||| ||||||| ||| ||||||| |  |||||||
Db     192010 ATATATATGCACATATATATACATATATAAATATATATACACATATATAAATCTATATAC 191951

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACAT-----ATATATA 354
              ||||||||| ||||| ||| ||||| ||  | |||||||| || ||||      ||||||
Db     191950 ATATATACATATATACACATATATAAATCTATATACATATATATACATATATAATATATT 191891

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              |||||| || | | ||| ||| | | ||||||| | ||| ||||| ||||| | | |||
Db     191890 TACATATATAAATATATACATATATAAATATACATATATAAATACATATATAAATATACA 191831

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              |||||| | ||||| | | ||| ||| ||||||| | ||| ||| | | ||||||||||||
Db     191830 TATATAAATATATACATATATAAATATACATATATAAATACATATATAAATATACATATA 191771

Qy        475 TACACATATATATA---ATATACACACATATATATACACATATATACACACATATATACA 531
              || | ||||| |||    | ||| ||| ||||| ||| | | ||||| | | ||| ||| |
Db     191770 TAAATATATACATATATAAATATACATATATAAATATATACATATATAAATATACATATA 191711

Qy        532 CATATATACACATATATATACACATATATACACATATATACATATATACACATATATATA 591
              | ||||| | ||||| ||| ||||||||| | ||||||||| ||||| | ||||| |||
Db     191710 TACATATATAAATATACATATACATATATAAATATATATACACATATATAAATATACATA 191651

Qy        592 CATATATACACATATATACATATATACACATATATACATATATACACACATATATACACA 651
              ||| ||| ||||||||| ||||||||| ||||| | ||| ||| | | ||||||| |||
Db     191650 TATAAATATACATATATAAATATATACATATATAAATATACATATATAAATATATATACA 191591

Qy        652 TACATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTA 711
              || ||| | | | | ||| | |   ||||| | | | ||| |||||   | ||||||
Db     191590 TATATAAATATATAAATATACAAGTATATACAAATATATACATATATAAACATATATACG 191531

Qy        712 TATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAA 771
              |||||| |||         ||  ||| |      ||  |   |||     |  ||| | | |
Db     191530 TATATACATGTATATACGTATATACGTATATACATGTATATACGTACGTATATACATGTA 191471

Qy        772 AAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAG 831
              | |    |  ||      |  ||    | | | | ||       |    | | |||
Db     191470 TATACGTATATACGTATATACATGTATATACGTATATACGTATATACATGTATATACGTA 191411

Qy        832 GGGAAATACATAAATGTGTAATAAAACCTATCT-GTAAGGTAGTTAAGAAGGTAACACTA 890
                |  || ||| ||   |    | || ||| |    |  ||  |||     ||    ||
Db     191410 TATACGTATATATATAAATGTATATACGTATATACGTATATATATAAATGTATATACGTA 191351

Qy        891 TATATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTAT 950
              |||| |||||||   |||    |  |||       |   || ||    |  |||| | ||
Db     191350 TATACATATATATATAAATATATAAAAAAACTTTGGCTGGGCACCTTTCCAAATCTCAT 191291

Qy        951 AGAAGA 956
              | | |
Db     191290 GGCACA 191285


RESULT 18
AC021955/c
LOCUS       AC021955              192139 bp    DNA     linear   PRI 22-MAR-2002
DEFINITION  Homo sapiens chromosome 8, clone RP11-238G17, complete sequence.
ACCESSION   AC021955
VERSION     AC021955.11  GI:19683766
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 192139)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 8, clone RP11-238G17
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 192139)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 192139)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
```

```
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C.,
            Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-FEB-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 192139)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 22, 2002 this sequence version replaced gi:18875268.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L5949
                Center clone name: 238_G_17
            --------------.
FEATURES             Location/Qualifiers
     source          1. .192139
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="RP11-238G17"
                     /clone_lib="RPCI-11 Human Male BAC"
     unsure          441. .445
                     /note="<30 qual SNGL region"
     unsure          455. .461
                     /note="<30 qual SNGL region"
     repeat_region   complement(1301. .1824)
                     /rpt_family="MLT1F"
     repeat_region   complement(1885. .2017)
                     /rpt_family="L2"
     repeat_region   2167. .2264
                     /rpt_family="GA-rich"
     repeat_region   2333. .2853
                     /rpt_family="(TA)n"
     repeat_region   complement(2862. .3156)
                     /rpt_family="AluSq"
     repeat_region   3916. .3974
                     /rpt_family="MLT1F"
     repeat_region   3977. .4182
                     /rpt_family="AluJo"
     repeat_region   complement(4553. .4782)
                     /rpt_family="MIR"
     repeat_region   5326. .5402
                     /rpt_family="LTR33"
     repeat_region   5570. .5756
                     /rpt_family="LTR33"
     repeat_region   complement(6278. .7151)
                     /rpt_family="MER4D"
     repeat_region   7424. .7644
                     /rpt_family="MER58A"
     repeat_region   7701. .7999
                     /rpt_family="AluY"
     repeat_region   9337. .9635
                     /rpt_family="AluSg"
     repeat_region   9723. .9878
                     /rpt_family="L1M4c"
     repeat_region   9877. .10902
                     /rpt_family="L1M4c"
     repeat_region   10903. .11141
                     /rpt_family="AluSc"
     repeat_region   11142. .11261
                     /rpt_family="L1M4c"
     repeat_region   11261. .11381
                     /rpt_family="L1MB4"
     repeat_region   11383. .11839
                     /rpt_family="MER51A"
     repeat_region   11956. .12100
                     /rpt_family="L1MB4"
     repeat_region   12101. .12132
                     /rpt_family="(CA)n"
     repeat_region   12133. .12496
                     /rpt_family="L1MB4"
     repeat_region   12994. .13074
                     /rpt_family="L2"
     repeat_region   13900. .13930
                     /rpt_family="AT_rich"
     repeat_region   14087. .14128
                     /rpt_family="(TG)n"
     repeat_region   complement(15008. .15212)
                     /rpt_family="MIR"
     repeat_region   15423. .15924
                     /rpt_family="MER87"
     repeat_region   16247. .16353
                     /rpt_family="MLT1I"
     repeat_region   17718. .17749
                     /rpt_family="(TTA)n"
     repeat_region   complement(18035. .18137)
                     /rpt_family="MIR3"
     repeat_region   18182. .18418
                     /rpt_family="MIR"
     repeat_region   18433. .18573
                     /rpt_family="L3"
     repeat_region   complement(18574. .18883)
                     /rpt_family="AluSg"
     repeat_region   18884. .18963
                     /rpt_family="L3"
     repeat_region   19497. .19592
                     /rpt_family="MIR"
     repeat_region   19909. .20082
                     /rpt_family="L3"
     repeat_region   complement(20319. .20627)
                     /rpt_family="AluSx"
     repeat_region   20785. .20932
```

```
                /rpt_family="MLT1A1"
     unsure     20910. .20953
                /note="<30 qual SNGL region"
     repeat_region 20933. .21252
                /rpt_family="AluSg"
     unsure     20944
                /note="probably G, possibly C"
     unsure     20965. .21026
                /note="<30 qual SNGL region"
     repeat_region 21253. .21466
                /rpt_family="MLT1A1"

  Query Match             8.1%;  Score 373.6;  DB 9;  Length 192139;
  Best Local Similarity   74.2%;  Pred. No. 2.7e-64;
  Matches  539;  Conservative    0;  Mismatches  179;  Indels    8;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAAT-ACAAAATT 59
              ||||| |||||  ||||  ||||||| ||||||||||||||||||||||| | ||||||
Db       3081 AGTTCGAGACCCACCTGGCCAACATGATGAAACCCCATCTCTACTAAAAATGAAAAAATT 3022

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGA-GGCCGAGGCAGGAGA 118
              ||  ||  |||||  |   || | ||||||||||||||  ||| ||| ||||||| |||
Db       3021 AGCTGGGTGTGGTGGCAGGTGCATCTAATCCCAGCTACTCTGGAGGGCTGAGGCAGAAGA 2962

Qy        119 ATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCC 177
              ||  ||||||||  | || |||||| ||||| | ||| | |||||  ||||||||||||
Db       2961 ATTGCTTGAACC--CAGGAGGTGGACGTTGCAGTGAGCCGAGATCGCGCCATTGCACTCC 2904

Qy        178 AGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAGGCATTTTATATATAT 237
              ||||  ||||||||||||  |||     ||||  |||||||||         |||||||||
Db       2903 AGCCTGGGCAACAAGAGCAAAACTCTGTCTCGAAAAAAAAAATTTTATAAATATATATAT 2844

Qy        238 ATATATATATATATACACACACACACATATATATATACACATATATATACACATATAT 297
              | ||||||||||| | | | | | | ||||| ||||| | | ||||||| | ||||||
Db       2843 AAATATATATATAAATATATATAAATATATATAAATATATATAAATATATATAAATATAT 2784

Qy        298 ACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACAC 357
              | ||| ||| | ||| ||| | ||||| ||| | ||| |||| || | | |||||  | |
Db       2783 ATATAAATATATATAAATATATATATAAATATATATAAATATATATATAAATATAAATAA 2724

Qy        358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417
              |||| |||| | ||| ||| | ||||| | |    ||||| | | ||| ||| | | ||||
Db       2723 ATATATATAAATATAAATAAATATATAAATATGTATATAAATATATAAATATATAAATAT 2664

Qy        418 ATACACATATATACACATATATACACATATACACATATATACACATATATACATATATAC 477
              || | | ||||| | ||| ||| | | |||    ||||| | | ||||| | |||||||
Db       2663 GTATATAAATATATATATAAATATATAAATATGTATATAAATATATATAAATATATATAA 2604

Qy        478 ACATATATATA---ATATACACACATATATATACACATATATACACACATATATACACAT 534
              | || ||||||   |||||  | | || |||||| | |||||||  | | |||||||  | |
Db       2603 ATATGTATATAAATATATATAAATATGTATATAAATATATATATAAAAATATATATAAAA 2544

Qy        535 ATATACACATATATATACACATATATACACATATATACATATATACACATATATATACAT 594
              |||||  | | |||||||  | |||||||  | |||||||  | |||||  | |||||||||  |
Db       2543 ATATATATAAATATATAAATATATATATAAATATATATAAATATATAAATATATATATAA 2484

Qy        595 ATATACACATATATACATATATACACATATATACATATATACACACATATATACACATAC 654
              |||||  | |  |||||  |||||  | |  |||||  | |||||||  | |  |||||||  | |||
Db       2483 ATATATATAAATATATATATAAATATATATAAATATATATAAATATATATATAAATATAT 2424

Qy        655 ATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATAT 714
              |||||  | |  |  |||  |  |||  |||||  |||  |  |  |||||||  |||||||||  ||  ||
Db       2423 ATATAAATATATATAAAAATATATATAAATATATAAATATATAAATATATATATATAAAT 2364

Qy        715 ATATAT 720
              ||||||
Db       2363 ATATAT 2358


RESULT 19
AC106873/c
LOCUS       AC106873               99596 bp    DNA     linear   PRI 31-JAN-2004
DEFINITION  Homo sapiens BAC clone RP11-374M7 from 7, complete sequence.

ACCESSION   AC106873
VERSION     AC106873.3  GI:18698877
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 99596)
  AUTHORS   Hillier,L.W., Fulton,R.S., Fulton,L.A., Graves,T.A., Pepin,K.H.,
            Wagner-McPherson,C., Layman,D., Maas,J., Jaeger,S., Walker,R.,
            Wylie,K., Sekhon,M., Becker,M.C., O'Laughlin,M.D., Schaller,M.E.,
            Fewell,G.A., Delehaunty,K.D., Miner,T.L., Nash,W.E., Cordes,M.,
            Du,H., Sun,H., Edwards,J., Bradshaw-Cordum,H., Ali,J., Andrews,S.,
            Isak,A., Vanbrunt,A., Nguyen,C., Du,F., Lamar,B., Courtney,L.,
            Kalicki,J., Ozersky,P., Bielicki,L., Scott,K., Holmes,A.,
            Harkins,R., Harris,A., Strong,C.M., Hou,S., Tomlinson,C.,
            Dauphin-Kohlberg,S., Kozlowicz-Reilly,A., Leonard,S., Rohlfing,T.,
            Rock,S.M., Tin-Wollam,A.M., Abbott,A., Minx,P., Maupin,R.,
            Strowmatt,C., Latreille,P., Miller,N., Johnson,D., Murray,J.,
            Woessner,J.P., Wendl,M.C., Yang,S.P., Schultz,B.R., Wallis,J.W.,
            Spieth,J., Bieri,T.A., Nelson,J.O., Berkowicz,N., Wohldmann,P.E.,
            Cook,L.L., Hickenbotham,M.T., Eldred,J., Williams,D., Bedell,J.A.,
            Mardis,E.R., Clifton,S.W., Chissoe,S.L., Marra,M.A., Raymond,C.,
            Haugen,E., Gillett,W., Zhou,Y., James,R., Phelps,K., Iadanoto,S.,
            Bubb,K., Simms,E., Levy,R., Clendenning,J., Kaul,R., Kent,W.J.,
            Furey,T.S., Baertsch,R.A., Brent,M.R., Keibler,E., Flicek,P.,
            Bork,P., Suyama,M., Bailey,J.A., Portnoy,M.E., Torrents,D.,
            Chinwalla,A.T., Gish,W.R., Eddy,S.R., McPherson,J.D., Olson,M.V.,
            Eichler,E.E., Green,E.D., Waterston,R.H. and Wilson,R.K.
  TITLE     The DNA sequence of human chromosome 7
  JOURNAL   Nature 424 (6945), 157-164 (2003)
  MEDLINE   22737999
   PUBMED   12853948
REFERENCE   2  (bases 1 to 99596)
  AUTHORS   Cotton,M. and Grewal,N.
  TITLE     The sequence of Homo sapiens BAC clone RP11-374M7
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 99596)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JAN-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 99596)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-FEB-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 99596)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-FEB-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   6  (bases 1 to 99596)
  AUTHORS   Wilson,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-JAN-2004) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Feb 17, 2002 this sequence version replaced gi:18158390.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0374M07
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.
```

```
          This sequence was finished as follows unless otherwise noted:
          all regions were double stranded, sequenced with an alternate
          chemistry, or covered by high quality data (i.e., phred quality >=
          30); an attempt was made to resolve all sequencing problems, such
          as compressions and repeats; all regions were covered by sequence
          from more than one subclone; and the assembly was confirmed by
          restriction digest.

          MAPPING INFORMATION:
          The sequence of this clone was established as part of a mapping and
          sequencing collaboration between the NHGRI Chromosome 7 Mapping
          Project (Eric D. Green, Director), John D. McPherson in the
          Department of Genetics (Washington University), and the Washington
          University Genome Sequencing Center. For additional information
          about the map position of this sequence, see
          http://www.nhgri.nih.gov/DIR/GTB/CHR7 , send
          mailto:egreen@nhgri.nih.gov , or see http://genome.wustl.edu

          SOURCE INFORMATION:
          The RPCI-11 human BAC library was made from the blood of one male
          donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
          Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
          approach for construction of bacterial artificial chromosome
          libraries. Genomics 51:1-8. The clone may be obtained either from
          Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
          and coworkers at http://www.chori.org
          VECTOR:  pBACe3.6.
FEATURES             Location/Qualifiers
     source          1. .99596
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7"
                     /clone="RP11-374M7"
                     /clone_lib="RPCI-11"
     repeat_region   5. .138
                     /rpt_family="Alu"
     repeat_region   399. .604
                     /rpt_family="L1"
     repeat_region   605. .919
                     /rpt_family="Alu"
     repeat_region   920. .1039
                     /rpt_family="L1"
     repeat_region   1045. .1325
                     /rpt_family="Alu"
     repeat_region   1330. .1444
                     /rpt_family="(GAAA)n"
     repeat_region   1552. .1572
                     /rpt_family="AT_rich"
     repeat_region   1699. .2061
                     /rpt_family="MER2_type"
     repeat_region   2062. .2284
                     /rpt_family="Alu"
     repeat_region   2285. .2351
                     /rpt_family="MER2_type"
     repeat_region   2949. .3225
                     /rpt_family="Alu"
     repeat_region   3237. .3413
                     /rpt_family="Alu"
     repeat_region   3438. .3548
                     /rpt_family="L1"
     repeat_region   3549. .3859
                     /rpt_family="Alu"
     repeat_region   3860. .3891
                     /rpt_family="L1"
     repeat_region   3936. .4009
                     /rpt_family="CR1"
     repeat_region   5320. .5362
                     /rpt_family="CR1"
     repeat_region   5365. .5446
                     /rpt_family="AcHobo"
     repeat_region   5491. .5521
                     /rpt_family="MIR"
     repeat_region   5526. .5571
                     /rpt_family="CR1"
     rRNA            complement(5771. .5860)
                     /product="rRNA-5S_rRNA-related"
     repeat_region   5771. .5860
                     /rpt_family="5S"
     repeat_region   6245. .6555
                     /rpt_family="Alu"
     repeat_region   6785. .6812
                     /rpt_family="AT_rich"
     repeat_region   7501. .7800
                     /rpt_family="Alu"
     repeat_region   7864. .7884
                     /rpt_family="AT_rich"
     repeat_region   8393. .8707
                     /rpt_family="Alu"
     repeat_region   9733. .10028
                     /rpt_family="Alu"
     repeat_region   10324. .10521
                     /rpt_family="MIR"
     repeat_region   11074. .11279
                     /rpt_family="L2"
     repeat_region   11338. .11642
                     /rpt_family="Alu"
     repeat_region   11690. .12111
                     /rpt_family="CT-rich"
     repeat_region   12118. .12392
                     /rpt_family="Alu"
     repeat_region   12924. .13569
                     /rpt_family="ERV1"
     repeat_region   14427. .14642
                     /rpt_family="MIR"
     misc_feature    14589. .15717
                     /note="CpG_island (%GC=70.4, o/e=0.91, #CpGs=143)"
     gene            14788. .15137
                     /gene="ST7"
     mRNA            14788. .15137
                     /gene="ST7"
     repeat_region   14814. .14913
                     /rpt_family="GC_rich"
     CDS             14987. .>15137
                     /gene="ST7"
                     /note="Homo sapiens suppression of tumorigenicity 7 (ST7),
                     transcript variant a, mRNA.; H_NH0374M07.1
                     This gene was based on gi(11761623 11761625)
                     Continues as H_RG114A06.2.1"
                     /codon_start=1
                     /product="unknown"
                     /protein_id="AAS07529.1"
                     /db_xref="GI:41473937"

  Query Match             8.1%;  Score 372;  DB 9;  Length 99596;
  Best Local Similarity   72.5%;  Pred. No. 6.2e-64;
  Matches  522;  Conservative    0;  Mismatches  195;  Indels    3;  Gaps    3;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              | ||| |||||||||||  ||||||||  ||||| ||||||||||||||||||| ||||||
Db      93552 AATTCGAGACCAGCCTGGTCAACATGGCTAAACCTCATCTCTACTAAAAATACAAAAATT 93493

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              || ||  | |||||  |   ||||||||||||||||||||  |||||| ||||||||||||
Db      93492 AGCCACACGTGGTGGCATTTGCCTGTAATCCCAGCTACTCAGGAGGCTGAGGCAGGAGAA 93433

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              |  ||||||||||  |  |  | || |   | |   ||| ||||  ||| ||||||||||
Db      93432 TTGCTTGAACCTG-GGAAGCAGAAGTTGCAGTGAGCCAAAATCACACCACTGCACTCCAG 93374

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  ||  ||||       | |     |  || ||||||||||      | ||||||| ||
Db      93373 CCTGGGTGACAAAGCAAGACTCCATTGCCAAAAAAAAAAAAAAAAATATATATATATAGAT 93314
```

```
Qy        240 ATATATATATATACACACACACACACATATATATATACACATATATATACACATATATAC 299
              | ||| ||| ||| | | | | | | | ||||| ||| | ||||| ||| |    |||||
Db      93313 AGATAGATAGATATAGATATAGATATAGATATAGATATAGATATAGATATATGCATATAT 93254

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACAT 359
              ||| | ||| ||||  | | ||||| | | | || ||||| || ||| |||||| || ||
Db      93253 ATACACACATATATGCATATATATACACATATATGCATATATATACACATATATGCATAT 93194

Qy        360 ATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATAT 419
              || || ||| ||||  | | ||||  | | | ||  ||| | | | ||| | |||||| |
Db      93193 ATATACACATATATGCATATATATGCATATATATGCATATATACACATATATACATATGT 93134

Qy        420 ACACATATATACACATATATACACATATACACATATATACACATATATACATATATACAC 479
                  ||||| | | ||||  ||| ||| ||||||||||||| |||||| || ||||||||
Db      93133 GTGTATATACATATATATGCACATATACACACATATATACATATATATGCACATATACAC 93074

Qy        480 ATATATATAATATACACACATATATATACACATATATACACACATATATACACATATATA 539
              | ||||||   ||| | |   | ||||||||| |||||  || |||||| ||||||||
Db      93073 ACATATAT-GCATATATATGCACATATACACACATATATGCATATATATGCACATATACA 93015

Qy        540 CACATATATATACACATATATACACATATATACATATATACACATATATATACATATATA 599
              |||||||| | | |||| ||| ||| | |||||||| || ||||||  |||||||
Db      93014 CACATATATGCATATATATGCACATATACACACATATATGCATATATATGCACATATACA 92955

Qy        600 CACATATATACATATATACACATATATACATATATACACACATATATACACATACATATA 659
              ||||||||| ||||||||  || |||||| | ||| |  |||||||  || | ||| |
Db      92954 CACATATATGCATATATATGCACATATACACACATATATGCATATATATGCACATATACA 92895

Qy        660 CACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATA 719
              |||| | ||  ||||| ||  | ||| ||| | ||||| |  |||||||| | |||||||
Db      92894 CACATATATGCATATATATGCACATATACACACATATATGCGTATATATGCACATATATA 92835


RESULT 20
BX901949
LOCUS       BX901949               38096 bp    DNA     linear   PRI 24-MAY-2004
DEFINITION  Human DNA sequence from clone WI2-86554F10 on chromosome X,
            complete sequence.
ACCESSION   BX901949
VERSION     BX901949.9  GI:47601709
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 38096)
  AUTHORS   Whitehead,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAY-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On May 24, 2004 this sequence version replaced gi:46559263.
            ------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome X, constructed by the Sanger Centre Chromosome X Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/ChrX
            WI2-86554F10 is from a Whitehead human fosmid library VECTOR:
            pEpiFos-5_Eco72I.
FEATURES              Location/Qualifiers
     source           1. .38096
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:9606"
                      /chromosome="X"
                      /clone="WI2-86554F10"
                      /clone_lib="Whitehead_fosmid"
ORIGIN

  Query Match             8.0%;  Score 371.4;  DB 9;  Length 38096;
  Best Local Similarity   77.9%;  Pred. No. 9.5e-64;
  Matches  540;  Conservative    0;  Mismatches  136;  Indels   17;  Gaps    7;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              |||||||||||||||||  |||  |||||||||||||||||| |||||||||| |||||
Db      35292 AGTTCAAGACCAGCCTGGCCAATGTGGTGAAACCCCATCTCAACTAAAAATACAAAAATC 35351

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   ||| |||||  |   ||||||||||||||||||||| |||||| ||||||||||||
Db      35352 AGCTGGGCGTGGTGGCAGGTGCCTGTAATCCCAGCTACTTGGGAGGCTGAGGCAGGAGAA 35411

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGC-AAGATCATGCCATTGCACTCCA 178
              |  ||||||||||  || ||| ||||||| | ||||  ||||    ||||||||| || |
Db      35412 TTGCTTGAACCTG--GGAGGTAGAGGTTGCAGTGAGCGGAGATTGCACCATTGCATTCTA 35469

Qy        179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCAT---TTTATATAT 235
              |||  || ||| | || |||||   |||| | |||.||| | |   ||   | |||||||
Db      35470 GCCTGGGTAACGAAAGTGAAATTCTATCTAAAAAACAAATATATATATACATATATATAT 35529

Qy        236 ATATATATATATATATACACACACACACAT---ATATATATACACATATATATACACA 292
              ||| |  | | | | |       | | ||| ||   |||||| ||| ||||| ||| | |
Db      35530 ATACACGTGTGTGTGTGTGTATATACATATACCATATATACACATATATACATATATA 35589

Qy        293 TATATACATATATACACATATATACACATATATATACACATACATATGTACACATATATA 352
              |||||||||||||||||||||||||||||||||||| ||| ||||| | || || ||| ||| |
Db      35590 CATATACATATATACACATATATACACATATACATATACATATACATATATACACATACA 35649

Qy        353 TACACATATGTATACACATATATACACATATATACACACATATATACACATATATACACA 412
              |||||||| || | | ||| ||| ||| ||||| |||||| || ||| |||| |
Db      35650 TACACATATATACATATATACATATACACATATATACACATGTACACATATAT------A 35703

Qy        413 CATATATACACATATATACACATATATACACATATACACATATATACACATATATACATA 472
              |||||||||||||||||||||| ||||| ||||||| ||| |||||||||| ||||| |
Db      35704 CATATATACACATATATACACACATATATACATATATACACATATACACATAGATACACA 35763

Qy        473 TATACACATATATATAATATACACACATATATATACACATATATACACACATATATACAC 532
              |||||| |||||||  | ||| | |||||||| ||| ||| ||| ||| ||||| |||
Db      35764 TATACAGATATATA-CACATATATACATATATACACATATACATATACATATATACACAT 35822

Qy        533 ATATATACACATATATATACACATATATACACATATATACATATATACACATATATATAC 592
              | ||| ||| ||||| ||| | | |||||||||||||||||| ||||||||||||||| |
Db      35823 ACATACACATATATACATATATACATATACACATATATACACATATACACATATATACAT 35882

Qy        593 ATATATACACATATATACATATATACACATATATACATATATACACACATATATACACAT 652
              ||||| |||||||||||||||||||||||||| ||||||| ||| | ||| | | | ||
Db      35883 ATATACACACATATATACATATATACACATATACACATATACACATAGATACACATATAT 35942

Qy        653 ACATATACACACACATAGATATACATATATATA 685
              ||||||| ||||| ||| |||      | | |||
Db      35943 ACATATATACACATATATATAATTTAAAAAATA 35975
```

```
RESULT 21
BX324168_0/c
WPCOMMENT
  Sequence split into 6 fragments  LOCUS BX324168 Accession BX324168
     Fragment Name      Begin     End
     BX324168_0             1     110000
     BX324168_1        100001     210000
     BX324168_2        200001     310000
     BX324168_3        300001     410000
     BX324168_4        400001     510000
     BX324168_5        500001     600124
LOCUS       BX324168               600124 bp    DNA     linear   HTG 13-JUN-2004
DEFINITION  Homo sapiens chromosome X clone XXyac-89H7, WORKING DRAFT SEQUENCE,
            31 unordered pieces.
ACCESSION   BX324168
VERSION     BX324168.8  GI:37777001
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 600124)
  AUTHORS   McLay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 20, 2003 this sequence version replaced gi:37650784.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: yX89H7
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 591433 bases at least Q40
            Consensus quality: 594175 bases at least Q30
            Consensus quality: 595562 bases at least Q20
            Insert size: 597124; sum-of-contigs
            Quality coverage: 6.16x in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 31 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    20165: contig of 20165 bp in length
            *    20166    20265: gap of 100 bp
            *    20266    44493: contig of 24228 bp in length
            *    44494    44593: gap of 100 bp
            *    44594    61605: contig of 17012 bp in length
            *    61606    61705: gap of 100 bp
            *    61706    67907: contig of 6202 bp in length
            *    67908    68007: gap of 100 bp
            *    68008    74311: contig of 6304 bp in length
            *    74312    74411: gap of 100 bp
            *    74412    79880: contig of 5469 bp in length
            *    79881    79980: gap of 100 bp
            *    79981    88557: contig of 8577 bp in length
            *    88558    88657: gap of 100 bp
            *    88658   112320: contig of 23663 bp in length
            *   112321   112420: gap of 100 bp
            *   112421   166393: contig of 53973 bp in length
            *   166394   166493: gap of 100 bp
            *   166494   227100: contig of 60607 bp in length
            *   227101   227200: gap of 100 bp
            *   227201   239356: contig of 12156 bp in length
            *   239357   239456: gap of 100 bp
            *   239457   242526: contig of 3070 bp in length
            *   242527   242626: gap of 100 bp
            *   242627   319013: contig of 76387 bp in length
            *   319014   319113: gap of 100 bp
            *   319114   327413: contig of 8300 bp in length
            *   327414   327513: gap of 100 bp
            *   327514   343545: contig of 16032 bp in length
            *   343546   343645: gap of 100 bp
            *   343646   391808: contig of 48163 bp in length
            *   391809   391908: gap of 100 bp
            *   391909   426746: contig of 34838 bp in length
            *   426747   426846: gap of 100 bp
            *   426847   440337: contig of 13491 bp in length
            *   440338   440437: gap of 100 bp
            *   440438   460971: contig of 20534 bp in length
            *   460972   461071: gap of 100 bp
            *   461072   465425: contig of 4354 bp in length
            *   465426   465525: gap of 100 bp
            *   465526   488190: contig of 22665 bp in length
            *   488191   488290: gap of 100 bp
            *   488291   497966: contig of 9676 bp in length
            *   497967   498066: gap of 100 bp
            *   498067   517262: contig of 19196 bp in length
            *   517263   517362: gap of 100 bp
            *   517363   567911: contig of 50549 bp in length
            *   567912   568011: gap of 100 bp
            *   568012   582171: contig of 14160 bp in length
            *   582172   582271: gap of 100 bp
            *   582272   584652: contig of 2381 bp in length
            *   584653   584752: gap of 100 bp
            *   584753   586940: contig of 2188 bp in length
            *   586941   587040: gap of 100 bp
            *   587041   589956: contig of 2916 bp in length
            *   589957   590056: gap of 100 bp
            *   590057   592460: contig of 2404 bp in length
            *   592461   592560: gap of 100 bp
            *   592561   597736: contig of 5176 bp in length
            *   597737   597836: gap of 100 bp
            *   597837   600124: contig of 2288 bp in length.
FEATURES             Location/Qualifiers
     source          1. .600124
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /clone="XXyac-89H7"
                     /clone_lib="Human-ICRF-YAC"
     misc_feature    1. .20165
                     /note="assembly_fragment:00356
                     fragment_chain:1"
     misc_feature    20266. .44493
                     /note="assembly_fragment:00589
                     fragment_chain:1"
     misc_feature    44594. .61605
                     /note="assembly_fragment:02258
                     fragment_chain:1"
     misc_feature    61706. .67907
                     /note="assembly_fragment:05479
                     fragment_chain:1"
     misc_feature    68008. .74311
                     /note="assembly_fragment:00274
                     fragment_chain:1"
     misc_feature    74412. .79880
                     /note="assembly_fragment:06012
                     fragment_chain:1"
     misc_feature    79981. .88557
                     /note="assembly_fragment:06208
                     fragment_chain:1"
     misc_feature    88658. .112320
                     /note="assembly_fragment:06075
                     fragment_chain:2"
     misc_feature    112421. .166393
```

```
                     /note="assembly_fragment:02111
                     fragment_chain:2"
     misc_feature    166494. .227100
                     /note="assembly_fragment:00544
                     fragment_chain:2"
     misc_feature    227201. .239356
                     /note="assembly_fragment:03674
                     fragment_chain:2"
     misc_feature    239457. .242526
                     /note="assembly_fragment:04283
                     fragment_chain:2"
     misc_feature    242627. .319013
                     /note="assembly_fragment:01306
                     fragment_chain:2"
     misc_feature    319114. .327413
                     /note="assembly_fragment:02688
                     fragment_chain:2"
     misc_feature    327514. .343545
                     /note="assembly_fragment:00658
                     fragment_chain:3"
     misc_feature    343646. .391808
                     /note="assembly_fragment:04025
                     fragment_chain:3"
     misc_feature    391909. .426746
                     /note="assembly_fragment:01126
                     fragment_chain:3"
     misc_feature    426847. .440337
                     /note="assembly_fragment:06618
                     fragment_chain:3"
     misc_feature    440438. .460971
                     /note="assembly_fragment:03859
                     fragment_chain:3"
     misc_feature    461072. .465425
                     /note="assembly_fragment:03896
                     fragment_chain:3"
     misc_feature    465526. .488190
                     /note="assembly_fragment:06799
                     fragment_chain:4"
     misc_feature    488291. .497966
                     /note="assembly_fragment:02743
                     fragment_chain:4"
     misc_feature    498067. .517262
                     /note="assembly_fragment:06602
                     fragment_chain:4"
     misc_feature    517363. .567911
                     /note="assembly_fragment:02958
                     fragment_chain:5"
     misc_feature    568012. .582171
                     /note="assembly_fragment:05773
                     fragment_chain:5"
     misc_feature    582272. .584652
                     /note="assembly_fragment:00466"
     misc_feature    584753. .586940
                     /note="assembly_fragment:02122"
     misc_feature    587041. .589956
                     /note="assembly_fragment:04085"

  Query Match             8.0%;  Score 371.4;  DB 2;  Length 110000;
  Best Local Similarity   77.9%;  Pred. No. 8e-64;
  Matches  540;  Conservative    0;  Mismatches  136;  Indels   17;  Gaps    7;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              |||||||||||||||||  |||  |||||||||||||||||| |||||||||| |||||
Db      82809 AGTTCAAGACCAGCCTGGCCAATGTGGTGAAACCCCATCTCAACTAAAAATACAAAAATC 82750

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   ||| |||||  |   ||||||||||||||||||| ||||||| ||||||||||||
Db      82749 AGCTGGGCGTGGTGGCAGGTGCCTGTAATCCCAGCTACTTGGGAGGCTGAGGCAGGAGAA 82690

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGC-AAGATCATGCCATTGCACTCCA 178
              |  ||||||||||  || ||| |||||||| | ||||  ||||    ||||||||| || |
Db      82689 TTGCTTGAACCTG--GGAGGTAGAGGTTGCAGTGAGCGGAGATTGCACCATTGCATTCTA 82632

Qy        179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAGGCAT---TTTATATAT 235
              |||  || ||| | || |||||   |||| | ||| ||| | |    ||   | |||||||
Db      82631 GCCTGGGTAACGAAAGTGAAATTCTATCTAAAAAACAAATATATATATACATATATATAT 82572

Qy        236 ATATATATATATATATACACACACACACAT---ATATATATACACATATATATACACA 292
              ||| |  | | | | |      | | ||| ||   |||||||| ||| ||||| ||| | |
Db      82571 ATACACGTGTGTGTGTGTGTATATACATATACCATATATACACATATATACATATATA 82512

Qy        293 TATATACATATATACACATATATACACATATATATACACATACATATGTACACATATATA 352
              ||||||||||||||||||||||||||||||||||| ||| ||||| | || || ||| ||| |
Db      82511 CATATACATATATACACATATATACACATATACATATACATATACATATATACACATACA 82452

Qy        353 TACACATATGTATACACATATATACACATATATACACACATATATACACATATATACACA 412
              |||||||| || | | ||| ||| ||| ||||| |||||| || ||| ||||      |
Db      82451 TACACATATATACATATATACATATACACATATATACACATGTACACATATAT------A 82398

Qy        413 CATATATACACATATATACACATATATACACATATACACATATATACACATATATACATA 472
              |||||||||||||||||||| ||||| ||||||| ||| |||||||||| ||||| |
Db      82397 CATATATACACATATATACACACATATATACATATATACACATATACACATAGATACACA 82338

Qy        473 TATACACATATATATAATATACACACATATATATACACATATATACACACATATATACAC 532
              |||||| |||||||  | ||| | |||||||||| ||| ||| ||| ||| ||||| |||
Db      82337 TATACAGATATATA-CACATATATACATATATACACATATACATATACATATATACACAT 82279

Qy        533 ATATATACACATATATATACACATATATACACATATATACATATATACACATATATATAC 592
              | ||| ||| ||||| ||| | | |||||||||||||||||| ||||||||||||||| |
Db      82278 ACATACACATATATACATATATACATATACACATATATACACATATACACATATATACAT 82219

Qy        593 ATATATACACATATATACATATATACACATATATACATATATACACACATATATACACAT 652
              ||||| ||||||||||||||||||||||||||| ||||||| ||| | ||| | | | ||
Db      82218 ATATACACACATATATACATATATACACATATACACATATACACATAGATACACATATAT 82159

Qy        653 ACATATACACACACATAGATATACATATATATA 685
              ||||||| ||||| ||| |||     | | |||
Db      82158 ACATATATACACATATATATAATTTAAAAAATA 82126


RESULT 22
AC097326
LOCUS       AC097326              252512 bp    DNA     linear   HTG 15-JUN-2002
DEFINITION  Pan troglodytes clone RP43-10G9, WORKING DRAFT SEQUENCE, 14
            unordered pieces.
ACCESSION   AC097326
VERSION     AC097326.3  GI:21397202
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 252512)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
```

```
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 252512)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-OCT-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 252512)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JUN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jun 13, 2002 this sequence version replaced gi:21392442.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: ZUAK
                Center clone name: RP43-10G9
            -------------- Summary Statistics
                Sequencing vector: Plasmid; M77789
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 241130 bases at least Q40
                Consensus quality: 243666 bases at least Q30
                Consensus quality: 245673 bases at least Q20
                Estimated insert size: 259407; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 7.4x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
            * NOTE: This sequence may represent more than one clone.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 14 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    55034: contig of 55034 bp in length
            *    55035    55134: gap of unknown length
            *    55135    92628: contig of 37494 bp in length
            *    92629    92728: gap of unknown length
            *    92729   119376: contig of 26648 bp in length
            *   119377   119476: gap of unknown length
            *   119477   142089: contig of 22613 bp in length
            *   142090   142189: gap of unknown length
            *   142190   165691: contig of 23502 bp in length
            *   165692   165791: gap of unknown length
            *   165792   184091: contig of 18300 bp in length
            *   184092   184191: gap of unknown length
            *   184192   198445: contig of 14254 bp in length
            *   198446   198545: gap of unknown length
            *   198546   211127: contig of 12582 bp in length
            *   211128   211227: gap of unknown length
            *   211228   223003: contig of 11776 bp in length
            *   223004   223103: gap of unknown length
            *   223104   234877: contig of 11774 bp in length
            *   234878   234977: gap of unknown length
            *   234978   239452: contig of 4475 bp in length
            *   239453   239552: gap of unknown length
            *   239553   246173: contig of 6621 bp in length
            *   246174   246273: gap of unknown length
            *   246274   250219: contig of 3946 bp in length
            *   250220   250319: gap of unknown length
            *   250320   252512: contig of 2193 bp in length.
FEATURES             Location/Qualifiers
     source          1. .252512
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="RP43-10G9"
ORIGIN

  Query Match             8.0%;  Score 370.4;  DB 2;  Length 252512;
  Best Local Similarity   72.5%;  Pred. No. 1.1e-63;
  Matches  535;  Conservative    0;  Mismatches  196;  Indels    7;  Gaps    4;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAATTA 60
              ||||  |||||||||||| |||||||||||||||| |||||||||||| ||||||||||
Db     249133 AGTTTGAGACCAGCCTGACCAACATGGTGAAACCCCGTCTCTACTAAAACTACAAAATTA 249192

Qy         61 GACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAAT 120
              | ||||| ||||| | ||||||||||  ||||||||| ||| || | |||| ||||||
Db     249193 GCCAGGCGTGGTGGTGCACGCCTGTAATATCAGCTACTTGGGAAGCTGGGGCAAGAGAAT 249252

Qy        121 CACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCCAG 179
              ||||||| |  | || |||||||| ||| | ||| |||||||   ||| ||||||||||
Db     249253 AACTTGAAGC--CGGGAGGTGGAGGCTGCAGTGAGCCAAGATCGCACCACTGCACTCCAG 249310

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  | |||||||||| |||   ||||||  | ||  ||        |     |||| |
Db     249311 CCTGGACAACAAGAGCAAAACTCCATCTCGAATAATTAATTTATTTATAATACTATAATT 249370

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              |     ||  |  |   |    |||   |  |||         |  | | | |||
Db     249371 AACCGGGCATGGAGGCGTGTGCCTGTATAGCCAGCTACTTGGGAGGCTGAAGAAAGATAA 249430

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACAT 359
              |  |  || | |||| | ||| ||||| | ||| |||| ||   |||||||| | ||
Db     249431 TCACTTGAACCTGTATATATATAGATATATAGATATATATATATCTATATATATATATAT 249490

Qy        360 ATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATAT 419
              || || | | ||||||| | ||||||| | | ||||||| | ||||||| | | ||||||
Db     249491 ATATAGATATATATATATATATATATATATATATATATATATATATATATATATATAT 249550

Qy        420 ACACATATATACACATATATACACATATACACATATATACACATATATACATATATACAC 479
              | | ||||||| |  |||||| | ||||| | ||||||| |  |||||| ||||||| |
Db     249551 ATATATATATATATCTATATATATATATATATATATATGTATATATATATATATAT 249610

Qy        480 ATATATATA---ATATACACACATATATATACACATATATACACACATATATACACATAT 536
              |||||||||   ||||| | | |||||||| | |||||| | | ||||||| | ||||
Db     249611 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 249670

Qy        537 ATACACATATATATACACATATATACACATATATACATATATACACA-TATATATACATA 595
              ||| | ||||||||| | |||||||| | |||||||| |||||||| | | |||||||| |||
Db     249671 ATATATATATATATATATATATATATATATATATATATATATATATAGTATATATATATA 249730

Qy        596 TATACACATATATACATATATACACATATATACATATATACACACATATATACACATACA 655
              |||| | ||||||| ||||||| | ||||||| ||||||| | | ||||||| | ||| |
Db     249731 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 249790

Qy        656 TATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATA 715
```

```
                |||| | | ||| |||||||||||||||| | ||||||||| | ||||||||||||||
Db     249791 TATATTTATATATATATATACATATATATATATATATATATATTTTTATATATGTATATC 249850

Qy        716 TATATGCTCCAGAGTTCA 733
              | | |  | |    ||||
Db     249851 TTTCTCTTTCCAGATTCA 249868


RESULT 23
AC123545/c
LOCUS       AC123545               170646 bp    DNA     linear   HTG 25-JUL-2003
DEFINITION  Pan troglodytes clone RP43-165N9, WORKING DRAFT SEQUENCE, 10
            ordered pieces.
ACCESSION   AC123545
VERSION     AC123545.3  GI:33235839
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 170646)
  AUTHORS   Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
            Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S., Cariaga,K.,
            Chu,G., Coleman,B., Coleman,H., Engle,J., Granite,S., Guan,X.,
            Gupta,J., Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Hu,P.,
            Hurle,B., Idol,J.R., Karlins,E., Kwong,P., Laric,P., Lee-Lin,S.-Q.,
            Legaspi,R., Maduro,Q.L., Maduro,V.B., Margulies,E.H., Masiello,C.,
            Maskeri,B., McDowell,J., Paguirigan,C., Pearson,R., Portnoy,M.E.,
            Prasad,A., Reddix-Dugue,N., Schandler,K., Schueler,M.G., Shah,K.,
            Sison,C., Stantripop,S., Thomas,J.W., Thomas,P.J., Tsipouri,V.,
            Vogt,J.L., Wetherby,K.D., Young,A. and Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 170646)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAY-2002) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE   3  (bases 1 to 170646)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUL-2003) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     On Jul 25, 2003 this sequence version replaced gi:24414279.
            -------------- Genome Center
                 Center: NIH Intramural Sequencing Center
                 Center code: NISC
                 Web site: http://www.nisc.nih.gov
                 Contact: nisc_zoo@nhgri.nih.gov
            -------------- Project Information
                 Center project name: cmz
                 Center clone name: 165N09

            The sequence data in this record represents an 'enhanced'
            version of a Phase 2 submission. Specifically, the indicated
            order and orientation of each sequence contig has been
            established using one or more of the following: read-pair
            data from individual subclones, overlaps with neighboring
            clones, alignment with available reference sequence (e.g.,
            human), and/or confirmation by PCR testing. In addition,
            the sequence assembly is based on at least 8X average
            coverage in Q20 bases and has been reviewed to rule out
            gross misassemblies, the low-quality ends of sequence
            contigs have been trimmed away, and each base is associated
            with a Phrap-derived quality score.
            -------------- Summary Statistics
                 Sequencing vector: plasmid; n/a; 100% of reads
                 Chemistry: Dye-terminator Big Dye; 100% of reads
                 Assembly program: Phrap; version 0.990319
                 Consensus quality: 167110 bases at least Q40
                 Consensus quality: 168559 bases at least Q30
                 Consensus quality: 169363 bases at least Q20
                 Insert size: 152000; agarose-fp
                 Insert size: 169746; sum-of-contigs
                 Quality coverage: 11.31x in Q20 bases; agarose-fp
                 Quality coverage: 10.13x in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1     6384: contig of 6384 bp in length
            *    6385     6484: gap of unknown length
            *    6485    37411: contig of 30927 bp in length
            *   37412    37511: gap of unknown length
            *   37512    46351: contig of 8840 bp in length
            *   46352    46451: gap of unknown length
            *   46452    51816: contig of 5365 bp in length
            *   51817    51916: gap of unknown length
            *   51917    67976: contig of 16060 bp in length
            *   67977    68076: gap of unknown length
            *   68077   114680: contig of 46604 bp in length
            *  114681   114780: gap of unknown length
            *  114781   121651: contig of 6871 bp in length
            *  121652   121751: gap of unknown length
            *  121752   125931: contig of 4180 bp in length
            *  125932   126031: gap of unknown length
            *  126032   128356: contig of 2325 bp in length
            *  128357   128456: gap of unknown length
            *  128457   170646: contig of 42190 bp in length.
FEATURES             Location/Qualifiers
     source          1. .170646
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="RP43-165N9"
                     /clone_lib="RP43"
     misc_feature    1. .6384
                     /note="assembly_fragment
                     missing T7 clone end on 5' end of insert"
     misc_feature    6485. .37411
                     /note="assembly_fragment"
     misc_feature    37512. .46351
                     /note="assembly_fragment"
     misc_feature    46452. .51816
                     /note="assembly_fragment"
     misc_feature    51917. .67976
                     /note="assembly_fragment"
     misc_feature    68077. .114680
                     /note="assembly_fragment"
     misc_feature    114781. .121651
                     /note="assembly_fragment"
     misc_feature    121752. .125931
                     /note="assembly_fragment"
     misc_feature    126032. .128356
                     /note="assembly_fragment"
     misc_feature    128457. .170646
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
ORIGIN

  Query Match             8.0%;  Score 368.2;  DB 2;  Length 170646;
  Best Local Similarity   74.9%;  Pred. No. 3.2e-63;
  Matches  543;  Conservative    0;  Mismatches  168;  Indels   14;  Gaps    6;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||| |||||||||||  ||| || | |||||||||||| ||| ||||| || ||||||
Db     103362 AGTTCGAGACCAGCCTGGCCAATATAGCGAAACCCCATCTGTACCAAAAACACAAAAATT 103303

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
```

```
Db     103302 AGCTGGGTGTGGTGACACATGCCTGTAATCCCAGGTACTCAGGAGGCTAAGGCAGGAGAA 103243

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179

Db     103242 TTGCTTGAACTGGGAGGCAGAGGTGGT-----GAGCCAAGATTGCACCAGTGTACTCCAG 103188

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239

Db     103187 GCTTGGTGAC-AGAGTGAAACTCCATCTCAAAAACAAACAAACAAAATACATATACATAA 103129

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATA- 298

Db     103128 ATATATATACACATATATATACATATATAAATATATATACCTATATATACATATATACAT 103069

Qy        299 ----CATATATACACATATATACACATATATATACACATACATATGTACACATATATATA 354

Db     103068 ATATAATATATATACATATATACATATATATA-TTATATATACATATATAAATATACA 103010

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414

Db     103009 TATATAAATATACATATAAAAATATATACATATAAAAATATATACATGTATAAATATACA 102950

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474

Db     102949 TGTATAAATATACATATATAAATATATACATATATACATATATAAATATACATATATAAA 102890

Qy        475 TACACATATATATAATATACACACATATATATACACATATATACACACATATATACACAT 534

Db     102889 TATATACATATATAA-ATATACATGTATAAATATACATATATAAATATATACATATATAA 102831

Qy        535 ATATACACATATATATACACATATATACACATATATACATATATACACATATATATACAT 594

Db     102830 ATATACATATATAAATATATAAATATATAAATATGTACAAATATATACATATATAAATAT 102771

Qy        595 ATATACACATATATACATATATACACATATATACATATATACACACATATATACACATAC 654

Db     102770 ATATACACATATATAGATATATAAATATATATACATATATACATATATAAATATATATAC 102711

Qy        655 ATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATAT 714

Db     102710 ATATATAAATATATATACATATATAAATATACATATATAAATATATATACATATATAAAT 102651

Qy        715 ATATA 719

Db     102650 ATGTA 102646
```

```
RESULT 24
HS323B6
LOCUS       HS323B6               112762 bp    DNA     linear   PRI 05-MAR-2003
DEFINITION  Human DNA sequence from clone RP3-323B6 on chromosome Xq11-12,
            complete sequence.
ACCESSION   Z83841
VERSION     Z83841.1  GI:1871528
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 112762)
  AUTHORS   McDougall,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAR-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Mar 10, 1997 this sequence version replaced gi:1772950.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome X, constructed by the Sanger Centre Chromosome X Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/ChrX
            RP3-323B6 is from the library RPCI-3 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pCYPAC2.
FEATURES             Location/Qualifiers
     source          1. .112762
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="RZPD:RPCIP704B06323"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="q11-12"
                     /clone="RP3-323B6"
                     /clone_lib="RPCI-3"
ORIGIN
```

```
  Query Match             7.9%;  Score 367;  DB 9;  Length 112762;
  Best Local Similarity   67.1%;  Pred. No. 5.9e-63;
  Matches  599;  Conservative    0;  Mismatches  280;  Indels   14;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC---AAAA 57

Db      39093 AGTTCAAGACCAGCTTGGCTAACACGGTGAAACCCCATCTCTACTATAAATACAAAAAAA 39152

Qy         58 TTAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAG 117

Db      39153 TTAGCCAGGCATGGTGGCATGCGCCTGTAATTCCAGTTGTTTGCGAGGCTGAGGCAGAAG 39212

Qy        118 AATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGA-GCAAGATCATGCCATTGCACTC 176

Db      39213 AATCGCTTGAACC--CAGGAGGTGGAGCTTGCAGTGAGGTGAGATCGCACCACTGCACTT 39270

Qy        177 CAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATA 236

Db      39271 CAGCCTGGATGAACAAA---AGAATTGAATTAATACTATTGCTTTTCACATATATATGTA 39327

Qy        237 TATATATATATATATACACACACACACATATATATATACACATATATATACACATATA 296

Db      39328 TGTGTATATATATATATGTGTGTGTGTGTATATATGTATATATATACATATACACATA 39387

Qy        297 TACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACA 356

Db      39388 TGTATATATATACATATATACACATATGTATATATATACATATATACACATATGTATATA 39447

Qy        357 CATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATA 416

Db      39448 TATACATATATACACATATGTATATATATACATATATACACATATGTATACATATACATA 39507

Qy        417 TATACACATATATACACATATATACACATATACACATATATACACATATATACATATATA 476
```

```
Db      39508 TATACACATATGTATATATATATACATATATACACATATGTATACATATACATATATACA 39567

Qy        477 CACATATATATAATATACACACATATATATACACATATATACACACATATATACACATAT 536
              || || |      ||| | | |||||||||| |||  |||| | | ||||||||||||||||
Db      39568 CATATGT-----ATACATATACATATATACACATGTATACATATACATATATACACATAT 39622

Qy        537 ATACACATATATATACACATATATACACATATATACATATATACACATATATATACATAT 596
              || ||||||| ||| | | | |||   ||| ||| | ||||| ||| |||| || | ||
Db      39623 GTATACATATACATATATACACATATGTATACATATACATATATACACATATGTATACAT 39682

Qy        597 ATACACATATATACATATATACACATATATACATATATACACACATATATACACATACAT 656
              ||||| ||||| |||||| || ||||||| | ||||| ||| |  |||| | | | | ||
Db      39683 ATACATATATACACATATGTATACATATACATATATACACATATGTATACATATACATAT 39742

Qy        657 ATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATAT 716
              ||||||| |  || | ||| | ||||||||||||||| ||||| |||| |||| || | ||
Db      39743 ATACACATATGTATACATATACATATATACACATATGTATACATATACATATATACACAT 39802

Qy        717 ATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAG 776
              ||||   |      |  |||         ||  | | ||    |  | ||| |   | | |
Db      39803 ATATATACATATATGTATATATATACACATATAAATATATATATGTATATACACACATAT 39862

Qy        777 AGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAA 836
              |  | | |     | |      | |   | || |      |      || |||       |
Db      39863 ATATACATATATATACACATATAAATATATATATGTATATACACACATATATATACACAT 39922

Qy        837 ATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACT 889
              |||.||| || | ||   | |  ||| | ||   |  ||  ||    ||  ||
Db      39923 ATATATACATATATATGTATATATATATATTTTTTTTTTGAGATGAAGTCT 39975


RESULT 25
AC108483/c
LOCUS       AC108483              163976 bp    DNA     linear   PRI 29-MAR-2003
DEFINITION  Homo sapiens chromosome 16 clone RP11-170P24, complete sequence.
ACCESSION   AC108483
VERSION     AC108483.3  GI:29366946
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 163976)
  AUTHORS   DOE Joint Genome Institute, Stanford Human Genome Center and Los
            Alamos National Laboratory.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 163976)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JAN-2002) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 163976)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-2002) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   4  (bases 1 to 163976)
  AUTHORS   DOE Joint Genome Institute, Stanford Human Genome Center and Los
            Alamos National Laboratory.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAR-2003) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Mar 29, 2003 this sequence version replaced gi:19073802.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center and Los Alamos
            National Laboratory
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 100% of Sequence;
            Estimated Total Number of Errors is 0.
FEATURES             Location/Qualifiers
     source          1. .163976
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone="RP11-170P24"
ORIGIN

  Query Match             7.9%;  Score 365.4;  DB 9;  Length 163976;
  Best Local Similarity   72.7%;  Pred. No. 1.1e-62;
  Matches  527;  Conservative    0;  Mismatches  191;  Indels    7;  Gaps    4;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC--AAAAT 58
              || |  |||||| ||||   |||||||||||||  || |||||  |||||||||  |||||
Db     146516 AGATTGAGACCATCCTGGCTAACATGGTGAAAGTCCGTCTCTCCTAAAAATACAAAAAAT 146457

Qy         59 TAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGA 118
              |||  ||   |||||  ||  ||||||| ||||||||||   |||||  |||||||||||
Db     146456 TAGCTAGTTGTGGTGGCACGTGCCTGTAGTCCCAGCTACTCGAGAGGCTGAGGCAGGAGA 146397

Qy        119 ATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCA 178
              ||  | ||||||    | ||  | || |   | |   ||||||  |||||  ||||||||
Db     146396 ATAGCCTGAACC-AAGGAGGCGGAAGTTGCAGTGAGCCAAGATTGTGCCACCGCACTCCA 146338

Qy        179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATA 238
              | ||  |||||||| | | | |           | || || || ||      | |||| | | |
Db     146337 G-CCTGGCAACAACAACAACAACAACAACAACAACAACAACAAATATATATATACACACA 146279

Qy        239 TATATATATATATACACACACACACATATATATATACACATATATATACACATATATA 298
              | | | | ||||| | | ||||||| ||||||| | ||||||||||||||||| | | |
Db     146278 CACACACACATATATATATACACACATATATATACACACACATATATATACACACACACA 146219

Qy        299 CATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACA 358
              |||||||  ||| | | | | ||||| | ||||| | |||| ||||||  | | | ||| |
Db     146218 TATATATATACACACACATATATATACACACACACATATATATACACACACACACACATA 146159

Qy        359 TATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATA 418
              ||| || ||||| | | | | ||||||||||||||||||| ||| | | ||| | |||||
Db     146158 TATATACACACACACACATATATATATACACACATATATATACACACACACATATATATA 146099

Qy        419 TACACATATATACACATATATACACATATACACATATATACACATATATACATATATACA 478
              |||||  | | | | ||||||||||| ||| | ||| | ||||| ||||| ||| | |||
Db     146098 CACACACACACATATATATATACACACATATATATACACACACACATATATATACACACA 146039

Qy        479 CATATATATA---ATATACACACATATATATACACATATATACACACATATATACACATA 535
              || | | |||   |||||||||||||||||||||||| | | ||| | ||||||||||| |
Db     146038 CACACACATATATATATACACACATATATATACACACACACACATATATATATACACACA 145979

Qy        536 TATACACATATATATACACATATATACACATATATACATATATACACATATATATACATA 595
              |||| | | | | | ||| |||||||||| | | |||||||| | | | | | | |||||
Db     145978 TATATATACACACACACATATATATACACACACACACATATATATATACACACACACATA 145919

Qy        596 TATACACATATATACATATATACACATATATACATATATACACACATATATACACATACA 655
              |||| | | | | | |||||||||| | ||| ||||| ||||||| ||||| | | |||
Db     145918 TATATATACACACATATATATACACACACATATATATACACACACACATATATATACACA 145859

Qy        656 TATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATA 715
              | ||| | | ||| | | |||||||||||||||| | |||||  |||| | |   | | |
Db     145858 CACACATATATATATACACACATATATATACACACACATATATATATACACACACACA 145799

Qy        716 TATAT 720
              |||||
Db     145798 TATAT 145794


RESULT 26
AC022166
LOCUS       AC022166              163998 bp    DNA     linear   PRI 21-JUL-2001
DEFINITION  Homo sapiens chromosome 16 clone RP11-525K10, complete sequence.
ACCESSION   AC022166
VERSION     AC022166.7  GI:14993688
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
```

```
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1  (bases 1 to 163998)
  AUTHORS  DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE    Direct Submission
  JOURNAL  Unpublished
REFERENCE  2  (bases 1 to 163998)
  AUTHORS  DOE Joint Genome Institute.
  TITLE    Direct Submission
  JOURNAL  Submitted (26-JAN-2000) Production Sequencing Facility, DOE Joint
           Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE  3  (bases 1 to 163998)
  AUTHORS  DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE    Direct Submission
  JOURNAL  Submitted (07-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell
           Drive, Walnut Creek, CA 94598, USA
REFERENCE  4  (bases 1 to 163998)
  AUTHORS  DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE    Direct Submission
  JOURNAL  Submitted (21-JUL-2001) DOE Joint Genome Institute, 2800 Mitchell
           Drive, Walnut Creek, CA 94598, USA
COMMENT    On Jul 21, 2001 this sequence version replaced gi:14327778.
           Draft Sequence Produced by DOE Joint Genome Institute
           www.jgi.doe.gov
           Finishing Completed at Stanford Human Genome Center
           www-shgc.stanford.edu
           Quality: Phrap Quality >=40 99.8% of Sequence;
           Estimated Total Number of Errors is 0.2.
           STS Content:
           WI-15858 G24147
           SHGC-79420 G52213.
FEATURES             Location/Qualifiers
     source          1. .163998
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone="RP11-525K10"
ORIGIN

  Query Match             7.9%;  Score 365.4;  DB 9;  Length 163998;
  Best Local Similarity   72.7%;  Pred. No. 1.1e-62;
  Matches  527;  Conservative    0;  Mismatches  191;  Indels    7;  Gaps    4;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC--AAAAT 58
              ||  |  |||||| ||||   ||||||||||||  || |||||  ||||||||||  |||||
Db     130747 AGATTGAGACCATCCTGGCTAACATGGTGAAAGTCCGTCTCTCCTAAAAATACAAAAAAT 130806

Qy         59 TAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGA 118
              |||  ||   |||||  ||  |||||||  ||||||||||||   |||||  ||||||||||
Db     130807 TAGCTAGTTGTGGTGGCACGTGCCTGTAGTCCCAGCTACTCGAGAGGCTGAGGCAGGAGA 130866

Qy        119 ATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCA 178
              ||  | ||||||    | ||  | ||  |   | |   ||||||  |||||  ||||||||
Db     130867 ATAGCCTGAACC-AAGGAGGCGGAAGTTGCAGTGAGCCAAGATTGTGCCACCGCACTCCA 130925

Qy        179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATA 238
              | || |||||||| | | | |       | || || || ||     | |||| | | |
Db     130926 G-CCTGGCAACAACAACAACAACAACAACAACAACAACAACAAATATATATATACACACA 130984

Qy        239 TATATATATATATACACACACACACATATATATATACACATATATATACACATATATA 298
              | | | | ||||| | | ||||||| ||||||| | ||||||||||||||||| | | |
Db     130985 CACACACACATATATATATACACACATATATATACACACACATATATATACACACACACA 131044

Qy        299 CATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACA 358
              ||||||| ||| | | | | ||||| | ||||| | |||| |||||| | | | ||| |
Db     131045 TATATATATACACACACATATATATACACACACACATATATATACACACACACACACATA 131104

Qy        359 TATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATA 418
              ||| || ||||| | | | | ||||||||||||||||||| ||| | | ||| | |||||
Db     131105 TATATACACACACACACATATATATATACACACATATATATACACACACACATATATATA 131164

Qy        419 TACACATATATACACATATATACACATATACACATATATACACATATATACATATATACA 478
              |||||  |  |  |  |||||||||||  ||| |  |||  |  |||||  |||||  |||  |  |||
Db     131165 CACACACACACATATATATATACACACATATATATACACACACACATATATATACACACA 131224

Qy        479 CATATATATA---ATATACACACATATATATACACATATATACACACATATATACACATA 535
              ||  |  |  |||   ||||||||||||||||||||||||  |  |  |||  |  ||||||||||  |
Db     131225 CACACACATATATATATACACACATATATATACACACACACACATATATATATACACACA 131284

Qy        536 TATACACATATATATACACATATATACACATATATACATATATACACATATATATACATA 595
              ||||  |  |  |  |  |||  ||||||||||  |  |  |||||||||  |  |  |  |  |||||
Db     131285 TATATATACACACACACATATATATACACACACACACATATATATATACACACACACATA 131344

Qy        596 TATACACATATATACATATATACACATATATACATATATACACACATATATACACATACA 655
              ||||  |  |  |  |  ||||||||||  |  |||  |||||  |||||||  |||||  |  |  |||
Db     131345 TATATATACACACATATATATACACACACATATATATACACACACACATATATATACACA 131404

Qy        656 TATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATA 715
              |  |||  |  |  |||  |  |  ||||||||||||||||  |  |||||  ||||  |  |  |  |  |  |
Db     131405 CACACATATATATATACACACATATATATACACACACATATATATATACACACACACACA 131464

Qy        716 TATAT 720
              |||||
Db     131465 TATAT 131469


RESULT 27
AC092863/c
LOCUS       AC092863               153201 bp    DNA     linear   HTG 11-DEC-2002
DEFINITION  Homo sapiens chromosome 12 clone RP11-355N1, WORKING DRAFT
            SEQUENCE, 5 unordered pieces.
ACCESSION   AC092863
VERSION     AC092863.10  GI:26449042
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 153201)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
```

```
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 153201)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-AUG-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 153201)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-DEC-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Dec 11, 2002 this sequence version replaced gi:19172705.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Drafting Center Code: BCM
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: HDJU
                Center clone name: RP11-355N1
            -------------- Summary Statistics
                Sequencing vector: Plasmid;
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 150350 bases at least Q40
                Consensus quality: 150698 bases at least Q30
                Consensus quality: 151119 bases at least Q20
                Estimated insert size: 144092; sum-of-contigs estimation
                Quality coverage: 14x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2895: contig of 2895 bp in length
            *     2896     2995: gap of unknown length
            *     2996    16175: contig of 13180 bp in length
            *    16176    16275: gap of unknown length
            *    16276    56442: contig of 40167 bp in length
            *    56443    56542: gap of unknown length
            *    56543    99843: contig of 43301 bp in length
            *    99844    99943: gap of unknown length
            *    99944   153201: contig of 53258 bp in length.
FEATURES             Location/Qualifiers
     source          1. .153201
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="12"
                     /clone="RP11-355N1"
ORIGIN

  Query Match             7.9%;  Score 364.2;  DB 2;  Length 153201;
  Best Local Similarity   74.5%;  Pred. No. 2e-62;
  Matches  543;  Conservative    0;  Mismatches  168;  Indels   18;  Gaps    6;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAATTA 60
              ||||  ||||||  ||||  ||||||||  ||||||||||| |||||       ||||||
Db       4255 AGTTTGAGACCATCCTGGCCAACATGGCAAAACCCCATCTTTACTA-------AAAATTG 4203

Qy         61 GACAGGCATGGTGATACACGCCTG-----TAATCCCAGCTACTTCGGAGGCCGAGGCAGG 115
              | | ||| ||||||  ||| |||||     ||||||||||||||  |||| |  ||||| |
Db       4202 GTCGGGCGTGGTGACACATGCCTGTAATCTAATCCCAGCTACTCAGGAGCCTAAGGCATG 4143

Qy        116 AGAATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCA-AGATCATGCCATTGCAC 174
              ||||||  ||||||||||   || ||||||||||||  | |||||  | |    |||| || ||
Db       4142 AGAATCGCTTGAACCTG--GGAGGTGGAGGTTGCAGTGAGCAGATAATGGGCCACTGTAC 4085

Qy        175 TCCAGCCCAGGCAAC--AAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATA 232
              |||||||  ||  ||  |  |      |||||    |  ||| |||| | ||| || ||
Db       4084 TCCAGCCTGGGTGACAGAGCAAGACTCTGTCAAAAAAACAAACAAAACATGCAGTTCATC 4025

Qy        233 TATATATATATATATATATACACACACACACATATATATATACACATATATATACACA 292
              | |   |  || ||||  | | |     | | |   | | ||| | || | | || | |
Db       4024 TTTCCCTCAGTAGATATCGAAAGAGTGGGATAAAAGGAAAAATATATATGTGTGTATATA 3965

Qy        293 TATATACATATATACACATATATACACATATATATACACATACATATGTACACATATATA 352
              |||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |||||||
Db       3964 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 3905

Qy        353 TACACATATGTATACACATATATACACATATATACACACATATATACACATATATACACA 412
              || | |||| |||| | ||||||| | ||||||| | | ||||||| | ||||||| | |
Db       3904 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 3845

Qy        413 CATATATACACATATATACACATATATACACATATACACATATATACACATATATACATA 472
              ||||||| | ||||||| | ||||||| | ||||| | ||||||| | ||||||| |||
Db       3844 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 3785

Qy        473 TATACACATATATATA-ATATACACACATATATATACACATATATACACACATATATACA 531
              |||| | ||||||||| ||||| | | ||||||||| | ||||||| | | ||||||| |
Db       3784 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 3725

Qy        532 CATATATACACATATATATACACATATATACACATATATACATATATACACATATATATA 591
              ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||||||||
Db       3724 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 3665

Qy        592 CATATATACACATATATACATATATACACATATATACATATATACACACATATATACACA 651
              ||||||| | ||||||| ||||||| | ||||||| ||||||| | | ||||||| | |
Db       3664 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 3605

Qy        652 TACATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTA 711
              || ||||| | | | ||| ||||| |||||||| | |||||||| |||||||| ||
Db       3604 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 3545

Qy        712 TATATATAT 720
              |||||||||
Db       3544 TATATATAT 3536


RESULT 28
AC092694/c
LOCUS       AC092694              151549 bp    DNA     linear   PRI 07-MAR-2002
DEFINITION  Homo sapiens 3q BAC RP11-172A10 (Roswell Park Cancer Institute
            Human BAC Library) complete sequence.
ACCESSION   AC092694
VERSION     AC092694.6  GI:18449730
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 151549)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
```

Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., He,X., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M., Neal,D., Nelson,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Naylor,S.L., Weinstock,G. and Gibbs,R.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 151549)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (20-JUL-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 151549)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (31-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 4 (bases 1 to 151549)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (01-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 5 (bases 1 to 151549)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (07-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Jan 31, 2002 this sequence version replaced gi:17977520.
INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE:Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL:
http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

QUALSTAT-REPORT.

FEATURES	Location/Qualifiers
source	1. .151549
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="3"
	/clone="RP11-172A10"
repeat_region	137. .646
	/rpt_family="L1M4c"
repeat_region	696. .1026
	/rpt_family="L1M4c"
repeat_region	complement(1065. .1364)
	/rpt_family="AluSx"
repeat_region	1365. .1966
	/rpt_family="L1M1"
repeat_region	1967. .2309
	/rpt_family="MLT1A1"
repeat_region	2310. .3436
	/rpt_family="L1M1"
repeat_region	3437. .3473
	/rpt_family="(TTTA)n"
repeat_region	complement(3474. .3802)
	/rpt_family="LTR17"
repeat_region	complement(3803. .4094)
	/rpt_family="HERV17"
repeat_region	complement(8054. .9082)
	/rpt_family="HERV17"
repeat_region	complement(8719. .9163)
	/rpt_family="Harlequin"
repeat_region	9166. .9611
	/rpt_family="L1M1"
repeat_region	complement(9605. .9798)
	/rpt_family="L1PA16"
repeat_region	complement(9805. .9857)
	/rpt_family="L1"
repeat_region	9880. .10350
	/rpt_family="L1M1"
repeat_region	10786. .11108
	/rpt_family="L1MA2"
repeat_region	11109. .11492
	/rpt_family="L1MA2"
repeat_region	11612. .12041
	/rpt_family="L1M1"
repeat_region	12053. .12130
	/rpt_family="(TA)n"
misc_feature	12121. .13100
	/function="unresolved tandem repeat"
repeat_region	12131. .12462
	/rpt_family="(CATATA)n"
repeat_region	12466. .13020

```
                     /rpt_family="(TATATG)n"
     repeat_region   13040. .13100
                     /rpt_family="(TATATG)n"
     repeat_region   complement(13101. .13387)
                     /rpt_family="AluY"
     repeat_region   13388. .13466
                     /rpt_family="(TATATG)n"
     repeat_region   13469. .13599
                     /rpt_family="L1M1"
     repeat_region   13611. .14584
                     /rpt_family="L1M1"
     repeat_region   14669. .15274
                     /rpt_family="L1M1"
     repeat_region   15275. .17378
                     /rpt_family="L1M1"
     repeat_region   17379. .17684
                     /rpt_family="AluSc"
     repeat_region   17685. .18217
                     /rpt_family="L1M1"
     repeat_region   18200. .18827
                     /rpt_family="L1"
     repeat_region   18829. .19120
                     /rpt_family="AluJb"
     repeat_region   19123. .19483
                     /rpt_family="L1"
     repeat_region   complement(19478. .19591)
                     /rpt_family="L1"
     repeat_region   20015. .20050
                     /rpt_family="(TA)n"
     repeat_region   20531. .20559
                     /rpt_family="AT_rich"
     repeat_region   20570. .20647
                     /rpt_family="AluSx"

  Query Match             7.9%;  Score 363.6;  DB 9;  Length 151549;
  Best Local Similarity   68.3%;  Pred. No. 2.6e-62;
  Matches  623;  Conservative    0;  Mismatches  274;  Indels   15;  Gaps    8;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              || ||||||||| ||||   |||| |||||||||||||||||||||||||||| ||||||
Db      13316 AGATCAAGACCATCCTGGCTAACACGGTGAAACCCCATCTCTACTAAAAATACAAAAATT 13257

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              || ||||||||||     | |||||| |||||||||||| |||||| ||||||||||||
Db      13256 AGCCAGGCATGGTGGCGGGCACCTGTAGTCCCAGCTACTTGGGAGGCTGAGGCAGGAGAA 13197

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGC-AAGATCATGCCATTGCACTCCA 178
              |  | ||||||||  || ||||||| |||| | |||| |||||   |||| |||||||||
Db      13196 TGGCCTGAACCTG--GGAGGTGGAGCTTGCAGTGAGCTAAGATGGCGCCACTGCACTCCA 13139

Qy        179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAGGCATTTTATATATATA 238
              |||  |||||| ||||| | | | |   |      |||||| | |      |  |||| ||||
Db      13138 GCCTGGGCAAC-AGAGCAAGAAGACTCCATCTCAAAAAATATATTATGTACATATGTATA 13080

Qy        239 TATATATATATATACACACACACACATATATATATACACATATATATACACATATATA 298
              | | | |||| |||   | | |||||||| |||| ||| | |  ||| |||| |||| ||
Db      13079 TGTGTCTATACATATGTATATACACACATGTATACATATATACTTATGTACATATATGTA 13020

Qy        299 CATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACA 358
              ||| ||| |||||| || | | ||||| | |  || |||  || |||||| |||| |||
Db      13019 TATACATATACATATGTATATACATATACATATGTATATACATATACATATGTATATACA 12960

Qy        359 TATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATA 418
              ||||||||    ||||| || ||| |||| | || ||||||||||||| | |    |||||
Db      12959 CATGTATAC----ATATATACTTATGTACATATATGTATACACATATACATATGTATATA 12904

Qy        419 TACACATATATACACATATATACACATATACACATATATACACATATATACATATATACA 478
              ||||||  ||||||||||||    |  || |||||| |||||||||||| | || |||| |
Db      12903 CACACATGTATACACATATACTTATGTACACACATGTATACACATATACATATGTATATA 12844

Qy        479 CATATATATA---ATATACACACATATATATACACATATATACACACATATATACACATA 535
              || | || ||   | ||| ||| || ||||||||||| || || ||||||||| | |  ||
Db      12843 CACACATGTATACACATATACATATGTATATACACACATGTATACACATATACATATGTA 12784

Qy        536 TATACACATATATATACACATATATACACATATATACATATATACACATATATATACATA 595
              |||||||| || |||||||||||| | |  |||||||| | ||  | | | |||||||||
Db      12783 TATACACACATGTATACACATATACATATGTATATACACACATGTATACACATATACATA 12724

Qy        596 TATACACATATATACATATATACACATATATACATATATACACACATATATACACATACA 655
              | || | | | |  |||| ||| ||| |||  ||||||||||||| ||||||||| |
Db      12723 TGTATATACACACATGTATACACATATACATATGTATATACACACATGTATACACATATA 12664

Qy        656 TATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATA 715
              |||   | | | | | ||  ||| | ||| ||||||| |||| || | | |||||| |
Db      12663 CATATGTATATACACACATGTATACACATATACATATGTATA--TACACACATGTATACA 12606

Qy        716 TATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGA 775
              |||| |    |  |  | |         |    |||   ||  ||| | | | | |
Db      12605 CATATACATATGTATATACACACATGTATACACATATACATATGTACATATATGTATATA 12546

Qy        776 GAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGA 835
              |  |  |||   ||  ||       |  |||| |    ||     || |||       |
Db      12545 CACATATACATATGTACATATATGTATATACATATATACTTATGTACATATATGTATATA 12486

Qy        836 AATACATAAATGTGTAATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTAACACTATATA 894
              |||  |  |||| ||   |  | | | |  | |  || ||| | |  ||    ||||||
Db      12485 CATATACATATGTATATACACCCATGTATACATATATACTTATGTACATATATGTATATA 12426

Qy        895 TATATATAGTGA 906
              ||||||| | |
Db      12425 CATATATACTTA 12414


RESULT 29
AC021193
LOCUS       AC021193               217985 bp    DNA     linear   PRI 21-FEB-2002
DEFINITION  Homo sapiens BAC clone RP11-774G5 from 4, complete sequence.
ACCESSION   AC021193
VERSION     AC021193.7  GI:15321574
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 217985)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
   PUBMED   9847074
REFERENCE   2  (bases 1 to 217985)
  AUTHORS   Paulson,E., Meyer,R., Elliott,G., Boyer,E. and Dixon,R.
  TITLE     The sequence of Homo sapiens BAC clone RP11-774G5
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 217985)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JAN-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 217985)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-AUG-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 217985)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JAN-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   6  (bases 1 to 217985)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-FEB-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
```

```
COMMENT     On Aug 28, 2001 this sequence version replaced gi:15022788.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0774G05
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. John D.
            McPherson, Department of Genetics, Washington University, St. Louis
            MO. For additional information about the map position of this
            sequence, see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries. Genomics 51:1-8. The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at http://www.chori.org
            VECTOR:  pBACe3.6

            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is AC079084.  Actual start of this
            clone is at base position 1 of RP11-774G5; actual end is at base
            position 217985 of RP11-774G5.
FEATURES              Location/Qualifiers
     source           1. .217985
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:9606"
                      /chromosome="4"
                      /map="4"
                      /clone="RP11-774G5"
                      /clone_lib="RPCI-11"
     repeat_region    1077. .1483
                      /rpt_family="MER2_type"
     repeat_region    1511. .1555
                      /rpt_family="AT_rich"
     repeat_region    1669. .1888
                      /rpt_family="MaLR"
     repeat_region    1889. .2198
                      /rpt_family="Alu"
     repeat_region    2172. .2191
                      /rpt_family="(CAAA)n"
     repeat_region    2192. .2219
                      /rpt_family="(CAAAA)n"
     repeat_region    2199. .2388
                      /rpt_family="MaLR"
     repeat_region    2668. .2947
                      /rpt_family="Alu"
     repeat_region    3099. .3125
                      /rpt_family="AT_rich"
     repeat_region    3100. .3462
                      /rpt_family="L1"
     repeat_region    3730. .3751
                      /rpt_family="AT_rich"
     repeat_region    3999. .4659
                      /rpt_family="L1"
     repeat_region    4460. .4551
                      /rpt_family="(TA)n"
     repeat_region    4698. .4897
                      /rpt_family="L1"
     repeat_region    4898. .5162
                      /rpt_family="L1"
     repeat_region    5377. .5405
                      /rpt_family="AT_rich"
     repeat_region    6918. .7257
                      /rpt_family="MaLR"
     repeat_region    8680. .8986
                      /rpt_family="MaLR"
     repeat_region    8989. .10092
                      /rpt_family="MER2_type"
     repeat_region    10216. .10313
                      /rpt_family="L1"
     repeat_region    11999. .12023
                      /rpt_family="AT_rich"
     repeat_region    12490. .12719
                      /rpt_family="MIR"
     repeat_region    13417. .13960
                      /rpt_family="L1"
     repeat_region    13945. .13969
                      /rpt_family="(A)n"
     repeat_region    14697. .15010
                      /rpt_family="Alu"
     repeat_region    14989. .15023
                      /rpt_family="AT_rich"
     repeat_region    16415. .16850
                      /rpt_family="MaLR"
     repeat_region    18007. .18299
                      /rpt_family="Alu"
     repeat_region    19150. .19172
                      /rpt_family="AT_rich"
     repeat_region    19177. .19375
                      /rpt_family="ERV1"
     repeat_region    19490. .19640
                      /rpt_family="ERV1"
     repeat_region    19679. .20025
                      /rpt_family="L1"
     repeat_region    20302. .20588
                      /rpt_family="ERV1"
     repeat_region    21525. .21574
                      /rpt_family="Mariner"
     repeat_region    22586. .22642
                      /rpt_family="(TG)n"
     repeat_region    22960. .22980
                      /rpt_family="AT_rich"
     repeat_region    23682. .23709
                      /rpt_family="(A)n"
     repeat_region    23892. .24010
                      /rpt_family="ERV1"
     repeat_region    24012. .24039
                      /rpt_family="AT_rich"
     repeat_region    24187. .24257
                      /rpt_family="(TTATA)n"
     repeat_region    24472. .24493
                      /rpt_family="(T)n"
     repeat_region    25473. .25523
                      /rpt_family="AT_rich"
     repeat_region    25899. .25993
                      /rpt_family="MIR"
     repeat_region    25994. .27011
                      /rpt_family="L1"
     repeat_region    26007. .26053
                      /rpt_family="AT_rich"
     repeat_region    26269. .26421
                      /rpt_family="T-rich"
     repeat_region    26617. .26647
                      /rpt_family="AT_rich"
```

```
     repeat_region   27102. .27425
                     /rpt_family="L1"
     repeat_region   27217. .27237
                     /rpt_family="AT_rich"
     repeat_region   27512. .27828
                     /rpt_family="L1"
     repeat_region   27595. .27630
                     /rpt_family="AT_rich"
     repeat_region   27923. .28308
                     /rpt_family="L1"
     repeat_region   28362. .29158
                     /rpt_family="L1"
     repeat_region   29254. .29527

  Query Match             7.7%;  Score 357.4;  DB 9;  Length 217985;
  Best Local Similarity   78.0%;  Pred. No. 4.1e-61;
  Matches  495;  Conservative    0;  Mismatches  131;  Indels    9;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC--AAAAT 58
              || || ||||||  |||   ||||||||||||||||| |||||||||||||| |  |||||
Db     146265 AGATCGAGACCATTCTGGCTAACATGGTGAAACCCCGTCTCTACTAAAAATTCCAAAAAT 146324

Qy         59 TAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGA 118
              ||| ||| | |||||    |||||||| ||||||| |||   ||||| ||||||||||||
Db     146325 TAGCCAGACGTGGTGGCGGGCGCCTGTAGTCCCAGCCACTCGAGAGGCTGAGGCAGGAGA 146384

Qy        119 ATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGC-AAGATCATGCCATTGCACTCC 177
              || | ||||||  | || || ||||  ||| | ||||  |||||  |||| ||||||||
Db     146385 ATGGCGTGAACC--CGGGAGGCGGAGCCTGCAGTGAGCTGAGATCGCGCCACTGCACTCC 146442

Qy        178 AGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATAT 237
              ||||  ||| || ||||||| |   | |||||  |||||||||||      | |||||||||
Db     146443 AGCCTGGGCGAC-AGAGCGAGACTCCGTCTCAAAAAAAAAAAAAAATATATATATATATAT 146501

Qy        238 ATATATATATATATACACACACACACACATATATATATACACATATATATACACATATAT 297
              | | |||||||||||||| | | | | ||| |||||||||| ||| ||||||| ||| ||||
Db     146502 ACACATATATATATACATATATATATACACATATATATATACACATATATATACACATAT 146561

Qy        298 ACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACAC 357
              | ||| ||| | ||| | | | ||| ||| | | ||||| || || | | ||||||||||
Db     146562 ATATACATATATATACACATATATACATACATATATACACATATATATACATATATACAC 146621

Qy        358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417
              |||| |||| ||||||||| ||| ||| | | | | ||||| | | ||||| ||| ||||
Db     146622 ATATATATATACATATATATACACATACATATATACATATATATACATATATACATATAT 146681

Qy        418 ATACACATATATACACATATATACACATATACACATATATACACATATATACATATATAC 477
              ||||| |||||||||| ||||| | | ||||||| ||||||||| ||||| ||||||||||
Db     146682 ATACATATATATACATATATACATATATATACATATATATACATATATACACATATATAC 146741

Qy        478 ACATATATATAATATACACACATATATATACACATATATACACACATATATACACATATA 537
              | |||||    ||||| ||| | |||||||||| ||||| ||| ||||||||||| |||||
Db     146742 ATATATA---CATATATACATACATATATACATATATACACATACATATATACATATATA 146798

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATACATATA 597
              |||  | ||||||| ||||||||| ||| ||| | ||||||||| | ||| ||| |||||
Db     146799 TACGTACATATATATACATATATATACACATACATATATATACATACATACATATATATA 146858

Qy        598 TACACATATATACATATATACACATATATACATAT 632
              || | |||||||| ||||||| | ||||||| || |
Db     146859 TATATATATATATATATATATATATATATATATGT 146893


RESULT 30
CR812471/c
LOCUS       CR812471               151424 bp    DNA     linear   HTG 27-SEP-2004
DEFINITION  Danio rerio clone CH211-237A4, *** SEQUENCING IN PROGRESS ***, 9
            unordered pieces.
ACCESSION   CR812471
VERSION     CR812471.2  GI:52748670
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 151424)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 27, 2004 this sequence version replaced gi:52694427.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zC237A4
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 147995 bases at least Q40
            Consensus quality: 148500 bases at least Q30
            Consensus quality: 148901 bases at least Q20
            Insert size: 150624; sum-of-contigs
            Insert size: 152063; 2.5% error; agarose-fp
            Quality coverage: 7.11x in Q20 bases; sum-of-contigs Quality
            coverage: 7.09x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    18105: contig of 18105 bp in length
            *    18106    18205: gap of 100 bp
            *    18206    20634: contig of 2429 bp in length
            *    20635    20734: gap of 100 bp
            *    20735    23916: contig of 3182 bp in length
            *    23917    24016: gap of 100 bp
            *    24017    35672: contig of 11656 bp in length
            *    35673    35772: gap of 100 bp
            *    35773    47596: contig of 11824 bp in length
            *    47597    47696: gap of 100 bp
            *    47697    63363: contig of 15667 bp in length
            *    63364    63463: gap of 100 bp
            *    63464   127690: contig of 64227 bp in length
            *   127691   127790: gap of 100 bp
            *   127791   133366: contig of 5576 bp in length
            *   133367   133466: gap of 100 bp
            *   133467   151424: contig of 17958 bp in length.
FEATURES             Location/Qualifiers
     source          1. .151424
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-237A4"
                     /clone_lib="CHORI-211"
     misc_feature    1. .18105
                     /note="assembly_fragment:00546
                     fragment_chain:1
                     clone_end:SP6
                     vector_side:left"
     misc_feature    18206. .20634
                     /note="assembly_fragment:00036
                     fragment_chain:1"
     misc_feature    20735. .23916
                     /note="assembly_fragment:00008
                     fragment_chain:1"
     misc_feature    24017. .35672
                     /note="assembly_fragment:00104
```

```
                     fragment_chain:2"
     misc_feature    35773. .47596
                     /note="assembly_fragment:00228
                     fragment_chain:2"
     misc_feature    47697. .63363
                     /note="assembly_fragment:00380
                     fragment_chain:2"
     misc_feature    63464. .127690
                     /note="assembly_fragment:01023
                     fragment_chain:2"
     misc_feature    127791. .133366
                     /note="assembly_fragment:00065"
     misc_feature    133467. .151424
                     /note="assembly_fragment:00782
                     clone_end:T7
                     vector_side:right"
ORIGIN

  Query Match             7.7%;  Score 357.2;  DB 2;  Length 151424;
  Best Local Similarity   65.2%;  Pred. No. 4.8e-61;
  Matches  468;  Conservative    0;  Mismatches  248;  Indels    2;  Gaps    2;

Qy        209 AGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACATA 268
              | ||||||||||||    || ||||||||||||||||||||||| | | | | | ||||
Db      21248 AAAAAAAAAAAAAGCATATTCATATATATATATATATATATATATATATATATATATA 21189

Qy        269 TATATATACACATATATATACACATATATACATATATACACATATATACACATATATATA 328
              |||||||| | ||||||||| | ||||||| ||||||| | ||||||| | |||||||||
Db      21188 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 21129

Qy        329 CACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACA 388
              | ||| |||| || | ||||||||| | |||| |||| | ||||||| | ||||||| |
Db      21128 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 21069

Qy        389 CACATATATACACATATATACACACATATATACACATATATACACATATATACACATATA 448
              | ||||||| | ||||||| | | ||||||| | ||||||| | ||||||| | |||||
Db      21068 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 21009

Qy        449 CACATATATACACATATATACATATATACACATATATATA-ATATACACACATATATATA 507
              | ||||||| | ||||||| ||||||| | ||||||||| ||||| | | |||||||||
Db      21008 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 20949

Qy        508 CACATATATACACACATATATACACATATATACACATATATATACACATATATACACATA 567
              | ||||||| | | ||||||| | ||||||| | ||||||||| | ||||||| | |||
Db      20948 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 20889

Qy        568 TATACATATATACACATATATATACATATATACACATATATACATATATACACATATATA 627
              |||| ||||||| | ||||||||| ||||||| | ||||||| ||||||| | |||||||
Db      20888 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 20829

Qy        628 CATATATACACACATATATACACATACATATACACACACATAGATATACATATATATACA 687
              ||||||| | | ||||||| | ||| ||||| | | | ||| ||||| |||||||||| |
Db      20828 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 20769

Qy        688 CATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGT 747
              |||||||||  ||||||||| ||||||||||
Db      20768 TATATATATATATATATATATATATATATATATANNNNNNNNNNNNNNNNNNNNNNNNN 20709

Qy        748 TGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACA 807

Db      20708 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 20649

Qy        808 TTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA 867
                          | |||         | ||| ||| || | ||   | |  ||| | ||
Db      20648 NNNNNNNNNNNNNTATATATATATATATATATATATATATATATATATATATATATATA 20589

Qy        868 -AGGTAGTTAAGAAGGTAACACTATATATATATATAGTGAAAGCAGTGTAAACCTAAA 924
               |  ||  ||   |  ||    |||||||||||||   | |    | || |  || |
Db      20588 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 20531


RESULT 31
HS141H5/c
'
LOCUS       HS141H5              121600 bp    DNA     linear   PRI 05-JUN-2003
DEFINITION  Human DNA sequence from clone RP6-141H5 on chromosome Xq22.1-23
            Contains the 3' end of the gene for a novel protein with von
            Willebrand factor type C domains, complete sequence.
ACCESSION   AL049176
VERSION     AL049176.3  GI:4808226
KEYWORDS    HTG; von Willebrand factor type C.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 121600)
  AUTHORS   Pearce,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JUN-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On May 12, 1999 this sequence version replaced gi:4495007.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome X, constructed by the Sanger Centre Chromosome X Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/ChrX
            RP6-141H5 is from the library RPCI-6 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pPAC4
            This sequence is the entire insert of clone RP6-141H5.
FEATURES             Location/Qualifiers
     source          1. .121600
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="RZPD:RPCIP709H05141"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="q22.1-23"
                     /clone="RP6-141H5"
                     /clone_lib="RPCI-6"
     repeat_region   1. .665
                     /note="L1P repeat: matches 2. .667 of consensus"
     repeat_region   825. .1034
                     /note="MER58A repeat: matches 2. .216 of consensus"
     repeat_region   1103. .1634
                     /note="L1ME repeat: matches 572. .1075 of consensus"
     repeat_region   1635. .2006
                     /note="MLT1A1 repeat: matches 1. .365 of consensus"
     repeat_region   2007. .2521
                     /note="L1ME repeat: matches 1075. .1707 of consensus"
     repeat_region   2526. .2715
```

```
                /note="L1MEc repeat: matches 1961. .2159 of consensus"
repeat_region   2761. .4193
                /note="Tigger1 repeat: matches 902. .2357 of consensus"
repeat_region   4195. .5277
                /note="L1PA16 repeat: matches 5057. .6133 of consensus"
repeat_region   5291. .6213
                /note="Tigger1 repeat: matches 1. .931 of consensus"
repeat_region   6576. .8350
                /note="L1MA8 repeat: matches 4428. .6279 of consensus"
repeat_region   8501. .11123
                /note="L1PA10 repeat: matches 3561. .6159 of consensus"
repeat_region   11119. .11399
                /note="L1M4 repeat: matches 2315. .2614 of consensus"
repeat_region   11408. .17541
                /note="L1PA6 repeat: matches 11. .6143 of consensus"
repeat_region   17545. .17592
                /note="24 copies 2 mer aa 77% conserved"
repeat_region   17716. .17752
                /note="L1PA16 repeat: matches 6117. .6153 of consensus"
repeat_region   18281. .18366
                /note="L1MB8 repeat: matches 6083. .6169 of consensus"
repeat_region   20394. .20549
                /note="MER97c repeat: matches 19. .177 of consensus"
repeat_region   20895. .21189
                /note="L2 repeat: matches 2428. .2709 of consensus"
repeat_region   21669. .21960
                /note="AluJo repeat: matches 1. .310 of consensus"
repeat_region   21977. .22111
                /note="MIR repeat: matches 107. .250 of consensus"
repeat_region   23514. .23619
                /note="53 copies 2 mer at 83% conserved"
repeat_region   23520. .23619
                /note="25 copies 4 mer atat 85% conserved"
repeat_region   23621. .23821
                /note="AluJo repeat: matches 92. .283 of consensus"
repeat_region   24261. .24306
                /note="L2 repeat: matches 2659. .2705 of consensus"
repeat_region   24403. .24543
                /note="MIR repeat: matches 116. .262 of consensus"
repeat_region   24579. .24823
                /note="L2 repeat: matches 2255. .2481 of consensus"
repeat_region   25586. .25873
                /note="MER34 repeat: matches 9. .297 of consensus"
repeat_region   25874. .26288
                /note="MSTB repeat: matches 1. .426 of consensus"
repeat_region   26289. .26372
                /note="MER34 repeat: matches 297. .380 of consensus"
repeat_region   26373. .26768
                /note="MSTA repeat: matches 1. .426 of consensus"
repeat_region   26769. .26932
                /note="MER34 repeat: matches 380. .543 of consensus"
repeat_region   27373. .27775
                /note="L2 repeat: matches 1058. .1475 of consensus"
repeat_region   28030. .28249
                /note="MIR repeat: matches 1. .243 of consensus"
gene            complement(28358. .117309)
                /gene="dA141H5.1"
mRNA            complement(join(28358. .30863,33838. .33927,35984. .36151,
                43096. .43305,48665. .48833,55156. .55223,74337. .74496,
                75887. .76032,114163. .114256,117197. .>117309))
                /gene="dA141H5.1"
                /product="dA141H5.1 (novel protein with von Willebrand
                factor type C domains)"
                /note="match: cDNAs: Em:AF039656
                match: ESTs: Em:AA302430 Em:AI246227 Em:H77289 Em:AA156742
                Em:AA724151 Em:AA843324 Em:AA057564 Em:AI337909 Em:N39848
                Em:AA555017 Em:AA135450 Em:N26819 Em:AI123481 Em:AI139649
                Em:AA018704 Em:R52421 Em:AA724185 Em:AI138345 Em:AA057563
                Em:AI004265 Em:AA150072 Em:AA244080 Em:AA854359
                Em:AI420184 Em:R59551 Em:W52860 Em:AA724504 Em:AA635439
                Em:AA018487 Em:Z42301 Em:AA327755 Em:C00874 Em:N43181
                Em:AA177053 Em:AA015694 Em:W96276 Em:AA029156 Em:AA463650
                Em:AA013040 Em:W28849 Em:AA918837"
                /evidence=not_experimental
polyA_signal    complement(28361. .28366)
                /gene="dA141H5.1"
polyA_site      complement(28384)
                /gene="dA141H5.1"
repeat_region   28960. .29007
                /note="24 copies 2 mer ca 100% conserved"
misc_feature    complement(28995. .29451)
                /gene="dA141H5.1"
                /note="match: GSS: Em:AQ442614"
misc_feature    29462. .29764
                /note="match: GSS: Em:AQ618076"
repeat_region   30162. .30217
                /note="28 copies 2 mer tg 89% conserved"
CDS             complement(join(30733. .30863,33838. .33927,35984. .36151,
                43096. .43305,48665. .48833,55156. .55223,74337. .74496,
                75887. .76032,114163. .114256,117197. .>117309))
                /gene="dA141H5.1"
                /note="continued from bA814C6.1.1 in Em:AL591489
                match: proteins: Tr:O57472 Sw:Q24025 Tr:O57465 Sw:Q91713"
                /codon_start=3
                /evidence=not_experimental
                /product="dA141H5.1 (novel protein with von Willebrand
                factor type C domains)"
                /protein_id="CAD92488.1"
                /db_xref="GI:31455411"
                /translation="SETYCMFQDKKYRVGERWHPYLEPYGLVYCVNCICSENGNVLCS
                RVRCPNVHCLSPVHIPHLCCPRCPEDSLPPVNNKVTSKSCEYNGTTYQHGELFVAEGL
                FQNRQPNQCTQCSCSVLSEAFTLILDASNLNSVSPWQEGNVYCGLKTCPKLTCAFPVS
                VPDSCCRVCRGDGELSWEHSDGDIFRQPANREARHSYHRSHYDPPPSRQAGGLSRFPG
                ARSHRGALMDSQQASGTIVQIVINNKHKHGQVCVSNGKTYSHGESWHPNLRAFGIVEC
                VLCTCNVTKQECKKIHCPNRYPCKYPQKIDGKCCKVCPGKKAKEELPGQSFDNKGYFC
                GEETMPVYESVFMEDGETTRKIALETERPPQVEVHVWTIRKGILQHFHIEKISKRMFE
                ELPHFKLVTRTTLSQWKIFTEGEAQISQMCSSRVCRTELEDLVKVLYLERSEKGHC"
misc_feature    complement(31798. .32170)
                /gene="dA141H5.1"
                /note="match: GSS: Em:AQ694976"
repeat_region   32245. .32729
                /note="L2 repeat: matches 1728. .2263 of consensus"
repeat_region   34978. .35165
                /note="MER5A repeat: matches 1. .189 of consensus"
misc_feature    35093. .35438
                /note="match: GSS: Em:AQ205639"
repeat_region   35552. .35734
                /note="MER5B repeat: matches 6. .177 of consensus"
misc_feature    37351. .37723
                /note="match: GSS: Em:B99326"
repeat_region   38099. .38585
                /note="LTR32 repeat: matches 15. .471 of consensus"
repeat_region   38934. .39248
                /note="L2 repeat: matches 2418. .2748 of consensus"
repeat_region   39760. .40064
                /note="AluJb repeat: matches 1. .301 of consensus"
repeat_region   40389. .40698
```

```
Query Match             7.7%;  Score 356.6;  DB 9;  Length 121600;
Best Local Similarity   74.7%;  Pred. No. 6.5e-61;
Matches  543;  Conservative    0;  Mismatches  169;  Indels   15;  Gaps    7;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAA--- 57
              || || |||||| ||||    ||||  |||||||||| |||||||||||||| |||||
Db      44776 AGATCGAGACCATCCTGGCTAACACAGTGAAACCCCGTCTCTACTAAAAATCCAAAAAAA 44717

Qy         58 --TTAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGG 115
                |||| | ||| |||||     | |||||| |||||||||||  |||||| ||||||||
Db      44716 AGTTAGCCGGGCGTGGTGGCGGGCACCTGTAGTCCCAGCTACTCTGGAGGCTGAGGCAGG 44657

Qy        116 AGAATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCAC 174
              |||||  | ||||| ||  || || |||| |||| | ||| | ||||   |||| |||||
Db      44656 AGAATGGCGTGAACTTG--GGAGGCGGAGCTTGCAGTGAGCCGAGATAGCGCCACTGCAC 44599

Qy        175 TCCAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATA 234
              ||||| |  ||| || ||||||| |   ||||||| |||||||| |      | ||||||
```

```
Db      44598 TCCAGTCTGGGCGAC-AGAGCGAGACTCCATCTCAAAAAAAAAATA----TATATATATA 44544

Qy        235 TATATATATATATATATACACACACACACACATATATATATACACATATATATACACATA 294
              || ||||||||| ||||||| | | | | | ||||||| ||| | | ||||||| | |||
Db      44543 TAGATATATATAGATATACAGATATATAGATATATATAGATATATAGATATATATAGATA 44484

Qy        295 TATACATATATACACATATATACACATATATATACACATACATATGTACACATATATATA 354
              |||| ||||||| | ||||||| | ||||||| | | ||| |||| || | ||||||| |
Db      44483 TATAGATATATATAGATATATAGATATATATAGATATATAGATATATATAGATATATAGA 44424

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              | ||||  ||| | | ||||| | | ||||| | | ||||| | | ||||| | | | |
Db      44423 TATATATAGATATATAGATATATATAGATATATAGATATATAGATATATATAGATATATA 44364

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              ||||| | | ||||| | ||||||| | ||||| | ||||||| | ||||||| |||||
Db      44363 GATATATATAGATATATAGATATATATAGATATATAGATATATAGATATATATAGATATA 44304

Qy        475 TACACATATATATAATATACACACATATATATACACATATATACACACATATATACACAT 534
              || |  |||| || ||||| | | ||||| ||| | ||| ||| | | ||||||| | ||
Db      44303 TAGATGTATAGAT-ATATATAGATATATAGATATATATAGATATATAGATATATAGATAT 44245

Qy        535 ATATACACATATATATACACATATATACACATATATACATATATACACATATATATACAT 594
              ||||| | ||||| ||| | ||| ||| | ||||| | ||||| | | ||| |||||
Db      44244 ATATAGATATATAGATATATATAGATATATATATAGATATATAGATATATAGATATATAG 44185

Qy        595 ATATACACATATATACATATATACA-CATATATACATATATACACACATATATACACATA 653
              ||||| | ||||||| ||||||| |  ||||||| ||||||| | | ||||| | | |||
Db      44184 ATATATAGATATATAGATATATAGATGATATATAGATATATATAGATATATAGATATATA 44125

Qy        654 CATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATA 713
              ||||| | | | ||||||||| ||| ||||| | ||||||| |  |||| |||| ||||
Db      44124 GATATATAGATATATAGATATATATAGATATATAGATATATAGATATATAGATATATATA 44065

Qy        714 TATATAT 720
              |||||||
Db      44064 GATATAT 44058


RESULT 32
CR788239/c
LOCUS       CR788239              127172 bp    DNA     linear   HTG 18-SEP-2004
DEFINITION  Danio rerio clone CH211-105M19, *** SEQUENCING IN PROGRESS ***, 26
            unordered pieces.
ACCESSION   CR788239
VERSION     CR788239.1  GI:52345234
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 127172)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zC105M19
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 116372 bases at least Q40
            Consensus quality: 119969 bases at least Q30
            Consensus quality: 122718 bases at least Q20
            Insert size: 124672; sum-of-contigs
            Insert size: 168914; 5.4% error; agarose-fp
            Quality coverage: 2.88x in Q20 bases; sum-of-contigs Quality
            coverage: 2.51x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 26 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     3262: contig of 3262 bp in length
            *    3263     3362: gap of 100 bp
            *    3363     5810: contig of 2448 bp in length
            *    5811     5910: gap of 100 bp
            *    5911     8259: contig of 2349 bp in length
            *    8260     8359: gap of 100 bp
            *    8360    11266: contig of 2907 bp in length
            *   11267    11366: gap of 100 bp
            *   11367    14728: contig of 3362 bp in length
            *   14729    14828: gap of 100 bp
            *   14829    18192: contig of 3364 bp in length
            *   18193    18292: gap of 100 bp
            *   18293    25841: contig of 7549 bp in length
            *   25842    25941: gap of 100 bp
            *   25942    28449: contig of 2508 bp in length
            *   28450    28549: gap of 100 bp
            *   28550    32988: contig of 4439 bp in length
            *   32989    33088: gap of 100 bp
            *   33089    37208: contig of 4120 bp in length
            *   37209    37308: gap of 100 bp
            *   37309    41754: contig of 4446 bp in length
            *   41755    41854: gap of 100 bp
            *   41855    48505: contig of 6651 bp in length
            *   48506    48605: gap of 100 bp
            *   48606    69764: contig of 21159 bp in length
            *   69765    69864: gap of 100 bp
            *   69865    77035: contig of 7171 bp in length
            *   77036    77135: gap of 100 bp
            *   77136    80534: contig of 3399 bp in length
            *   80535    80634: gap of 100 bp
            *   80635    85265: contig of 4631 bp in length
            *   85266    85365: gap of 100 bp
            *   85366    87432: contig of 2067 bp in length
            *   87433    87532: gap of 100 bp
            *   87533    92877: contig of 5345 bp in length
            *   92878    92977: gap of 100 bp
            *   92978   100746: contig of 7769 bp in length
            *  100747   100846: gap of 100 bp
            *  100847   103064: contig of 2218 bp in length
            *  103065   103164: gap of 100 bp
            *  103165   106145: contig of 2981 bp in length
            *  106146   106245: gap of 100 bp
            *  106246   108396: contig of 2151 bp in length
            *  108397   108496: gap of 100 bp
            *  108497   114275: contig of 5779 bp in length
            *  114276   114375: gap of 100 bp
            *  114376   121458: contig of 7083 bp in length
            *  121459   121558: gap of 100 bp
            *  121559   124719: contig of 3161 bp in length
            *  124720   124819: gap of 100 bp
            *  124820   127172: contig of 2353 bp in length.
FEATURES             Location/Qualifiers
     source          1. .127172
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-105M19"
                     /clone_lib="CHORI-211"
     misc_feature    1. .3262
                     /note="assembly_fragment:00259
                     fragment_chain:1"
     misc_feature    3363. .5810
```

```
                     /note="assembly_fragment:00160
                     fragment_chain:1
                     clone_end:T7
                     vector_side:left"
     misc_feature    5911. .8259
                     /note="assembly_fragment:00081
                     fragment_chain:2"
     misc_feature    8360. .11266
                     /note="assembly_fragment:00224
                     fragment_chain:2"
     misc_feature    11367. .14728
                     /note="assembly_fragment:00278
                     fragment_chain:2"
     misc_feature    14829. .18192
                     /note="assembly_fragment:00208
                     fragment_chain:2"
     misc_feature    18293. .25841
                     /note="assembly_fragment:00560
                     fragment_chain:2"
     misc_feature    25942. .28449
                     /note="assembly_fragment:00192
                     fragment_chain:2"
     misc_feature    28550. .32988
                     /note="assembly_fragment:00144
                     fragment_chain:2"
     misc_feature    33089. .37208
                     /note="assembly_fragment:00176
                     fragment_chain:2"
     misc_feature    37309. .41754
                     /note="assembly_fragment:00343
                     fragment_chain:2"
     misc_feature    41855. .48505
                     /note="assembly_fragment:00443
                     fragment_chain:2"
     misc_feature    48606. .69764
                     /note="assembly_fragment:00603
                     fragment_chain:2"
     misc_feature    69865. .77035
                     /note="assembly_fragment:00417
                     fragment_chain:3"
     misc_feature    77136. .80534
                     /note="assembly_fragment:00367
                     fragment_chain:3"
     misc_feature    80635. .85265
                     /note="assembly_fragment:00391
                     fragment_chain:3"
     misc_feature    85366. .87432
                     /note="assembly_fragment:00132
                     fragment_chain:3"
     misc_feature    87533. .92877
                     /note="assembly_fragment:00321
                     fragment_chain:3"
     misc_feature    92978. .100746
                     /note="assembly_fragment:00479
                     fragment_chain:3"
     misc_feature    100847. .103064
                     /note="assembly_fragment:00049
                     fragment_chain:3"
     misc_feature    103165. .106145
                     /note="assembly_fragment:00241
                     fragment_chain:3"
     misc_feature    106246. .108396
                     /note="assembly_fragment:00090
                     fragment_chain:3"
     misc_feature    108497. .114275
                     /note="assembly_fragment:00299
                     fragment_chain:4"
     misc_feature    114376. .121458
                     /note="assembly_fragment:00518
                     fragment_chain:4"
     misc_feature    121559. .124719
                     /note="assembly_fragment:00110"
     misc_feature    124820. .127172
                     /note="assembly_fragment:00100
                     clone_end:SP6
                     vector_side:right"
ORIGIN

  Query Match             7.7%;  Score 356.6;  DB 2;  Length 127172;
  Best Local Similarity   64.9%;  Pred. No. 6.5e-61;
  Matches  454;  Conservative    0;  Mismatches  244;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db      42349 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 42290

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db      42289 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 42230

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db      42229 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 42170

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db      42169 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 42110

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db      42109 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 42050

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db      42049 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 41990

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db      41989 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 41930

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db      41929 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 41870

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||
Db      41869 TATATATATATATATNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 41810

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
                                                                  | |
Db      41809 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTTATA 41750

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885
              ||       | ||| ||| || | ||   | |  ||| | ||   ||  ||   |  ||
Db      41749 TATATATATATATATATATATATATATATATATATATATATATATTATATATATATATAT 41690

Qy        886 CACTATATATATATAGTGAAAGCAGTGTAAACCTAAA 924
                |||||||||||||   | |     | || |  || |
Db      41689 ATATATATATATATATATATATATATATATATATATA 41651


RESULT 33
CR753865
LOCUS       CR753865              282240 bp    DNA     linear   HTG 04-SEP-2004
DEFINITION  Danio rerio clone DKEY-172B2, WORKING DRAFT SEQUENCE, 17 unordered
            pieces.
ACCESSION   CR753865
VERSION     CR753865.4  GI:51890091
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 282240)
  AUTHORS   Burton,J.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (03-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 4, 2004 this sequence version replaced gi:51870573.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK172B2
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 274515 bases at least Q40
            Consensus quality: 275937 bases at least Q30
            Consensus quality: 276861 bases at least Q20
            Insert size: 280640; sum-of-contigs
            Insert size: 261465; 4.0% error; agarose-fp
            Quality coverage: 9.23x in Q20 bases; sum-of-contigs Quality
            coverage: 10.00x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 17 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     6899: contig of 6899 bp in length
            *     6900     6999: gap of 100 bp
            *     7000    20575: contig of 13576 bp in length
            *    20576    20675: gap of 100 bp
            *    20676    38280: contig of 17605 bp in length
            *    38281    38380: gap of 100 bp
            *    38381    76364: contig of 37984 bp in length
            *    76365    76464: gap of 100 bp
            *    76465    95438: contig of 18974 bp in length
            *    95439    95538: gap of 100 bp
            *    95539   114600: contig of 19062 bp in length
            *   114601   114700: gap of 100 bp
            *   114701   134153: contig of 19453 bp in length
            *   134154   134253: gap of 100 bp
            *   134254   167898: contig of 33645 bp in length
            *   167899   167998: gap of 100 bp
            *   167999   171913: contig of 3915 bp in length
            *   171914   172013: gap of 100 bp
            *   172014   176193: contig of 4180 bp in length
            *   176194   176293: gap of 100 bp
            *   176294   206625: contig of 30332 bp in length
            *   206626   206725: gap of 100 bp
            *   206726   209648: contig of 2923 bp in length
            *   209649   209748: gap of 100 bp
            *   209749   213682: contig of 3934 bp in length
            *   213683   213782: gap of 100 bp
            *   213783   230655: contig of 16873 bp in length
            *   230656   230755: gap of 100 bp
            *   230756   239434: contig of 8679 bp in length
            *   239435   239534: gap of 100 bp
            *   239535   241595: contig of 2061 bp in length
            *   241596   241695: gap of 100 bp
            *   241696   282240: contig of 40545 bp in length.
FEATURES             Location/Qualifiers
     source          1. .282240
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-172B2"
                     /clone_lib="DanioKey"
     misc_feature    1. .6899
                     /note="assembly_fragment:00130
                     fragment_chain:1"
     misc_feature    7000. .20575
                     /note="assembly_fragment:00501
                     fragment_chain:1"
     misc_feature    20676. .38280
                     /note="assembly_fragment:00282
                     fragment_chain:1"
     misc_feature    38381. .76364
                     /note="assembly_fragment:02629
                     fragment_chain:1"
     misc_feature    76465. .95438
                     /note="assembly_fragment:00723
                     fragment_chain:1"
     misc_feature    95539. .114600
                     /note="assembly_fragment:01220
                     fragment_chain:1"
     misc_feature    114701. .134153
                     /note="assembly_fragment:01489
                     fragment_chain:1"
     misc_feature    134254. .167898
                     /note="assembly_fragment:01785
                     fragment_chain:1"
     misc_feature    167999. .171913
                     /note="assembly_fragment:00062
                     fragment_chain:1"
     misc_feature    172014. .176193
                     /note="assembly_fragment:00105
                     fragment_chain:1"
     misc_feature    176294. .206625
                     /note="assembly_fragment:02143
                     fragment_chain:1"
     misc_feature    206726. .209648
                     /note="assembly_fragment:00050
                     fragment_chain:2"
     misc_feature    209749. .213682
                     /note="assembly_fragment:00082
                     fragment_chain:2"
     misc_feature    213783. .230655
                     /note="assembly_fragment:00958
                     fragment_chain:2"
     misc_feature    230756. .239434
                     /note="assembly_fragment:00171
                     fragment_chain:2"
     misc_feature    239535. .241595
                     /note="assembly_fragment:00042"
     misc_feature    241696. .282240
                     /note="assembly_fragment:03238.0"
ORIGIN

  Query Match             7.7%;  Score 355.4;  DB 2;  Length 282240;
  Best Local Similarity   43.6%;  Pred. No. 9.8e-61;
  Matches  564;  Conservative    0;  Mismatches  728;  Indels    2;  Gaps    2;

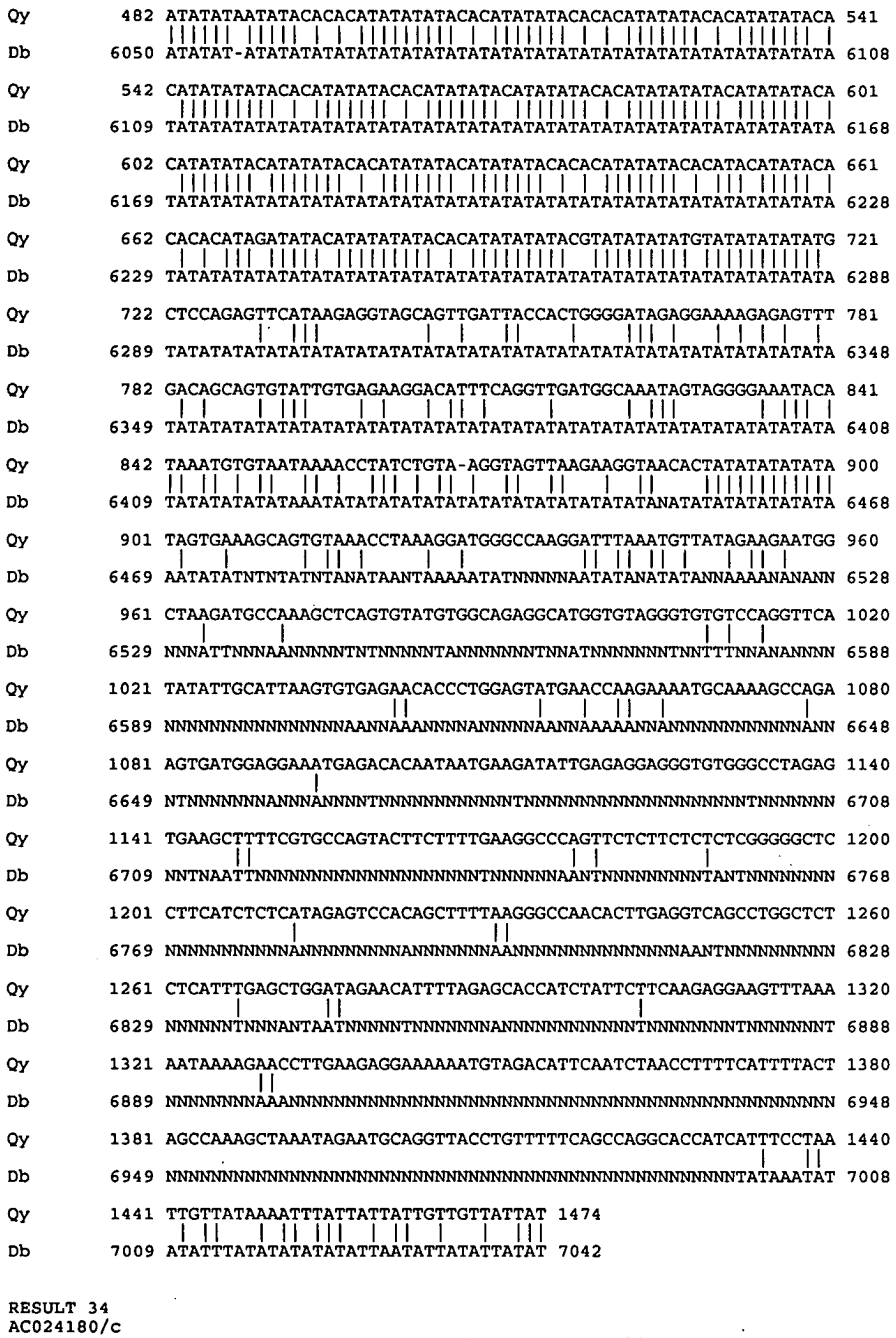
Qy        182 CAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATAT 241
              || |||| || |   ||||        | || || || ||     | ||||||||||||||
Db       5750 CATGCAAAAAAAAATAAATAAATAAAAATAATAATAATAATAATATATATATATATATAT 5809

Qy        242 ATATATATATACACACACACACATATATATATACACATATATATACACATATATACAT 301
              ||||||||||| | | | | | ||||||||||| | ||||||||| | ||||||| ||
Db       5810 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 5869

Qy        302 ATATACACATATATACACATATATATACACATACATATGTACACATATATATACACATAT 361
              ||||| | ||||||| | ||||||||| | ||| |||| || | ||||||||| | ||||
Db       5870 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 5929

Qy        362 GTATACACATATATACACATATATACACACATATATACACATATATACACACATATATAC 421
               |||| | ||||||| | ||||||| | | ||||||| | ||||||| | | |||||||
Db       5930 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 5989

Qy        422 ACATATATACACATATATACACATATACACATATATACACATATATACATATATACACAT 481
              | ||||||| | ||||||| | ||||| | ||||||| | ||||||| ||||||| | ||
Db       5990 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 6049
```



```
Qy       482 ATATATAATATACACACATATATATACACATATATACACACATATATACACATATATACA 541
Db      6050 ATATAT-ATATATATATATATATATATATATATATATATATATATATATATATATATA 6108

Qy       542 CATATATATACACATATATACACATATATACATATATACACATATATATACATATATACA 601
Db      6109 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 6168

Qy       602 CATATATACATATATACACATATATACATATATACACACATATATACACATACATATACA 661
Db      6169 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 6228

Qy       662 CACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATATG 721
Db      6229 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 6288

Qy       722 CTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTT 781
Db      6289 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 6348

Qy       782 GACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACA 841
Db      6349 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 6408

Qy       842 TAAATGTGTAATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTAACACTATATATATATA 900
Db      6409 TATATATATATAAATATATATATATATATATATATATATATATANATATATATATATATA 6468

Qy       901 TAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTATAGAAGAATGG 960
Db      6469 AATATATNTNTATNTANATAANTAAAAATATNNNNNAATATANATATANNAAAANANANN 6528

Qy       961 CTAAGATGCCAAAGCTCAGTGTATGTGGCAGAGGCATGGTGTAGGGTGTGTCCAGGTTCA 1020
Db      6529 NNNATTNNNAANNNNNTNTNNNNNTANNNNNNNTNNATNNNNNNNTNNTTTNNANANNNN 6588

Qy      1021 TATATTGCATTAAGTGTGAGAACACCCTGGAGTATGAACCAAGAAAATGCAAAAGCCAGA 1080
Db      6589 NNNNNNNNNNNNNNNAANNAAANNNNANNNNNAANNAAAAANNANNNNNNNNNNNANN 6648

Qy      1081 AGTGATGGAGGAAATGAGACACAATAATGAAGATATTGAGAGGAGGGTGTGGGCCTAGAG 1140
Db      6649 NTNNNNNNANNNANNNNTNNNNNNNNNNNNTNNNNNNNNNNNNNNNNNNNNTNNNNNN 6708

Qy      1141 TGAAGCTTTTCGTGCCAGTACTTCTTTTGAAGGCCCAGTTCTCTTCTCTCTCGGGGGCTC 1200
Db      6709 NNTNAATTNNNNNNNNNNNNNNNNNNNTNNNNNNAANTNNNNNNNNNTANTNNNNNNNN 6768

Qy      1201 CTTCATCTCTCATAGAGTCCACAGCTTTTAAGGGCCAACACTTGAGGTCAGCCTGGCTCT 1260
Db      6769 NNNNNNNNNNNANNNNNNNNNANNNNNNNAANNNNNNNNNNNNNNNAANTNNNNNNNNN 6828

Qy      1261 CTCATTTGAGCTGGATAGAACATTTTAGAGCACCATCTATTCTTCAAGAGGAAGTTTAAA 1320
Db      6829 NNNNNNTNNNANTAATNNNNNTNNNNNNNANNNNNNNNNNNNTNNNNNNNTNNNNNNT 6888

Qy      1321 AATAAAAGAACCTTGAAGAGGAAAAAATGTAGACATTCAATCTAACCTTTTCATTTTACT 1380
Db      6889 NNNNNNNNAAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 6948

Qy      1381 AGCCAAAGCTAAATAGAATGCAGGTTACCTGTTTTTCAGCCAGGCACCATCATTTCCTAA 1440
Db      6949 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTATAAATAT 7008

Qy      1441 TTGTTATAAAATTTATTATTATTGTTGTTATTAT 1474
Db      7009 ATATTTATATATATATTAATATTATATTATAT 7042
```

RESULT 34
AC024180/c
LOCUS AC024180 149969 bp DNA linear HTG 01-SEP-2000
DEFINITION Homo sapiens chromosome 3 clone RP11-172A10, WORKING DRAFT
SEQUENCE, 17 unordered pieces.
ACCESSION AC024180
VERSION AC024180.3 GI:8569838
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 149969)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 149969)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (25-FEB-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Jun 16, 2000 this sequence version replaced gi:7630874.

---------- Genome Center ----------
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site:http://genome.wustl.edu/gsc/index.shtml
---------- Project Information ----------
Center project name: H_NH0172A10
---------- Summary Statistics ----------
Sequencing vector: M13; 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-primer ET; 97% of reads
Chemistry: Dye-terminator Big Dye; 3% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 141584 bases at least Q40
Consensus quality: 144367 bases at least Q30
Consensus quality: 145803 bases at least Q20
Insert size: 150000; agarose-fp
Insert size: 148369; sum-of-contigs
Quality coverage: 3.76 in Q20 bases; agarose-fp
Quality coverage: 4.04 in Q20 bases; sum-of-contigs

```
* NOTE: This is a 'working draft' sequence. It currently
* consists of 17 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*        1     1731: contig of 1731 bp in length
*     1732     1831: gap of unknown length
*     1832     4275: contig of 2444 bp in length
*     4276     4375: gap of unknown length
*     4376     6562: contig of 2187 bp in length
*     6563     6662: gap of unknown length
*     6663     9713: contig of 3051 bp in length
*     9714     9813: gap of unknown length
*     9814    13035: contig of 3222 bp in length
*    13036    13135: gap of unknown length
*    13136    17265: contig of 4130 bp in length
*    17266    17365: gap of unknown length
*    17366    23146: contig of 5781 bp in length
*    23147    23246: gap of unknown length
*    23247    28187: contig of 4941 bp in length
*    28188    28287: gap of unknown length
*    28288    35880: contig of 7593 bp in length
*    35881    35980: gap of unknown length
*    35981    46076: contig of 10096 bp in length
*    46077    46176: gap of unknown length
*    46177    56492: contig of 10316 bp in length
*    56493    56592: gap of unknown length
*    56593    67536: contig of 10944 bp in length
*    67537    67636: gap of unknown length
*    67637    79604: contig of 11968 bp in length
```

```
            *    79605    79704: gap of unknown length
            *    79705    94913: contig of 15209 bp in length
            *    94914    95013: gap of unknown length
            *    95014   112048: contig of 17035 bp in length
            *   112049   112148: gap of unknown length
            *   112149   130002: contig of 17854 bp in length
            *   130003   130102: gap of unknown length
            *   130103   149969: contig of 19867 bp in length.
FEATURES             Location/Qualifiers
     source          1. .149969
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-172A10"
     misc_feature    1. .1731
                     /note="assembly_name:Contig10"
     misc_feature    1832. .4275
                     /note="assembly_name:Contig11"
     misc_feature    4376. .6562
                     /note="assembly_name:Contig12
                     clone_end:T7
                     vector_side:left"
     misc_feature    6663. .9713
                     /note="assembly_name:Contig13"
     misc_feature    9814. .13035
                     /note="assembly_name:Contig14"
     misc_feature    13136. .17265
                     /note="assembly_name:Contig15"
     misc_feature    17366. .23146
                     /note="assembly_name:Contig16"
     misc_feature    23247. .28187
                     /note="assembly_name:Contig17
                     clone_end:SP6
                     vector_side:left"
     misc_feature    28288. .35880
                     /note="assembly_name:Contig18"
     misc_feature    35981. .46076
                     /note="assembly_name:Contig19"
     misc_feature    46177. .56492
                     /note="assembly_name:Contig20"
     misc_feature    56593. .67536
                     /note="assembly_name:Contig21"
     misc_feature    67637. .79604
                     /note="assembly_name:Contig22"
     misc_feature    79705. .94913
                     /note="assembly_name:Contig23"
     misc_feature    95014. .112048
                     /note="assembly_name:Contig24"
     misc_feature    112149. .130002
                     /note="assembly_name:Contig25"
     misc_feature    130103. .149969
                     /note="assembly_name:Contig26"
ORIGIN

  Query Match             7.7%;  Score 355.2;  DB 2;  Length 149969;
  Best Local Similarity   68.3%;  Pred. No. 1.2e-60;
  Matches  596;  Conservative    0;  Mismatches  263;  Indels   13;  Gaps    7;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              || ||||||||| ||||   |||| ||||||||||||||||||||||||||| ||||||
Db      74523 AGATCAAGACCATCCTGGCTAACACGGTGAAACCCCATCTCTACTAAAAATACAAAAATT 74464

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              || ||||||||||     | ||||| |||||||||||| |||||| ||||||||||||
Db      74463 AGCCAGGCATGGTGGCGGGCACCTGTAGTCCCAGCTACTTGGGAGGCTGAGGCAGGAGAA 74404

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGC-AAGATCATGCCATTGCACTCCA 178
              |  | ||||||||  || |||||| |||| | |||| |||||   |||| |||||||||
Db      74403 TGGCCTGAACCTG--GGAGGTGGAGCTTGCAGTGAGCTAAGATGGCGCCACTGCACTCCA 74346

Qy        179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATA 238
              |||  |||||| ||||| | | | |  |     |||||| | |     |  |||| ||||

Db      74345 GCCTGGGCAAC-AGAGCAAGAAGACTCCATCTCAAAAAATATATTATGTACATATGTATA 74287

Qy        239 TATATATATATATACACACACACACATATATATATACACATATATATACACATATATA 298
              | | | |||| |||   | | |||||||| |||| ||| | |  ||| |||| |||| ||
Db      74286 TGTGTCTATACATATGTATATACACACATGTATACATATATACTTATGTACATATATGTA 74227

Qy        299 CATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACA 358
              ||| ||| |||||| || | | ||||| | |   || |||  || |||||| |||| |||
Db      74226 TATACATATACATATGTATATACATATACATATGTATATACATATACATATGTATATACA 74167

Qy        359 TATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATA 418
              ||||||||     |||||  || ||| |||| | || ||||||||||||| | |    |||||
Db      74166 CATGTATAC----ATATATACTTATGTACATATATGTATACACATATACATATGTATATA 74111

Qy        419 TACACATATATACACATATATACACATATACACATATATACACATATATACATATATACA 478
              ||||||  ||||||||||||    |  || ||||||  ||||||||||||  | ||  |||| |
Db      74110 CACACATGTATACACATATACTTATGTACACACATGTATACACATATACATATGTATATA 74051

Qy        479 CATATATATA---ATATACACACATATATATACACATATATACACACATATATACACATA 535
              || | || ||    | ||| ||| || ||||||||||| || || |||||||||  | |  ||
Db      74050 CACACATGTATACACATATACATATGTATATACACACATGTATACACATATACATATGTA 73991

Qy        536 TATACACATATATATACACATATATACACATATATACATATATACACATATATATACATA 595
              |||||||| || |||||||||||| | |  |||||||| | ||  | | | |||||||||
Db      73990 TATACACACATGTATACACATATACATATGTATATACACACATGTATACACATATACATA 73931

Qy        596 TATACACATATATACATATATACACATATATACATATATACACACATATATACACATACA 655
              | || | | | |  |||| ||| ||| |||  || |||||   |||||  ||| |||||
Db      73930 TGTATATACACACATGTATACACATATACATATGTACATACATGTATATACACATATACA 73871

Qy        656 TATACACACACATAGATATACATATATATACAC-ATATATATACGTATATATATGTATAT 714
              |||  ||| | ||  |||||||||| |||   ||  || ||||| |||||| || || ||
Db      73870 TATGTACATATATGTATATACATACATACTTACTGTACATATATGTATATACATATACAT 73811

Qy        715 ATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAG 774
              || |||    |     | |||        |   ||   |    ||   ||| |   | |
Db      73810 ATGTATATACACCCATGTATACATATATACTTATGTACATATATGTATATACATATATAC 73751

Qy        775 AGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGG 834
                | | | |  | ||    |    || | |  | |  |   | | |||
Db      73750 TTATGTACATATATGAATACATATATGTAGATACATGTGTATATGTGTAGATACATATAT 73691

Qy        835 AAATACATAAATGTGTAATAAAACCTATCTGT 866
              ||||||||  ||| ||   |    ||| |||
Db      73690 GAATACATATGTGTATATAGATGTGTATGTGT 73659


RESULT 35
CR759776/c
LOCUS       CR759776              151047 bp    DNA     linear   HTG 08-SEP-2004
DEFINITION  Danio rerio clone DKEY-205B10, *** SEQUENCING IN PROGRESS ***, 5
            unordered pieces.
ACCESSION   CR759776
VERSION     CR759776.3  GI:51965222
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 151047)
  AUTHORS   McLay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 9, 2004 this sequence version replaced gi:51890366.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
```

```
          Center project name: zK205B10
          -------------- Summary Statistics
          Assembly program: XGAP4; version 4.5
          Chemistry: Dye-terminator; 100% of reads
          Consensus quality: 149837 bases at least Q40
          Consensus quality: 150168 bases at least Q30
          Consensus quality: 150366 bases at least Q20
          Insert size: 150647; sum-of-contigs
          Insert size: 157117; 1.3% error; agarose-fp
          Quality coverage: 8.56x in Q20 bases; sum-of-contigs Quality
          coverage: 8.40x in Q20 bases; agarose-fp
          --------------.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 5 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *       1    21342: contig of 21342 bp in length
          *   21343    21442: gap of 100 bp
          *   21443    23542: contig of 2100 bp in length
          *   23543    23642: gap of 100 bp
          *   23643    56824: contig of 33182 bp in length
          *   56825    56924: gap of 100 bp
          *   56925   140273: contig of 83349 bp in length
          *  140274   140373: gap of 100 bp
          *  140374   151047: contig of 10674 bp in length.
FEATURES             Location/Qualifiers
     source          1. .151047
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-205B10"
                     /clone_lib="DanioKey"
     misc_feature    1. .21342
                     /note="assembly_fragment:00167
                     fragment_chain:1"
     misc_feature    21443. .23542
                     /note="assembly_fragment:00022
                     fragment_chain:1"
     misc_feature    23643. .56824
                     /note="assembly_fragment:00416
                     fragment_chain:1"
     misc_feature    56925. .140273
                     /note="assembly_fragment:00819
                     fragment_chain:1"
     misc_feature    140374. .151047
                     /note="assembly_fragment:00054"
ORIGIN

  Query Match             7.7%;  Score 355.2;  DB 2;  Length 151047;
  Best Local Similarity   64.9%;  Pred. No. 1.2e-60;
  Matches  470;  Conservative    0;  Mismatches  248;  Indels    6;  Gaps    2;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db     140869 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 140810

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db     140809 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 140750

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db     140749 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 140690

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db     140689 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 140630

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||||| | ||| |||
Db     140629 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 140570

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db     140569 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 140510

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db     140509 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 140450

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | ||||| ||||| ||||||||| | |||||||||  ||||||
Db     140449 TATATATATATATATATATACATATATATATATATATATATATATATATATATATATATA 140390

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||
Db     140389 TATATATATATATATANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 140330

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
                                                                    |||
Db     140329 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNAAAA 140270

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA-----AGGTAGTTAAGAA 880
              | |    | ||| ||| || | | | | ||| | ||     |  ||  ||   |
Db     140269 AAAAAAATTATATATATATATTTATATTTATATATATATATATATATATATATATATA 140210

Qy        881 GGTAACACTATATATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATT 940
                ||    |||||||||||||   | || | |  |||  | |||    |  |  | ||
Db     140209 TATATATATATATATATATATATATAGTACTTAAAAAATGAAGACATGAAGACAGCTT 140150

Qy        941 TAAA 944
              ||||
Db     140149 TAAA 140146
```

```
RESULT 36
AC006441/c
LOCUS       AC006441               162996 bp    DNA     linear   PRI 14-MAY-1999
DEFINITION  Homo sapiens chromosome 17, clone hRPC.56_K_13, complete sequence.
ACCESSION   AC006441
VERSION     AC006441.13  GI:4760435
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 162996)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone hRPC.56_K_13
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 162996)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 162996)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
```

```
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-MAY-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 8, 1999 this sequence version replaced gi:4755216.
            All repeats were identified using RepeatMasker: Smit, A.F.A. &
            Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html.
FEATURES             Location/Qualifiers
     source          1. .162996
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="hRPC.56_K_13"
                     /clone_lib="Peter de Jong/ human PAC library"
     repeat_region   90. .376
                     /rpt_family="AluSx"
     repeat_region   1084. .1390
                     /rpt_family="AluSq"
     unsure          1092. .1093
                     /note="Unresolved compression. Probably GC."
     repeat_region   1424. .1557
                     /rpt_family="AluSg"
     repeat_region   complement(2260. .2414)
                     /rpt_family="AluSc"
     repeat_region   2562. .2857
                     /rpt_family="AluY"
     repeat_region   3102. .3129
                     /rpt_family="(CAAAAA)n"
     repeat_region   complement(3499. .3790)
                     /rpt_family="AluSg"
     repeat_region   complement(3931. .4252)
                     /rpt_family="LTR28"
     repeat_region   complement(4253. .4693)
                     /rpt_family="Tigger1"
     repeat_region   4694. .4996
                     /rpt_family="AluSq"
     repeat_region   complement(4997. .5181)
                     /rpt_family="Tigger1"
     repeat_region   complement(5186. .5466)
                     /rpt_family="AluSg"
     repeat_region   complement(5472. .6215)
                     /rpt_family="Tigger1"
     repeat_region   7119. .7436
                     /rpt_family="AluSq"
     repeat_region   7451. .7521
                     /rpt_family="MADE1"
     repeat_region   7945. .8251
                     /rpt_family="AluSq"
     repeat_region   complement(8354. .8748)
                     /rpt_family="MER21A"
     repeat_region   8743. .9125
                     /rpt_family="LTR1B"
     repeat_region   complement(9126. .9419)
                     /rpt_family="AluSg"
     repeat_region   9420. .9581
                     /rpt_family="LTR1B"
     repeat_region   complement(9784. .9915)
                     /rpt_family="MER21B"
     repeat_region   complement(9916. .10213)
                     /rpt_family="AluSg"
     repeat_region   complement(10214. .10249)
                     /rpt_family="MER21B"
     repeat_region   complement(10305. .10409)
                     /rpt_family="MER21A"
     repeat_region   complement(10410. .11116)
                     /rpt_family="LTR8"
     repeat_region   complement(11117. .11155)
                     /rpt_family="MER21A"
     repeat_region   11215. .11433
                     /rpt_family="MER33"
     repeat_region   11434. .11718
                     /rpt_family="AluSx"
     repeat_region   11719. .11827
                     /rpt_family="MER33"
     repeat_region   complement(11851. .12153)
                     /rpt_family="AluSx"
     repeat_region   complement(12155. .12281)
                     /rpt_family="MIR"
     repeat_region   12282. .12619
                     /rpt_family="MER1B"
     repeat_region   complement(12620. .12649)
                     /rpt_family="MIR"
     repeat_region   12656. .12809
                     /rpt_family="L1MA2"
     repeat_region   complement(12941. .13659)
                     /rpt_family="LTR1B"
     repeat_region   13662. .13779
                     /rpt_family="L1MA3"
     repeat_region   13780. .14075
                     /rpt_family="AluJb"
     repeat_region   14076. .14240
                     /rpt_family="L1MA3"
     repeat_region   complement(14333. .14627)
                     /rpt_family="AluY"
     repeat_region   14628. .14651
                     /rpt_family="AT_rich"
     repeat_region   complement(14691. .14761)
                     /rpt_family="MIR"
     repeat_region   15905. .16202
                     /rpt_family="AluJo"
     repeat_region   complement(16333. .16724)
                     /rpt_family="MSTC"
     repeat_region   complement(16777. .16993)
                     /rpt_family="MIR"
     repeat_region   complement(17020. .17125)
                     /rpt_family="L1MC5"
     repeat_region   complement(17141. .17428)
                     /rpt_family="AluSg1"
     repeat_region   complement(17432. .17570)
                     /rpt_family="AluSq/x"
     repeat_region   complement(17582. .17653)
                     /rpt_family="L1MC5"
     repeat_region   17954. .18218
                     /rpt_family="AluSp"
     repeat_region   18219. .18289
                     /rpt_family="(GGGAA)n"
     repeat_region   18359. .18405
                     /rpt_family="(TG)n"
     repeat_region   18639. .18676
                     /rpt_family="(TG)n"
     repeat_region   18878. .19179
                     /rpt_family="AluJb"
     repeat_region   19202. .19503
                     /rpt_family="AluSx"
     repeat_region   complement(20063. .20356)
                     /rpt_family="AluSx"
     repeat_region   20835. .20877
                     /rpt_family="L2"
     repeat_region   20878. .20952
                     /rpt_family="GA-rich"
     repeat_region   complement(21223. .21533)
```

```
                     /rpt_family="AluSx"
     repeat_region   complement(21991. .22245)
                     /rpt_family="L2"
     repeat_region   complement(22402. .22474)
                     /rpt_family="MIR"
     repeat_region   complement(22472. .22790)
                     /rpt_family="L2"
     repeat_region   complement(22924. .23524)
                     /rpt_family="L2"
     repeat_region   23525. .23829
                     /rpt_family="AluJb"
     repeat_region   complement(23830. .23861)
                     /rpt_family="L2"
     repeat_region   23936. .24244
                     /rpt_family="AluSx"
     repeat_region   24247. .24443
                     /rpt_family="AluSg/x"
     repeat_region   25664. .25971
                     /rpt_family="AluY"

  Query Match             7.7%;  Score 355.2;  DB 9;  Length 162996;
  Best Local Similarity   66.4%;  Pred. No. 1.2e-60;
  Matches  616;  Conservative    0;  Mismatches  303;  Indels    9;  Gaps    7;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC--AAAAT 58
              || || |||||| ||||   |||| |||||||||| |||||||||| ||||  |||||
Db      79561 AGATCGAGACCAACCTGGCTAACACGGTGAAACCCTGTCTCTACTAAATATACAAAAAAT 79502

Qy         59 TAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGA 118
               || |  ||  || ||      |||||||| |||||||||||| |||| | ||||||||||
Db      79501 AAGCCGGGAGTGATGGCGGGCGCCTGTAGTCCCAGCTACTTGGGAGACTGAGGCAGGAGA 79442

Qy        119 ATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCC 177
              ||  | | ||||  | || ||||||| |||  | ||| ||||||| ||||| ||||||||
Db      79441 ATGGCGTTAACC--CGGGAGGTGGAGCTTGTAGTGAGCCAAGATCGTGCCACTGCACTCC 79384

Qy        178 AGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATAT 237
              ||||  ||  || ||||||| |      |||| ||||||||||     | |||||||||
Db      79383 AGCCTGGGTGAC-AGAGCGAGACTCTGGCTCAAAAAAAAAAAAAAATATATATATATATAT 79325

Qy        238 ATATATATATATATACACACACACACATATATATATACACATATATATACACATATAT 297
              | ||||||| ||||| | ||| | | | | ||||||||| | | ||| ||| | | ||||
Db      79324 AAATATATAAATATATATACAAATATATAAATATATATAAATACATAAATATAAAAATAT 79265

Qy        298 ACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACAC 357
              | | ||||| | | ||||| | | | ||||| | |||||||  || | || |||||| |
Db      79264 AAAAATATATATAAATATATATA-AAATATATAAATACATAAATATATATTTATATATAA 79206

Qy        358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417
              |||  || | | |||| || | ||| ||||| | ||| ||| | ||| ||| | | ||||
Db      79205 ATACATAAATATATATTTATATATAAATACATAAATAAATATATATACATATATAAATAT 79146

Qy        418 ATACACATATATACACATATATACACATATACACATATATACACATATATACATATATAC 477
              ||| | | ||| | | ||||| | | ||| | | ||||||| | ||||| | |||||||
Db      79145 ATAAATACATAAAAATATATAAATATATAAAAATATATATAAATATATAAAAATATATAT 79086

Qy        478 ACATATATATAATATACACACATATATATACACATATATACACACATATATACACATATA 537
              | ||||||| || | | | | ||| ||||| ||| ||||| | | ||||| | | || ||
Db      79085 AAATATATACAA-AAATATATATAAATATATACAAATATATATAAATATAAAAATATGTA 79027

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATACATATA 597
              || | ||||||| | | ||||||| | ||||| | ||||||| | ||||||| | |||||
Db      79026 TAAATATATATAAAAATATATATAAATATATAAAAATATATATAAATATATAAATATATA 78967

Qy        598 TACACATATATACATATATACACATATATACATATATACACACATATATACACATACATA 657
              || | ||||| | ||||||| | ||||| | ||||||| | | ||| ||| | ||| | |
Db      78966 TAAATATATAAATATATATAAATATATAAATATATATAAATATATAGATATATATATAAA 78907

Qy        658 TACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATA 717
              || | | |  || ||||| ||||||| | | |||||||    |||| |||| |||  |||
Db      78906 TATATATATCTATATATAAATATATAGATATATATATAACTATATAAATATATATGAATA 78847

Qy        718 TATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGA 777
              |||    |   |  |||      |  | |         |    ||| | ||| | | |
Db      78846 TATATAACTATATAAATATATATGAATATATATAACTATATAAATATATATGAATATATA 78787

Qy        778 GTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAA 837
                | |  | | |||    | |   | | |         |   |    ||
Db      78786 TAACTATATAAATATATATGAATATATATAACTATATAAATATATATGAATATATATAAC 78727

Qy        838 TACATAAATGTGTAATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTAACACTATATATA 896
              || |||||| | ||  || |  || || || |  ||  | |  |  ||  |||||||| |
Db      78726 TATATAAATATATATGAATATATAACTATATAAATATATGAATATATATAACTATATAAA 78667

Qy        897 TATATAGTGAAAGCAGTGTAAACCTAAA 924
              ||||||   | |      || |  ||||
Db      78666 TATATATGAATATATAAATATATATAAA 78639


RESULT 37
AL590381/c
LOCUS       AL590381              182430 bp    DNA     linear   PRI 16-MAY-2001
DEFINITION  Human DNA sequence from clone RP11-373O10 on chromosome 9, complete
            sequence.
ACCESSION   AL590381
VERSION     AL590381.4  GI:13872470
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 182430)
  AUTHORS   Griffiths,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Apr 27, 2001 this sequence version replaced gi:13752046.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr9
            RP11-373O10 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            This sequence is the entire insert of clone RP11-373O10 The true
            left end of clone RP11-82L2 is at 128987 in this sequence. The true
            right end of clone RP11-346G21 is at 9227 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .182430
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /clone="RP11-373O10"
                     /clone_lib="RPCI-11.2"
     repeat_region   8. .317
```

```
                     /note="L1M4 repeat: matches 2333. .2656 of consensus"
     repeat_region   348. .683
                     /note="L1M4 repeat: matches 2747. .3082 of consensus"
     repeat_region   763. .1062
                     /note="AluSc repeat: matches 1. .296 of consensus"
     repeat_region   1568. .1826
                     /note="L1M4c repeat: matches 1694. .1953 of consensus"
     repeat_region   2178. .2434
                     /note="L1MB8 repeat: matches 5132. .5403 of consensus"
     repeat_region   2454. .2584
                     /note="MLT1E repeat: matches 35. .157 of consensus"
     repeat_region   2638. .2888
                     /note="MLT1E repeat: matches 312. .561 of consensus"
     repeat_region   2895. .3648
                     /note="L1MB8 repeat: matches 5403. .6171 of consensus"
     repeat_region   4203. .4252
                     /note="L1MC2 repeat: matches 6263. .6312 of consensus"
     repeat_region   4262. .4578
                     /note="L1M4 repeat: matches 2683. .3032 of consensus"
     repeat_region   4866. .5228
                     /note="L1PB3 repeat: matches 5785. .6149 of consensus"
     repeat_region   5844. .5891
                     /note="24 copies 2 mer ta 75% conserved"
     repeat_region   5893. .6306
                     /note="23 copies 18 mer 82% conserved"
     repeat_region   5909. .6298
                     /note="13 copies 30 mer 83% conserved"
     repeat_region   6301. .6597
                     /note="AluY repeat: matches 1. .296 of consensus"
     repeat_region   6743. .6839
                     /note="L1M4 repeat: matches 4041. .4137 of consensus"
     repeat_region   7150. .7519
                     /note="THE1C repeat: matches 1. .371 of consensus"
     repeat_region   8154. .8217
                     /note="THE1B repeat: matches 1. .62 of consensus"
     repeat_region   8237. .8467
                     /note="MIR repeat: matches 2. .253 of consensus"
     repeat_region   8679. .8814
                     /note="AluJo/FRAM repeat: matches 159. .295 of consensus"
     repeat_region   9551. .9910
                     /note="THE1B repeat: matches 1. .363 of consensus"
     repeat_region   11221. .11700
                     /note="L1MC4 repeat: matches 7390. .7847 of consensus"
     repeat_region   13929. .14162
                     /note="L1PA4 repeat: matches 1430. .1662 of consensus"
     repeat_region   14156. .18641
                     /note="L1PA4 repeat: matches 1666. .6146 of consensus"
     repeat_region   18690. .19243
                     /note="L1ME3A repeat: matches 5523. .6091 of consensus"
     repeat_region   19861. .20102
                     /note="L1MEc repeat: matches 1817. .2070 of consensus"
     repeat_region   20203. .20330
                     /note="L1MEc repeat: matches 2167. .2303 of consensus"
     repeat_region   20430. .21121
                     /note="L1 repeat: matches 2894. .3643 of consensus"
     repeat_region   21332. .22100
                     /note="L1M4 repeat: matches 3912. .4721 of consensus"
     repeat_region   22457. .22551
                     /note="L1MC/D repeat: matches 5374. .5466 of consensus"
     repeat_region   23344. .23992
                     /note="L1PA4 repeat: matches 5496. .6144 of consensus"
     repeat_region   26070. .26233
                     /note="MLT1G repeat: matches 12. .179 of consensus"
     repeat_region   26946. .27101
                     /note="LTR1 repeat: matches 631. .785 of consensus"
     repeat_region   27115. .27210
                     /note="LTR1 repeat: matches 1. .95 of consensus"
     repeat_region   27634. .27945
                     /note="AluSx repeat: matches 1. .312 of consensus"
     repeat_region   29014. .29302
                     /note="AluJb repeat: matches 1. .290 of consensus"
     repeat_region   30332. .30492
                     /note="L1MEc repeat: matches 1749. .1897 of consensus"
     repeat_region   31684. .32009
                     /note="MLT1A2 repeat: matches 34. .374 of consensus"
     repeat_region   32010. .32313
                     /note="AluY repeat: matches 1. .305 of consensus"
     repeat_region   33320. .33629
                     /note="AluSx repeat: matches 1. .308 of consensus"
     repeat_region   33683. .33975
                     /note="AluSc repeat: matches 1. .295 of consensus"
     repeat_region   33986. .34132
                     /note="MIR repeat: matches 100. .243 of consensus"
     repeat_region   34156. .34309
                     /note="L2 repeat: matches 2575. .2734 of consensus"
     repeat_region   34739. .34862
                     /note="MIR repeat: matches 2. .139 of consensus"
     repeat_region   35497. .35578
                     /note="41 copies 2 mer tt 67% conserved"
     repeat_region   36076. .36464
                     /note="L1M4c repeat: matches 1599. .1989 of consensus"
     repeat_region   36471. .36974
                     /note="MLT1D repeat: matches 3. .499 of consensus"
     repeat_region   36977. .37504
                     /note="L1M4c repeat: matches 1001. .1604 of consensus"
     repeat_region   37507. .38001
                     /note="L1ME2 repeat: matches 5652. .6156 of consensus"
     repeat_region   38998. .39313
                     /note="AluSx repeat: matches 1. .289 of consensus"
     repeat_region   39775. .40319
                     /note="MLT2D repeat: matches 1. .553 of consensus"
     repeat_region   41011. .41583
                     /note="MLT2D repeat: matches 1. .545 of consensus"
     repeat_region   42023. .42458
                     /note="MLT2G repeat: matches 1. .386 of consensus"
     repeat_region   42910. .42964
                     /note="11 copies 5 mer atata 72% conserved"
     repeat_region   43137. .43251
                     /note="L1MB6 repeat: matches 5937. .6053 of consensus"
     repeat_region   43240. .43428
                     /note="L1MB6 repeat: matches 5584. .5776 of consensus"
     repeat_region   43431. .44672
                     /note="L1MA4 repeat: matches 5053. .6289 of consensus"
     repeat_region   44661. .44816
                     /note="L1MA4 repeat: matches 4931. .5090 of consensus"
     repeat_region   44810. .45261
                     /note="L1 repeat: matches 3313. .3738 of consensus"
     repeat_region   45270. .45407
                     /note="L1M4 repeat: matches 5474. .5595 of consensus"
     repeat_region   45415. .45634
                     /note="AluSg repeat: matches 1. .301 of consensus"
     repeat_region   45635. .45677
                     /note="AluSg repeat: matches 257. .299 of consensus"
     repeat_region   45680. .45891
                     /note="L1M4 repeat: matches 5254. .5470 of consensus"
     repeat_region   46033. .47277
                     /note="L1PA14 repeat: matches 4885. .6149 of consensus"
     repeat_region   47274. .47631
                     /note="L1MB8 repeat: matches 5816. .6169 of consensus"
     repeat_region   47654. .47961
                     /note="AluSx repeat: matches 1. .308 of consensus"
     repeat_region   47964. .49294
                     /note="L1MA4 repeat: matches 4951. .6299 of consensus"
     repeat_region   49298. .49946
                     /note="L1MD3 repeat: matches 6922. .7588 of consensus"
     repeat_region   51174. .51342
                     /note="MER90 repeat: matches 4. .169 of consensus"
     repeat_region   52038. .52075
                     /note="19 copies 2 mer tt 81% conserved"
     repeat_region   52894. .53118
                     /note="MIR repeat: matches 14. .245 of consensus"
     repeat_region   53389. .53773
                     /note="L1ME1 repeat: matches 5778. .6159 of consensus"
     repeat_region   54054. .54383
                     /note="L2 repeat: matches 2370. .2709 of consensus"
```

```
  Query Match             7.7%;  Score 354.8;  DB 9;  Length 182430;
  Best Local Similarity   73.8%;  Pred. No. 1.4e-60;
  Matches  549;  Conservative    0;  Mismatches  177;  Indels   18;  Gaps    7;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAA---A 57
              || |  |||||| |||    ||||  |||||||||| ||||||| ||||||| ||    |
Db       6524 AGATTGAGACCAACCTAGCGAACACTGTGAAACCCCGTCTCTACCAAAAATACCAAAATA 6465

Qy         58 TTAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAG 117
              ||||   |||  ||||     |||||||  ||| |||||| |||||| ||||||||||
Db       6464 TTAGCTGGGCGCGGTGGCCGGCGCCTGTAGGCCCGGCTACTTGGGAGGCTGAGGCAGGAG 6405

Qy        118 AATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTC 176
              |||    ||||||||  || ||  ||| |||| | ||| | |||    |||| |||||||
Db       6404 AATGGTGTGAACCTG--GGAGGCAGAGCTTGCAGTGAGCCGAGACTGCGCCACTGCACTC 6347

Qy        177 CAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATA 236
              |||||  |||||||      | | |      | ||||||||||     | |||| |||
Db       6346 CAGCCTGGGCAACAGAGCAAGACTCTGTCTCAAAAAAAAAAAAAAAATATATATATACATA 6287

Qy        237 TATATATATATATATACACACACACACATATATATATACACATATATATACACATATA 296
              |||||||| | ||||| | ||||| | | ||||| | ||| | ||| | ||| | ||| |
Db       6286 TATATATACACATATATATACACATATATATATACACATATATATACACATATATATACA 6227

Qy        297 TACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACA 356
              |||||||| ||| ||||| | | ||||||||||  |||||| || ||||||||||||||
Db       6226 CACATATATATACACATATATATACATATATACAC--ACATATATATACATATATATACA 6169

Qy        357 CATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATA 416
              ||||| |||||||| |||||||||||||| ||| ||||||||| ||||||||| | |||
Db       6168 CATATATATACACACATATACACATATATATACATATATATACATATATATACATATATA 6109

Qy        417 TATACACATATATACACATATATACACATATACAC------ATATATACACATATATACA 470
              | ||| ||||| ||| |||||||||||||| ||      ||||| ||| |||||||||
Db       6108 CACACATATATACACATATATATACACATATATACACATATATATACACATATATATACA 6049

Qy        471 TATATACACATATATATA-ATATACACACATATATATACACATATATACACACATATATA 529
              |||||||||||||||||| ||||| | ||| ||||||||||||||| ||| | ||||| |
Db       6048 TATATACACATATATATACATATATATACACATATATACACATATACACATATATATACA 5989

Qy        530 CACATATATACACATATATATACACATATATACACATATATACATATATACACATATATA 589
              | |||| |||| |||||| |||| ||||||||||||||||  | |||||   | | |||
Db       5988 TATATATGTACATATATATGTACATATATATACACATATATGTACATATATGTACACATA 5929

Qy        590 TACATATATACACATATATACATATATACACATATATACATATATACACACATATATACA 649
              || | ||||| |||||||||||||||||| |||||   ||||||||||| ||||| | |
Db       5928 TATACATATATACATATATACATATATACATATATA---ATATATACACATATATACATA 5872

Qy        650 CATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATATG 709
              ||| ||||||| | | | ||| | | ||    |    ||||  || |  |    |||
Db       5871 TATATATATACAGAGAGAGAGAGAGAGATGGTGAATTGCATAAGTAAATTTTCTGATAAT 5812

Qy        710 TATATATATATGCTCCAGAGTTCA 733
                  ||   | | | |||| | |
Db       5811 GGATAATTATTTCACTAGAGATAA 5788


RESULT 38
CR753861/c
LOCUS       CR753861               252450 bp    DNA     linear   HTG 11-OCT-2004
DEFINITION  Danio rerio clone DKEY-192K11, WORKING DRAFT SEQUENCE, 9 unordered
            pieces.
ACCESSION   CR753861
VERSION     CR753861.4  GI:54035031
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 252450)
  AUTHORS   McLaren,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 11, 2004 this sequence version replaced gi:51861477.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK192K11
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 247462 bases at least Q40
            Consensus quality: 248434 bases at least Q30
            Consensus quality: 249070 bases at least Q20
            Insert size: 251650; sum-of-contigs
            Insert size: 239884; 3.5% error; agarose-fp
            Quality coverage: 10.19x in Q20 bases; sum-of-contigs Quality
            coverage: 10.69x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     5351: contig of 5351 bp in length
            *     5352     5451: gap of 100 bp
            *     5452    19343: contig of 13892 bp in length
            *    19344    19443: gap of 100 bp
            *    19444    58067: contig of 38624 bp in length
            *    58068    58167: gap of 100 bp
            *    58168    75897: contig of 17730 bp in length
            *    75898    75997: gap of 100 bp
            *    75998   166203: contig of 90206 bp in length
            *   166204   166303: gap of 100 bp
            *   166304   216140: contig of 49837 bp in length
            *   216141   216240: gap of 100 bp
            *   216241   224264: contig of 8024 bp in length
            *   224265   224364: gap of 100 bp
            *   224365   228110: contig of 3746 bp in length
            *   228111   228210: gap of 100 bp
            *   228211   252450: contig of 24240 bp in length.
FEATURES             Location/Qualifiers
     source          1. .252450
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-192K11"
                     /clone_lib="DanioKey"
     misc_feature    1. .5351
                     /note="assembly_fragment:00110
                     fragment_chain:1"
     misc_feature    5452. .19343
                     /note="assembly_fragment:00179
                     fragment_chain:1"
     misc_feature    19444. .58067
                     /note="assembly_fragment:00986
                     fragment_chain:1"
     misc_feature    58168. .75897
                     /note="assembly_fragment:00384
                     fragment_chain:1"
     misc_feature    75998. .166203
                     /note="assembly_fragment:02209
                     fragment_chain:2"
     misc_feature    166304. .216140
                     /note="assembly_fragment:01459
                     fragment_chain:2"
     misc_feature    216241. .224264
```

```
                     /note="assembly_fragment:00050
                     fragment_chain:2"
     misc_feature    224365. .228110
                     /note="assembly_fragment:00015"
     misc_feature    228211. .252450
                     /note="assembly_fragment:00628.0"
ORIGIN

  Query Match             7.7%;  Score 354;  DB 2;  Length 252450;
  Best Local Similarity   64.7%;  Pred. No. 1.9e-60;
  Matches  452;  Conservative    0;  Mismatches  246;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db     166799 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 166740

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db     166739 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 166680

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db     166679 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 166620

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db     166619 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 166560

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db     166559 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 166500

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db     166499 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 166440

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||| | ||||||| | ||||||| ||||||| | |  ||||
Db     166439 TATATATATATATATATATATACATATATATATATATATATATATATATATTTATA 166380

Qy        646 TACACATACATATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || ||||| ||||| | | | ||| ||||| |||||||||| | |||||||||  ||||||
Db     166379 TATACATATATATAGATATATATATATATATATATATATATATATATATATATATATATA 166320

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||  ||||||||||||
Db     166319 TAAATATATATATATGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 166260

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
                                                                      |||
Db     166259 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNAAAA 166200

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885
              |       | | | ||||||   ||   | |  ||| | ||   |  |||   |  ||
Db     166199 AATATTTATATAAAAATAAATANATATATATATATTTATATTTTTTTTATTTATATAT 166140

Qy        886 CACTATATATATATATAGTGAAAGCAGTGTAAACCTAAA 924
                |||||||||||||   | |    | || |  || |
Db     166139 ATATATATATATATATATATATATATATATATATATA 166101


RESULT 39
CR790381
LOCUS       CR790381              163651 bp    DNA     linear   HTG 24-SEP-2004
DEFINITION  Danio rerio clone DKEY-202K24, *** SEQUENCING IN PROGRESS ***, 10
            unordered pieces.
ACCESSION   CR790381
VERSION     CR790381.3  GI:52673626
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 163651)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 24, 2004 this sequence version replaced gi:52626686.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK202K24
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 159309 bases at least Q40
            Consensus quality: 160146 bases at least Q30
            Consensus quality: 160896 bases at least Q20
            Insert size: 162751; sum-of-contigs
            Insert size: 165246; 2.0% error; agarose-fp
            Quality coverage: 8.55x in Q20 bases; sum-of-contigs Quality
            coverage: 8.63x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     4775: contig of 4775 bp in length
            *     4776     4875: gap of 100 bp
            *     4876    26055: contig of 21180 bp in length
            *    26056    26155: gap of 100 bp
            *    26156    28584: contig of 2429 bp in length
            *    28585    28684: gap of 100 bp
            *    28685    34192: contig of 5508 bp in length
            *    34193    34292: gap of 100 bp
            *    34293    48345: contig of 14053 bp in length
            *    48346    48445: gap of 100 bp
            *    48446    55079: contig of 6634 bp in length
            *    55080    55179: gap of 100 bp
            *    55180   119903: contig of 64724 bp in length
            *   119904   120003: gap of 100 bp
            *   120004   140199: contig of 20196 bp in length
            *   140200   140299: gap of 100 bp
            *   140300   150280: contig of 9981 bp in length
            *   150281   150380: gap of 100 bp
            *   150381   163651: contig of 13271 bp in length.
FEATURES             Location/Qualifiers
     source          1. .163651
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-202K24"
                     /clone_lib="DanioKey"
     misc_feature    1. .4775
                     /note="assembly_fragment:00038
                     fragment_chain:1"
     misc_feature    4876. .26055
                     /note="assembly_fragment:00583
                     fragment_chain:1"
     misc_feature    26156. .28584
                     /note="assembly_fragment:00027
                     fragment_chain:1"
     misc_feature    28685. .34192
                     /note="assembly_fragment:00076
                     fragment_chain:1"
     misc_feature    34293. .48345
```

```
                     /note="assembly_fragment:00435
                     fragment_chain:1"
     misc_feature    48446. .55079
                     /note="assembly_fragment:00118
                     fragment_chain:1"
     misc_feature    55180. .119903
                     /note="assembly_fragment:01052
                     fragment_chain:1"
     misc_feature    120004. .140199
                     /note="assembly_fragment:00806
                     fragment_chain:1"
     misc_feature    140300. .150280
                     /note="assembly_fragment:00174
                     fragment_chain:1"
     misc_feature    150381. .163651
                     /note="assembly_fragment:00293.0"
ORIGIN

  Query Match          7.7%;  Score 353.6;  DB 2;  Length 163651;
  Best Local Similarity   68.5%;  Pred. No. 2.4e-60;
  Matches  502;  Conservative    0;  Mismatches  230;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||||| | | | | | | |||||||||||| | |||||||
Db     139185 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 139244

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db     139245 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 139304

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db     139305 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 139364

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db     139365 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 139424

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || |||||||| | ||||||||| ||||| | | |||||||||| | ||||||| | | |||
Db     139425 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 139484

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db     139485 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 139544

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db     139545 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 139604

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db     139605 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 139664

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||          | |||          | |   ||    |   |||
Db     139665 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 139724

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
              |   | | | |  | | |     | |||     | |   | || |      |       | |
Db     139725 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 139784

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885
              ||        | ||| || | ||    | |  ||| | ||     |  ||  ||  ||
Db     139785 TATATATATATATATATATATATATATATATATATATATATATAATATATCAATATAT 139844

Qy        886 CACTATATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAAT 945
              | ||||||||||||||   | |   | ||||  ||||          |  || ||||
Db     139845 AAATATATATATATATATATATATATATATAAAAATAAATATATATATATAAATNTAAAA 139904

Qy        946 GTTATAGAAGAAT 958
              || | |  |||
Db     139905 AATAAATATTAAT 139917
```

```
RESULT 40
AC141288/c
LOCUS       AC141288              164994 bp    DNA     linear   HTG 11-MAR-2003
DEFINITION  Homo sapiens chromosome 16 clone RP11-357A20, WORKING DRAFT
            SEQUENCE, 2 unordered pieces.
ACCESSION   AC141288
VERSION     AC141288.1  GI:28913068
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 164994)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 164994)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-MAR-2003) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 543473
            Center clone name: RPCI-11_357A20
            -------
            Summary Statistics
            Consensus quality: 164710 bases at least Q40
            Consensus quality: 164791 bases at least Q30
            Consensus quality: 164821 bases at least Q20
            Estimated insert size: 175000; agarose-fp estimation
            Estimated insert size: 164894; sum-of-contigs estimation
            Quality coverage:  8.51 in Q20 bases; agarose-fp estimation
            Quality coverage:  9.03 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    66545: contig of 66545 bp in length
            *    66546    66645: gap of unknown length
            *    66646   164994: contig of 98349 bp in length.
FEATURES             Location/Qualifiers
     source          1. .164994
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone="RP11-357A20"
                     /clone_lib="RPCI human BAC library 11"
ORIGIN

  Query Match          7.7%;  Score 353.6;  DB 2;  Length 164994;
  Best Local Similarity   72.6%;  Pred. No. 2.4e-60;
  Matches  527;  Conservative    0;  Mismatches  189;  Indels   10;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAA--- 57
              || || |||||| ||||    ||||  |||||||||| |||||||||||||| |||||
Db     149586 AGATCGAGACCATCCTGGCTAACACAGTGAAACCCCGTCTCTACTAAAAATCCAAAAAAA 149527

Qy         58 --TTAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGG 115
                |||| | ||| |||||     | |||||| |||||||||||  |||||| ||||||||
Db     149526 AGTTAGCCGGGCGTGGTGGCGGGCACCTGTAGTCCCAGCTACTCTGGAGGCTGAGGCAGG 149467
```

```
Qy        116 AGAATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCAC 174
              ||||| | ||||| ||  || ||  ||| |||| | ||| | ||||  ||||| |||||
Db     149466 AGAATGGCGTGAACTTG--GGAGGCAGAGCTTGCAGTGAGCCGAGATGGTGCCACTGCAC 149409

Qy        175 TCCAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATA 234
              ||||| |  ||| || |||||||  |   |||||||  ||||||  | |       | |||| |
Db     149408 TCCAGTCTGGGCGAC-AGAGCGAGACTCCATCTCAAAAAAAATATATATATATATATAGA 149350

Qy        235 TATATATATATATATATACACACACACACACATATATATATACACATATATATACACATA 294
              ||||||||  |||||||  | | | | | | | |||  |||||||  | |||||||  | | |||
Db     149349 TATATATAGATATATAGATATATATAGATATATAGATATATATAGATATATAGATATATA 149290

Qy        295 TATACATATATACACATATATACACATATATATACACATACATATGTACACATATATATA 354
              || | ||||| | | |||||||  | |||||||  | |  ||| | ||  || | |||||||  |
Db     149289 TAGATATATAGATATATATATATAGATATATAGATATATATAGATATATAGATATATAGA 149230

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
               | ||||  |||  | | ||||| | |||||||  | |  ||| ||| | |||||||  | | |
Db     149229 TATATATAGATATATAGATATATAGATATATAGATATATAGATATAGATATATAGATATA 149170

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              |||| | | |||  ||| | | |||||  | |  ||| | |  ||||| | |||||||  |   ||
Db     149169 TATAGATATATAGATATATAGATATATATAGATATATAGATATATAGATATATATAGTTA 149110

Qy        475 TACACATATATATAATATACACACATATATATACACATATATACACACATATATACACAT 534
              || | |||||||  |||||  | | ||||| | ||| ||| ||| | | | |||||  | ||
Db     149109 TATAGATATATA-GATATATATAGATATAGAGACATATAGATATATATAGATATAGAGAT 149051

Qy        535 ATATACACATATATATACACATATATACACATATATACATATATACACATATATATACAT 594
              ||||| | ||||| | | | ||| ||| | | ||||| | | ||| | ||| |||||  |
Db     149050 ATATAGATATATAGAGATATATAGATATATAGATATATAGAGATATATATAGATATATAG 148991

Qy        595 ATATACACATATATACATATATACACATATATACATATATACACACATATATACACATAC 654
              ||||| | ||||||| |||||||  | |||||||  |||||||  | | |||||    | |||
Db     148990 ATATATAGATATATAGATATATATAGATATATAGATATATAGATATATATAGTTATATAG 148931

Qy        655 ATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATAT 714
              ||||| | | | ||| | ||| | | | ||| | |||||||||  |||| | || ||||
Db     148930 ATATATAGATATATATAGATATAGAGACATATAGATATATATAGATATAGAGATATATAG 148871

Qy        715 ATATAT 720
              ||||||
Db     148870 ATATAT 148865
```

RESULT 41
CR376741/c

```
LOCUS       CR376741              165979 bp    DNA     linear   HTG 24-SEP-2004
DEFINITION  Danio rerio clone DKEY-114B16, *** SEQUENCING IN PROGRESS ***, 12
            unordered pieces.
ACCESSION   CR376741
VERSION     CR376741.5  GI:52673643
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 165979)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 24, 2004 this sequence version replaced gi:52546082.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK114B16
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 159153 bases at least Q40
            Consensus quality: 160217 bases at least Q30
            Consensus quality: 161345 bases at least Q20
            Insert size: 164879; sum-of-contigs
            Insert size: 169971; 1.8% error; agarose-fp
            Quality coverage: 8.37x in Q20 bases; sum-of-contigs Quality
            coverage: 8.39x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 12 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     9842: contig of 9842 bp in length
            *    9843     9942: gap of 100 bp
            *    9943    22622: contig of 12680 bp in length
            *   22623    22722: gap of 100 bp
            *   22723    45138: contig of 22416 bp in length
            *   45139    45238: gap of 100 bp
            *   45239    70145: contig of 24907 bp in length
            *   70146    70245: gap of 100 bp
            *   70246    73805: contig of 3560 bp in length
            *   73806    73905: gap of 100 bp
            *   73906    77192: contig of 3287 bp in length
            *   77193    77292: gap of 100 bp
            *   77293   101185: contig of 23893 bp in length
            *  101186   101285: gap of 100 bp
            *  101286   127529: contig of 26244 bp in length
            *  127530   127629: gap of 100 bp
            *  127630   143030: contig of 15401 bp in length
            *  143031   143130: gap of 100 bp
            *  143131   145591: contig of 2461 bp in length
            *  145592   145691: gap of 100 bp
            *  145692   159780: contig of 14089 bp in length
            *  159781   159880: gap of 100 bp
            *  159881   165979: contig of 6099 bp in length.
FEATURES             Location/Qualifiers
     source          1. .165979
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-114B16"
                     /clone_lib="DanioKey"
     misc_feature    1. .9842
                     /note="assembly_fragment:00238
                     fragment_chain:1
                     clone_end:T7
                     vector_side:left"
     misc_feature    9943. .22622
                     /note="assembly_fragment:00377
                     fragment_chain:1"
     misc_feature    22723. .45138
                     /note="assembly_fragment:00739
                     fragment_chain:2"
     misc_feature    45239. .70145
                     /note="assembly_fragment:01657
                     fragment_chain:2"
     misc_feature    70246. .73805
                     /note="assembly_fragment:00125
                     fragment_chain:2"
     misc_feature    73906. .77192
                     /note="assembly_fragment:00093
                     fragment_chain:2"
     misc_feature    77293. .101185
                     /note="assembly_fragment:01273
                     fragment_chain:2"
     misc_feature    101286. .127529
                     /note="assembly_fragment:02046
```

```
                     fragment_chain:3"
     misc_feature    127630. .143030
                     /note="assembly_fragment:00558
                     fragment_chain:3"
     misc_feature    143131. .145591
                     /note="assembly_fragment:00062
                     fragment_chain:3"
     misc_feature    145692. .159780
                     /note="assembly_fragment:00999
                     fragment_chain:3"
     misc_feature    159881. .165979
                     /note="assembly_fragment:00165.0"
ORIGIN

  Query Match             7.6%;  Score 353;  DB 2;  Length 165979;
  Best Local Similarity   68.8%;  Pred. No. 3.2e-60;
  Matches  499;  Conservative    0;  Mismatches  225;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db      74825 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 74766

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db      74765 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 74706

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | |||||||| | ||||||| | | ||||||| | |||||
Db      74705 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 74646

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db      74645 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 74586

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db      74585 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 74526

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | |||||||||| | ||||||| | ||||||| ||||||| | |||
Db      74525 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 74466

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db      74465 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 74406

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db      74405 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 74346

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||          |  |||          |  |   ||     |   |||
Db      74345 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 74286

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
              |  | | | |  |  | |  | |||   | |  | || |     |       | |
Db      74285 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 74226

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885
              ||       | ||| |||||| | ||    |    ||| | ||      ||||  |  |||
Db      74225 TATATATATATATATATAAATATATATATATTTATATATATATATATATTAATTATATAA 74166

Qy        886 CACTATATATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAAT 945
              ||||||||||||||| | | |  | | || |  |||| |      |  |   ||||
Db      74165 ATATATATATATATATAATTATAAAATTATATAAATAAATAAAATTAATAAAAAAAAAAT 74106

Qy        946 GTTAT 950
              | ||
Db      74105 ATAAT 74101


RESULT 42
CR627491/c
```

```
LOCUS       CR627491              248167 bp    DNA     linear   HTG 07-AUG-2004
DEFINITION  Danio rerio clone DKEY-163B6, *** SEQUENCING IN PROGRESS ***, 11
            unordered pieces.
ACCESSION   CR627491
VERSION     CR627491.2  GI:51036473
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 248167)
  AUTHORS   McLay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-AUG-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 7, 2004 this sequence version replaced gi:50978439.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK163B6
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 244397 bases at least Q40
            Consensus quality: 245102 bases at least Q30
            Consensus quality: 245733 bases at least Q20
            Insert size: 247167; sum-of-contigs
            Insert size: 234721; 4.3% error; agarose-fp
            Quality coverage: 5.83x in Q20 bases; sum-of-contigs Quality
            coverage: 6.34x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     5680: contig of 5680 bp in length
            *     5681     5780: gap of 100 bp
            *     5781    34320: contig of 28540 bp in length
            *    34321    34420: gap of 100 bp
            *    34421    48965: contig of 14545 bp in length
            *    48966    49065: gap of 100 bp
            *    49066    71965: contig of 22900 bp in length
            *    71966    72065: gap of 100 bp
            *    72066    76303: contig of 4238 bp in length
            *    76304    76403: gap of 100 bp
            *    76404   114940: contig of 38537 bp in length
            *   114941   115040: gap of 100 bp
            *   115041   163084: contig of 48044 bp in length
            *   163085   163184: gap of 100 bp
            *   163185   176987: contig of 13803 bp in length
            *   176988   177087: gap of 100 bp
            *   177088   189934: contig of 12847 bp in length
            *   189935   190034: gap of 100 bp
            *   190035   220399: contig of 30365 bp in length
            *   220400   220499: gap of 100 bp
            *   220500   248167: contig of 27668 bp in length.
FEATURES             Location/Qualifiers
     source          1. .248167
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-163B6"
                     /clone_lib="DanioKey"
     misc_feature    1. .5680
```

```
                     /note="assembly_fragment:00062
                     fragment_chain:1"
     misc_feature    5781. .34320
                     /note="assembly_fragment:01271
                     fragment_chain:1"
     misc_feature    34421. .48965
                     /note="assembly_fragment:00097
                     fragment_chain:1"
     misc_feature    49066. .71965
                     /note="assembly_fragment:00433
                     fragment_chain:1"
     misc_feature    72066. .76303
                     /note="assembly_fragment:00035
                     fragment_chain:2"
     misc_feature    76404. .114940
                     /note="assembly_fragment:01575
                     fragment_chain:2"
     misc_feature    115041. .163084
                     /note="assembly_fragment:01979
                     fragment_chain:2"
     misc_feature    163185. .176987
                     /note="assembly_fragment:00318
                     fragment_chain:3"
     misc_feature    177088. .189934
                     /note="assembly_fragment:00207
                     fragment_chain:3"
     misc_feature    190035. .220399
                     /note="assembly_fragment:00968
                     fragment_chain:3"
     misc_feature    220500. .248167
                     /note="assembly_fragment:00686.0"
ORIGIN

  Query Match             7.6%;  Score 353;  DB 2;  Length 248167;
  Best Local Similarity   65.2%;  Pred. No. 3e-60;
  Matches  462;  Conservative    0;  Mismatches  245;  Indels    2;  Gaps    2;

Qy        218 AAAAGGCATTTTATATATATATATATATATATACACACACACACATATATATATAC 277
              ||||  ||  |||||||||||||||||||||||| | | | | | ||||||||||
Db     115544 AAAAAACAGCATATATATATATATATATATATATATATATATATATATATATATAT 115485

Qy        278 ACATATATATACACATATATACATATATACACATATATACACATATATATACACATACAT 337
              | ||||||||| | ||||||| ||||||| | ||||||| | ||||||||| | ||| ||
Db     115484 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 115425

Qy        338 ATGTACACATATATATACACATATGTATACACATATATACACATATATACACACATATAT 397
              || || | ||||||||| | |||| |||| | ||||||| | ||||||| | | ||||||
Db     115424 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 115365

Qy        398 ACACATATATACACACATATATACACATATATACACATATATACACATATACACATATAT 457
              | | ||||||| | | ||||||| | ||||||| | ||||||| | |||||| | ||||||
Db     115364 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 115305

Qy        458 ACACATATATACATATATACACATATATATA-ATATACACACATATATATACACATATAT 516
              | | ||||||| ||||||| | ||||||||| ||||| | | ||||||||| | ||||||
Db     115304 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 115245

Qy        517 ACACACATATATACACATATATACACATATATATACACATATATACACATATATACATAT 576
              | | | ||||||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||
Db     115244 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 115185

Qy        577 ATACACATATATATACATATATACACATATATACATATATACACATATATACATATATAC 636
              ||| | ||||||||| ||||||| | ||||||| ||||||| | ||||||| |||||||
Db     115184 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 115125

Qy        637 ACACATATATACACATACATATACACACACATAGATATACATATATATACACATATATAT 696
              | | ||||||| | ||| ||||| | | | ||| ||||| ||||||||| | ||||||||
Db     115124 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 115065

Qy        697 ACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCA 756
              | ||||||||| ||||||||||
Db     115064 ATATATATATATATATATATATATNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 115005

Qy        757 CTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTT 816

Db     115004 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 114945

Qy        817 GATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA-AGGTAGTT 875
                  |||||       | ||| |||||| | ||   | |  ||| | || |  ||  |
Db     114944 NNNNAAAAATATATATATATATATATAAATATATATATATATATATATATATATAT 114885

Qy        876 AAGAAGGTAACACTATATATATATATAGTGAAAGCAGTGTAAACCTAAA 924
              |   |  ||    |||||||||||||   | |    | || |  || |
Db     114884 ATATATATATATATATATATATATATATATATATATATATATATATATATA 114836


RESULT 43
AL731541
LOCUS       AL731541              178254 bp    DNA     linear   PRI 20-JUN-2002
DEFINITION  Human DNA sequence from clone RP11-472K8 on chromosome 10, complete
            sequence.
ACCESSION   AL731541 AC016823
VERSION     AL731541.6  GI:21537521
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 178254)
  AUTHORS   Heath,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAY-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jun 21, 2002 this sequence version replaced gi:21065395.
            Draft Sequence Produced by Genome Therapeutics Corp, 100 Beaver
            Street, Waltham, MA 02453, USA
            http://www.genomecorp.com
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 10, constructed by the Sanger Centre Chromosome 10
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr10
            RP11-472K8 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6.
FEATURES             Location/Qualifiers
     source          1. .178254
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP11-472K8"
                     /clone_lib="RPCI-11.2"
ORIGIN

  Query Match             7.6%;  Score 352.4;  DB 9;  Length 178254;
  Best Local Similarity   70.2%;  Pred. No. 4.1e-60;
```

```
  Matches  559;  Conservative    0;  Mismatches  226;  Indels   11;  Gaps    6;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||  ||||||||||  ||||||||||||||||||| ||| ||||||||||| ||||||
Db     104371 AGTTTGAGACCAGCCTGGCCAACATGGTGAAACCCCAGCTCCACTAAAAATACAAAAATT 104430

Qy         60 AGACAGGCATGGTGATACACGCCTGTA-----ATCCCAGCTACTTCGGAGGCCGAGGCAG 114
              || ||||||| ||| | ||||||||||     |||| |||||||  |||||| |||||||
Db     104431 AGCCAGGCATTGTGGTGCACGCCTGTAATCTCATCCTAGCTACTCGGGAGGCTGAGGCAG 104490

Qy        115 GAGAATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCA 173
               |||||| ||||||  | | || || ||||||||| | ||| | |||||  |||| ||||
Db     104491 AAGAATCGCTTGAA--TCCGGGAGGCGGAGGTTGCAGTGAGCCCAGATCGCGCCACTGCA 104548

Qy        174 CTCCAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAGGCATTTTATAT 233
              ||||||||  || ||| |||| ||||      | ||||||| | | |          | |
Db     104549 CTCCAGCCTGGGTAAC-AGAGTGAAACCGTGTTTCAGAAAATATACACACACACACACAC 104607

Qy        234 ATATATATATATATATATACACACACACACACATATATATATACACATATATATACACAT 293
              | | | | | ||||| | |||||||||||| | ||||| | | ||||| ||||| |||||
Db     104608 ACACACACACATATACACACACACACACATATATATACACACACACACATATACACACAC 104667

Qy        294 ATATACATATATACACATATATACACATATATATACACATACATATGTACACATATATAT 353
              | ||| ||| | ||||| ||||| ||| | | | | | ||||| |   ||| |||||||
Db     104668 ACATATATACACACACACATATATACACACACACATATATACACACACACATATATATAC 104727

Qy        354 ACACATATGTATACACATATATACACATATATACACACATATATACACATATATACACAC 413
              ||||| || |||| ||| | | ||| ||||||||||||||| | ||||| ||||||||||||
Db     104728 ACACACATATATATACACACACACATATATATACACACACACATACATATATATACACAC 104787

Qy        414 ATATATACACATATATACACATATATACACATATACACATATATACACATATATACATAT 473
              | | ||| | | | | ||||| ||||| ||| | |||||||||||| ||| | | ||||||
Db     104788 ACACATATATACACACACACACATATATACACACACACATATATATATACACACACATAT 104847

Qy        474 ATACACATATATATAATATACACACATATATATACACATATATACACACATATATACACA 533
              ||| ||| | | || ||| ||||||| ||| ||||| |||| |    | ||||||  | |
Db     104848 ATATACACACACAT-ATACACACACACATACATACATATATGTGTGTATATATATGTATA 104906

Qy        534 TATATACACATATATATACACATATATACACATATATACATATATACACATATATATACA 593
              || ||| | || | | || | |||| || | | ||||| || | |  | |||||| || |
Db     104907 TACATATATATGTGTGTATATATATGTATATACATATATATGTGTGTATATATATGTATA 104966

Qy        594 TATATACACATATATACATATATACACATATATACATATATACACACATATATACACATA 653
              || ||| | || | |  ||||||    |||||||  |||||||| | |||| ||   |||
Db     104967 TACATATATATGTGTGTATATATGTGTATATATATGTATATACATATATATGTATGTATA 105026

Qy        654 CATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATA 713
               ||||  | | ||||| ||||  || ||||||    |||||||| ||||||| |   | |
Db     105027 TATATGTATATACATATATATGTATGTATATATGTGTATATATATGTATATACACACACA 105086

Qy        714 TATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAA 773
              |||||||          | | | |||          | |  |  |||  |   || |   |
Db     105087 TATATATATATATATATCCCTGAGATAATAATACTTAAAAAGACTTAGTCCAGTGCCCAG 105146

Qy        774 GAGAGTTTGACAGCAG 789
               | |||  ||   |||
Db     105147 CACAGTGGGAACTCAG 105162


RESULT 44
CR752655
LOCUS       CR752655               204486 bp    DNA     linear   HTG 22-SEP-2004
DEFINITION  Danio rerio clone DKEY-107O6, WORKING DRAFT SEQUENCE, 11 unordered
            pieces.
ACCESSION   CR752655
VERSION     CR752655.5  GI:52546093
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 204486)
  AUTHORS   Burton,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 23, 2004 this sequence version replaced gi:52353233.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK107O6
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 201681 bases at least Q40
            Consensus quality: 202171 bases at least Q30
            Consensus quality: 202651 bases at least Q20
            Insert size: 203486; sum-of-contigs
            Insert size: 204129; 3.5% error; agarose-fp
            Quality coverage: 11.12x in Q20 bases; sum-of-contigs Quality
            coverage: 11.09x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    10267: contig of 10267 bp in length
            *    10268    10367: gap of 100 bp
            *    10368    14914: contig of 4547 bp in length
            *    14915    15014: gap of 100 bp
            *    15015    25928: contig of 10914 bp in length
            *    25929    26028: gap of 100 bp
            *    26029    68196: contig of 42168 bp in length
            *    68197    68296: gap of 100 bp
            *    68297   108045: contig of 39749 bp in length
            *   108046   108145: gap of 100 bp
            *   108146   114266: contig of 6121 bp in length
            *   114267   114366: gap of 100 bp
            *   114367   118079: contig of 3713 bp in length
            *   118080   118179: gap of 100 bp
            *   118180   139935: contig of 21756 bp in length
            *   139936   140035: gap of 100 bp
            *   140036   151139: contig of 11104 bp in length
            *   151140   151239: gap of 100 bp
            *   151240   189955: contig of 38716 bp in length
            *   189956   190055: gap of 100 bp
            *   190056   204486: contig of 14431 bp in length.
FEATURES             Location/Qualifiers
     source          1. .204486
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-107O6"
                     /clone_lib="DanioKey"
     misc_feature    1. .10267
                     /note="assembly_fragment:00170
                     fragment_chain:1"
     misc_feature    10368. .14914
                     /note="assembly_fragment:00029
                     fragment_chain:1"
     misc_feature    15015. .25928
                     /note="assembly_fragment:00569
                     fragment_chain:1"
     misc_feature    26029. .68196
                     /note="assembly_fragment:03030
                     fragment_chain:1"
     misc_feature    68297. .108045
```

```
                     /note="assembly_fragment:02194
                     fragment_chain:1"
     misc_feature    108146. .114266
                     /note="assembly_fragment:00286
                     fragment_chain:2"
     misc_feature    114367. .118079
                     /note="assembly_fragment:00088
                     fragment_chain:2"
     misc_feature    118180. .139935
                     /note="assembly_fragment:00956
                     fragment_chain:2"
     misc_feature    140036. .151139
                     /note="assembly_fragment:00410
                     fragment_chain:2"
     misc_feature    151240. .189955
                     /note="assembly_fragment:01419
                     fragment_chain:2"
     misc_feature    190056. .204486
                     /note="assembly_fragment:00735.0"
ORIGIN

  Query Match             7.6%;  Score 352.4;  DB 2;  Length 204486;
  Best Local Similarity   69.8%;  Pred. No. 4.1e-60;
  Matches  504;  Conservative    0;  Mismatches  216;  Indels    2;  Gaps    2;

Qy        209 AGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACACATA 268
              ||||||||||||||      | |||||||||||||||||||||||| | | | | | |||
Db     107299 AGAAAAAAAAAAAAAAATATATATATATATATATATATATATATATATATATATATATA 107358

Qy        269 TATATATACACATATATATACACATATATACATATATACACATATATACACATATATATA 328
              |||||||| | ||||||||| | ||||||| ||||||| | ||||||| | |||||||||
Db     107359 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 107418

Qy        329 CACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACA 388
              | ||| |||| || | ||||||||| | |||| |||| | ||||||| | ||||||| |
Db     107419 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 107478

Qy        389 CACATATATACACATATATACACACATATATACACATATATACACATATATACACATATA 448
              | |||||||| | ||||||| | | ||||||| | ||||||| | ||||||| | |||||
Db     107479 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 107538

Qy        449 CACATATATACACATATATACATATATACACATATATATA-ATATACACACATATATATA 507
              | ||||||| | ||||||| ||||||| | ||||||||| ||||| | | |||||||||
Db     107539 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 107598

Qy        508 CACATATATACACACATATATACACATATATACACATATATATACACATATATACACATA 567
              | ||||||| | | ||||||| | ||||||| | ||||||||| | ||||||| | |||
Db     107599 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 107658

Qy        568 TATACATATATACACATATATATACATATATACACATATATACATATATACACATATATA 627
              |||| ||||||| | ||||||||| ||||||| | ||||||| ||||||| | |||||||
Db     107659 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 107718

Qy        628 CATATATACACACATATATACACATACATATACACACACATAGATATACATATATATACA 687
              ||||||| | | ||||||| | ||| ||||| | | | ||| ||||| |||||||||
Db     107719 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 107778

Qy        688 CATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGT 747
              |||||||||  |||||||||| ||||||||||          |  |||          |
Db     107779 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 107838

Qy        748 TGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACA 807
              |   ||    |    ||| |   | | | |  | | |    | |||    | |   | |
Db     107839 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 107898

Qy        808 TTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA 867
              | |    |    | |||      | ||| ||| || | ||   | |  ||| | ||
Db     107899 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 107958

Qy        868 -AGGTAGTTAAGAAGGTAACACTATATATATATATAGTGAAAGCAGTGTAAACCTAAAGG 926
              | ||  ||  ||  ||    |||||| |||||||   ||| || | || |  || |
Db     107959 TATATATATATAAATATATATATATATACATATATATATAAATCAATATATATATATATA 108018

Qy        927 AT 928
              ||
Db     108019 AT 108020


RESULT 45
CR392006/c
LOCUS       CR392006              178427 bp    DNA     linear   HTG 23-AUG-2004
DEFINITION  Danio rerio clone DKEYP-57F11, *** SEQUENCING IN PROGRESS ***, 9
            unordered pieces.
ACCESSION   CR392006
VERSION     CR392006.3  GI:51510714
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 178427)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-AUG-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 23, 2004 this sequence version replaced gi:46879107.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zKp57F11
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 173955 bases at least Q40
            Consensus quality: 174654 bases at least Q30
            Consensus quality: 175292 bases at least Q20
            Insert size: 177627; sum-of-contigs
            Insert size: 177796; 2.7% error; agarose-fp
            Quality coverage: 7.97x in Q20 bases; sum-of-contigs Quality
            coverage: 8.05x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    11505: contig of 11505 bp in length
            *    11506    11605: gap of 100 bp
            *    11606    24931: contig of 13326 bp in length
            *    24932    25031: gap of 100 bp
            *    25032    31819: contig of 6788 bp in length
            *    31820    31919: gap of 100 bp
            *    31920    50297: contig of 18378 bp in length
            *    50298    50397: gap of 100 bp
            *    50398    83841: contig of 33444 bp in length
            *    83842    83941: gap of 100 bp
            *    83942   122559: contig of 38618 bp in length
            *   122560   122659: gap of 100 bp
            *   122660   166383: contig of 43724 bp in length
            *   166384   166483: gap of 100 bp
            *   166484   175663: contig of 9180 bp in length
            *   175664   175763: gap of 100 bp
            *   175764   178427: contig of 2664 bp in length.
FEATURES             Location/Qualifiers
     source          1. .178427
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
```

```
                     /clone="DKEYP-57F11"
                     /clone_lib="DanioKeypilot"
     misc_feature    1. .11505
                     /note="assembly_fragment:00415
                     fragment_chain:1"
     misc_feature    11606. .24931
                     /note="assembly_fragment:00266
                     fragment_chain:1"
     misc_feature    25032. .31819
                     /note="assembly_fragment:00041
                     fragment_chain:1"
     misc_feature    31920. .50297
                     /note="assembly_fragment:00579
                     fragment_chain:1"
     misc_feature    50398. .83841
                     /note="assembly_fragment:00848
                     fragment_chain:1"
     misc_feature    83942. .122559
                     /note="assembly_fragment:01340
                     fragment_chain:1"
     misc_feature    122660. .166383
                     /note="assembly_fragment:01877
                     fragment_chain:1"
     misc_feature    166484. .175663
                     /note="assembly_fragment:00124
                     fragment_chain:1"
     misc_feature    175764. .178427
                     /note="assembly_fragment:00006"
ORIGIN

  Query Match             7.6%;  Score 352.2;  DB 2;  Length 178427;
  Best Local Similarity   60.3%;  Pred. No. 4.5e-60;
  Matches  502;  Conservative    0;  Mismatches  328;  Indels    3;  Gaps    1;

Qy        194 GCGAAATGTCATCTCAGAAAAAAAAAAAGGCATTTTATATATATATATATATATATATAC 253
              || ||||  |   | |||  ||||||||   |  ||||||||||||||||||||||||
Db      12310 GCAAAATAACTGGTAAGATGAAAAAAAAAAAAAAATATATATATATATATATATATATAT 12251

Qy        254 ACACACACACACATATATATATACACATATATATACACATATATACATATATACACATAT 313
              | | | | | | ||||||||||| | ||||||||| | ||||||| ||||||| | ||||
Db      12250 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 12191

Qy        314 ATACACATATATATACACATACATATGTACACATATATATACACATATGTATACACATAT 373
              ||| | |||||||||| | ||| |||| || | |||||||||| | |||| |||| | ||||
Db      12190 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 12131

Qy        374 ATACACATATATACACACATATATACACATATATACACACATATATACACATATATACAC 433
              ||| | ||||||| | | ||||||| | ||||||| | | ||||||| | ||||||| |
Db      12130 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 12071

Qy        434 ATATATACACATATACACATATATACACATATATACATATATACACATATATATAATATA 493
              ||||||| | ||||| | |||||||| | ||||||| ||||||| | ||||||   |||||
Db      12070 ATATATATATATATATATATATATATATATATATATATATATATATATATATAT---ATATA 12014

Qy        494 CACACATATATATACACATATATACACACATATATACACATATATACACATATATATACA 553
              | | ||||||||| | ||||||| | | ||||||| | ||||||| | ||||||||| |
Db      12013 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 11954

Qy        554 CATATATACACATATATACATATATACACATATATATACATATATACACATATATACATA 613
              |||||||  | |||||||  |||||||  | |||||||||  |||||||  | |||||||  |||
Db      11953 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 11894

Qy        614 TATACACATATATACATATATACACACATATATACACATACATATACACACACATAGATA 673
              ||||  | |||||||  |||||||  | |  ||||||||  | |||  |||||  | |  | |||  |||
Db      11893 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 11834

Qy        674 TACATATATATACACATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCA 733
              || |||||||||  | |||||||||  |||||||||  ||||||||||        |  |
Db      11833 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 11774

Qy        734 TAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTA 793
              ||          |  |   ||    |   ||| |   | | | |  |  | |    | ||
Db      11773 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 11714

Qy        794 TTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAAT 853
              |    | |   | || |     |       | |||       | ||| ||| || | ||
Db      11713 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 11654

Qy        854 AAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATATATATAGTGAAAGCAGT 913
              | |  ||||    |  ||  ||   |  ||    ||||||||||||
Db      11653 TATATATATCATATATATATATATATATATATATATATATATNNNNNNNNNNN 11594

Qy        914 GTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTATAGAAGAATGGCTAAGATGCCAAA 973

Db      11593 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 11534

Qy        974 GCTCAGTGTATGTGGCAGAGGCATGGTGTAGGGTGTGTCCAGGTTCATATATT 1026
                                      |      | |  |  |  |||||||
Db      11533 NNNNNNNNNNNNNNNNNNNNNNNNNNNTTTATAATATATAATTATATATATT 11481


RESULT 46
CR548632/c
LOCUS       CR548632              184630 bp    DNA     linear   HTG 08-JUL-2004
DEFINITION  Danio rerio clone DKEY-175H17, WORKING DRAFT SEQUENCE, 8 unordered
            pieces.
ACCESSION   CR548632
VERSION     CR548632.3  GI:50080050
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 184630)
  AUTHORS   Burton,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUL-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 8, 2004 this sequence version replaced gi:50057714.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK175H17
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 181205 bases at least Q40
            Consensus quality: 181924 bases at least Q30
            Consensus quality: 182467 bases at least Q20
            Insert size: 183930; sum-of-contigs
            Insert size: 181082; 1.5% error; agarose-fp
            Quality coverage: 10.04x in Q20 bases; sum-of-contigs Quality
            coverage: 10.36x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    14714: contig of 14714 bp in length
            *    14715    14814: gap of 100 bp
            *    14815    95419: contig of 80605 bp in length
            *    95420    95519: gap of 100 bp
            *    95520   132623: contig of 37104 bp in length
            *   132624   132723: gap of 100 bp
            *   132724   165602: contig of 32879 bp in length
            *   165603   165702: gap of 100 bp
            *   165703   169699: contig of 3997 bp in length
```

```
            *   169700   169799: gap of 100 bp
            *   169800   174998: contig of 5199 bp in length
            *   174999   175098: gap of 100 bp
            *   175099   177397: contig of 2299 bp in length
            *   177398   177497: gap of 100 bp
            *   177498   184630: contig of 7133 bp in length.
FEATURES             Location/Qualifiers
     source          1. .184630
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-175H17"
                     /clone_lib="DanioKey"
     misc_feature    1. .14714
                     /note="assembly_fragment:00296
                     fragment_chain:1"
     misc_feature    14815. .95419
                     /note="assembly_fragment:01430
                     fragment_chain:1"
     misc_feature    95520. .132623
                     /note="assembly_fragment:00810
                     fragment_chain:1"
     misc_feature    132724. .165602
                     /note="assembly_fragment:00444
                     fragment_chain:1"
     misc_feature    165703. .169699
                     /note="assembly_fragment:00028
                     fragment_chain:1"
     misc_feature    169800. .174998
                     /note="assembly_fragment:00068
                     fragment_chain:2"
     misc_feature    175099. .177397
                     /note="assembly_fragment:00017
                     fragment_chain:2"
     misc_feature    177498. .184630
                     /note="assembly_fragment:00176.0"
ORIGIN

  Query Match             7.6%;  Score 352.2;  DB 2;  Length 184630;
  Best Local Similarity   69.2%;  Pred. No. 4.5e-60;
  Matches  491;  Conservative    0;  Mismatches  218;  Indels    1;  Gaps    1;

Qy        216 AAAAAAGGCATTTTATATATATATATATATATATACACACACACACATATATATAT 275
              ||| ||   ||||||||||||||||||||||||||| | | | | | ||||||||||
Db        744 AAATAAAAGATTTTATATATATATATATATATATATATATATATATATATATATAT 685

Qy        276 ACACATATATATACACATATATACATATATACACATATATACACATATATATACACATAC 335
              | | ||||||||| | ||||||| ||||||| | ||||||| | ||||||||| | |||
Db        684 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 625

Qy        336 ATATGTACACATATATATACACATATGTATACACATATATACACATATATACACACATAT 395
              |||| || | ||||||||| | |||| |||| | ||||||| | ||||||| | | ||||
Db        624 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 565

Qy        396 ATACACATATATACACACATATATACACATATATACACATATATACACATATACACATAT 455
              ||| | ||||||| | | ||||||| | ||||||| | ||||||| | ||||| | ||||
Db        564 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 505

Qy        456 ATACACATATATACATATATACACATATATATA-ATATACACACATATATATACACATAT 514
              ||| | ||||||| ||||||| | ||||||||| ||||| | | ||||||||| | ||||
Db        504 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 445

Qy        515 ATACACACATATATACACATATATACACATATATATACACATATATACACATATATACAT 574
              ||| | | ||||||| | ||||||| | ||||||||| | ||||||| | ||||||| ||
Db        444 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 385

Qy        575 ATATACACATATATATACATATATACACATATATACATATATACACATATATACATATAT 634
              ||||| | ||||||||| ||||||| | ||||||| ||||||| | ||||||| ||||||
Db        384 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 325

Qy        635 ACACACATATATACACATACATATACACACACATAGATATACATATATATACACATATAT 694
              | | | ||||||| | ||| ||||| | | | ||| ||||| ||||||||| | ||||||
Db        324 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 265

Qy        695 ATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTAC 754
              |||  |||||||| ||||||||||          | |||         | |   ||
Db        264 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 205

Qy        755 CACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGG 814
              |    ||| |   | | | |  | | |    | |||    | |   | || |
Db        204 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 145

Qy        815 TTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGT 874
              |       | |||       | ||| ||| || | ||   | |  ||| | | |  ||
Db        144 ATATATATATATATATATATATATATATATATATATATATATATATNTATTATATATA 85

Qy        875 TAAGAAGGTAACACTATATATATATATAGTGAAAGCAGTGTAAACCTAAA 924
              |||  |  ||    |||| ||||||  |   | |    | || |  || |
Db         84 TAAATATATATATATATANATATATNNATATATATATATATATATNTATA 35


RESULT 47
CR762385
LOCUS       CR762385              195644 bp    DNA     linear   HTG 19-SEP-2004
DEFINITION  Danio rerio clone DKEY-245L10, *** SEQUENCING IN PROGRESS ***, 7
            unordered pieces.
ACCESSION   CR762385
VERSION     CR762385.3  GI:52346294
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 195644)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 19, 2004 this sequence version replaced gi:51988239.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK245L10
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 193251 bases at least Q40
            Consensus quality: 193792 bases at least Q30
            Consensus quality: 194212 bases at least Q20
            Insert size: 195044; sum-of-contigs
            Insert size: 196613; 1.8% error; agarose-fp
            Quality coverage: 8.31x in Q20 bases; sum-of-contigs Quality
            coverage: 8.42x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    38651: contig of 38651 bp in length
            *    38652    38751: gap of 100 bp
            *    38752    58158: contig of 19407 bp in length
            *    58159    58258: gap of 100 bp
            *    58259   117292: contig of 59034 bp in length
            *   117293   117392: gap of 100 bp
            *   117393   123058: contig of 5666 bp in length
            *   123059   123158: gap of 100 bp
            *   123159   150654: contig of 27496 bp in length
```

```
            *   150655   150754: gap of 100 bp
            *   150755   160809: contig of 10055 bp in length
            *   160810   160909: gap of 100 bp
            *   160910   195644: contig of 34735 bp in length.
FEATURES             Location/Qualifiers
     source          1. .195644
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-245L10"
                     /clone_lib="DanioKey"
     misc_feature    1. .38651
                     /note="assembly_fragment:01153
                     fragment_chain:1"
     misc_feature    38752. .58158
                     /note="assembly_fragment:00183
                     fragment_chain:1"
     misc_feature    58259. .117292
                     /note="assembly_fragment:01608
                     fragment_chain:1"
     misc_feature    117393. .123058
                     /note="assembly_fragment:00020
                     fragment_chain:2"
     misc_feature    123159. .150654
                     /note="assembly_fragment:00406
                     fragment_chain:2"
     misc_feature    150755. .160809
                     /note="assembly_fragment:00070.0"
     misc_feature    160910. .195644
                     /note="assembly_fragment:00721"
ORIGIN

  Query Match             7.6%;  Score 352.2;  DB 2;  Length 195644;
  Best Local Similarity   59.1%;  Pred. No. 4.5e-60;
  Matches  518;  Conservative    0;  Mismatches  358;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||||| | | | | | | |||||||||||| | |||||||
Db     122359 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 122418

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db     122419 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 122478

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | |||||||| | ||||||| | | ||||||| | |||||
Db     122479 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 122538

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db     122539 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 122598

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db     122599 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 122658

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db     122659 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 122718

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db     122719 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 122778

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db     122779 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 122838

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||          | |||          | |   ||    |    |||
Db     122839 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 122898

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
              |   | | | |  | | |   | |||   | |   | || |     |     | |
Db     122899 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 122958

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885
              ||       | ||| ||| || | ||   | |  ||| |   |  ||  ||  |    |
Db     122959 TATATATAAATATAAATATATATATATATATATATATAATATATATATATAATAAATA 123018

Qy        886 CACTATATATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAAT 945
              | |||||||||||||| | | |    | || |  ||
Db     123019 AAATATATATATATATAATTATATATATATATAAATATTTNNNNNNNNNNNNNNNNNNN 123078

Qy        946 GTTATAGAAGAATGGCTAAGATGCCAAAGCTCAGTGTATGTGGCAGAGGCATGGTGTAGG 1005

Db     123079 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 123138

Qy       1006 GTGTGTCCAGGTTCATATATTGCATTAAGTGTGAGAACACCCTGGAGTATGAACCAAGAA 1065
                                 |  ||| |  | | |    | | |  | |  |  | ||
Db     123139 NNNNNNNNNNNNNNNNNNNCGTTTTATGGATAATATTTTGCGGAACAAGGCCCTTAAAA 123198

Qy       1066 AATGCAAAAGCCAGAAGTGATGGAGGAAATGAGACAC 1102
              | ||   | |  | |  | || | |   | |  || |
Db     123199 AGTGGGGAGGGAACACCTTATTGGGTCTAGGTAACCC 123235


RESULT 48
CR749751/c
LOCUS       CR749751              224393 bp    DNA     linear   HTG 27-AUG-2004
DEFINITION  Danio rerio clone DKEY-24H18, WORKING DRAFT SEQUENCE, 10 unordered
            pieces.
ACCESSION   CR749751
VERSION     CR749751.4  GI:51591749
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 224393)
  AUTHORS   Burton,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-AUG-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 27, 2004 this sequence version replaced gi:51510864.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK24H18
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 221304 bases at least Q40
            Consensus quality: 221921 bases at least Q30
            Consensus quality: 222377 bases at least Q20
            Insert size: 223493; sum-of-contigs
            Insert size: 225057; 1.0% error; agarose-fp
            Quality coverage: 10.31x in Q20 bases; sum-of-contigs Quality
            coverage: 10.28x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2873: contig of 2873 bp in length
            *     2874     2973: gap of 100 bp
            *     2974    66389: contig of 63416 bp in length
```

```
            *    66390    66489: gap of 100 bp
            *    66490   106977: contig of 40488 bp in length
            *   106978   107077: gap of 100 bp
            *   107078   145207: contig of 38130 bp in length
            *   145208   145307: gap of 100 bp
            *   145308   153395: contig of 8088 bp in length
            *   153396   153495: gap of 100 bp
            *   153496   167269: contig of 13774 bp in length
            *   167270   167369: gap of 100 bp
            *   167370   175339: contig of 7970 bp in length
            *   175340   175439: gap of 100 bp
            *   175440   179204: contig of 3765 bp in length
            *   179205   179304: gap of 100 bp
            *   179305   200083: contig of 20779 bp in length
            *   200084   200183: gap of 100 bp
            *   200184   224393: contig of 24210 bp in length.
FEATURES             Location/Qualifiers
     source          1. .224393
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-24H18"
                     /clone_lib="DanioKey"
     misc_feature    1. .2873
                     /note="assembly_fragment:00089
                     fragment_chain:1"
     misc_feature    2974. .66389
                     /note="assembly_fragment:02891
                     fragment_chain:1"
     misc_feature    66490. .106977
                     /note="assembly_fragment:01374
                     fragment_chain:1"
     misc_feature    107078. .145207
                     /note="assembly_fragment:02083
                     fragment_chain:1"
     misc_feature    145308. .153395
                     /note="assembly_fragment:00265
                     fragment_chain:1"
     misc_feature    153496. .167269
                     /note="assembly_fragment:00404
                     fragment_chain:2"
     misc_feature    167370. .175339
                     /note="assembly_fragment:00155
                     fragment_chain:2"
     misc_feature    175440. .179204
                     /note="assembly_fragment:00108
                     fragment_chain:2"
     misc_feature    179305. .200083
                     /note="assembly_fragment:00671"
     misc_feature    200184. .224393
                     /note="assembly_fragment:00942.0"
ORIGIN

  Query Match             7.6%;  Score 352.2;  DB 2;  Length 224393;
  Best Local Similarity   79.4%;  Pred. No. 4.4e-60;
  Matches  417;  Conservative    0;  Mismatches  108;  Indels    0;  Gaps    0;

Qy        196 GAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACAC 255
              | ||| | || ||   |   ||||||   | | ||||||||||||||||||||||| |
Db      19100 GTAATTTTATTTCTTTAGTTAAAAAAAAAATATATATATATATATATATATATATAT 19041

Qy        256 ACACACACACATATATATATACACATATATATACACATATATACATATATACACATATAT 315
              | | | | | |||||||||||| | ||||||||| | ||||||| ||||||| | ||||||
Db      19040 ATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 18981

Qy        316 ACACATATATATACACATACATATGTACACATATATATACACATATGTATACACATATAT 375
              | | ||||||||| | ||| |||| || | ||||||||| | |||| |||| | ||||||
Db      18980 ATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 18921

Qy        376 ACACATATATACACACATATATACACATATATACACACATATATACACATATATACACAT 435
              | | ||||||| | | ||||||| | ||||||| | | ||||||| | ||||||| | ||
Db      18920 ATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 18861

Qy        436 ATATACACATATACACATATATACACATATATACATATATACACATATATATAATATACA 495
              ||||| | ||||| | ||||||| | ||||||| |||||||| |||||||| ||||| |
Db      18860 ATATATATATATATATATATATATATATATATATATATACATATATATATCATATATA 18801

Qy        496 CACATATATATACACATATATACACACATATATACACATATATACACATATATATACACA 555
              | |||||||||| | ||||||| | |||||||||||| ||||||| | ||||||| | | |
Db      18800 TATATATATATATATATATATATATACATATATACATATATATATATATATACATATA 18741

Qy        556 TATATACACATATATACATATATACACATATATATACATATATACACATATATACATATA 615
              |||||| | ||||||| ||||||| | ||||||||| ||||||| |  |||||| |||||
Db      18740 TATATATATATATATATATATATATATATATATATATATATATATATGTATATATATATA 18681

Qy        616 TACACATATATACATATATACACACATATATACACATACATATACACACACATAGATATA 675
              || | ||||||| ||||||| | | ||||||| | ||| ||||| | | | ||| |||||
Db      18680 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 18621

Qy        676 CATATATATACACATATATATACGTATATATATGTATATATATAT 720
              ||||||||| | |||||||||  |||||||| ||||||||||
Db      18620 TATATATATATATATATATATATATATATATATATATATATATAT 18576


RESULT 49
CR762494
LOCUS       CR762494              238804 bp    DNA     linear   HTG 10-OCT-2004
DEFINITION  Danio rerio clone DKEYP-78B11, WORKING DRAFT SEQUENCE, 7 unordered
            pieces.
ACCESSION   CR762494
VERSION     CR762494.3  GI:52345306
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 238804)
  AUTHORS   McLaren,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 18, 2004 this sequence version replaced gi:52214011.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zKp78B11
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 236234 bases at least Q40
            Consensus quality: 236747 bases at least Q30
            Consensus quality: 237190 bases at least Q20
            Insert size: 238204; sum-of-contigs
            Insert size: 231499; 2.2% error; agarose-fp
            Quality coverage: 11.55x in Q20 bases; sum-of-contigs Quality
            coverage: 11.97x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    21508: contig of 21508 bp in length
            *    21509    21608: gap of 100 bp
            *    21609    47490: contig of 25882 bp in length
            *    47491    47590: gap of 100 bp
            *    47591    54540: contig of 6950 bp in length
            *    54541    54640: gap of 100 bp
```

```
            *    54641   140186: contig of 85546 bp in length
            *   140187   140286: gap of 100 bp
            *   140287   163133: contig of 22847 bp in length
            *   163134   163233: gap of 100 bp
            *   163234   166432: contig of 3199 bp in length
            *   166433   166532: gap of 100 bp
            *   166533   238804: contig of 72272 bp in length.
FEATURES             Location/Qualifiers
     source          1. .238804
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEYP-78B11"
                     /clone_lib="DanioKeypilot"
     misc_feature    1. .21508
                     /note="assembly_fragment:00516
                     fragment_chain:1
                     clone_end:SP6
                     vector_side:left"
     misc_feature    21609. .47490
                     /note="assembly_fragment:00888
                     fragment_chain:1"
     misc_feature    47591. .54540
                     /note="assembly_fragment:00086
                     fragment_chain:2"
     misc_feature    54641. .140186
                     /note="assembly_fragment:01354
                     fragment_chain:2"
     misc_feature    140287. .163133
                     /note="assembly_fragment:00178
                     fragment_chain:2"
     misc_feature    163234. .166432
                     /note="assembly_fragment:00035
                     fragment_chain:3"
     misc_feature    166533. .238804
                     /note="assembly_fragment:02624
                     fragment_chain:3"
ORIGIN

  Query Match             7.6%;  Score 352;  DB 2;  Length 238804;
  Best Local Similarity   68.8%;  Pred. No. 4.7e-60;
  Matches  512;  Conservative    0;  Mismatches  230;  Indels    2;  Gaps    2;

Qy        182 CAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATAT 241
              ||   || || |   |||   |||  || ||||||||||||   |   ||||||||||||
Db     165682 CAAAAAAAAAAAAAAAAAAAACATGACATAAAAAAAAAAAAAAAAAAAATATATATATAT 165741

Qy        242 ATATATATATACACACACACACATATATATATACACATATATATACACATATATACAT 301
              |||||||||| | | | | | | |||||||||| | ||||||||| | ||||||| ||
Db     165742 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 165801

Qy        302 ATATACACATATATACACATATATATACACATACATATGTACACATATATATACACATAT 361
              ||||| | ||||||| | ||||||||| | ||| |||| || | ||||||||| | ||||
Db     165802 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 165861

Qy        362 GTATACACATATATACACATATATACACACATATATACACATATATACACACATATATAC 421
              |||| | ||||||| | ||||||| | | ||||||| | ||||||| | | |||||||
Db     165862 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 165921

Qy        422 ACATATATACACATATATACACATATACACATATATACACATATATACATATATACACAT 481
              | ||||||| | ||||||| | ||||| | ||||||| | ||||||| ||||||| | ||
Db     165922 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 165981

Qy        482 ATATATAATATACACACATATATATACACATATATACACACATATATACACATATATACA 541
              |||||| ||||| | | ||||||||| | ||||||| | | ||||||| | ||||||| |
Db     165982 ATATAT-ATATATATATATATATATATATATATATATATATATATATATATATATATA 166040

Qy        542 CATATATATACACATATATACACATATATACATATATACACATATATATACATATATACA 601
              ||||||||| | ||||||| | ||||||| ||||||| | ||||||||| ||||||| |
Db     166041 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 166100

Qy        602 CATATATACATATATACACATATATACATATATACACACATATATACACATACATATACA 661
              ||||||| ||||||| | ||||||| ||||||| | | ||||||| | ||| ||||| |

Db     166101 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 166160

Qy        662 CACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATATG 721
              | | ||| ||||| ||||||||| | |||||||||  |||||||| ||||||||||
Db     166161 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 166220

Qy        722 CTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTT 781
                  |  |||       |  |   ||    |    ||| |   | | | |  |
Db     166221 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 166280

Qy        782 GACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACA 841
              | |    | |||    | |   | || |     |       | |||       | ||| |
Db     166281 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 166340

Qy        842 TAAATGTGTAATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTAACACTATATATATATA 900
              || || | ||   | |  ||| | || |  ||  ||   |  ||    ||||||||||||
Db     166341 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 166400

Qy        901 TAGTGAAAGCAGTGTAAACCTAAA 924
              ||   |||    | || |  || |
Db     166401 TATATAAATATATATATATATA 166424


RESULT 50
CR790372/c
LOCUS       CR790372              299106 bp    DNA     linear   HTG 05-OCT-2004
DEFINITION  Danio rerio clone DKEY-70K13, *** SEQUENCING IN PROGRESS ***, 25
            unordered pieces.
ACCESSION   CR790372
VERSION     CR790372.4  GI:53828302
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish).
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 299106)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 5, 2004 this sequence version replaced gi:53755940.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK70K13
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 289331 bases at least Q40
            Consensus quality: 291856 bases at least Q30
            Consensus quality: 293478 bases at least Q20
            Insert size: 296706; sum-of-contigs
            Insert size: 225769; 2.9% error; agarose-fp
            Quality coverage: 6.22x in Q20 bases; sum-of-contigs Quality
            coverage: 9.16x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 25 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     3928: contig of 3928 bp in length
            *     3929     4028: gap of 100 bp
            *     4029     8417: contig of 4389 bp in length
            *     8418     8517: gap of 100 bp
```

```
           *     8518    12283: contig of 3766 bp in length
           *    12284    12383: gap of 100 bp
           *    12384    16946: contig of 4563 bp in length
           *    16947    17046: gap of 100 bp
           *    17047    19834: contig of 2788 bp in length
           *    19835    19934: gap of 100 bp
           *    19935    26863: contig of 6929 bp in length
           *    26864    26963: gap of 100 bp
           *    26964    29135: contig of 2172 bp in length
           *    29136    29235: gap of 100 bp
           *    29236    43722: contig of 14487 bp in length
           *    43723    43822: gap of 100 bp
           *    43823   109813: contig of 65991 bp in length
           *   109814   109913: gap of 100 bp
           *   109914   140749: contig of 30836 bp in length
           *   140750   140849: gap of 100 bp
           *   140850   156362: contig of 15513 bp in length
           *   156363   156462: gap of 100 bp
           *   156463   165292: contig of 8830 bp in length
           *   165293   165392: gap of 100 bp
           *   165393   243285: contig of 77893 bp in length
           *   243286   243385: gap of 100 bp
           *   243386   245562: contig of 2177 bp in length
           *   245563   245662: gap of 100 bp
           *   245663   250094: contig of 4432 bp in length
           *   250095   250194: gap of 100 bp
           *   250195   256405: contig of 6211 bp in length
           *   256406   256505: gap of 100 bp
           *   256506   258788: contig of 2283 bp in length
           *   258789   258888: gap of 100 bp
           *   258889   262780: contig of 3892 bp in length
           *   262781   262880: gap of 100 bp
           *   262881   265140: contig of 2260 bp in length
           *   265141   265240: gap of 100 bp
           *   265241   269635: contig of 4395 bp in length
           *   269636   269735: gap of 100 bp
           *   269736   272991: contig of 3256 bp in length
           *   272992   273091: gap of 100 bp
           *   273092   291385: contig of 18294 bp in length
           *   291386   291485: gap of 100 bp
           *   291486   293629: contig of 2144 bp in length
           *   293630   293729: gap of 100 bp
           *   293730   296461: contig of 2732 bp in length
           *   296462   296561: gap of 100 bp
           *   296562   299106: contig of 2545 bp in length.
FEATURES             Location/Qualifiers
     source          1. .299106
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-70K13"
                     /clone_lib="DanioKey"
     misc_feature    1. .3928
                     /note="assembly_fragment:00025
                     fragment_chain:1"
     misc_feature    4029. .8417
                     /note="assembly_fragment:00178
                     fragment_chain:1"
     misc_feature    8518. .12283
                     /note="assembly_fragment:00051
                     fragment_chain:1"
     misc_feature    12384. .16946
                     /note="assembly_fragment:00068
                     fragment_chain:1"
     misc_feature    17047. .19834
                     /note="assembly_fragment:03721
                     fragment_chain:1"
     misc_feature    19935. .26863
                     /note="assembly_fragment:00206
                     fragment_chain:1"
     misc_feature    26964. .29135
                     /note="assembly_fragment:03693
                     fragment_chain:1"
     misc_feature    29236. .43722
                     /note="assembly_fragment:00405
                     fragment_chain:2"
     misc_feature    43823. .109813
                     /note="assembly_fragment:01445
                     fragment_chain:2"
     misc_feature    109914. .140749
                     /note="assembly_fragment:01030
                     fragment_chain:2"
     misc_feature    140850. .156362
                     /note="assembly_fragment:00596
                     fragment_chain:3"
     misc_feature    156463. .165292
                     /note="assembly_fragment:00282
                     fragment_chain:3"
     misc_feature    165393. .243285
                     /note="assembly_fragment:02358
                     fragment_chain:3"
     misc_feature    243386. .245562
                     /note="assembly_fragment:03730
                     fragment_chain:4"
     misc_feature    245663. .250094
                     /note="assembly_fragment:00104
                     fragment_chain:4"
     misc_feature    250195. .256405
                     /note="assembly_fragment:00151
                     fragment_chain:4"
     misc_feature    256506. .258788
                     /note="assembly_fragment:00124
                     fragment_chain:5"
     misc_feature    258889. .262780
                     /note="assembly_fragment:00234
                     fragment_chain:5"
     misc_feature    262881. .265140
                     /note="assembly_fragment:03713
                     fragment_chain:6"
     misc_feature    265241. .269635
                     /note="assembly_fragment:00038
                     fragment_chain:6"
     misc_feature    269736. .272991
                     /note="assembly_fragment:00085"
     misc_feature    273092. .291385
                     /note="assembly_fragment:00801.0"
     misc_feature    291486. .293629
                     /note="assembly_fragment:03699"
     misc_feature    293730. .296461
                     /note="assembly_fragment:03706"
     misc_feature    296562. .299106
                     /note="assembly_fragment:03750"
ORIGIN

  Query Match             7.6%;  Score 351.8;  DB 2;  Length 299106;
  Best Local Similarity   66.8%;  Pred. No. 5e-60;
  Matches  531;  Conservative    0;  Mismatches  262;  Indels    2;  Gaps    2;

Qy        171 GCACTCCAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTA 230
              | | ||||    | || |||   | |   ||   |   | ||||||||||||    | ||
Db     166223 GTATTCCAATTTAAGCTACAGATGGGGCCTGCGCTAAAAAAAAAAAAAAAAAATATATATA 166164

Qy        231 TATATATATATATATATATATACACACACACACATATATATATACACATATATATACA 290
              |||||||||||||||||||| | | | | | |||||||||||| | ||||||||| |
Db     166163 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 166104

Qy        291 CATATATACATATATACACATATATACACATATATATACACATACATATGTACACATATA 350
              ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |||||
Db     166103 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 166044

Qy        351 TATACACATATGTATACACATATATACACATATATACACACATATATACACATATATACA 410
              |||| | |||| |||| | ||||||| | ||||||| | | ||||||| | ||||||| |
Db     166043 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 165984

Qy        411 CACATATATACACATATATACACATATATACACATATACACATATATACACATATATACA 470
              | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | ||||||| |
```

```
Db      165983 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 165924

Qy         471 TATATACACATATATATA-ATATACACACATATATATACACATATATACACACATATATA 529
               |||||| | |||||||||| ||||| | | |||||||||| | ||||||| | | |||||||
Db      165923 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 165864

Qy         530 CACATATATACACATATATATACACATATATACACATATATACATATATACACATATATA 589
               | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||||||
Db      165863 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 165804

Qy         590 TACATATATACACATATATACATATATACACATATATACATATATACACACATATATACA 649
               || ||||||| | ||||||| ||||||| | ||||||| ||||||| | | ||||||| |
Db      165803 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 165744

Qy         650 CATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATATG 709
               ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  |||||||||
Db      165743 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 165684

Qy         710 TATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGG 769
               ||||||||||          | |||         | |   ||    |    ||| |
Db      165683 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 165624

Qy         770 AAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGT 829
               | | | |  | | |    | |||    | |   | || |     |       | |||
Db      165623 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 165564

Qy         830 AGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTAACAC 888
               | ||| ||| || | ||   | |  ||| | || |  ||  ||   |  ||
Db      165563 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 165504

Qy         889 TATATATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTT 948
               |||||||||||||    | |   | || |  || |       |   || ||||| |
Db      165503 TATATATATATATATATATATATATATATATATATATATATATATATATATATAAATATA 165444

Qy         949 ATAGAAGAATGGCTA 963
               || |   ||   ||
Db      165443 TTATATTTATAAATA 165429


RESULT 51
CR762407/c
LOCUS       CR762407              149106 bp    DNA     linear   HTG 13-SEP-2004
DEFINITION  Danio rerio clone DKEY-262N17, *** SEQUENCING IN PROGRESS ***, 5
            unordered pieces.
ACCESSION   CR762407
VERSION     CR762407.2  GI:52000335
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 149106)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 13, 2004 this sequence version replaced gi:51988183.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK262N17
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 147385 bases at least Q40
            Consensus quality: 147795 bases at least Q30
            Consensus quality: 148053 bases at least Q20
            Insert size: 148706; sum-of-contigs
            Insert size: 147741; 3.4% error; agarose-fp
            Quality coverage: 9.30x in Q20 bases; sum-of-contigs Quality
            coverage: 9.36x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    36781: contig of 36781 bp in length
            *    36782    36881: gap of 100 bp
            *    36882    62357: contig of 25476 bp in length
            *    62358    62457: gap of 100 bp
            *    62458    68981: contig of 6524 bp in length
            *    68982    69081: gap of 100 bp
            *    69082    89839: contig of 20758 bp in length
            *    89840    89939: gap of 100 bp
            *    89940   149106: contig of 59167 bp in length.
FEATURES             Location/Qualifiers
     source          1. .149106
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-262N17"
                     /clone_lib="DanioKey"
     misc_feature    1. .36781
                     /note="assembly_fragment:00719
                     fragment_chain:1"
     misc_feature    36882. .62357
                     /note="assembly_fragment:00059
                     fragment_chain:1"
     misc_feature    62458. .68981
                     /note="assembly_fragment:00007"
     misc_feature    69082. .89839
                     /note="assembly_fragment:00379.0"
     misc_feature    89940. .149106
                     /note="assembly_fragment:01240"
ORIGIN

  Query Match             7.6%;  Score 351.6;  DB 2;  Length 149106;
  Best Local Similarity   68.8%;  Pred. No. 6.1e-60;
  Matches  497;  Conservative    0;  Mismatches  224;  Indels    1;  Gaps    1;

Qy         197 AAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACA 256
               || |||  | ||| ||||||||||||   |    | |||||||||||||||||||| | |
Db       90716 AACTGTATTTTCAAAAAAAAAAAAAAAAAAAAAAAAAATATATATATATATATATATATA 90657

Qy         257 CACACACACATATATATATACACATATATATACACATATATACATATATACACATATATA 316
               | | | | |||||||||||| | ||||||||| | ||||||| ||||||| | |||||||
Db       90656 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 90597

Qy         317 CACATATATATACACATACATATGTACACATATATATACACATATGTATACACATATATA 376
               | ||||||||| | ||| |||| || | ||||||||| | |||| |||| | |||||||
Db       90596 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 90537

Qy         377 CACATATATACACACATATATACACATATATACACACATATATACACATATATACACATA 436
               | ||||||| | | ||||||| | ||||||| | | ||||||| | ||||||| | |||
Db       90536 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 90477

Qy         437 TATACACATATACACATATATACACATATATACATATATACACATATATATAATATACAC 496
               |||| | ||||| | ||||||| | ||||||| ||||||| | |||||||| ||||| |
Db       90476 TATATATATATATATATATATATATATATATATATATATATATATATATATATAT-ATATATAT 90418

Qy         497 ACATATATATACACATATATACACACATATATACACATATATACACATATATATACACAT 556
               | ||||||||| | ||||||| | | ||||||| | ||||||| | ||||||||  | ||
Db       90417 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 90358

Qy         557 ATATACACATATATACATATATACACATATATATACATATATACACATATATACATATAT 616
               ||||| | ||||||| ||||||| | ||||||||| ||||||| | ||||||| ||||||
Db       90357 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 90298
```

```
Qy        617 ACACATATATACATATATACACACATATATACACATACATATACACACACATAGATATAC 676
              | | ||||||| ||||||| | | ||||||| | ||| ||||| | | | ||| |||||
Db      90297 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 90238

Qy        677 ATATATATACACATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAA 736
              ||||||||| | |||||||||  |||||||| |||||||||||      |  |||
Db      90237 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 90178

Qy        737 GAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTG 796
                |  |   ||    |    ||| |   | | | |  | |    | |||
Db      90177 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 90118

Qy        797 TGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAA 856
                | |   | || |     |       | |||       | ||| ||| || | ||   |
Db      90117 ATATATATATATATATATATATATATAAATATATATATATATATATATATATATATATAT 90058

Qy        857 ACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATATATATAGTGAAAGCAGTGTA 916
              |  ||| |   |  ||  ||   |  ||    |||||| ||| |||   | |  | | ||
Db      90057 ATATATATAATATATATATATATATATATATACATATATACATAAATACACATAAAATTATA 89998

Qy        917 AA 918
              ||
Db      89997 TA 89996


RESULT 52
CR762414
LOCUS       CR762414             171302 bp    DNA     linear   HTG 13-SEP-2004
DEFINITION  Danio rerio clone DKEY-197K18, *** SEQUENCING IN PROGRESS ***, 6
            unordered pieces.
ACCESSION   CR762414
VERSION     CR762414.2  GI:52000341
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 171302)
  AUTHORS   McLay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 13, 2004 this sequence version replaced gi:51988190.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK197K18
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 168888 bases at least Q40
            Consensus quality: 169406 bases at least Q30
            Consensus quality: 169716 bases at least Q20
            Insert size: 170802; sum-of-contigs
            Insert size: 170040; 4.9% error; agarose-fp
            Quality coverage: 7.32x in Q20 bases; sum-of-contigs Quality
            coverage: 7.59x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    15887: contig of 15887 bp in length
            *    15888    15987: gap of 100 bp
            *    15988    42962: contig of 26975 bp in length
            *    42963    43062: gap of 100 bp
            *    43063    58072: contig of 15010 bp in length
            *    58073    58172: gap of 100 bp
            *    58173    60197: contig of 2025 bp in length
            *    60198    60297: gap of 100 bp
            *    60298   112076: contig of 51779 bp in length
            *   112077   112176: gap of 100 bp
            *   112177   171302: contig of 59126 bp in length.
FEATURES             Location/Qualifiers
     source          1. .171302
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-197K18"
                     /clone_lib="DanioKey"
     misc_feature    1. .15887
                     /note="assembly_fragment:00045
                     fragment_chain:1"
     misc_feature    15988. .42962
                     /note="assembly_fragment:00332
                     fragment_chain:1"
     misc_feature    43063. .58072
                     /note="assembly_fragment:00177
                     fragment_chain:2"
     misc_feature    58173. .60197
                     /note="assembly_fragment:00025
                     fragment_chain:2"
     misc_feature    60298. .112076
                     /note="assembly_fragment:00601.0"
     misc_feature    112177. .171302
                     /note="assembly_fragment:01241"
ORIGIN

  Query Match             7.6%;  Score 351.4;  DB 2;  Length 171302;
  Best Local Similarity   69.0%;  Pred. No. 6.6e-60;
  Matches  486;  Conservative    0;  Mismatches  217;  Indels    1;  Gaps    1;

Qy        226 TTTTATATATATATATATATATATATACACACACACACACATATATATATACACATATAT 285
              || |||||||||||||||||||||||| | | | | | |||||||||| | ||||||
Db      42255 TTATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 42314

Qy        286 ATACACATATATACATATATACACATATATACACATATATATACACATACATATGTACAC 345
              ||| | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || |
Db      42315 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 42374

Qy        346 ATATATATACACATATGTATACACATATATACACATATATACACACATATATACACATAT 405
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | ||||
Db      42375 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 42434

Qy        406 ATACACACATATATACACATATATACACATATATACACATATACACATATATACACATAT 465
              ||| | | ||||||| | |||||||| | ||||| | ||||||| | ||||
Db      42435 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 42494

Qy        466 ATACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACAT 524
              ||| ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | ||
Db      42495 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 42554

Qy        525 ATATACACATATATACACATATATATACACATATATACACATATATACACATATATACACAT 584
              ||||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | ||
Db      42555 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 42614

Qy        585 ATATATACATATATACACATATATACATATATACACATATATACATATATACACACATAT 644
              ||||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | ||||
Db      42615 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 42674

Qy        645 ATACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATAT 704
              ||| | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  |||||
Db      42675 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 42734

Qy        705 ATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGAT 764
              |||| ||||||||||||         | |||         | |   ||    |    ||
```

```
Db      42735 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 42794

Qy        765 AGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAA 824

Db      42795 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 42854

Qy        825 ATAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTA 884

Db      42855 ATATAAAAAAATATATATAAATATATATAAATATATATATNTATATAAAATANAAANNAA 42914

Qy        885 ACACTATATATATATATAGTGAAAGCAGTGTAAACCTAAAGGAT 928

Db      42915 TAAATATAAATTAANATAATAAANNAAATAATNNATTATAATAT 42958


RESULT 53
CR749761
LOCUS       CR749761              216259 bp    DNA     linear   HTG 10-OCT-2004
DEFINITION  Danio rerio clone CH211-215F19, WORKING DRAFT SEQUENCE, 10
            unordered pieces.
ACCESSION   CR749761
VERSION     CR749761.4  GI:54021792
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 216259)
  AUTHORS   McLaren,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 10, 2004 this sequence version replaced gi:51988273.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zC215F19
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 212820 bases at least Q40
            Consensus quality: 213560 bases at least Q30
            Consensus quality: 214140 bases at least Q20
            Insert size: 215359; sum-of-contigs
            Insert size: 205140; 3.4% error; agarose-fp
            Quality coverage: 7.94x in Q20 bases; sum-of-contigs Quality
            coverage: 8.53x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    48219: contig of 48219 bp in length
            *    48220    48319: gap of 100 bp
            *    48320    65856: contig of 17537 bp in length
            *    65857    65956: gap of 100 bp
            *    65957    86815: contig of 20859 bp in length
            *    86816    86915: gap of 100 bp
            *    86916   108880: contig of 21965 bp in length
            *   108881   108980: gap of 100 bp
            *   108981   118104: contig of 9124 bp in length
            *   118105   118204: gap of 100 bp
            *   118205   185027: contig of 66823 bp in length
            *   185028   185127: gap of 100 bp
            *   185128   193579: contig of 8452 bp in length
            *   193580   193679: gap of 100 bp
            *   193680   207482: contig of 13803 bp in length
            *   207483   207582: gap of 100 bp
            *   207583   211897: contig of 4315 bp in length
            *   211898   211997: gap of 100 bp
            *   211998   216259: contig of 4262 bp in length.
FEATURES             Location/Qualifiers
     source          1. .216259
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-215F19"
                     /clone_lib="CHORI-211"
     misc_feature    1. .48219
                     /note="assembly_fragment:01191
                     fragment_chain:1
                     clone_end:T7
                     vector_side:left"
     misc_feature    48320. .65856
                     /note="assembly_fragment:00391
                     fragment_chain:1
                     clone_end:SP6
                     vector_side:right"
     misc_feature    65957. .86815
                     /note="assembly_fragment:00649
                     fragment_chain:2"
     misc_feature    86916. .108880
                     /note="assembly_fragment:00908
                     fragment_chain:2"
     misc_feature    108981. .118104
                     /note="assembly_fragment:00069
                     fragment_chain:2"
     misc_feature    118205. .185027
                     /note="assembly_fragment:01823
                     fragment_chain:2"
     misc_feature    185128. .193579
                     /note="assembly_fragment:00271
                     fragment_chain:2"
     misc_feature    193680. .207482
                     /note="assembly_fragment:00167
                     fragment_chain:3"
     misc_feature    207583. .211897
                     /note="assembly_fragment:00018
                     fragment_chain:3"
     misc_feature    211998. .216259
                     /note="assembly_fragment:00040
                     fragment_chain:3"
ORIGIN

  Query Match             7.6%;  Score 351.4;  DB 2;  Length 216259;
  Best Local Similarity   69.0%;  Pred. No. 6.3e-60;
  Matches  492;  Conservative    0;  Mismatches  220;  Indels    1;  Gaps    1;

Qy        191 AGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATA 250

Db     215544 AGAGATTAATAAGTCTGTAAAAAAAAAAAAAATATATATATATATATATATATATATATA 215603

Qy        251 TACACACACACACATATATATATACACATATATATACACATATATACATATATACACA 310

Db     215604 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 215663

Qy        311 TATATACACATATATATACACATACATATGTACACATATATATACACATATGTATACACA 370

Db     215664 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 215723

Qy        371 TATATACACATATATACACACATATATACACATATATACACACATATATACACATATATA 430

Db     215724 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 215783

Qy        431 CACATATATACACATATACACATATATACACATATATACATATATACACATATATATA-A 489

Db     215784 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 215843
```

```
Qy        490 TATACACACATATATATACACATATATACACACATATATACACATATATACACATATATA 549
              |||| | | ||||||||| | ||||||| | | ||||||| | ||||||| | |||||||
Db     215844 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 215903

Qy        550 TACACATATATACACATATATACATATATACACATATATATACATATATACACATATATA 609
              || | ||||||| | ||||||| ||||||| | ||||||||| ||||||| | |||||||
Db     215904 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 215963

Qy        610 CATATATACACATATATACATATATACACACATATATACACATACATATACACACACATA 669
              ||||||| | ||||||| ||||||| | | ||||||| | ||| ||||| | | | |||
Db     215964 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 216023

Qy        670 GATATACATATATATACACATATATATACGTATATATATGTATATATATATGCTCCAGAG 729
              ||||| ||||||||| | |||||||||  ||||||||| ||||||||||
Db     216024 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 216083

Qy        730 TTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAG 789
              |  |||       |  |   ||    |   ||| |  | | | |  |  | |
Db     216084 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 216143

Qy        790 TGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTG 849
              | |||     | |   | || |      |       | |||       | ||| ||| |  |
Db     216144 TATATATATATATATATATATATATATATATATATATATATATATATATATATAAATA 216203

Qy        850 TAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATATATATATA 902
              || | | |||| | ||    | |  |     ||  | ||||| || |||||
Db     216204 TATTTATATATATTTATATANAAATATAANNANTATAAATATATTTAAATATA 216256


RESULT 54
CR548622
LOCUS       CR548622              200752 bp    DNA     linear   HTG 06-JUL-2004
DEFINITION  Danio rerio clone DKEY-189A9, WORKING DRAFT SEQUENCE, 8 unordered
            pieces.
ACCESSION   CR548622
VERSION     CR548622.2  GI:50057705
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 200752)
  AUTHORS   Burton,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUL-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 8, 2004 this sequence version replaced gi:49658985.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK189A9
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 198295 bases at least Q40
            Consensus quality: 198730 bases at least Q30
            Consensus quality: 199141 bases at least Q20
            Insert size: 200052; sum-of-contigs
            Insert size: 193380; 3.4% error; agarose-fp
            Quality coverage: 9.52x in Q20 bases; sum-of-contigs Quality
            coverage: 9.92x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     5799: contig of 5799 bp in length
            *    5800     5899: gap of 100 bp
            *    5900    61323: contig of 55424 bp in length
            *   61324    61423: gap of 100 bp
            *   61424    94914: contig of 33491 bp in length
            *   94915    95014: gap of 100 bp
            *   95015   169245: contig of 74231 bp in length
            *  169246   169345: gap of 100 bp
            *  169346   171607: contig of 2262 bp in length
            *  171608   171707: gap of 100 bp
            *  171708   174537: contig of 2830 bp in length
            *  174538   174637: gap of 100 bp
            *  174638   183583: contig of 8946 bp in length
            *  183584   183683: gap of 100 bp
            *  183684   200752: contig of 17069 bp in length.
FEATURES             Location/Qualifiers
     source          1. .200752
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-189A9"
                     /clone_lib="DanioKey"
     misc_feature    1. .5799
                     /note="assembly_fragment:00060
                     fragment_chain:1"
     misc_feature    5900. .61323
                     /note="assembly_fragment:00964
                     fragment_chain:1"
     misc_feature    61424. .94914
                     /note="assembly_fragment:00439
                     fragment_chain:1"
     misc_feature    95015. .169245
                     /note="assembly_fragment:01748
                     fragment_chain:2"
     misc_feature    169346. .171607
                     /note="assembly_fragment:00009
                     fragment_chain:2"
     misc_feature    171708. .174537
                     /note="assembly_fragment:00025"
     misc_feature    174638. .183583
                     /note="assembly_fragment:00114"
     misc_feature    183684. .200752
                     /note="assembly_fragment:00228.0"
ORIGIN

  Query Match             7.6%;  Score 351.2;  DB 2;  Length 200752;
  Best Local Similarity   63.0%;  Pred. No. 7e-60;
  Matches  466;  Conservative    0;  Mismatches  273;  Indels    1;  Gaps    1;

Qy        186 CAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATAT 245
              ||| || |   | |  | |    | ||||||||||||   | | ||||||||||||||||||
Db      60737 CAAAAATAATAATAAATAAATAAAAAAAAAAAAAAAAAAAAAATATATATATATATATATAT 60796

Qy        246 ATATATACACACACACACATATATATATACACATATATATACACATATATACATATAT 305
              ||||||| | | | | | | |||||||||||| | ||||||||| | ||||||| ||||||
Db      60797 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 60856

Qy        306 ACACATATATACACATATATATACACATACATATGTACACATATATATACACATATGTAT 365
              | | ||||||| | ||||||||| | ||| |||| || | ||||||||| | |||| |||
Db      60857 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 60916

Qy        366 ACACATATATACACATATATACACACATATATACACATATATACACACATATATACACAT 425
              | | ||||||| | ||||||| | | ||||||| | ||||||| | | ||||||| | ||
Db      60917 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 60976

Qy        426 ATATACACATATATACACATATACACATATATACACATATATACATATATACACATATAT 485
              ||||| | ||||||| | ||||| | ||||||| | ||||||| ||||||| | ||||||
Db      60977 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 61036

Qy        486 ATA-ATATACACACATATATATACACATATATACACACATATATACACATATATACACAT 544
```

```
Db      61037 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 61096

Qy        545 ATATATACACATATATACACATATATACATATATACACATATATATACATATATACACAT 604

Db      61097 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 61156

Qy        605 ATATACATATATACACATATATACATATATACACACATATATACACATACATATACACAC 664

Db      61157 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 61216

Qy        665 ACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATATGCTC 724

Db      61217 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 61276

Qy        725 CAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGAC 784

Db      61277 ATATATATATATATATATATATATATATATATATATATATATATATATATANNNNNNNNNNNNN 61336

Qy        785 AGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAA 844

Db      61337 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 61396

Qy        845 ATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATATATATATAGT 904

Db      61397 NNNNNNNNNNNNNNNNNNNNNNNNNNNATATATATATATATATATATATATATATATA 61456

Qy        905 GAAAGCAGTGTAAACCTAAA 924

Db      61457 TATATATATATATATATATA 61476


RESULT 55
CR786577/c
LOCUS       CR786577              156578 bp    DNA     linear   HTG 17-SEP-2004
DEFINITION  Danio rerio clone CH211-285C10, *** SEQUENCING IN PROGRESS ***, 11
            unordered pieces.
ACCESSION   CR786577
VERSION     CR786577.1  GI:52313179
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 156578)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zC285C10
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 152392 bases at least Q40
            Consensus quality: 153620 bases at least Q30
            Consensus quality: 154402 bases at least Q20
            Insert size: 155578; sum-of-contigs
            Insert size: 156602; 1.5% error; agarose-fp
            Quality coverage: 6.30x in Q20 bases; sum-of-contigs Quality
            coverage: 6.26x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    28552: contig of 28552 bp in length
            *    28553    28652: gap of 100 bp
            *    28653    32448: contig of 3796 bp in length
            *    32449    32548: gap of 100 bp
            *    32549    46601: contig of 14053 bp in length
            *    46602    46701: gap of 100 bp
            *    46702    54418: contig of 7717 bp in length
            *    54419    54518: gap of 100 bp
            *    54519    79156: contig of 24638 bp in length
            *    79157    79256: gap of 100 bp
            *    79257    90677: contig of 11421 bp in length
            *    90678    90777: gap of 100 bp
            *    90778    95922: contig of 5145 bp in length
            *    95923    96022: gap of 100 bp
            *    96023   100602: contig of 4580 bp in length
            *   100603   100702: gap of 100 bp
            *   100703   125015: contig of 24313 bp in length
            *   125016   125115: gap of 100 bp
            *   125116   150124: contig of 25009 bp in length
            *   150125   150224: gap of 100 bp
            *   150225   156578: contig of 6354 bp in length.
FEATURES             Location/Qualifiers
     source          1. .156578
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-285C10"
                     /clone_lib="CHORI-211"
     misc_feature    1. .28552
                     /note="assembly_fragment:01088
                     fragment_chain:1
                     clone_end:SP6
                     vector_side:left"
     misc_feature    28653. .32448
                     /note="assembly_fragment:00002
                     fragment_chain:1"
     misc_feature    32549. .46601
                     /note="assembly_fragment:00430
                     fragment_chain:1"
     misc_feature    46702. .54418
                     /note="assembly_fragment:00210
                     fragment_chain:1"
     misc_feature    54519. .79156
                     /note="assembly_fragment:00576
                     fragment_chain:1"
     misc_feature    79257. .90677
                     /note="assembly_fragment:00303
                     fragment_chain:1"
     misc_feature    90778. .95922
                     /note="assembly_fragment:00082
                     fragment_chain:1"
     misc_feature    96023. .100602
                     /note="assembly_fragment:00039
                     fragment_chain:1"
     misc_feature    100703. .125015
                     /note="assembly_fragment:00813
                     fragment_chain:1"
     misc_feature    125116. .150124
                     /note="assembly_fragment:01376"
     misc_feature    150225. .156578
                     /note="assembly_fragment:00130
                     clone_end:T7
                     vector_side:right"
ORIGIN

  Query Match             7.6%;  Score 351;  DB 2;  Length 156578;
  Best Local Similarity   69.8%;  Pred. No. 8e-60;
  Matches  488;  Conservative    0;  Mismatches  210;  Indels    1;  Gaps    1;
```

```
Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||||| | | | | | | |||||||||||| | |||||||
Db     100582 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 100523

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db     100522 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 100463

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              ||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db     100462 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 100403

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | |. ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db     100402 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 100343

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db     100342 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 100283

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db     100282 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 100223

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db     100222 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 100163

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db     100162 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 100103

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||          | |||          | |   ||    |    |||
Db     100102 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 100043

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
              |   | | | |  |  | |   | |||    | |   | || |     |       | |
Db     100042 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 99983

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885
              ||       | ||| ||| || | ||   | |  ||| | ||      |||| ||||||
Db      99982 TATATATATATATATATATATATATATATATATATATATATATATATATTAATAAGGTAT 99923

Qy        886 CACTATATATATATATAGTGAAAGCAGTGTAAACCTAAA 924
              ||||||||||||||   | |    | || |  || |
Db      99922 ATATATATATATATATATATATATATATATATATATATA 99884
```

```
RESULT 56
CR762422/c
LOCUS       CR762422             159603 bp    DNA     linear   HTG 13-SEP-2004
DEFINITION  Danio rerio clone CH211-184I23, *** SEQUENCING IN PROGRESS ***, 13
            unordered pieces.
ACCESSION   CR762422
VERSION     CR762422.2  GI:52000347
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 159603)
  AUTHORS   McLay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 13, 2004 this sequence version replaced gi:51988198.
            ------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            ------------- Project Information
            Center project name: zC184I23
            ------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 153793 bases at least Q40
            Consensus quality: 154824 bases at least Q30
            Consensus quality: 155824 bases at least Q20
            Insert size: 158403; sum-of-contigs
            Insert size: 158676; 6.9% error; agarose-fp
            Quality coverage: 8.08x in Q20 bases; sum-of-contigs Quality
            coverage: 8.13x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 13 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    33117: contig of 33117 bp in length
            *    33118    33217: gap of 100 bp
            *    33218    38465: contig of 5248 bp in length
            *    38466    38565: gap of 100 bp
            *    38566    47730: contig of 9165 bp in length
            *    47731    47830: gap of 100 bp
            *    47831    83298: contig of 35468 bp in length
            *    83299    83398: gap of 100 bp
            *    83399    86467: contig of 3069 bp in length
            *    86468    86567: gap of 100 bp
            *    86568    88634: contig of 2067 bp in length
            *    88635    88734: gap of 100 bp
            *    88735   103656: contig of 14922 bp in length
            *   103657   103756: gap of 100 bp
            *   103757   120853: contig of 17097 bp in length
            *   120854   120953: gap of 100 bp
            *   120954   132618: contig of 11665 bp in length
            *   132619   132718: gap of 100 bp
            *   132719   146433: contig of 13715 bp in length
            *   146434   146533: gap of 100 bp
            *   146534   152937: contig of 6404 bp in length
            *   152938   153037: gap of 100 bp
            *   153038   156530: contig of 3493 bp in length
            *   156531   156630: gap of 100 bp
            *   156631   159603: contig of 2973 bp in length.
FEATURES             Location/Qualifiers
     source          1. .159603
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-184I23"
                     /clone_lib="CHORI-211"
     misc_feature    1. .33117
                     /note="assembly_fragment:01129
                     fragment_chain:1
                     clone_end:SP6
                     vector_side:left"
     misc_feature    33218. .38465
                     /note="assembly_fragment:00128
                     fragment_chain:1"
     misc_feature    38566. .47730
                     /note="assembly_fragment:00252
                     fragment_chain:1"
     misc_feature    47831. .83298
                     /note="assembly_fragment:01535
                     fragment_chain:1"
     misc_feature    83399. .86467
                     /note="assembly_fragment:00044
                     fragment_chain:1"
     misc_feature    86568. .88634
                     /note="assembly_fragment:00029
```

```
                     fragment_chain:2"
     misc_feature    88735. .103656
                     /note="assembly_fragment:00682
                     fragment_chain:2"
     misc_feature    103757. .120853
                     /note="assembly_fragment:00894"
     misc_feature    120954. .132618
                     /note="assembly_fragment:00349
                     fragment_chain:3"
     misc_feature    132719. .146433
                     /note="assembly_fragment:00481
                     fragment_chain:3"
     misc_feature    146534. .152937
                     /note="assembly_fragment:00172
                     fragment_chain:3"
     misc_feature    153038. .156530
                     /note="assembly_fragment:00068
                     fragment_chain:3"
     misc_feature    156631. .159603
                     /note="assembly_fragment:00094
                     fragment_chain:3
                     clone_end:T7
                     vector_side:right"
ORIGIN

  Query Match             7.6%;  Score 350.8;  DB 2;  Length 159603;
  Best Local Similarity   70.0%;  Pred. No. 8.7e-60;
  Matches  486;  Conservative    0;  Mismatches  207;  Indels    1;  Gaps    1;

Qy        209 AGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACACATA 268
              |||||||||||    | | ||||||||||||||||||||| | | | | | |||
Db     121963 AAAAAAAAAAAAAAAAAAATATATATATATATATATATATATATATATATATATATATA 121904

Qy        269 TATATATACACATATATATACACATATATACATATATACACATATATACACATATATATA 328
              |||||||| | ||||||||| | ||||||| ||||||| | ||||||| | |||||||||
Db     121903 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 121844

Qy        329 CACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACA 388
              | ||| |||| || | ||||||||| | |||| |||| | ||||||| | ||||||| |
Db     121843 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 121784

Qy        389 CACATATATACACATATATACACACATATATACACATATATACACATATATACACATATA 448
              | ||||||| | ||||||| | | ||||||| | ||||||| | ||||||| | |||||
Db     121783 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 121724

Qy        449 CACATATATACACATATATACATATATACACATATATATA-ATATACACACATATATATA 507
              | ||||||| | ||||||| ||||||| | ||||||||| ||||| | | |||||||||
Db     121723 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 121664

Qy        508 CACATATATACACACATATATACACATATATACACATATATATACACATATATACACATA 567
              | ||||||| | | ||||||| | ||||||| | ||||||||| | ||||||| | |||
Db     121663 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 121604

Qy        568 TATACATATATACACATATATATACATATATACACATATATACATATATACACATATATA 627
              |||| ||||||| | ||||||||| ||||||| | ||||||| ||||||| | |||||||
Db     121603 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 121544

Qy        628 CATATATACACACATATATACACATACATATACACACACATAGATATACATATATATACA 687
              ||||||| | | ||||||| | ||| ||||| | | | ||| ||||| |||||||||
Db     121543 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 121484

Qy        688 CATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGT 747
              |||||||||  |||||||||| ||||||||||           |||          |
Db     121483 TATATATATATATATATATATATATATATATATATATATATAAATATATATATATATATA 121424

Qy        748 TGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACA 807
              |  ||    |   ||| |  | | | |  | | |    | |||   | |   | |
Db     121423 TATATATATATATATATATATATACATATATATATATAAATATATATATATATATATAAA 121364

Qy        808 TTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA 867
              | |     |     | |||       | ||| |||||| | |||||| |  ||| | ||
Db     121363 TATATATATATATATATATATATATATATATATAAATATATAATAATAAATATTTATA 121304
```

```
Qy        868 AGGTAGTTAAGAAGGTAACACTATATATATATAT 901
              |  |  |  | ||  ||  |    |||   | ||
Db     121303 ATATTATATATAATTTATAATATAATAATAAAAT 121270


RESULT 57
CR762498/c
LOCUS       CR762498              200607 bp    DNA     linear   HTG 10-OCT-2004
DEFINITION  Danio rerio clone DKEY-4J1, WORKING DRAFT SEQUENCE, 9 unordered
            pieces.
ACCESSION   CR762498
VERSION     CR762498.4  GI:52346306
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 200607)
  AUTHORS   McLaren,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 19, 2004 this sequence version replaced gi:52345308.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK4J1
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 196926 bases at least Q40
            Consensus quality: 197654 bases at least Q30
            Consensus quality: 198332 bases at least Q20
            Insert size: 199807; sum-of-contigs
            Insert size: 210583; 10.0% error; agarose-fp
            Quality coverage: 10.81x in Q20 bases; sum-of-contigs Quality
            coverage: 10.33x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    18795: contig of 18795 bp in length
            *    18796    18895: gap of 100 bp
            *    18896    73470: contig of 54575 bp in length
            *    73471    73570: gap of 100 bp
            *    73571    76405: contig of 2835 bp in length
            *    76406    76505: gap of 100 bp
            *    76506    89061: contig of 12556 bp in length
            *    89062    89161: gap of 100 bp
            *    89162   100153: contig of 10992 bp in length
            *   100154   100253: gap of 100 bp
            *   100254   157131: contig of 56878 bp in length
            *   157132   157231: gap of 100 bp
            *   157232   161046: contig of 3815 bp in length
            *   161047   161146: gap of 100 bp
            *   161147   167068: contig of 5922 bp in length
            *   167069   167168: gap of 100 bp
            *   167169   200607: contig of 33439 bp in length.
FEATURES             Location/Qualifiers
     source          1. .200607
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
```

```
                     /clone="DKEY-4J1"
                     /clone_lib="DanioKey"
     misc_feature    1. .18795
                     /note="assembly_fragment:00545
                     fragment_chain:1"
     misc_feature    18896. .73470
                     /note="assembly_fragment:02336
                     fragment_chain:1"
     misc_feature    73571. .76405
                     /note="assembly_fragment:00024
                     fragment_chain:1"
     misc_feature    76506. .89061
                     /note="assembly_fragment:00349
                     fragment_chain:1"
     misc_feature    89162. .100153
                     /note="assembly_fragment:00174
                     fragment_chain:2"
     misc_feature    100254. .157131
                     /note="assembly_fragment:01357
                     fragment_chain:2"
     misc_feature    157232. .161046
                     /note="assembly_fragment:00051.0"
     misc_feature    161147. .167068
                     /note="assembly_fragment:00099"
     misc_feature    167169. .200607
                     /note="assembly_fragment:00747"
ORIGIN

  Query Match             7.6%;  Score 350.6;  DB 2;  Length 200607;
  Best Local Similarity   62.3%;  Pred. No. 9.2e-60;
  Matches  472;  Conservative    0;  Mismatches  285;  Indels    1;  Gaps    1;

Qy        197 AAATGTCATCTCAGAAAAAAAAAAAGGCATTTTATATATATATATATATATATATACACA 256
              |||||| |    | ||||||| |      | ||||||||||||||||||||||| | |
Db      77043 AAATGTAAAAAAAAAAAAAAAATATATATATATATATATATATATATATATATATATA 76984

Qy        257 CACACACACATATATATATACACATATATATACACATATATACATATATACACATATATA 316
              | | | | |||||||||||| | ||||||||| | ||||||| ||||||| | |||||||
Db      76983 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 76924

Qy        317 CACATATATATACACATACATATGTACACATATATATACACATATGTATACACATATATA 376
              | ||||||||| | ||| |||| || | ||||||||| | |||| |||| | |||||||
Db      76923 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 76864

Qy        377 CACATATATACACACATATATACACATATATACACACATATATACACATATATACACATA 436
              | ||||||| | | ||||||| | ||||||| | | ||||||| | ||||||| | |||
Db      76863 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 76804

Qy        437 TATACACATATACACATATATACACATATATACATATATACACATATATATAATATACAC 496
              |||| | ||||| | ||||||| | ||||||| ||||||| | |||||||| ||||| |
Db      76803 TATATATATATATATATATATATATATATATATATATATATATATATATATAT-ATATATAT 76745

Qy        497 ACATATATATACACATATATACACACATATATACACATATATACACATATATATACACAT 556
              | ||||||||| | ||||||| | | ||||||| | ||||||| | ||||||||| | ||
Db      76744 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 76685

Qy        557 ATATACACATATATACATATATACACATATATATACATATATACACATATATACATATAT 616
              ||||| | ||||||| ||||||| | ||||||||| ||||||| | ||||||| ||||||
Db      76684 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 76625

Qy        617 ACACATATATACATATATACACACATATATACACATACATATACACACACATAGATATAC 676
              | | ||||||| ||||||| | | ||||||| | ||| ||||| | | | ||| |||||
Db      76624 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 76565

Qy        677 ATATATATACACATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAA 736
              ||||||||| | |||||||||  |||||||||  ||||||||||          |  |||
Db      76564 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAN 76505

Qy        737 GAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTG 796

Db      76504 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 76445

Qy        797 TGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAA 856
```

```
                                                        ||| ||| || | ||   |
Db      76444 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTATATATATATATATATAT 76385

Qy        857 ACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATATATATATAGTGAAAGCAGTGTA 916
              |  ||| | ||       |  | |  |     | | | | || | |  |  ||   |
Db      76384 ATATATATATATATATATATATATTTTTTTTTTTTTTTNTCTTTACAATATATACAAATTT 76325

Qy        917 AACCTAAAGGATGGGCCAAGGATTTAAATGTTATAGAA 954
              |    | |          |   || |||  | |||| ||
Db      76324 ACAAAATATATATATATATATATATAATGGATATATAA 76287


RESULT 58
CR589875/c
LOCUS       CR589875              227441 bp    DNA     linear   HTG 19-JUL-2004
DEFINITION  Danio rerio clone DKEY-15K16, WORKING DRAFT SEQUENCE, 17 unordered
            pieces.
ACCESSION   CR589875
VERSION     CR589875.2  GI:50399773
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 227441)
  AUTHORS   Burton,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JUL-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK15K16
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 221759 bases at least Q40
            Consensus quality: 222952 bases at least Q30
            Consensus quality: 223650 bases at least Q20
            Insert size: 225841; sum-of-contigs
            Insert size: 225442; 2.6% error; agarose-fp
            Quality coverage: 7.98x in Q20 bases; sum-of-contigs Quality
            coverage: 8.44x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 17 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     7750: contig of 7750 bp in length
            *     7751     7850: gap of 100 bp
            *     7851    17808: contig of 9958 bp in length
            *    17809    17908: gap of 100 bp
            *    17909    26300: contig of 8392 bp in length
            *    26301    26400: gap of 100 bp
            *    26401    37193: contig of 10793 bp in length
            *    37194    37293: gap of 100 bp
            *    37294    42790: contig of 5497 bp in length
            *    42791    42890: gap of 100 bp
            *    42891    75037: contig of 32147 bp in length
            *    75038    75137: gap of 100 bp
            *    75138    77164: contig of 2027 bp in length
            *    77165    77264: gap of 100 bp
            *    77265    93266: contig of 16002 bp in length
            *    93267    93366: gap of 100 bp
```

```
            *    93367   109748: contig of 16382 bp in length
            *   109749   109848: gap of 100 bp
            *   109849   156408: contig of 46560 bp in length
            *   156409   156508: gap of 100 bp
            *   156509   168441: contig of 11933 bp in length
            *   168442   168541: gap of 100 bp
            *   168542   193830: contig of 25289 bp in length
            *   193831   193930: gap of 100 bp
            *   193931   212755: contig of 18825 bp in length
            *   212756   212855: gap of 100 bp
            *   212856   215964: contig of 3109 bp in length
            *   215965   216064: gap of 100 bp
            *   216065   218399: contig of 2335 bp in length
            *   218400   218499: gap of 100 bp
            *   218500   221549: contig of 3050 bp in length
            *   221550   221649: gap of 100 bp
            *   221650   227441: contig of 5792 bp in length.
FEATURES             Location/Qualifiers
     source          1. .227441
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-15K16"
                     /clone_lib="DanioKey"
     misc_feature    1. .7750
                     /note="assembly_fragment:00294
                     fragment_chain:1"
     misc_feature    7851. .17808
                     /note="assembly_fragment:00581
                     fragment_chain:1"
     misc_feature    17909. .26300
                     /note="assembly_fragment:00382
                     fragment_chain:1"
     misc_feature    26401. .37193
                     /note="assembly_fragment:00471
                     fragment_chain:1"
     misc_feature    37294. .42790
                     /note="assembly_fragment:00222
                     fragment_chain:1"
     misc_feature    42891. .75037
                     /note="assembly_fragment:02032
                     fragment_chain:1"
     misc_feature    75138. .77164
                     /note="assembly_fragment:00092
                     fragment_chain:1"
     misc_feature    77265. .93266
                     /note="assembly_fragment:01346
                     fragment_chain:1"
     misc_feature    93367. .109748
                     /note="assembly_fragment:00863
                     fragment_chain:1"
     misc_feature    109849. .156408
                     /note="assembly_fragment:02587
                     fragment_chain:1"
     misc_feature    156509. .168441
                     /note="assembly_fragment:00719
                     fragment_chain:1"
     misc_feature    168542. .193830
                     /note="assembly_fragment:01629
                     fragment_chain:1"
     misc_feature    193931. .212755
                     /note="assembly_fragment:01104
                     fragment_chain:2"
     misc_feature    212856. .215964
                     /note="assembly_fragment:00130
                     fragment_chain:2"
     misc_feature    216065. .218399
                     /note="assembly_fragment:00077"
     misc_feature    218500. .221549
                     /note="assembly_fragment:00107.0"
     misc_feature    221650. .227441
                     /note="assembly_fragment:00155.0"
ORIGIN
```

```
  Query Match             7.6%;  Score 350.4;  DB 2;  Length 227441;
  Best Local Similarity   64.4%;  Pred. No. 9.9e-60;
  Matches  466;  Conservative    0;  Mismatches  256;  Indels    2;  Gaps    2;

Qy        227 TTTATATATATATATATATATATATACACACACACACATATATATATACACATATATA 286
              | |||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db      18405 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 18346

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db      18345 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 18286

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db      18285 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 18226

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db      18225 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 18166

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db      18165 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 18106

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db      18105 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 18046

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db      18045 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 17986

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db      17985 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 17926

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||||
Db      17925 TATATATATATATATAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 17866

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
                                                                     |||
Db      17865 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNAAA 17806

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTA 884
              ||       | ||| ||| || | ||   | |  ||| | || |  ||  ||   |  ||
Db      17805 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 17746

Qy        885 ACACTATATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAA 944
                |||| |||||||| |  | |  || | |  |       || |||   | ||   |
Db      17745 TATATATATGTATATATATTCTATGATGTTTTAGTCATTTTTGTGTGCCTCTGTTTCTCA 17686

Qy        945 TGTT 948
              | ||
Db      17685 TATT 17682


RESULT 59
CR788315
LOCUS       CR788315             153947 bp    DNA     linear   HTG 24-SEP-2004
DEFINITION  Danio rerio clone DKEY-64E1, *** SEQUENCING IN PROGRESS ***, 14
            unordered pieces.
ACCESSION   CR788315
VERSION     CR788315.3  GI:52673623
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 153947)
  AUTHORS   Sims,S.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (23-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 24, 2004 this sequence version replaced gi:52626672.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK64E1
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 150089 bases at least Q40
            Consensus quality: 150989 bases at least Q30
            Consensus quality: 151632 bases at least Q20
            Insert size: 152647; sum-of-contigs
            Insert size: 169881; 7.1% error; agarose-fp
            Quality coverage: 10.12x in Q20 bases; sum-of-contigs Quality
            coverage: 9.68x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 14 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     9129: contig of 9129 bp in length
            *     9130     9229: gap of 100 bp
            *     9230    14190: contig of 4961 bp in length
            *    14191    14290: gap of 100 bp
            *    14291    46311: contig of 32021 bp in length
            *    46312    46411: gap of 100 bp
            *    46412    54860: contig of 8449 bp in length
            *    54861    54960: gap of 100 bp
            *    54961    62361: contig of 7401 bp in length
            *    62362    62461: gap of 100 bp
            *    62462    70465: contig of 8004 bp in length
            *    70466    70565: gap of 100 bp
            *    70566    85672: contig of 15107 bp in length
            *    85673    85772: gap of 100 bp
            *    85773    90034: contig of 4262 bp in length
            *    90035    90134: gap of 100 bp
            *    90135   104444: contig of 14310 bp in length
            *   104445   104544: gap of 100 bp
            *   104545   111290: contig of 6746 bp in length
            *   111291   111390: gap of 100 bp
            *   111391   123791: contig of 12401 bp in length
            *   123792   123891: gap of 100 bp
            *   123892   125954: contig of 2063 bp in length
            *   125955   126054: gap of 100 bp
            *   126055   129586: contig of 3532 bp in length
            *   129587   129686: gap of 100 bp
            *   129687   153947: contig of 24261 bp in length.
FEATURES             Location/Qualifiers
     source          1. .153947
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-64E1"
                     /clone_lib="DanioKey"
     misc_feature    1. .9129
                     /note="assembly_fragment:00638
                     fragment_chain:1"
     misc_feature    9230. .14190
                     /note="assembly_fragment:00181
                     fragment_chain:1"
     misc_feature    14291. .46311
                     /note="assembly_fragment:01886
                     fragment_chain:1"
     misc_feature    46412. .54860
                     /note="assembly_fragment:00323
                     fragment_chain:1"
     misc_feature    54961. .62361
                     /note="assembly_fragment:00252
                     fragment_chain:1"
     misc_feature    62462. .70465
                     /note="assembly_fragment:00518
                     fragment_chain:1"
     misc_feature    70566. .85672
                     /note="assembly_fragment:00985
                     fragment_chain:1"
     misc_feature    85773. .90034
                     /note="assembly_fragment:00086
                     fragment_chain:1"
     misc_feature    90135. .104444
                     /note="assembly_fragment:00797
                     fragment_chain:1"
     misc_feature    104545. .111290
                     /note="assembly_fragment:00417
                     fragment_chain:2"
     misc_feature    111391. .123791
                     /note="assembly_fragment:01185
                     fragment_chain:2"
     misc_feature    123892. .125954
                     /note="assembly_fragment:00065"
     misc_feature    126055. .129586
                     /note="assembly_fragment:00133.0"
     misc_feature    129687. .153947
                     /note="assembly_fragment:01427"
ORIGIN

  Query Match             7.6%;  Score 350.2;  DB 2;  Length 153947;
  Best Local Similarity   64.4%;  Pred. No. 1.2e-59;
  Matches  450;  Conservative    0;  Mismatches  248;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db     123283 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 123342

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db     123343 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 123402

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db     123403 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 123462

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db     123463 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 123522

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db     123523 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 123582

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db     123583 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 123642

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db     123643 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 123702

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db     123703 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 123762

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              |||  ||||||||||          |  ||
Db     123763 TATAAATATATATATATAAATATATATATNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 123822
```

```
Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825

Db     123823 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 123882

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885
                  |   || || | || |||||  ||| | ||   || ||    |  |||

Db     123883 NNNNNNNNNTTTTTTTTTATATATATATTAAAAAATATATATATATTATTTTTATATATAA 123942

Qy        886 CACTATATATATATATAGTGAAAGCAGTGTAAACCTAAA 924
                | ||||||||||||   | |    | || |  || |

Db     123943 ATAAAAATATATATATATATATATATATATATATATA 123981


RESULT 60
CR788302/c
LOCUS       CR788302              175303 bp    DNA     linear   HTG 24-SEP-2004
DEFINITION  Danio rerio clone DKEYP-38A10, *** SEQUENCING IN PROGRESS ***, 9
            unordered pieces.
ACCESSION   CR788302
VERSION     CR788302.3  GI:52673620
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 175303)
  AUTHORS   McLay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 24, 2004 this sequence version replaced gi:52420996.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zKp38A10
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 173044 bases at least Q40
            Consensus quality: 173386 bases at least Q30
            Consensus quality: 173663 bases at least Q20
            Insert size: 174503; sum-of-contigs
            Insert size: 174235; 1.3% error; agarose-fp
            Quality coverage: 8.55x in Q20 bases; sum-of-contigs Quality
            coverage: 8.62x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     6846: contig of 6846 bp in length
            *     6847     6946: gap of 100 bp
            *     6947    39633: contig of 32687 bp in length
            *    39634    39733: gap of 100 bp
            *    39734    44270: contig of 4537 bp in length
            *    44271    44370: gap of 100 bp
            *    44371    81545: contig of 37175 bp in length
            *    81546    81645: gap of 100 bp
            *    81646    84574: contig of 2929 bp in length
            *    84575    84674: gap of 100 bp
            *    84675   103084: contig of 18410 bp in length
            *   103085   103184: gap of 100 bp
            *   103185   122912: contig of 19728 bp in length
            *   122913   123012: gap of 100 bp
            *   123013   130331: contig of 7319 bp in length
            *   130332   130431: gap of 100 bp
            *   130432   175303: contig of 44872 bp in length.
FEATURES             Location/Qualifiers
     source          1. .175303
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEYP-38A10"
                     /clone_lib="DanioKeypilot"
     misc_feature    1. .6846
                     /note="assembly_fragment:00101
                     fragment_chain:1"
     misc_feature    6947. .39633
                     /note="assembly_fragment:01279
                     fragment_chain:1"
     misc_feature    39734. .44270
                     /note="assembly_fragment:00052
                     fragment_chain:1"
     misc_feature    44371. .81545
                     /note="assembly_fragment:00801
                     fragment_chain:1"
     misc_feature    81646. .84574
                     /note="assembly_fragment:00024
                     fragment_chain:1"
     misc_feature    84675. .103084
                     /note="assembly_fragment:00248
                     fragment_chain:2"
     misc_feature    103185. .122912
                     /note="assembly_fragment:00522
                     fragment_chain:2"
     misc_feature    123013. .130331
                     /note="assembly_fragment:00171.0"
     misc_feature    130432. .175303
                     /note="assembly_fragment:01761"
ORIGIN

  Query Match             7.6%;  Score 350.2;  DB 2;  Length 175303;
  Best Local Similarity   63.5%;  Pred. No. 1.1e-59;
  Matches  459;  Conservative    0;  Mismatches  263;  Indels    1;  Gaps    1;

Qy        202 TCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACAC 261
                | | ||   ||||||||||||     | ||||||||||||||||||||||||| | | | |

Db     123573 TTACCTTTAAAAAAAAAAAAATATATATATATATATATATATATATATATATATATAT 123514

Qy        262 ACACATATATATATACACATATATATACACATATATACATATATACACATATATACACAT 321
                | | |||||||||||| | ||||||||| | ||||||| ||||||| | ||||||| | ||

Db     123513 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 123454

Qy        322 ATATATACACATACATATGTACACATATATATACACATATGTATACACATATATACACAT 381
                ||||||| | ||| |||| || | ||||||||| | |||| |||| | ||||||| | ||

Db     123453 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 123394

Qy        382 ATATACACACATATATACACATATATACACACATATATACACATATATACACATATATAC 441
                ||||| | | ||||||| | ||||||| | |.||||||| | ||||||| | |||||||

Db     123393 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 123334

Qy        442 ACATATACACATATATACACATATATACATATATACACATATATATAATATACACACATA 501
                | ||||| | ||||||| | ||||||| ||||||| | |||||||| ||||| | | |||

Db     123333 ATATATATATATATATATATATATATATATATATATATATATATATAT-ATATATATATA 123275

Qy        502 TATATACACATATATACACACATATATACACATATATACACATATATATACACATATATA 561
                |||||| | ||||||| | | ||||||| | ||||||| | ||||||||| | |||||||

Db     123274 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 123215

Qy        562 CACATATATACATATATACACATATATATACATATATACACATATATACATATATACACA 621
                | ||||||| ||||||| | ||||||||| ||||||| | ||||||| ||||||| | |

Db     123214 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 123155

Qy        622 TATATACATATATACACACATATATACACATACATATACACACACATAGATATACATATA 681
                |||||| ||||||| | | ||||||| | ||| ||||| | | | ||| ||||| |||||

Db     123154 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 123095
```

```
Qy        682 TATACACATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGT 741
              |||| | |||||||||  |||||||| ||||||||||          |  |||  |
Db     123094 TATATATATATATATATATATATATATATATATATATATATATATATATATATAAATA 123035

Qy        742 AGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGA 801
                |  |   ||    |   |
Db     123034 TATATATATATATATATAAAAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 122975

Qy        802 AGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTA 861

Db     122974 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 122915

Qy        862 TCTGTAAGGTAGTTAAGAAGGTAACACTATATATATATATAGTGAAAGCAGTGTAAACCT 921
              | |||  |  ||    |  ||   |||||||||||||   | |    | || |  |
Db     122914 NNTTTAATATTATTTTTTATATATATATATATATATATATATATATATATATATATAT 122855

Qy        922 AAA 924
              |||
Db     122854 AAA 122852


RESULT 61
CR753838
LOCUS       CR753838              204662 bp    DNA     linear   HTG 10-OCT-2004
DEFINITION  Danio rerio clone DKEY-189H5, WORKING DRAFT SEQUENCE, 10 unordered
            pieces.
ACCESSION   CR753838
VERSION     CR753838.3  GI:54021757
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 204662)
  AUTHORS   McLaren,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 10, 2004 this sequence version replaced gi:51592199.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK189H5
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 200213 bases at least Q40
            Consensus quality: 201277 bases at least Q30
            Consensus quality: 202286 bases at least Q20
            Insert size: 203762; sum-of-contigs
            Insert size: 192359; 2.6% error; agarose-fp
            Quality coverage: 8.93x in Q20 bases; sum-of-contigs Quality
            coverage: 10.24x in Q20 bases; agarose-fp
            ----------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    54213: contig of 54213 bp in length
            *    54214    54313: gap of 100 bp
            *    54314    67044: contig of 12731 bp in length
            *    67045    67144: gap of 100 bp
            *    67145   118416: contig of 51272 bp in length
            *   118417   118516: gap of 100 bp
            *   118517   132097: contig of 13581 bp in length
            *   132098   132197: gap of 100 bp
            *   132198   135006: contig of 2809 bp in length
            *   135007   135106: gap of 100 bp
            *   135107   137155: contig of 2049 bp in length
            *   137156   137255: gap of 100 bp
            *   137256   139284: contig of 2029 bp in length
            *   139285   139384: gap of 100 bp
            *   139385   141755: contig of 2371 bp in length
            *   141756   141855: gap of 100 bp
            *   141856   146655: contig of 4800 bp in length
            *   146656   146755: gap of 100 bp
            *   146756   204662: contig of 57907 bp in length.
FEATURES             Location/Qualifiers
     source          1. .204662
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-189H5"
                     /clone_lib="DanioKey"
     misc_feature    1. .54213
                     /note="assembly_fragment:01901
                     fragment_chain:1"
     misc_feature    54314. .67044
                     /note="assembly_fragment:00280
                     fragment_chain:1"
     misc_feature    67145. .118416
                     /note="assembly_fragment:00441
                     fragment_chain:1"
     misc_feature    118517. .132097
                     /note="assembly_fragment:00148
                     fragment_chain:1"
     misc_feature    132198. .135006
                     /note="assembly_fragment:00046
                     fragment_chain:2"
     misc_feature    135107. .137155
                     /note="assembly_fragment:00042
                     fragment_chain:2"
     misc_feature    137256. .139284
                     /note="assembly_fragment:00024"
     misc_feature    139385. .141755
                     /note="assembly_fragment:00059"
     misc_feature    141856. .146655
                     /note="assembly_fragment:00081.0"
     misc_feature    146756. .204662
                     /note="assembly_fragment:01145"
ORIGIN

  Query Match             7.6%;  Score 350.2;  DB 2;  Length 204662;
  Best Local Similarity   68.7%;  Pred. No. 1.1e-59;
  Matches  497;  Conservative    0;  Mismatches  223;  Indels    3;  Gaps    1;

Qy        183 AGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATA 242
              | | ||||| |   |||     ||   | || ||||  |       | ||||||||||||||
Db     203771 ATGAAACAATAAAAAAAAACAATAAAAAAATAAAATTATATATATATATATATATATATA 203830

Qy        243 TATATATATACACACACACACATATATATATACACATATATATACACATATATACATA 302
              |||||||||| | | | | | |||||||||||| | |||||||| | ||||||| |||
Db     203831 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 203890

Qy        303 TATACACATATATACACATATATATACACATACATATGTACACATATATATACACATATG 362
              |||| | ||||||| | |||||||| | ||| |||| || | |||||||| | ||||
Db     203891 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 203950

Qy        363 TATACACATATATACACATATATACACACATATATACACATATATACACACATATATACA 422
              |||| | ||||||| | ||||||| | | ||||||| | ||||||| | | ||||||| |
Db     203951 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 204010

Qy        423 CATATATACACATATATACACATATACACATATATACACATATATACATATATACACATA 482
              ||||||| | ||||||| | ||||| | ||||||| | ||||||| ||||||| | |||
Db     204011 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 204070

Qy        483 TATATA---ATATACACACATATATATACACATATATACACACATATATACACATATATA 539
```

```
Db     204071 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 204130

Qy        540 CACATATATATACACATATATACACATATATACATATATACACATATATATACATATATA 599

Db     204131 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 204190

Qy        600 CACATATATACATATATACACATATATACATATATACACACATATATACACATACATATA 659

Db     204191 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 204250

Qy        660 CACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATA 719

Db     204251 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 204310

Qy        720 TGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGT 779

Db     204311 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 204370

Qy        780 TTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATA 839

Db     204371 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 204430

Qy        840 CATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATATATAT 899

Db     204431 TATATATATATATAAAAATATAAATATAATATATATTTTAATATATATATAAATATAT 204490

Qy        900 ATA 902

Db     204491 ATA 204493
```

```
RESULT 62
CR626907
LOCUS       CR626907              218701 bp    DNA     linear   HTG 30-JUL-2004
DEFINITION  Danio rerio clone DKEY-52D15, WORKING DRAFT SEQUENCE, 7 unordered
            pieces.
ACCESSION   CR626907
VERSION     CR626907.2  GI:50871854
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 218701)
  AUTHORS   Burton,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUL-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 30, 2004 this sequence version replaced gi:50787983.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK52D15
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 216083 bases at least Q40
            Consensus quality: 216675 bases at least Q30
            Consensus quality: 216971 bases at least Q20
            Insert size: 218101; sum-of-contigs
            Insert size: 195570; 5.6% error; agarose-fp
            Quality coverage: 8.49x in Q20 bases; sum-of-contigs Quality
            coverage: 9.59x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     6174: contig of 6174 bp in length
            *     6175     6274: gap of 100 bp
            *     6275    10497: contig of 4223 bp in length
            *    10498    10597: gap of 100 bp
            *    10598    49931: contig of 39334 bp in length
            *    49932    50031: gap of 100 bp
            *    50032   118150: contig of 68119 bp in length
            *   118151   118250: gap of 100 bp
            *   118251   146821: contig of 28571 bp in length
            *   146822   146921: gap of 100 bp
            *   146922   186335: contig of 39414 bp in length
            *   186336   186435: gap of 100 bp
            *   186436   218701: contig of 32266 bp in length.
FEATURES             Location/Qualifiers
     source          1. .218701
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-52D15"
                     /clone_lib="DanioKey"
     misc_feature    1. .6174
                     /note="assembly_fragment:00016
                     fragment_chain:1"
     misc_feature    6275. .10497
                     /note="assembly_fragment:00055
                     fragment_chain:1"
     misc_feature    10598. .49931
                     /note="assembly_fragment:00497
                     fragment_chain:1"
     misc_feature    50032. .118150
                     /note="assembly_fragment:01902
                     fragment_chain:1"
     misc_feature    118251. .146821
                     /note="assembly_fragment:00934"
     misc_feature    146922. .186335
                     /note="assembly_fragment:01398"
     misc_feature    186436. .218701
                     /note="assembly_fragment:00096
                     clone_end:SP6
                     vector_side:right"
ORIGIN
```

```
  Query Match             7.6%;  Score 350.2;  DB 2;  Length 218701;
  Best Local Similarity   44.5%;  Pred. No. 1.1e-59;
  Matches  554;  Conservative    0;  Mismatches  689;  Indels    2;  Gaps    2;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286

Db       5167 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 5226

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346

Db       5227 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 5286

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406

Db       5287 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 5346

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466

Db       5347 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 5406

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525

Db       5407 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 5466

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
```

```
Db       5467 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 5526

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db       5527 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 5586

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | ||||||||  ||||||
Db       5587 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 5646

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||||          | |||          |  |   ||    |    |||
Db       5647 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 5706

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
              |  | | | | | | | | | |||    | |  | || |     |     | | |
Db       5707 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 5766

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTA 884
              ||       | ||| ||| || | ||   | |  ||| |  | |  ||  |||  |   |
Db       5767 TATATATATATATATATATATATATATATATANATATATAAATATATATATAAANANANA 5826

Qy        885 ACACTATATATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAA 944
              |||||||||  ||      ||  | |    |    |          |       ||
Db       5827 TATATATATATATNTAANNATNAAAAAATNNNTATNNNANAANNNNNNNNANNNNNNNAA 5886

Qy        945 TGTTATAGAAGAATGGCTAAGATGCCAAAGCTCAGTGTATGTGGCAGAGGCATGGTGTAG 1004
              |  |                |    |  |                |     |   |
Db       5887 ANNNNTTTANNNNNNANNNNTNTNTTNNANNTTNNNNNNNNNNNTAANNNNNTNNNTTTN 5946

Qy       1005 GGTGTGTCCAGGTTCATATATTGCATTAAGTGTGAGAACACCCTGGAGTATGAACCAAGA 1064
                  | ||     ||   |    |         | | |     |
Db       5947 TNNANNNAATAAAAANNNTTTTANNNNAATNNTNNTTATNNANNNNAATNTTNTNTANNN 6006

Qy       1065 AAATGCAAAAGCCAGAAGTGATGGAGGAAATGAGACACAATAATGAAGATATTGAGAGGA 1124
               |   |                    |   |   |                  |  |
Db       6007 TNAAAAANNNNNNNNNNNNANNTNNNNNNNTNNNANNNNANNNNNNNNNNNNNNNANAA 6066

Qy       1125 GGGTGTGGGCCTAGAGTGAAGCTTTTCGTGCCAGTACTTCTTTTGAAGGCCCAGTTCTCT 1184
                   |          ||||      | || || |              |
Db       6067 AATATTNNTTTNTTNNNNNNNTTTTTTNNNTAATTATTTTTNNNTTNNNNTAATNNNNNN 6126

Qy       1185 TCTCTCTCGGGGGCTCCTTCATCTCTCATAGAGTCCACAGCTTTTAAGGGCCAACACTTG 1244
               |                    |                    || ||
Db       6127 NATNANNNTNNTAANAAANAAANNNNNNNNNNNNNNNNNNNTNTTNAAANNNNNNNNNNN 6186

Qy       1245 AGGTCAGCCTGGCTCTCTCATTTGAGCTGGATAGAACATTTTAGAGCACCATCTATTCTT 1304

Db       6187 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 6246

Qy       1305 CAAGAGGAAGTTTAAAAATAAAAGAACCTTGAAGAGGAAAAAATGTAGACATTCAATCTA 1364
                                      ||          |   |  | |     || |
Db       6247 NNNNNNNNNNNNNNNNNNNNNNNNNNNNTTANNNNNNTATTNTANATANNNNTNATTTT 6306

Qy       1365 ACCTTTTCATTTTACTAGCCAAAGCTAAATAGAATGCAGGTTACCTGTTTTTCAGCCAGG 1424
               |    ||              |
Db       6307 TNNTNNNNATNNNTTNNNNANANNNNNNNNNNNNNNNNNNANNNNNNNNNNNNNNNNNN 6366

Qy       1425 CACCATCATTTCCTAATTGTTATAAAATTTATTATTATTGTTGTT 1469
                        |   | || |    ||| ||  |||     ||
Db       6367 NNNNNNNNNNNNTTNNTTTTTTTTTTTTTTTTTNNTATNNNNNTT 6411


RESULT 63
CR769779
LOCUS       CR769779              219074 bp    DNA     linear   HTG 18-SEP-2004
DEFINITION  Danio rerio clone DKEY-18C12, *** SEQUENCING IN PROGRESS ***, 5
            unordered pieces.
ACCESSION   CR769779
VERSION     CR769779.3  GI:52345314
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 219074)
  AUTHORS   McLay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 18, 2004 this sequence version replaced gi:52214022.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK18C12
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 217273 bases at least Q40
            Consensus quality: 217586 bases at least Q30
            Consensus quality: 217870 bases at least Q20
            Insert size: 218674; sum-of-contigs
            Insert size: 217168; 4.7% error; agarose-fp
            Quality coverage: 8.00x in Q20 bases; sum-of-contigs Quality
            coverage: 8.05x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    31758: contig of 31758 bp in length
            *    31759    31858: gap of 100 bp
            *    31859    46986: contig of 15128 bp in length
            *    46987    47086: gap of 100 bp
            *    47087    62942: contig of 15856 bp in length
            *    62943    63042: gap of 100 bp
            *    63043   117588: contig of 54546 bp in length
            *   117589   117688: gap of 100 bp
            *   117689   219074: contig of 101386 bp in length.
FEATURES             Location/Qualifiers
     source          1. .219074
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-18C12"
                     /clone_lib="DanioKey"
     misc_feature    1. .31758
                     /note="assembly_fragment:00384
                     fragment_chain:1"
     misc_feature    31859. .46986
                     /note="assembly_fragment:00008
                     fragment_chain:1"
     misc_feature    47087. .62942
                     /note="assembly_fragment:00176.0"
     misc_feature    63043. .117588
                     /note="assembly_fragment:00812"
     misc_feature    117689. .219074
                     /note="assembly_fragment:01508"
ORIGIN

  Query Match             7.6%;  Score 350.2;  DB 2;  Length 219074;
  Best Local Similarity   67.3%;  Pred. No. 1.1e-59;
  Matches  524;  Conservative    0;  Mismatches  253;  Indels    2;  Gaps    2;

Qy        187 AACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATA 246
              | ||| |   |||    |    | ||||||||||||   | | ||||||||||||||||||
Db     218052 ACCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATATATATATATATATATATA 218111
```

```
Qy        247 TATATACACACACACACATATATATATACACATATATATACACATATATACATATATA 306
              |||||| | | | | | | |||||||||||| | ||||||||| | ||||||| |||||||
Db     218112 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 218171

Qy        307 CACATATATACACATATATATACACATACATATGTACACATATATATACACATATGTATA 366
              | ||||||| | ||||||||| | ||| |||| || | ||||||||| | |||| ||||
Db     218172 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 218231

Qy        367 CACATATATACACATATATACACACATATATACACATATATACACACATATATACACATA 426
              | ||||||| | ||||||| | | ||||||| | ||||||| | | ||||||| | |||
Db     218232 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 218291

Qy        427 TATACACATATATACACATATACACATATATACACATATATACATATATACACATATATA 486
              |||| | ||||||| | ||||| | ||||||| | ||||||| ||||||| | |||||||
Db     218292 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 218351

Qy        487 TAATATACACACATATATATACACATATATACACACATATATACACATATATACACATAT 546
              | ||||| | | ||||||||| | ||||||| | | ||||||| | ||||||| | ||||
Db     218352 T-ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 218410

Qy        547 ATATACACATATATACACATATATACATATATACACATATATATACATATATACACATAT 606
              ||||| | ||||||| | ||||||| ||||||| | ||||||||| ||||||| | ||||
Db     218411 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 218470

Qy        607 ATACATATATACACATATATACATATATACACACATATATACACATACATATACACACAC 666
              ||| ||||||| | ||||||| ||||||| | | ||||||| | ||| ||||| | | |
Db     218471 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 218530

Qy        667 ATAGATATACATATATATACACATATATATACGTATATATATGTATATATATATGCTCCA 726
              ||| ||||| |||||||||| | ||||||||| |||||||| ||||||||||
Db     218531 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 218590

Qy        727 GAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAG 786
              |  |||  |    | |  ||    |   ||| |   | | | |  | | |
Db     218591 ATATATATATAAATATATATATATATATATATATATATATATATATATATATATATATAT 218650

Qy        787 CAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAAT 846
              | |||   | |  | || |     |      | |||       | ||| ||| ||
Db     218651 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 218710

Qy        847 GTGTAATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTAACACTATATATATATATAGTG 905
              | |||  | |  ||| | || |  ||    |  |  ||    || ||||| |||||
Db     218711 ATATAAATATATATATATATATATATAAAATAATATATATATTTAAATATAAATATAAAT 218770

Qy        906 AAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTATAGAAGAATGGCTAA 964
              | |      |||    |   ||     ||  || ||||| ||  |  | |||   |||
Db     218771 ATAAATTATAAAAATATATATATATATTAATAATATAAATATTTAAATAAAATATATAA 218829


RESULT 64
CR749742
LOCUS       CR749742              194235 bp    DNA     linear   HTG 23-AUG-2004
DEFINITION  Danio rerio clone DKEYP-90H9, *** SEQUENCING IN PROGRESS ***, 9
            unordered pieces.
ACCESSION   CR749742
VERSION     CR749742.3  GI:51510860
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 194235)
  AUTHORS   McLay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-AUG-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 23, 2004 this sequence version replaced gi:51491327.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zKp90H9
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 190869 bases at least Q40
            Consensus quality: 191868 bases at least Q30
            Consensus quality: 192427 bases at least Q20
            Insert size: 193435; sum-of-contigs
            Insert size: 191728; 8.4% error; agarose-fp
            Quality coverage: 8.95x in Q20 bases; sum-of-contigs Quality
            coverage: 9.11x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2462: contig of 2462 bp in length
            *     2463     2562: gap of 100 bp
            *     2563    19346: contig of 16784 bp in length
            *    19347    19446: gap of 100 bp
            *    19447    64582: contig of 45136 bp in length
            *    64583    64682: gap of 100 bp
            *    64683    99973: contig of 35291 bp in length
            *    99974   100073: gap of 100 bp
            *   100074   123327: contig of 23254 bp in length
            *   123328   123427: gap of 100 bp
            *   123428   130685: contig of 7258 bp in length
            *   130686   130785: gap of 100 bp
            *   130786   142959: contig of 12174 bp in length
            *   142960   143059: gap of 100 bp
            *   143060   157214: contig of 14155 bp in length
            *   157215   157314: gap of 100 bp
            *   157315   194235: contig of 36921 bp in length.
FEATURES             Location/Qualifiers
     source          1. .194235
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEYP-90H9"
                     /clone_lib="DanioKeypilot"
     misc_feature    1. .2462
                     /note="assembly_fragment:00014
                     fragment_chain:1"
     misc_feature    2563. .19346
                     /note="assembly_fragment:00268
                     fragment_chain:1"
     misc_feature    19447. .64582
                     /note="assembly_fragment:02137
                     fragment_chain:1"
     misc_feature    64683. .99973
                     /note="assembly_fragment:01580
                     fragment_chain:1"
     misc_feature    100074. .123327
                     /note="assembly_fragment:00761
                     fragment_chain:1"
     misc_feature    123428. .130685
                     /note="assembly_fragment:00032
                     fragment_chain:1"
     misc_feature    130786. .142959
                     /note="assembly_fragment:00136.0"
     misc_feature    143060. .157214
                     /note="assembly_fragment:00500"
     misc_feature    157315. .194235
                     /note="assembly_fragment:01065"
ORIGIN
```

```
  Query Match             7.6%;  Score 350;  DB 2;  Length 194235;
  Best Local Similarity   69.2%;  Pred. No. 1.2e-59;
  Matches  504;  Conservative    0;  Mismatches  222;  Indels    2;  Gaps    2;

Qy        175 TCCAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATA 234
              | ||  |||    |   |||   | |   || ||   ||||||||||    |   ||||||
Db     122601 TTCATTCCATATATTCAGACTTAGAATCCAACTTGAAAAAAAAAAAAAAAAAAAATATATA 122660

Qy        235 TATATATATATATATATACACACACACACACATATATATATACACATATATATACACATA 294
              ||||||||||||||||| | | | | | | |||||||||| | ||||||||| | |||
Db     122661 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 122720

Qy        295 TATACATATATACACATATATACACATATATATACACATACATATGTACACATATATATA 354
              |||| ||||||| | ||||||| | |||||||| | ||| |||| || | |||||||||
Db     122721 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 122780

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              | |||| |||| | ||||||| | ||||||| | | ||||||| | ||||||| | | |
Db     122781 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 122840

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              |||||| | ||||||| | ||||||| | ||||| | ||||||| | ||||||| |||||
Db     122841 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 122900

Qy        475 TACACATATATATAATATACACACATATATATACACATATATACACACATATATACACAT 534
              || | |||||||| |||| | | ||||||||| | ||||||| | | ||||||| | ||
Db     122901 TATATATATATAT-ATATATATATATATATATATATATATATATATATATATATATAT 122959

Qy        535 ATATACACATATATATACACATATATACACATATATACATATATACACATATATATACAT 594
              ||||| | |||||||| | ||||||| | ||||||| ||||||| | |||||||| ||
Db     122960 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 123019

Qy        595 ATATACACATATATACATATATACACATATATACATATATACACACATATATACACATAC 654
              ||||| | ||||||| ||||||| | ||||||| ||||||| | | ||||||| | |||
Db     123020 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 123079

Qy        655 ATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATAT 714
              ||||| | | | ||| ||||| |||||||||  | |||||||||  |||||||||  |||||
Db     123080 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 123139

Qy        715 ATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAG 774
              ||||||          |  |||          |  |    ||     |    ||| |   | |
Db     123140 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 123199

Qy        775 AGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGG 834
              | | | | | |  | |||   | |  | || |      |        | |||
Db     123200 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 123259

Qy        835 AAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATA 894
              | ||| ||| || | ||   | |  ||| | | |  ||  ||  ||  |||   ||||||
Db     123260 ATATAAATATATATATATATATATATATATAT-ANATATNTATAAATATAAATTTATATA 123318

Qy        895 TATATATA 902
              ||||||||
Db     123319 TATATATA 123326


RESULT 65
AL513530/c
LOCUS       AL513530               50612 bp    DNA     linear   PRI 15-MAY-2001
DEFINITION  Human DNA sequence from clone RP11-447M22 on chromosome 6, complete
            sequence.
ACCESSION   AL513530
VERSION     AL513530.8  GI:14133035
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 50612)
  AUTHORS   Sycamore,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On May 16, 2001 this sequence version replaced gi:13751553.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            RP11-447M22 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-447M22 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true left end of clone RP4-708F5 is at 50513 in this sequence.
            The true right end of clone RP4-682J15 is at 100 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .50612
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP11-447M22"
                     /clone_lib="RPCI-11.2"
     repeat_region   1485. .1570
                     /note="MIR repeat: matches 88. .174 of consensus"
     repeat_region   1745. .1774
                     /note="15 copies 2 mer tt 86% conserved"
     repeat_region   3540. .3811
                     /note="MLT1E repeat: matches 265. .561 of consensus"
     repeat_region   3800. .4122
                     /note="MLT1D repeat: matches 1. .322 of consensus"
     repeat_region   4169. .4648
                     /note="L1M4 repeat: matches 3589. .4085 of consensus"
     repeat_region   5046. .5161
                     /note="58 copies 2 mer aa 59% conserved"
     repeat_region   6172. .6350
                     /note="MER5A repeat: matches 1. .189 of consensus"
     repeat_region   6352. .6435
                     /note="42 copies 2 mer ta 63% conserved"
     repeat_region   6783. .8591
                     /note="TIGGER2 repeat: matches 895. .2718 of consensus"
     repeat_region   8592. .8882
                     /note="AluSx repeat: matches 1. .291 of consensus"
     repeat_region   8883. .9301
                     /note="TIGGER2 repeat: matches 474. .895 of consensus"
     repeat_region   9302. .9598
                     /note="AluJb repeat: matches 1. .291 of consensus"
     repeat_region   9599. .10107
                     /note="TIGGER2 repeat: matches 4. .474 of consensus"
     repeat_region   10470. .10756
                     /note="AluSc repeat: matches 1. .288 of consensus"
     repeat_region   13276. .13732
                     /note="MLT1D repeat: matches 1. .505 of consensus"
     repeat_region   13898. .14119
```

```
                     /note="MLT1-INTERNAL repeat: matches 951. .1182 of
                     consensus"
     repeat_region   14554. .15103
                     /note="MLT1-INTERNAL repeat: matches 36. .592 of consensus"
     repeat_region   15130. .15428
                     /note="MLT2E repeat: matches 136. .397 of consensus"
     repeat_region   15401. .15553
                     /note="MLT1D repeat: matches 2. .154 of consensus"
     repeat_region   16010. .16185
                     /note="L2 repeat: matches 2191. .2372 of consensus"
     repeat_region   16302. .16343
                     /note="L1MB8 repeat: matches 6117. .6163 of consensus"
     repeat_region   16353. .16656
                     /note="AluSx repeat: matches 1. .304 of consensus"
     repeat_region   16841. .16908
                     /note="L2 repeat: matches 2641. .2709 of consensus"
     repeat_region   17903. .17997
                     /note="L2 repeat: matches 2575. .2680 of consensus"
     repeat_region   18264. .19239
                     /note="488 copies 2 mer ta 87% conserved"
     misc_feature    18959
                     /note="Tandem repeat. Forced join. Gap size estimated to
                     be approximately 250bp by restriction digest data."
     repeat_region   19240. .19524
                     /note="AluY repeat: matches 1. .285 of consensus"
     repeat_region   19735. .20045
                     /note="AluY repeat: matches 1. .308 of consensus"
     repeat_region   20344. .20626
                     /note="L2 repeat: matches 2002. .2294 of consensus"
     repeat_region   20849. .21066
                     /note="MIR repeat: matches 34. .244 of consensus"
     repeat_region   21538. .21606
                     /note="MER5A repeat: matches 38. .103 of consensus"
     repeat_region   23617. .23800
                     /note="MER5A repeat: matches 2. .186 of consensus"
     repeat_region   23953. .24005
                     /note="MADE1 repeat: matches 28. .80 of consensus"
     repeat_region   24271. .24453
                     /note="MER5A repeat: matches 2. .189 of consensus"
     repeat_region   26159. .26323
                     /note="MIR repeat: matches 78. .249 of consensus"
     repeat_region   26313. .26567
                     /note="L2 repeat: matches 397. .654 of consensus"
     repeat_region   26731. .26876
                     /note="L2 repeat: matches 879. .1028 of consensus"
     repeat_region   27304. .27596
                     /note="AluY repeat: matches 1. .292 of consensus"
     repeat_region   28752. .29042
                     /note="L1 repeat: matches 2111. .2413 of consensus"
     repeat_region   29124. .30099
                     /note="L1 repeat: matches 2504. .3534 of consensus"
     repeat_region   30174. .31422
                     /note="L1M4 repeat: matches 3279. .4581 of consensus"
     repeat_region   31424. .32086
                     /note="L1MC3 repeat: matches 7067. .7739 of consensus"
     repeat_region   32092. .32407
                     /note="AluSq repeat: matches 1. .306 of consensus"
     repeat_region   32417. .32806
                     /note="L1M4 repeat: matches 4603. .5002 of consensus"
     repeat_region   32818. .33181
                     /note="L1MA1 repeat: matches -1388. .-1025 of consensus"
     repeat_region   33182. .33493
                     /note="AluSp repeat: matches 1. .313 of consensus"
     repeat_region   33494. .34874
                     /note="L1MA1 repeat: matches -1025. .356 of consensus"
     repeat_region   34876. .34981
                     /note="L1P repeat: matches 3001. .3103 of consensus"
     repeat_region   34966. .35806
                     /note="L1PA2 repeat: matches 5275. .6144 of consensus"
     repeat_region   35825. .41028
                     /note="L1MA1 repeat: matches 347. .5615 of consensus"
     repeat_region   41029. .41334
                     /note="AluSq repeat: matches 6. .310 of consensus"
     repeat_region   41335. .41728
                     /note="L1MA1 repeat: matches 5615. .6005 of consensus"
     repeat_region   41727. .41963
                     /note="L1MC2 repeat: matches 5825. .6063 of consensus"
     repeat_region   41976. .42095
                     /note="60 copies 2 mer at 85% conserved"
     repeat_region   42097. .42330
                     /note="L1MC1 repeat: matches 6081. .6322 of consensus"
     repeat_region   42547. .42818
                     /note="L1MB7 repeat: matches 5778. .6033 of consensus"
     repeat_region   42819. .43022
                     /note="AluY repeat: matches 102. .303 of consensus"
     repeat_region   43023. .43157
                     /note="L1MB7 repeat: matches 6033. .6171 of consensus"
     repeat_region   43283. .43474
                     /note="MIR repeat: matches 21. .247 of consensus"
     repeat_region   45140. .45454
                     /note="AluY repeat: matches 1. .311 of consensus"
     repeat_region   45522. .45828
                     /note="AluSg repeat: matches 1. .307 of consensus"
     repeat_region   47590. .47773
                     /note="HAL1 repeat: matches 1309. .1491 of consensus"
     repeat_region   47984. .48257
                     /note="HAL1 repeat: matches 630. .931 of consensus"
     repeat_region   49161. .49464
                     /note="AluSq repeat: matches 1. .305 of consensus"
     repeat_region   50177. .50264
                     /note="22 copies 4 mer agag 70% conserved"
     repeat_region   50186. .50263
                     /note="39 copies 2 mer ga 74% conserved"
     repeat_region   50465. .50490
                     /note="13 copies 2 mer gt 96% conserved"
ORIGIN

  Query Match             7.6%;  Score 349.6;  DB 9;  Length 50612;
  Best Local Similarity   72.2%;  Pred. No. 1.8e-59;
  Matches  523;  Conservative    0;  Mismatches  194;  Indels    7;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC--AAAAT 58
              ||  |  |||||| ||||   |||| |||||||||| ||||||||||||||  |||||
Db      19451 AGATTGAGACCATCCTGGCTAACACGGTGAAACCCCGTCTCTACTAAAAATACAAAAAAT 19392

Qy         59 TAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGA 118
              |||   ||| |||||  ||  ||||||| ||||||||||  ||||||  ||||||||||
Db      19391 TAGCAGGGCGTGGTGGCACGTGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGA 19332

Qy        119 ATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCC 177
              ||| ||||||||||  || ||||| |||||||| ||| |||||||  || | ||||||||
Db      19331 ATCGCTTGAACCTG--GGAGGTGGGGGTTGCGGTGAGCCAAGATCGCGCTACTGCACTCC 19274

Qy        178 AGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATAT 237
              ||||  ||| || ||||| | |    | ||||| || | | | |       | |||||  |||
Db      19273 AGCCTGGGCGAC-AGAGCAAGACCCCGTCTCAAAATATATATATATTTATATATATTTAT 19215

Qy        238 ATATATATATATATACACACACACACATATATATATACACATATATATACACATATATAT 297
              |||||||| ||||||   | | | |   |||||||||  | ||||||  || | ||  |||
Db      19214 ATATATATTTATATATTTATATATATTTATATATATATTTATATATATTTATATATTTAT 19155

Qy        298 ACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACAC 357
              | |||| || | ||||||  | ||||||  || | |||  |||  || | | || ||||||
Db      19154 ATATATTTATATATATATTTATATATATTTATATATATTTATATATATATTTATATATTT 19095

Qy        358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417
              |||| |||  | ||||||  | |||| || | | || |||| |  ||||||  | |  |||
Db      19094 ATATATATTTATATATATTTATATATTTATATATATTTATATATTTATATATATATTTAT 19035

Qy        418 ATACACATATATACACATATATACACATATACACATATATACACATATATACATATATAC 477
              |||   ||||||| |  |||||| |  |||| |  ||||||   |||| || ||||||
Db      19034 ATATTTATATATATATTTATATATATCTATATATTTATATATGTATATTTATATATATCT 18975

Qy        478 ACATATATATAATATACACACATATATATACACATATATACACACATATATACACATATA 537
              | |||| ||| ||||| | | |||| |||| | || |||| |   ||||||  | |||||
Db      18974 ATATATTTAT-ATATATATATATTTATATATATTTATATATTTATATATTTATATATA 18916
```

```
Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATACATATA 597
              |  | ||||||||  | |||| || | |||| || |||||| || ||||||||  |||||
Db      18915 TTTATATATATATTTATATATTTATATATATTTATATATATTCATATATATATTTATATA 18856

Qy        598 TACACATATATACATATATACACATATATACATATATACACACATATATACACATACATA 657
              |  ||||||  |  ||||||  |||||||  ||||||  ||||||  ||  |||  ||
Db      18855 TTTACATATGTGTTTATATATTTATATATATTTATATATTTACATATGTATTTATATATT 18796

Qy        658 TACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATA 717
              || | | |  || ||||  ||||||  ||  ||||||  ||  |||||  |||  ||| ||||
Db      18795 TATATATATTTATATATTTATATATGTATTTATATATTTATATATATTTATATATTTATA 18736

Qy        718 TATG 721
              ||||
Db      18735 TATG 18732


RESULT 66
AF438327_3
WPCOMMENT
  Sequence split into 4 fragments  LOCUS AF438327 Accession AF438327
     Fragment Name      Begin     End
     AF438327_0             1    110000
     AF438327_1        100001    210000
     AF438327_2        200001    310000
     AF438327_3        300001    381696
Continuation (4 of 4) of AF438327 from base 300001 (AF438327 Homo sapiens alpha 1 type X

  Query Match             7.6%;  Score 349.6;  DB 9;  Length 81696;
  Best Local Similarity   72.2%;  Pred. No. 1.7e-59;
  Matches  523;  Conservative    0;  Mismatches  194;  Indels    7;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC--AAAAT 58
              || |  |||||| ||||   |||| ||||||||||| ||||||||||||||  |||||
Db       8385 AGATTGAGACCATCCTGGCTAACACGGTGAAACCCCGTCTCTACTAAAAATACAAAAAAT 8444

Qy         59 TAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGA 118
              |||   ||| |||||  ||  ||||||| ||||||||||  ||||||  ||||||||||
Db       8445 TAGCAGGGCGTGGTGGCACGTGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGA 8504

Qy        119 ATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCC 177
              ||| ||||||||||  || ||||| ||||||| ||| |||||||  || | ||||||||
Db       8505 ATCGCTTGAACCTG--GGAGGTGGGGGTTGCGGTGAGCCAAGATCGCGCTACTGCACTCC 8562

Qy        178 AGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATAT 237
              ||||  ||| || ||||| | |   | ||||| || | | | |       | ||||| |||
Db       8563 AGCCTGGGCGAC-AGAGCAAGACCCCGTCTCAAAATATATATATATTTATATATATTTAT 8621

Qy        238 ATATATATATATATACACACACACACATATATATATACACATATATATACACATATATAT 297
              |||||||| ||||||   | | | |   |||||||||  | |||||| || | || |||
Db       8622 ATATATATTTATATATTTATATATATTTATATATATATTTATATATATTTATATATTTAT 8681

Qy        298 ACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACAC 357
              | |||| || | ||||||  | ||||||  || | |||  ||| || | || ||||||
Db       8682 ATATATTTATATATATATTTATATATATTTATATATATTTATATATATATTTATATATTT 8741

Qy        358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417
              |||| |||  | ||||||  | |||| || | | || |||| |  ||||||  | |  |||
Db       8742 ATATATATTTATATATATTTATATATTTATATATATTTATATATTTATATATATATTTAT 8801

Qy        418 ATACACATATATACACATATATACACATATACACATATATACACATATATACATATATAC 477
              |||   |||||||  |  ||||||  |  ||||  |  ||||||   |||| || ||||||
Db       8802 ATATTTATATATATATTTATATATATCTATATATTTATATATGTATATTTATATATATCT 8861

Qy        478 ACATATATATAATATACACACATATATATACACATATATACACACATATATACACACATATA 537
              | |||| ||| ||||| | | |||| |||| | || |||| |   ||||||  | |||||
Db       8862 ATATATTTAT-ATATATATATATATTTATATATATTTATATATTTATATATTTATATATA 8920

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATACATATA 597
              |  | ||||||||  | |||| || | |||| || |||||| || ||||||||  |||||
Db       8921 TTTATATATATATTTATATATTTATATATATTTATATATATTCATATATATATTTATATA 8980
```

```
Qy        598 TACACATATATACATATATACACATATATACATATATACACACATATATACACATACATA 657
              |  ||||||  |  ||||||  |||||||  ||||||  ||||||  ||  |||  ||
Db       8981 TTTACATATGTGTTTATATATTTATATATATTTATATATTTACATATGTATTTATATATT 9040

Qy        658 TACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATA 717
              || | | |  || ||||  ||||||  ||  ||||||  ||  |||||  |||  ||| ||||
Db       9041 TATATATATTTATATATTTATATATGTATTTATATATTTATATATATTTATATATTTATA 9100

Qy        718 TATG 721
              ||||
Db       9101 TATG 9104


RESULT 67
AF438327_2
WPCOMMENT
  Sequence split into 4 fragments  LOCUS AF438327 Accession AF438327
     Fragment Name      Begin     End
     AF438327_0             1    110000
     AF438327_1        100001    210000
     AF438327_2        200001    310000
     AF438327_3        300001    381696
Continuation (3 of 4) of AF438327 from base 200001 (AF438327 Homo sapiens alpha 1 type X

  Query Match             7.6%;  Score 349.6;  DB 9;  Length 110000;
  Best Local Similarity   72.2%;  Pred. No. 1.6e-59;
  Matches  523;  Conservative    0;  Mismatches  194;  Indels    7;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC--AAAAT 58
              || |  |||||| ||||   |||| ||||||||||| ||||||||||||||  |||||
Db     108385 AGATTGAGACCATCCTGGCTAACACGGTGAAACCCCGTCTCTACTAAAAATACAAAAAAT 108444

Qy         59 TAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGA 118
              |||   ||| |||||  ||  ||||||| ||||||||||  ||||||  ||||||||||
Db     108445 TAGCAGGGCGTGGTGGCACGTGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGA 108504

Qy        119 ATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCC 177
              ||| ||||||||||  || ||||| ||||||| ||| |||||||  || | ||||||||
Db     108505 ATCGCTTGAACCTG--GGAGGTGGGGGTTGCGGTGAGCCAAGATCGCGCTACTGCACTCC 108562

Qy        178 AGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATAT 237
              ||||  ||| || ||||| | |   | ||||| || | | | |       | ||||| |||
Db     108563 AGCCTGGGCGAC-AGAGCAAGACCCCGTCTCAAAATATATATATATTTATATATATTTAT 108621

Qy        238 ATATATATATATATACACACACACACATATATATATACACATATATATACACATATATAT 297
              |||||||| ||||||   | | | |   |||||||||  | |||||| || | || |||
Db     108622 ATATATATTTATATATTTATATATATTTATATATATATTTATATATATTTATATATTTAT 108681

Qy        298 ACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACAC 357
              | |||| || | ||||||  | ||||||  || | |||  ||| || | || ||||||
Db     108682 ATATATTTATATATATATTTATATATATTTATATATATTTATATATATATTTATATATTT 108741

Qy        358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417
              |||| |||  | ||||||  | |||| || | | || |||| |  ||||||  | |  |||
Db     108742 ATATATATTTATATATATTTATATATTTATATATATTTATATATTTATATATATATTTAT 108801

Qy        418 ATACACATATATACACATATATACACATATACACATATATACACATATATACACATATAC 477
              |||   |||||||  |  ||||||  |  ||||  |  ||||||   |||| || ||||||
Db     108802 ATATTTATATATATATTTATATATATCTATATATTTATATATGTATATTTATATATATCT 108861

Qy        478 ACATATATATAATATACACACATATATATACACATATATACACACATATATACACATATA 537
              | |||| ||| ||||| | | |||| |||| | || |||| |   ||||||  | |||||
Db     108862 ATATATTTAT-ATATATATATATATTTATATATATTTATATATTTATATATTTATATATA 108920

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATACATATA 597
              |  | ||||||||  | |||| || | |||| || |||||| || ||||||||  |||||
Db     108921 TTTATATATATATTTATATATTTATATATATTTATATATATTCATATATATATTTATATA 108980

Qy        598 TACACATATATACATATATACACATATATACATATATACACACATATATACACATACATA 657
              |  ||||||  |  ||||||  |||||||  ||||||  ||||||  ||  |||  ||
Db     108981 TTTACATATGTGTTTATATATTTATATATATTTATATATTTACATATGTATTTATATATT 109040

Qy        658 TACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATA 717
```

```
Db     109041 TATATATATTTATATATTTATATATGTATTTATATATTTATATATATTTATATATTTATA 109100

Qy        718 TATG 721
              ||||
Db     109101 TATG 109104


RESULT 68
CR759826/c
LOCUS       CR759826              157042 bp    DNA     linear   HTG 05-SEP-2004
DEFINITION  Danio rerio clone CH211-266K13, *** SEQUENCING IN PROGRESS ***, 10
            unordered pieces.
ACCESSION   CR759826
VERSION     CR759826.1  GI:51890341
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 157042)
  AUTHORS   McLay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zC266K13
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 153933 bases at least Q40
            Consensus quality: 154487 bases at least Q30
            Consensus quality: 154866 bases at least Q20
            Insert size: 156142; sum-of-contigs
            Insert size: 159217; 0.8% error; agarose-fp
            Quality coverage: 6.01x in Q20 bases; sum-of-contigs Quality
            coverage: 5.90x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    44418: contig of 44418 bp in length
            *    44419    44518: gap of 100 bp
            *    44519    52984: contig of 8466 bp in length
            *    52985    53084: gap of 100 bp
            *    53085    65205: contig of 12121 bp in length
            *    65206    65305: gap of 100 bp
            *    65306    67571: contig of 2266 bp in length
            *    67572    67671: gap of 100 bp
            *    67672   110374: contig of 42703 bp in length
            *   110375   110474: gap of 100 bp
            *   110475   142025: contig of 31551 bp in length
            *   142026   142125: gap of 100 bp
            *   142126   146693: contig of 4568 bp in length
            *   146694   146793: gap of 100 bp
            *   146794   151219: contig of 4426 bp in length
            *   151220   151319: gap of 100 bp
            *   151320   154853: contig of 3534 bp in length
            *   154854   154953: gap of 100 bp
            *   154954   157042: contig of 2089 bp in length.
FEATURES             Location/Qualifiers
     source          1. .157042
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-266K13"
                     /clone_lib="CHORI-211"
     misc_feature    1. .44418
                     /note="assembly_fragment:00615
                     fragment_chain:1"
     misc_feature    44519. .52984
                     /note="assembly_fragment:00139
                     fragment_chain:1"
     misc_feature    53085. .65205
                     /note="assembly_fragment:00214
                     fragment_chain:1"
     misc_feature    65306. .67571
                     /note="assembly_fragment:00005
                     fragment_chain:1"
     misc_feature    67672. .110374
                     /note="assembly_fragment:01059
                     fragment_chain:1"
     misc_feature    110475. .142025
                     /note="assembly_fragment:00318
                     fragment_chain:2"
     misc_feature    142126. .146693
                     /note="assembly_fragment:00071
                     fragment_chain:2"
     misc_feature    146794. .151219
                     /note="assembly_fragment:00104"
     misc_feature    151320. .154853
                     /note="assembly_fragment:00042
                     fragment_chain:3"
     misc_feature    154954. .157042
                     /note="assembly_fragment:00020
                     fragment_chain:3
                     clone_end:SP6
                     vector_side:right"
ORIGIN

  Query Match             7.6%;  Score 349.6;  DB 2;  Length 157042;
  Best Local Similarity   63.1%;  Pred. No. 1.5e-59;
  Matches  462;  Conservative    0;  Mismatches  269;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db      68181 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 68122

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | |||||| ||||||| | ||||||| | |||||||||| | ||| |||| || | |
Db      68121 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 68062

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db      68061 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 68002

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db      68001 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 67942

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db      67941 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 67882

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db      67881 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 67822

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db      67821 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 67762

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||  ||||| | | | ||| ||||| ||||||||| | |||||||||  || |||
```

```
Db      67761 TATATATTTATATATATATATATATATATATATATATATATATATATATATATACATA 67702

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| |||||||||||| ||||    |  |
Db      67701 TATATATATATATATGATCCATCTCTACTCNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 67642

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825

Db      67641 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 67582

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885
                  |  | | || | ||   | |  ||| | ||   |  ||       ||
Db      67581 NNNNNNNNNTTTTTAAATATATATATATATATATATATATTTTTTTTTTTTTTATAT 67522

Qy        886 CACTATATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAAT 945
              |||||||||||||   | |    | || |  || |       |   || || ||
Db      67521 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 67462

Qy        946 GTTATAGAAGAA 957
                | | || ||
Db      67461 AAAAAAAAAAA 67450


RESULT 69
CR589883/c
LOCUS       CR589883              180560 bp    DNA     linear   HTG 27-JUL-2004
DEFINITION  Danio rerio clone CH211-51N18, WORKING DRAFT SEQUENCE, 6 unordered
            pieces.
ACCESSION   CR589883
VERSION     CR589883.2  GI:50724901
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 180560)
  AUTHORS   Burton,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JUL-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 27, 2004 this sequence version replaced gi:50399732.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zC51N18
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 177647 bases at least Q40
            Consensus quality: 178065 bases at least Q30
            Consensus quality: 178456 bases at least Q20
            Insert size: 180060; sum-of-contigs
            Insert size: 184777; 2.8% error; agarose-fp
            Quality coverage: 8.72x in Q20 bases; sum-of-contigs Quality
            coverage: 8.60x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     7987: contig of 7987 bp in length
            *     7988     8087: gap of 100 bp
            *     8088    33232: contig of 25145 bp in length
            *    33233    33332: gap of 100 bp
            *    33333    70114: contig of 36782 bp in length
            *    70115    70214: gap of 100 bp
            *    70215   126158: contig of 55944 bp in length
            *   126159   126258: gap of 100 bp
            *   126259   162306: contig of 36048 bp in length
            *   162307   162406: gap of 100 bp
            *   162407   180560: contig of 18154 bp in length.
FEATURES             Location/Qualifiers
     source          1. .180560
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-51N18"
                     /clone_lib="CHORI-211"
     misc_feature    1. .7987
                     /note="assembly_fragment:00018
                     clone_end:SP6
                     vector_side:left"
     misc_feature    8088. .33232
                     /note="assembly_fragment:00362.0"
     misc_feature    33333. .70114
                     /note="assembly_fragment:01176
                     fragment_chain:1"
     misc_feature    70215. .126158
                     /note="assembly_fragment:01644
                     fragment_chain:1"
     misc_feature    126259. .162306
                     /note="assembly_fragment:00726
                     fragment_chain:1"
     misc_feature    162407. .180560
                     /note="assembly_fragment:00143
                     fragment_chain:1
                     clone_end:T7
                     vector_side:right"
ORIGIN

  Query Match             7.6%;  Score 349.6;  DB 2;  Length 180560;
  Best Local Similarity   43.8%;  Pred. No. 1.5e-59;
  Matches  560;  Conservative    0;  Mismatches  717;  Indels    2;  Gaps    2;

Qy        205 TCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACA 264
              | | || | | | | |       | |||||||||||||||||||||||| | | | | | |
Db     127006 TTTAAGTATATATATATATATATATATATATATATATATATATATATATATATATATA 126947

Qy        265 CATATATATATACACATATATATACACATATATACATATATACACATATATACACATATA 324
              ||||||||||| | |||||||||| | ||||||| ||||||| | ||||||| | |||||
Db     126946 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 126887

Qy        325 TATACACATACATATGTACACATATATATACACATATGTATACACATATATACACATATA 384
              |||| | ||| |||| || | ||||||||| | |||| |||| | ||||||| | |||||
Db     126886 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 126827

Qy        385 TACACACATATATACACATATATACACACATATATACACATATATACACATATATACACA 444
              || | | ||||||| | ||||||| | | ||||||| | ||||||| | ||||||| | |
Db     126826 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 126767

Qy        445 TATACACATATATACACATATATACATATATACACATATATATAATATACACACATATAT 504
              |||| | ||||||| | ||||||| ||||||| | |||||||| ||||| | | ||||||
Db     126766 TATATATATATATATATATATATATATATATATATATATATATATAT-ATATATATATATAT 126708

Qy        505 ATACACATATATACACACATATATACACATATATACACATATATATACACATATATACAC 564
              ||| | ||||||| | | ||||||| | ||||||| | |||||||||| | ||||||| |
Db     126707 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 126648

Qy        565 ATATATACATATATACACATATATATACATATATACACATATATACATATATACACATAT 624
              ||||||| ||||||| | ||||||||| ||||||| | ||||||| ||||||| | ||||
Db     126647 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 126588

Qy        625 ATACATATATACACACATATATACACATACATATACACACACATAGATATACATATATAT 684
              ||| ||||||| | | ||||||| | ||| ||||| | | | ||| ||||| ||||||||
Db     126587 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 126528

Qy        685 ACACATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGC 744
```

```
Db     126527 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 126468

Qy        745 AGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGG 804

Db     126467 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 126408

Qy        805 ACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCT 864

Db     126407 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 126348

Qy        865 GTA-AGGTAGTTAAGAAGGTAACACTATATATATATATAGTGAAAGCAGTGTAAACCTAA 923

Db     126347 ATATATATATATATATATATATATATATATATATATATATATATATATCTATATGAATAT 126288

Qy        924 AGGATGGGCCAAGGATTTAAATGTTATAGAAGAATGGCTAAGATGCCAAAGCTCAGTGTA 983

Db     126287 ATATATATATATATATATATATATATANNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 126228

Qy        984 TGTGGCAGAGGCATGGTGTAGGGTGTGTCCAGGTTCATATATTGCATTAAGTGTGAGAAC 1043

Db     126227 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 126168

Qy       1044 ACCCTGGAGTATGAACCAAGAAAATGCAAAAGCCAGAAGTGATGGAGGAAATGAGACACA 1103

Db     126167 NNNNNNNNNTAAAAANNNNNNNANNATNNNNNNNNNNNANNNNNNNNNNTNNNNNNNTN 126108

Qy       1104 ATAATGAAGATATTGAGAGGAGGGTGTGGGCCTAGAGTGAAGCTTTTCGTGCCAGTACTT 1163

Db     126107 TTNNNTNNNNNNNNNNNNANNTNNNNNTNNNNNNNNNNAAANNNNNNNNNNNNNNNNN 126048

Qy       1164 CTTTTGAAGGCCCAGTTCTCTTCTCTCTCGGGGGCTCCTTCATCTCTCATAGAGTCCACA 1223

Db     126047 NNANNNNNNNNTANNNNNTNNANNTNNNTNNNNNNNNNNNNNNNNNNANNNNNNNNNN 125988

Qy       1224 GCTTTTAAGGGCCAACACTTGAGGTCAGCCTGGCTCTCTCATTTGAGCTGGATAGAACAT 1283

Db     125987 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 125928

Qy       1284 TTTAGAGCACCATCTATTCTTCAAGAGGAAGTTTAAAAATAAAAGAACCTTGAAGAGGAA 1343

Db     125927 NNNNNNNTTNNNTTTTTTTTTNTATAANANATANTAAAAANNNNNANNNNNNANNNNANT 125868

Qy       1344 AAAATGTAGACATTCAATCTAACCTTTTCATTTTACTAGCCAAAGCTAAATAGAATGCAG 1403

Db     125867 TNNNNNTTNNNNNNNNNNNNNNNTAAANNNNNNNNANNANNATTAAATTANTTAAAAT 125808

Qy       1404 GTTACCTGTTTTTCAGCCAGGCACCATCATTTCCTAATTGTTATAAAATTTATTATTATT 1463

Db     125807 ATAANTNNNTTTTTTTTNTNTNNNNNTNTNTNNNNNTTTTTTTNTNTNTNTTTNNNTNNNT 125748

Qy       1464 GTTGTTATTATTATTATTT 1482

Db     125747 NTTNNNTTTNNNTTTNTNT 125729


RESULT 70
CR556714
LOCUS       CR556714               210384 bp    DNA     linear   HTG 12-JUL-2004
DEFINITION  Danio rerio clone DKEY-29P13, WORKING DRAFT SEQUENCE, 5 unordered
            pieces.
ACCESSION   CR556714
VERSION     CR556714.2  GI:50251122
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 210384)
  AUTHORS   Burton,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUL-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 12, 2004 this sequence version replaced gi:50234128.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK29P13
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 209301 bases at least Q40
            Consensus quality: 209711 bases at least Q30
            Consensus quality: 209847 bases at least Q20
            Insert size: 209984; sum-of-contigs
            Insert size: 209994; 2.1% error; agarose-fp
            Quality coverage: 10.01x in Q20 bases; sum-of-contigs Quality
            coverage: 10.13x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    19750: contig of 19750 bp in length
            *    19751    19850: gap of 100 bp
            *    19851    37399: contig of 17549 bp in length
            *    37400    37499: gap of 100 bp
            *    37500   112088: contig of 74589 bp in length
            *   112089   112188: gap of 100 bp
            *   112189   196376: contig of 84188 bp in length
            *   196377   196476: gap of 100 bp
            *   196477   210384: contig of 13908 bp in length.
FEATURES             Location/Qualifiers
     source          1. .210384
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-29P13"
                     /clone_lib="DanioKey"
     misc_feature    1. .19750
                     /note="assembly_fragment:00422
                     fragment_chain:1"
     misc_feature    19851. .37399
                     /note="assembly_fragment:00191
                     fragment_chain:1"
     misc_feature    37500. .112088
                     /note="assembly_fragment:00690
                     fragment_chain:2"
     misc_feature    112189. .196376
                     /note="assembly_fragment:01732
                     fragment_chain:2"
     misc_feature    196477. .210384
                     /note="assembly_fragment:00009.0"
ORIGIN

  Query Match             7.6%;  Score 349.6;  DB 2;  Length 210384;
  Best Local Similarity   68.8%;  Pred. No. 1.4e-59;
  Matches  509;  Conservative    0;  Mismatches  229;  Indels    2;  Gaps    2;

Qy        187 AACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATA 246

Db     111340 AAGTAAAAAAAAAGTAAAACAAAATAAAATAAAAAGTAAATATATATATATATATATATA 111399

Qy        247 TATATACACACACACACATATATATATACACATATATATACACATATATACATATATA 306

Db     111400 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 111459
```

```
Qy        307 CACATATATACACATATATATACACATACATATGTACACATATATATACACATATGTATA 366
              | ||||||| | ||||||||| | ||| |||| || | ||||||||| | |||| ||||
Db     111460 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 111519

Qy        367 CACATATATACACATATATACACACATATATACACATATATACACACATATATACACATA 426
              | ||||||| | ||||||| | | ||||||| | ||||||| | | ||||||| | |||
Db     111520 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 111579

Qy        427 TATACACATATATACACATATACACATATATACACATATATACATATATACACATATATA 486
              |||| | ||||||| | ||||| | ||||||| | ||||||| ||||||| | |||||||
Db     111580 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 111639

Qy        487 TA-ATATACACACATATATATACACATATATACACACATATATACACATATATACACATA 545
              || ||||| | | ||||||||| | ||||||| | | ||||||| | ||||||| | |||
Db     111640 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 111699

Qy        546 TATATACACATATATACACATATATACATATATACACATATATATACATATATACACATA 605
              |||||| | ||||||| | ||||||| ||||||| | ||||||||| ||||||| | |||
Db     111700 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 111759

Qy        606 TATACATATATACACATATATACATATATACACACATATATACACATACATATACACACA 665
              |||| ||||||| | ||||||| ||||||| | | ||||||| | ||| ||||| | | |
Db     111760 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 111819

Qy        666 CATAGATATACATATATATACACATATATATACGTATATATATGTATATATATATGCTCC 725
              ||| ||||| ||||||||| | |||||||||  ||||||||| ||||||||||
Db     111820 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 111879

Qy        726 AGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACA 785
                  |  |||          |  |   ||    |   ||| |   | | | |  | | |
Db     111880 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 111939

Qy        786 GCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAA 845
                | |||    | |  | || |    |     | |||     | ||| ||| |
Db     111940 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 111999

Qy        846 TGTGTAATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTAACACTATATATATATATAGT 904
              | | ||   | |  ||| | || |  ||  ||   |  |||  ||||||||||||||
Db     112000 TATATATATATATATATATATATATATATATATATATATAAATATATATATATATATATA 112059

Qy        905 GAAAGCAGTGTAAACCTAAA 924
              | |  | | ||  |  || |
Db     112060 TATATAAATATATATATATA 112079
```

```
RESULT 71
CR790372
LOCUS       CR790372              299106 bp    DNA     linear   HTG 05-OCT-2004
DEFINITION  Danio rerio clone DKEY-70K13, *** SEQUENCING IN PROGRESS ***, 25
            unordered pieces.
ACCESSION   CR790372
VERSION     CR790372.4  GI:53828302
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 299106)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 5, 2004 this sequence version replaced gi:53755940.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK70K13
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 289331 bases at least Q40
            Consensus quality: 291856 bases at least Q30
            Consensus quality: 293478 bases at least Q20
            Insert size: 296706; sum-of-contigs
            Insert size: 225769; 2.9% error; agarose-fp
            Quality coverage: 6.22x in Q20 bases; sum-of-contigs Quality
            coverage: 9.16x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 25 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     3928: contig of 3928 bp in length
            *     3929     4028: gap of 100 bp
            *     4029     8417: contig of 4389 bp in length
            *     8418     8517: gap of 100 bp
            *     8518    12283: contig of 3766 bp in length
            *    12284    12383: gap of 100 bp
            *    12384    16946: contig of 4563 bp in length
            *    16947    17046: gap of 100 bp
            *    17047    19834: contig of 2788 bp in length
            *    19835    19934: gap of 100 bp
            *    19935    26863: contig of 6929 bp in length
            *    26864    26963: gap of 100 bp
            *    26964    29135: contig of 2172 bp in length
            *    29136    29235: gap of 100 bp
            *    29236    43722: contig of 14487 bp in length
            *    43723    43822: gap of 100 bp
            *    43823   109813: contig of 65991 bp in length
            *   109814   109913: gap of 100 bp
            *   109914   140749: contig of 30836 bp in length
            *   140750   140849: gap of 100 bp
            *   140850   156362: contig of 15513 bp in length
            *   156363   156462: gap of 100 bp
            *   156463   165292: contig of 8830 bp in length
            *   165293   165392: gap of 100 bp
            *   165393   243285: contig of 77893 bp in length
            *   243286   243385: gap of 100 bp
            *   243386   245562: contig of 2177 bp in length
            *   245563   245662: gap of 100 bp
            *   245663   250094: contig of 4432 bp in length
            *   250095   250194: gap of 100 bp
            *   250195   256405: contig of 6211 bp in length
            *   256406   256505: gap of 100 bp
            *   256506   258788: contig of 2283 bp in length
            *   258789   258888: gap of 100 bp
            *   258889   262780: contig of 3892 bp in length
            *   262781   262880: gap of 100 bp
            *   262881   265140: contig of 2260 bp in length
            *   265141   265240: gap of 100 bp
            *   265241   269635: contig of 4395 bp in length
            *   269636   269735: gap of 100 bp
            *   269736   272991: contig of 3256 bp in length
            *   272992   273091: gap of 100 bp
            *   273092   291385: contig of 18294 bp in length
            *   291386   291485: gap of 100 bp
            *   291486   293629: contig of 2144 bp in length
            *   293630   293729: gap of 100 bp
            *   293730   296461: contig of 2732 bp in length
            *   296462   296561: gap of 100 bp
            *   296562   299106: contig of 2545 bp in length.
FEATURES             Location/Qualifiers
     source          1. .299106
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-70K13"
```

```
                     /clone_lib="DanioKey"
     misc_feature    1. .3928
                     /note="assembly_fragment:00025
                     fragment_chain:1"
     misc_feature    4029. .8417
                     /note="assembly_fragment:00178
                     fragment_chain:1"
     misc_feature    8518. .12283
                     /note="assembly_fragment:00051
                     fragment_chain:1"
     misc_feature    12384. .16946
                     /note="assembly_fragment:00068
                     fragment_chain:1"
     misc_feature    17047. .19834
                     /note="assembly_fragment:03721
                     fragment_chain:1"
     misc_feature    19935. .26863
                     /note="assembly_fragment:00206
                     fragment_chain:1"
     misc_feature    26964. .29135
                     /note="assembly_fragment:03693
                     fragment_chain:1"
     misc_feature    29236. .43722
                     /note="assembly_fragment:00405
                     fragment_chain:2"
     misc_feature    43823. .109813
                     /note="assembly_fragment:01445
                     fragment_chain:2"
     misc_feature    109914. .140749
                     /note="assembly_fragment:01030
                     fragment_chain:2"
     misc_feature    140850. .156362
                     /note="assembly_fragment:00596
                     fragment_chain:3"
     misc_feature    156463. .165292
                     /note="assembly_fragment:00282
                     fragment_chain:3"
     misc_feature    165393. .243285
                     /note="assembly_fragment:02358
                     fragment_chain:3"
     misc_feature    243386. .245562
                     /note="assembly_fragment:03730
                     fragment_chain:4"
     misc_feature    245663. .250094
                     /note="assembly_fragment:00104
                     fragment_chain:4"
     misc_feature    250195. .256405
                     /note="assembly_fragment:00151
                     fragment_chain:4"
     misc_feature    256506. .258788
                     /note="assembly_fragment:00124
                     fragment_chain:5"
     misc_feature    258889. .262780
                     /note="assembly_fragment:00234
                     fragment_chain:5"
     misc_feature    262881. .265140
                     /note="assembly_fragment:03713
                     fragment_chain:6"
     misc_feature    265241. .269635
                     /note="assembly_fragment:00038
                     fragment_chain:6"
     misc_feature    269736. .272991
                     /note="assembly_fragment:00085"
     misc_feature    273092. .291385
                     /note="assembly_fragment:00801.0"
     misc_feature    291486. .293629
                     /note="assembly_fragment:03699"
     misc_feature    293730. .296461
                     /note="assembly_fragment:03706"
     misc_feature    296562. .299106
                     /note="assembly_fragment:03750"
ORIGIN
```

```
  Query Match             7.6%;  Score 349.6;  DB 2;  Length 299106;
  Best Local Similarity   60.4%;  Pred. No. 1.4e-59;
  Matches  492;  Conservative    0;  Mismatches  321;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db     242607 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 242666

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | |||||||||| | ||| |||| || | |
Db     242667 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 242726

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db     242727 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 242786

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db     242787 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 242846

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | |||||||||| ||||| | | |||||||||| | ||||||| | | |||
Db     242847 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 242906

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db     242907 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 242966

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db     242967 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 243026

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db     243027 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 243086

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||          | |||       |  |   ||   |    |||
Db     243087 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 243146

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
              |   | | | |  | | |    | |||   | |   | || |     |      | |
Db     243147 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 243206

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885
              ||        | ||| ||| || | |||  | |  ||| | ||  || |||| |   ||
Db     243207 TATATATATATATATATATATATATAAATATATATATATTTATTGTTGTTATATANANAA 243266

Qy        886 CACTATATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAAT 945
              |  | | |||||||||
Db     243267 AAAAAAAAATATATATATANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 243326

Qy        946 GTTATAGAAGAATGGCTAAGATGCCAAAGCTCAGTGTATGTGGCAGAGGCATGGTGTAGG 1005

Db     243327 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNC 243386

Qy       1006 GTGTGTCCAGGTTCATATATTGCATTAAGTGTGA 1039
                  | |    ||| | |  || ||||||
Db     243387 TGAGCAAAGGCTGTGAATACTTCTGTACGTGTGA 243420


RESULT 72
CR790384/c
LOCUS       CR790384              182223 bp    DNA     linear   HTG 08-OCT-2004
DEFINITION  Danio rerio clone DKEY-208I6, *** SEQUENCING IN PROGRESS ***, 7
            unordered pieces.
ACCESSION   CR790384
VERSION     CR790384.3  GI:54019829
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
```

```
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 182223)
  AUTHORS   McLay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 9, 2004 this sequence version replaced gi:52694441.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK208I6
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 180226 bases at least Q40
            Consensus quality: 180685 bases at least Q30
            Consensus quality: 180956 bases at least Q20
            Insert size: 181623; sum-of-contigs
            Insert size: 177450; 3.6% error; agarose-fp
            Quality coverage: 17.50x in Q20 bases; sum-of-contigs Quality
            coverage: 18.50x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    27274: contig of 27274 bp in length
            *   27275    27374: gap of 100 bp
            *   27375    66205: contig of 38831 bp in length
            *   66206    66305: gap of 100 bp
            *   66306   114861: contig of 48556 bp in length
            *  114862   114961: gap of 100 bp
            *  114962   118489: contig of 3528 bp in length
            *  118490   118589: gap of 100 bp
            *  118590   147487: contig of 28898 bp in length
            *  147488   147587: gap of 100 bp
            *  147588   158792: contig of 11205 bp in length
            *  158793   158892: gap of 100 bp
            *  158893   182223: contig of 23331 bp in length.
FEATURES             Location/Qualifiers
     source          1. .182223
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-208I6"
                     /clone_lib="DanioKey"
     misc_feature    1. .27274
                     /note="assembly_fragment:01073
                     fragment_chain:1"
     misc_feature    27375. .66205
                     /note="assembly_fragment:01954
                     fragment_chain:1"
     misc_feature    66306. .114861
                     /note="assembly_fragment:03886
                     fragment_chain:1"
     misc_feature    114962. .118489
                     /note="assembly_fragment:00044
                     fragment_chain:2"
     misc_feature    118590. .147487
                     /note="assembly_fragment:02902
                     fragment_chain:2"
     misc_feature    147588. .158792
                     /note="assembly_fragment:00096.0"
     misc_feature    158893. .182223
                     /note="assembly_fragment:00385"

ORIGIN

  Query Match             7.6%;  Score 349.4;  DB 2;  Length 182223;
  Best Local Similarity   68.3%;  Pred. No. 1.6e-59;
  Matches  499;  Conservative    0;  Mismatches  231;  Indels    1;  Gaps    1;

Qy        199 ATGTCATCTCAGAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACA 258
              || | | |    |||||||||       | ||||||||||||||||||||||||| | | |
Db      67159 ATTTAAGCAAGTAAAAAAAAAATATATATATATATATATATATATATATATATATATA 67100

Qy        259 CACACACATATATATATACACATATATATACACATATATACATATATACACATATATACA 318
              | | | |||||||||||| | ||||||||| | ||||||| ||||||| | ||||||| |
Db      67099 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 67040

Qy        319 CATATATATACACATACATATGTACACATATATATACACATATGTATACACATATATACA 378
              |||||||| | ||| |||| || | |||||||| | |||| |||| | ||||||| |
Db      67039 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 66980

Qy        379 CATATATACACACATATATACACATATATACACACATATATACACATATATACACATATA 438
              ||||||| | | ||||||| | ||||||| | | ||||||| | ||||||| | |||||
Db      66979 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 66920

Qy        439 TACACATATACACATATATACACATATATACATATATACACATATATATA-ATATACACA 497
              || | ||||| | ||||||| | ||||||| ||||||| | ||||||||| ||||| | |
Db      66919 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 66860

Qy        498 CATATATATACACATATATACACACATATATACACATATATACACATATATATACACATA 557
              |||||||| | ||||||| | | ||||||| | ||||||| | ||||||||| | |||
Db      66859 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 66800

Qy        558 TATACACATATATACATATATACACATATATATACATATATACACATATATACATATATA 617
              |||| | ||||||| ||||||| | ||||||||| ||||||| | ||||||| |||||||
Db      66799 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 66740

Qy        618 CACATATATACATATATACACACATATATACACATACATATACACACACATAGATATACA 677
              | ||||||| ||||||| | | ||||||| | ||| ||||| | | | ||| ||||| |
Db      66739 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 66680

Qy        678 TATATATACACATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAG 737
              |||||||| | |||||||||  |||||||| ||||||||||          |  |||
Db      66679 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 66620

Qy        738 AGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGT 797
                    |    ||     | |    ||| | ||  | | |  |  | |    | |||
Db      66619 TATATATATAATATATATAATATATATATATATAAATAAATATATAAATATATATATA 66560

Qy        798 GAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAA 857
              | |   | |||| |  |  ||    |  |       | ||| | |||      ||  ||
Db      66559 TATATATATATTTTAAATATATATATATAATATATAAATATATAAAAAATAAATATTTAA 66500

Qy        858 CCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATATATATATAGTGAAAGCAGTGTAA 917
              ||   | |   | |||  |  ||| |  || ||  ||||   | ||  |   |||
Db      66499 ATAATAATTTATAAATATAAATATATAATAAAATTTAATTATAATATAAATTAATAATAA 66440

Qy        918 ACCTAAAGGAT 928
              |  | ||  ||
Db      66439 ATATTAAATAT 66429


RESULT 73
CR790362
LOCUS       CR790362              223166 bp    DNA     linear   HTG 05-OCT-2004
DEFINITION  Danio rerio clone DKEY-219H14, *** SEQUENCING IN PROGRESS ***, 13
            unordered pieces.
ACCESSION   CR790362
VERSION     CR790362.3  GI:53828288
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 223166)
```

```
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 5, 2004 this sequence version replaced gi:52626676.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK219H14
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 218162 bases at least Q40
            Consensus quality: 219035 bases at least Q30
            Consensus quality: 219848 bases at least Q20
            Insert size: 221966; sum-of-contigs
            Insert size: 231484; 3.1% error; agarose-fp
            Quality coverage: 9.07x in Q20 bases; sum-of-contigs Quality
            coverage: 9.03x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 13 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *      1    45629: contig of 45629 bp in length
            *  45630    45729: gap of 100 bp
            *  45730    47743: contig of 2014 bp in length
            *  47744    47843: gap of 100 bp
            *  47844    71492: contig of 23649 bp in length
            *  71493    71592: gap of 100 bp
            *  71593    80196: contig of 8604 bp in length
            *  80197    80296: gap of 100 bp
            *  80297    94270: contig of 13974 bp in length
            *  94271    94370: gap of 100 bp
            *  94371   105308: contig of 10938 bp in length
            * 105309   105408: gap of 100 bp
            * 105409   113062: contig of 7654 bp in length
            * 113063   113162: gap of 100 bp
            * 113163   125638: contig of 12476 bp in length
            * 125639   125738: gap of 100 bp
            * 125739   129703: contig of 3965 bp in length
            * 129704   129803: gap of 100 bp
            * 129804   143077: contig of 13274 bp in length
            * 143078   143177: gap of 100 bp
            * 143178   148174: contig of 4997 bp in length
            * 148175   148274: gap of 100 bp
            * 148275   180179: contig of 31905 bp in length
            * 180180   180279: gap of 100 bp
            * 180280   223166: contig of 42887 bp in length.
FEATURES             Location/Qualifiers
     source          1. .223166
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-219H14"
                     /clone_lib="DanioKey"
     misc_feature    1. .45629
                     /note="assembly_fragment:02641
                     fragment_chain:1"
     misc_feature    45730. .47743
                     /note="assembly_fragment:00088
                     fragment_chain:1"
     misc_feature    47844. .71492
                     /note="assembly_fragment:01125
                     fragment_chain:1"
     misc_feature    71593. .80196
                     /note="assembly_fragment:00289
                     fragment_chain:1"
     misc_feature    80297. .94270
                     /note="assembly_fragment:00715
                     fragment_chain:1"
     misc_feature    94371. .105308
                     /note="assembly_fragment:00568
                     fragment_chain:2"
     misc_feature    105409. .113062
                     /note="assembly_fragment:00191
                     fragment_chain:2"
     misc_feature    113163. .125638
                     /note="assembly_fragment:00919
                     fragment_chain:2"
     misc_feature    125739. .129703
                     /note="assembly_fragment:00071
                     fragment_chain:3"
     misc_feature    129804. .143077
                     /note="assembly_fragment:00426
                     fragment_chain:3"
     misc_feature    143178. .148174
                     /note="assembly_fragment:00118.0"
     misc_feature    148275. .180179
                     /note="assembly_fragment:01481"
     misc_feature    180280. .223166
                     /note="assembly_fragment:01968"
ORIGIN

  Query Match             7.6%;  Score 349.4;  DB 2;  Length 223166;
  Best Local Similarity   79.2%;  Pred. No. 1.6e-59;
  Matches  427;  Conservative    0;  Mismatches  111;  Indels    1;  Gaps    1;

Qy        188 ACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATAT 247
              | || |  ||||| |  | | ||||||||| |||   ||| ||||||||||||||||||
Db     201702 AAAAAAAAGAAATTTTTTTTTAGAAAAAAATAAATAAATTATATATATATATATATATAT 201761

Qy        248 ATATACACACACACACATATATATATACACATATATATACACATATATACATATATAC 307
              ||||| | | | | | ||||||||||| | ||||||||| | ||||||| |||||||
Db     201762 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 201821

Qy        308 ACATATATACACATATATATACACATACATATGTACACATATATATACACATATGTATAC 367
              | ||||||| | ||||||||| | ||| |||| || | ||||||||| | |||| ||||
Db     201822 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 201881

Qy        368 ACATATATACACATATATACACACATATATACACATATATACACACATATATACACATAT 427
              | ||||||| | ||||||| | | ||||||| | ||||||| | | ||||||| | ||||
Db     201882 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 201941

Qy        428 ATACACATATATACACATATACACATATATACACATATATACATATATACACATATATAT 487
              ||| | ||||||| | ||||| | ||||||| | ||||||| ||||||| | ||||||||
Db     201942 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 202001

Qy        488 A-ATATACACACATATATATACACATATATACACACATATATACACATATATACACATAT 546
              | ||||| | | |||||||||| | ||||||| | | ||||||| | ||||||| | ||||
Db     202002 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 202061

Qy        547 ATATACACATATATACACATATATACATATATACACATATATATACATATATACACATAT 606
              ||||| | ||||||| | ||||||| ||||||| | ||||||||| ||||||| | ||||
Db     202062 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 202121

Qy        607 ATACATATATACACATATATACATATATACACACATATATACACATACATATACACACAC 666
              ||| || |||| | ||||||| ||||||| | | ||||||| | ||| |||| | | |
Db     202122 ATATATCTATATATATATATATATATATATATATATATATATATATATATATATATAT 202181

Qy        667 ATAGATATACATATATATACACATATATATACGTATATATATGTATATATATATGCTCC 725
              ||| ||||| ||||||||| | ||||||||| ||||||||| |||||||||| || |
Db     202182 ATATATATATATATATATATATATATATATATATATATATATATATATATATAACCTTC 202240

RESULT 74
CR759946
LOCUS       CR759946              232824 bp    DNA     linear   HTG 11-SEP-2004
```

```
DEFINITION  Danio rerio clone DKEY-79C1, *** SEQUENCING IN PROGRESS ***, 8
            unordered pieces.
ACCESSION   CR759946
VERSION     CR759946.2  GI:51988227
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 232824)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 11, 2004 this sequence version replaced gi:51965771.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK79C1
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 229932 bases at least Q40
            Consensus quality: 230704 bases at least Q30
            Consensus quality: 231171 bases at least Q20
            Insert size: 232124; sum-of-contigs
            Insert size: 219964; 3.3% error; agarose-fp
            Quality coverage: 8.73x in Q20 bases; sum-of-contigs Quality
            coverage: 9.26x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    33972: contig of 33972 bp in length
            *    33973    34072: gap of 100 bp
            *    34073    37589: contig of 3517 bp in length
            *    37590    37689: gap of 100 bp
            *    37690    45659: contig of 7970 bp in length
            *    45660    45759: gap of 100 bp
            *    45760    54370: contig of 8611 bp in length
            *    54371    54470: gap of 100 bp
            *    54471    87298: contig of 32828 bp in length
            *    87299    87398: gap of 100 bp
            *    87399   111650: contig of 24252 bp in length
            *   111651   111750: gap of 100 bp
            *   111751   170131: contig of 58381 bp in length
            *   170132   170231: gap of 100 bp
            *   170232   232824: contig of 62593 bp in length.
FEATURES             Location/Qualifiers
     source          1. .232824
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-79C1"
                     /clone_lib="DanioKey"
     misc_feature    1. .33972
                     /note="assembly_fragment:00912
                     fragment_chain:1"
     misc_feature    34073. .37589
                     /note="assembly_fragment:00025
                     fragment_chain:1"
     misc_feature    37690. .45659
                     /note="assembly_fragment:00053
                     fragment_chain:1"
     misc_feature    45760. .54370
                     /note="assembly_fragment:00127
                     fragment_chain:1"
     misc_feature    54471. .87298
                     /note="assembly_fragment:00512
                     fragment_chain:2"
     misc_feature    87399. .111650
                     /note="assembly_fragment:00209
                     fragment_chain:2"
     misc_feature    111751. .170131
                     /note="assembly_fragment:01393"
     misc_feature    170232. .232824
                     /note="assembly_fragment:02188.0"
ORIGIN

  Query Match             7.6%;  Score 349.4;  DB 2;  Length 232824;
  Best Local Similarity   59.8%;  Pred. No. 1.6e-59;
  Matches  502;  Conservative    0;  Mismatches  336;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACATATATATATACACATATATA 286
              | |||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db     169438 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 169497

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db     169498 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 169557

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db     169558 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 169617

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db     169618 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 169677

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db     169678 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 169737

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db     169738 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 169797

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db     169798 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 169857

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db     169858 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 169917

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||          |  |||          |  |   ||    |   |||
Db     169918 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 169977

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
              |   | | | |  | | |   | |||   | |  | || |   |    |    | ||
Db     169978 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 170037

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885
              ||       | ||| ||| || | ||   | |  ||| | ||   ||  ||      ||
Db     170038 TATATATATATATATATATATATATATATATATATATATATATAATATATATATTTATAT 170097

Qy        886 CACTATATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAAT 945
                |||||||||||||   | |      || |
Db     170098 ATATATATATATATATATATATATACATATATNNNNNNNNNNNNNNNNNNNNNNNNN 170157

Qy        946 GTTATAGAAGAATGGCTAAGATGCCAAAGCTCAGTGTATGTGGCAGAGGCATGGTGTAGG 1005

Db     170158 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 170217

Qy       1006 GTGTGTCCAGGTTCATATATTGCATTAAGTGTGAGAACACCCTGGAGTATGAACCAAGA 1064
```

```
                     ||| |||   | |      | ||   |||   || |||  | | |
Db     170218 NNNNNNNNNNNNNNATACATTTTGTCACTAAAAACAATGTCCTCTGGTGTGACACTATA 170276


RESULT 75
CR318636
LOCUS       CR318636               233296 bp    DNA     linear   HTG 31-AUG-2004
DEFINITION  Danio rerio clone DKEY-249N13, *** SEQUENCING IN PROGRESS ***, 10
            unordered pieces.
ACCESSION   CR318636
VERSION     CR318636.3  GI:51861411
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 233296)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-AUG-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 2, 2004 this sequence version replaced gi:44909414.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK249N13
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 229955 bases at least Q40
            Consensus quality: 230652 bases at least Q30
            Consensus quality: 231088 bases at least Q20
            Insert size: 232396; sum-of-contigs
            Insert size: 229559; 5.6% error; agarose-fp
            Quality coverage: 7.04x in Q20 bases; sum-of-contigs Quality
            coverage: 7.26x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2974: contig of 2974 bp in length
            *     2975     3074: gap of 100 bp
            *     3075    13370: contig of 10296 bp in length
            *    13371    13470: gap of 100 bp
            *    13471    23811: contig of 10341 bp in length
            *    23812    23911: gap of 100 bp
            *    23912    37932: contig of 14021 bp in length
            *    37933    38032: gap of 100 bp
            *    38033    54166: contig of 16134 bp in length
            *    54167    54266: gap of 100 bp
            *    54267   102290: contig of 48024 bp in length
            *   102291   102390: gap of 100 bp
            *   102391   125681: contig of 23291 bp in length
            *   125682   125781: gap of 100 bp
            *   125782   142646: contig of 16865 bp in length
            *   142647   142746: gap of 100 bp
            *   142747   165218: contig of 22472 bp in length
            *   165219   165318: gap of 100 bp
            *   165319   233296: contig of 67978 bp in length.
FEATURES             Location/Qualifiers
     source          1. .233296
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-249N13"
                     /clone_lib="DanioKey"
     misc_feature    1. .2974
                     /note="assembly_fragment:00024
                     fragment_chain:1"
     misc_feature    3075. .13370
                     /note="assembly_fragment:00163
                     fragment_chain:1"
     misc_feature    13471. .23811
                     /note="assembly_fragment:00045
                     fragment_chain:1"
     misc_feature    23912. .37932
                     /note="assembly_fragment:00283
                     fragment_chain:2"
     misc_feature    38033. .54166
                     /note="assembly_fragment:00602
                     fragment_chain:2"
     misc_feature    54267. .102290
                     /note="assembly_fragment:01245
                     fragment_chain:2"
     misc_feature    102391. .125681
                     /note="assembly_fragment:01003
                     fragment_chain:3"
     misc_feature    125782. .142646
                     /note="assembly_fragment:00432
                     fragment_chain:3"
     misc_feature    142747. .165218
                     /note="assembly_fragment:00787
                     fragment_chain:3"
     misc_feature    165319. .233296
                     /note="assembly_fragment:01715.0"
ORIGIN

  Query Match             7.6%;  Score 349.4;  DB 2;  Length 233296;
  Best Local Similarity   69.7%;  Pred. No. 1.6e-59;
  Matches  487;  Conservative    0;  Mismatches  211;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db     164502 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 164561

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db     164562 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 164621

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db     164622 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 164681

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db     164682 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 164741

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db     164742 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 164801

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db     164802 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 164861

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db     164862 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 164921

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db     164922 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 164981

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||||          | |||          |  |   ||    |   |||
Db     164982 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 165041
```

```
Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
              |   | | | |  | | |    | |||    | |   | || |      |       | |
Db     165042 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 165101

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885
              |       | ||| ||| || | ||   | |  ||| | ||   ||  |||  |  |||
Db     165102 AATATATATATATATATATATATATATATATATATATATAATATATAAATATATAA 165161

Qy        886 CACTATATATATATATAGTGAAAGCAGTGTAAACCTAAA 924
              | |||| ||||||||  | |  | | || |  ||||
Db     165162 AAATATAATTATATATATATATATAAATATATATATAAA 165200


RESULT 76
CR631122
LOCUS       CR631122              162977 bp    DNA     linear   HTG 13-AUG-2004
DEFINITION  Danio rerio clone DKEYP-36H11, *** SEQUENCING IN PROGRESS ***, 7
            unordered pieces.
ACCESSION   CR631122
VERSION     CR631122.2  GI:51241630
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 162977)
  AUTHORS   McLay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-AUG-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 13, 2004 this sequence version replaced gi:51127401.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zKp36H11
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 159934 bases at least Q40
            Consensus quality: 160525 bases at least Q30
            Consensus quality: 161089 bases at least Q20
            Insert size: 162377; sum-of-contigs
            Insert size: 159156; 13.8% error; agarose-fp
            Quality coverage: 7.39x in Q20 bases; sum-of-contigs Quality
            coverage: 7.54x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    11615: contig of 11615 bp in length
            *    11616    11715: gap of 100 bp
            *    11716    30875: contig of 19160 bp in length
            *    30876    30975: gap of 100 bp
            *    30976    47841: contig of 16866 bp in length
            *    47842    47941: gap of 100 bp
            *    47942   115881: contig of 67940 bp in length
            *   115882   115981: gap of 100 bp
            *   115982   134454: contig of 18473 bp in length
            *   134455   134554: gap of 100 bp
            *   134555   137147: contig of 2593 bp in length
            *   137148   137247: gap of 100 bp
            *   137248   162977: contig of 25730 bp in length.

FEATURES             Location/Qualifiers
     source          1. .162977
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEYP-36H11"
                     /clone_lib="DanioKeypilot"
     misc_feature    1. .11615
                     /note="assembly_fragment:00030
                     fragment_chain:1"
     misc_feature    11716. .30875
                     /note="assembly_fragment:00862
                     fragment_chain:1"
     misc_feature    30976. .47841
                     /note="assembly_fragment:00397
                     fragment_chain:1"
     misc_feature    47942. .115881
                     /note="assembly_fragment:01121
                     fragment_chain:1"
     misc_feature    115982. .134454
                     /note="assembly_fragment:00177
                     fragment_chain:1"
     misc_feature    134555. .137147
                     /note="assembly_fragment:00008"
     misc_feature    137248. .162977
                     /note="assembly_fragment:00618.0"
ORIGIN

  Query Match             7.6%;  Score 349.2;  DB 2;  Length 162977;
  Best Local Similarity   69.9%;  Pred. No. 1.8e-59;
  Matches  485;  Conservative    0;  Mismatches  208;  Indels    1;  Gaps    1;

Qy        209 AGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACACATA 268
              ||| |||  | |      | ||||||||||||||||||||||| | | | | | | |||
Db     115076 AGAGAAACTATATATATATATATATATATATATATATATATATATATATATATATATATA 115135

Qy        269 TATATATACACATATATATACACATATATACATATATACACATATATACACATATATATA 328
              |||||||| | ||||||||| | ||||||| ||||||| | ||||||| | |||||||||
Db     115136 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 115195

Qy        329 CACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACA 388
              | ||| |||| || | ||||||||| | |||| |||| | ||||||| | ||||||| |
Db     115196 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 115255

Qy        389 CACATATATACACATATATACACACATATATACACATATATACACATATATACACATATA 448
              | ||||||| | ||||||| | | ||||||| | ||||||| | ||||||| | |||||
Db     115256 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 115315

Qy        449 CACATATATACACATATATACATATATACACATATATATAATATACACACATATATATAC 508
              | ||||||| | ||||||| ||||||| | |||||||| ||||| | | |||||||||
Db     115316 TATATATATATATATATATATATATATATATATATATAT-ATATATATATATATATAT 115374

Qy        509 ACATATATACACACATATATACACATATATACACATATATATACACATATATACACATAT 568
              | ||||||| | | ||||||| | ||||||| | ||||||||| | ||||||| | ||||
Db     115375 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 115434

Qy        569 ATACATATATACACATATATATACATATATACACATATATACATATATACACATATATAC 628
              ||| ||||||| | ||||||||| ||||||| | ||||||| ||||||| | |||||||
Db     115435 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 115494

Qy        629 ATATATACACACATATATACACATACATATACACACACATAGATATACATATATATACAC 688
              ||||||| | | ||||||| | ||| ||||| | | | ||| ||||| |||||||||| |
Db     115495 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 115554

Qy        689 ATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTT 748
              |||||||||  ||||||||| ||||||||||          | |||          | |
Db     115555 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 115614

Qy        749 GATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACAT 808
                ||    |   ||| |   | | | |  | | |   | |||   | |   | ||
Db     115615 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 115674

Qy        809 TTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAA 868
```

```
Db     115675 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 115734

Qy        869 GGTAGTTAAGAAGGTAACACTATATATATATATA 902
              ||  ||   |  ||| | ||||||||||||||||
Db     115735 TATATATATATATATAATATTATATATATATATA 115768


RESULT 77
AC021590/c
LOCUS       AC021590               171427 bp    DNA     linear   PRI 01-JAN-2002
DEFINITION  Homo sapiens chromosome 8, clone RP11-143H8, complete sequence.
ACCESSION   AC021590
VERSION     AC021590.10  GI:17432882
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 171427)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 8, clone RP11-143H8
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 171427)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 171427)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JAN-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Dec 10, 2001 this sequence version replaced gi:17402804.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L5777
                Center clone name: 143_H_8
            --------------.
FEATURES             Location/Qualifiers
     source          1. .171427
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="RP11-143H8"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   complement(2020. .2137)
                     /rpt_family="L2"
     repeat_region   complement(2160. .2295)
                     /rpt_family="L2"
     repeat_region   complement(2349. .2479)
                     /rpt_family="FLAM_C"
     repeat_region   complement(2483. .2835)
                     /rpt_family="L2"
     repeat_region   complement(3975. .4107)
                     /rpt_family="FLAM_C"
     repeat_region   complement(4639. .4747)
                     /rpt_family="L2"
     repeat_region   complement(5060. .5878)
                     /rpt_family="L1PA16"
     repeat_region   complement(6034. .6096)
                     /rpt_family="MIR"
     repeat_region   6248. .6555
                     /rpt_family="AluJb"
     repeat_region   6716. .7017
                     /rpt_family="L2"
     repeat_region   7035. .7070
                     /rpt_family="(CA)n"
     repeat_region   7333. .7572
                     /rpt_family="MIR"
     repeat_region   7627. .7903
                     /rpt_family="L2"
     repeat_region   8277. .8484
                     /rpt_family="AluJo"
     repeat_region   8718. .8877
                     /rpt_family="L3"
     repeat_region   complement(9247. .9483)
                     /rpt_family="MER46A"
     repeat_region   10206. .10263
                     /rpt_family="AT_rich"
     repeat_region   10421. .10493
                     /rpt_family="AT_rich"
     repeat_region   complement(10891. .11060)
                     /rpt_family="L2"
     repeat_region   11256. .11305
                     /rpt_family="AT_rich"
     repeat_region   11809. .12090
                     /rpt_family="AluSx"
     repeat_region   12091. .12189
                     /rpt_family="(TAGA)n"
     repeat_region   12406. .12512
                     /rpt_family="FRAM"
     repeat_region   12513. .12539
                     /rpt_family="AT_rich"
     repeat_region   12566. .12755
                     /rpt_family="MER61A"
     repeat_region   12883. .18519
                     /rpt_family="L1PA6"
     repeat_region   18929. .18994
                     /rpt_family="MIR"
```

```
  repeat_region   20160. .20182
                  /rpt_family="(TGGA)n"
  repeat_region   complement(20466. .20667)
                  /rpt_family="MIR"
  repeat_region   20671. .20850
                  /rpt_family="FAM"
  repeat_region   20880. .20909
                  /rpt_family="AT_rich"
  repeat_region   complement(20910. .21037)
                  /rpt_family="FRAM"
  repeat_region   21121. .21419
                  /rpt_family="AluJo"
  repeat_region   complement(21424. .21508)
                  /rpt_family="L1ME"
  repeat_region   21934. .22046
                  /rpt_family="MIR3"
  repeat_region   23344. .23388
                  /rpt_family="AT_rich"
  repeat_region   23449. .23483
                  /rpt_family="(TA)n"
  repeat_region   complement(23484. .23779)
                  /rpt_family="AluSx"
  repeat_region   24093. .24190
                  /rpt_family="L2"
  repeat_region   24985. .25019
                  /rpt_family="L1MA1"
  repeat_region   25020. .25319
                  /rpt_family="AluSx"
  repeat_region   25320. .25557
                  /rpt_family="L1MA1"
  repeat_region   25558. .25581
                  /rpt_family="AT_rich"
  repeat_region   25777. .25802
                  /rpt_family="AT_rich"
  repeat_region   complement(26663. .26756)
                  /rpt_family="L1ME4A"
  repeat_region   26764. .26817
                  /rpt_family="(TAAAAA)n"
  repeat_region   27050. .27082
                  /rpt_family="AT_rich"
  repeat_region   complement(27455. .27626)
                  /rpt_family="L1ME4A"
  repeat_region   30048. .30108
                  /rpt_family="AluJ/FRAM"
  repeat_region   30249. .30269
                  /rpt_family="AT_rich"
  repeat_region   complement(30270. .30568)
                  /rpt_family="AluSq"
  repeat_region   31382. .31690
                  /rpt_family="AluSx"
  repeat_region   complement(32052. .32214)
                  /rpt_family="L2"
  repeat_region   32452. .32743
                  /rpt_family="AluSc"
  repeat_region   32839. .32969
                  /rpt_family="L2"
  repeat_region   complement(33055. .33190)
                  /rpt_family="FLAM_C"
  repeat_region   complement(33266. .33548)
                  /rpt_family="L2"
  repeat_region   33600. .33795
                  /rpt_family="L2"
  repeat_region   33939. .34117
                  /rpt_family="L1MDa"

 Query Match             7.6%;  Score 349.2;  DB 9;  Length 171427;
 Best Local Similarity   65.2%;  Pred. No. 1.8e-59;
 Matches  609;  Conservative    0;  Mismatches  313;  Indels   12;  Gaps    6;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC--AAAAT 58
              || ||||||||| ||||   |||| |||||||||||| ||||||||||||||  |||||
Db      74773 AGATCAAGACCATCCTGGCTAACACGGTGAAACCCCGTCTCTACTAAAAATACAAAAAT 74714

Qy         59 TAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGA 118
              |||   |||||||||      |||||||| ||||||||||  ||||||  ||||||||||
Db      74713 TAGCAGGGCATGGTGGCGGGCGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGA 74654

Qy        119 ATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCC 177
              ||  |  |||  | | || || |||| |||| | ||| | ||||   |||| ||||||||
Db      74653 ATGGCGGGAA--TCCGGGAGGCGGAGCTTGCAGTGAGCCGAGATAGCGCCACTGCACTCC 74596

Qy        178 AGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAGGCATTT-----TATA 232
              ||||  ||  |||        | | |       | ||||||||  ||  |||  |      || |
Db      74595 AGCCTGGGGGACAGAGCGAGACTCTGTCTCAAAAAAAAAAAGAAAACATATAAATATAAA 74536

Qy        233 TATATATATATATATATATACACACACACACATATATATATACACATATATATACACA 292
              ||| |||||| ||||||||| |||   | | |||| |||||||   | |||||  ||| |||
Db      74535 TATGTATATAAATATATATATACATTTATATACATTTATATATTTATATATAAATATACA 74476

Qy        293 TATATACATATATACACATATATACACATATATATACACATACATATGTACACATATATA 352
              ||||||  | ||||   ||||||    ||||||  || | ||   ||| ||   ||||||
Db      74475 TATATATGTGTATATTTATATATGTTTATATATTTATATATGTTTATATATTTATATATG 74416

Qy        353 TACACATATGTATACACATATATACACATATATACACACATATATACACATATATACACA 412
              | | |||| |||| | ||||  | | || | || | |  ||||| | ||||||| | |
Db      74415 TTTATATATTTATATATATATTAATATATGTTTATATATTTATATATAAATATATATAAA 74356

Qy        413 CATATATACACATATATACACATATATACACATATACACATATATACACATATATACATA 472
              ||| |||| | | ||||| | ||||||| | | ||  | ||||| | | ||| ||| |||
Db      74355 CATTTATATATAAATATATAAATATATATAAACATTTATATATAAATATATAAATATATA 74296

Qy        473 TATACACATATATAT-AATATACACACATATATATACACATATATACACACATATATACA 531
              || |||  ||||||| |||||| | | ||||||| |||  |||||| | | ||||| | |
Db      74295 TAAACATTTATATATAAATATATAAATATATATAAACATTTATATATAAATATATAAATA 74236

Qy        532 CATATATACACATATATATACACATATATACACATATATACATATATACACATATATATA 591
              |||||  |||  |||||||| | |||||| |||| |||| ||| ||| | | |||||||
Db      74235 TATATAAACATTTATATATAAATATATATAAACATTTATATATAAATATATAAATATATA 74176

Qy        592 CATATATACACATATATACATATATACACATATATACATATATACACACATATATACACA 651
              | | ||  | ||||| | ||||| | | ||| | |  |||||| | | ||||| | | |
Db      74175 TAAACATTTATATATAAATATATAAATATATAAACATTTATATATAAATATATAAATATA 74116

Qy        652 TACATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTA 711
              || | ||  | | | | | ||||| ||||||| |||  |||||||  |||||||   |
Db      74115 TAAACATTTATATATAAATATATAAATATATAAACATTTATATATAAATATATATAAACA 74056

Qy        712 TATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAA 771
              | |||||||          || |||          | | |  |    |  |  | | |  |
Db      74055 TTTATATATAAATATATATTTATATATAAATATATATAAACATTTATATATAAATATATA 73996

Qy        772 AAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAG 831
              || | |  || | |     ||||    | |  | ||||    |      | |||
Db      73995 AATATATATTTATATATACGTATATATAAATATATATTTATATATACGTATATATAAATA 73936

Qy        832 GGGAAATACATAAATGTGTAATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTAACACTA 890
              |  ||  |||||  | ||  || |  ||| | || | |||    |   |  |||   ||
Db      73935 TATATATGTATAAACATATATAAATATATATTTATATATGTATAAACATATATAAATATA 73876

Qy        891 TATATATATATAGTGAAAGCAGTGTAAACCTAAA 924
              ||| |||||||    |||    | ||||  || |
Db      73875 TATTTATATATGTATAAACATATATAAATATATA 73842


RESULT 78
CR759788/c
LOCUS       CR759788              195311 bp    DNA     linear   HTG 07-SEP-2004
DEFINITION  Danio rerio clone CH211-153F13, *** SEQUENCING IN PROGRESS ***, 11
            unordered pieces.
ACCESSION   CR759788
VERSION     CR759788.3  GI:51950464
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
```

```
          Cypriniformes; Cyprinidae; Danio.
REFERENCE 1  (bases 1 to 195311)
  AUTHORS McLay,K.
  TITLE   Direct Submission
  JOURNAL Submitted (06-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
          Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
          zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT   On Sep 9, 2004 this sequence version replaced gi:51890035.
          -------------- Genome Center
          Center: Wellcome Trust Sanger Institute
          Center code: SC
          Web site: http://www.sanger.ac.uk
          Contact: zfish-help@sanger.ac.uk
          -------------- Project Information
          Center project name: zC153F13
          -------------- Summary Statistics
          Assembly program: XGAP4; version 4.5
          Chemistry: Dye-terminator; 100% of reads
          Consensus quality: 191941 bases at least Q40
          Consensus quality: 192830 bases at least Q30
          Consensus quality: 193638 bases at least Q20
          Insert size: 194311; sum-of-contigs
          Insert size: 184045; 2.4% error; agarose-fp
          Quality coverage: 6.41x in Q20 bases; sum-of-contigs Quality
          coverage: 6.89x in Q20 bases; agarose-fp
          --------------.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 11 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *       1     4825: contig of 4825 bp in length
          *    4826     4925: gap of 100 bp
          *    4926    14966: contig of 10041 bp in length
          *   14967    15066: gap of 100 bp
          *   15067    17900: contig of 2834 bp in length
          *   17901    18000: gap of 100 bp
          *   18001    22764: contig of 4764 bp in length
          *   22765    22864: gap of 100 bp
          *   22865    40985: contig of 18121 bp in length
          *   40986    41085: gap of 100 bp
          *   41086    65693: contig of 24608 bp in length
          *   65694    65793: gap of 100 bp
          *   65794   118569: contig of 52776 bp in length
          *  118570   118669: gap of 100 bp
          *  118670   131148: contig of 12479 bp in length
          *  131149   131248: gap of 100 bp
          *  131249   172432: contig of 41184 bp in length
          *  172433   172532: gap of 100 bp
          *  172533   178335: contig of 5803 bp in length
          *  178336   178435: gap of 100 bp
          *  178436   195311: contig of 16876 bp in length.
FEATURES             Location/Qualifiers
     source          1. .195311
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-153F13"
                     /clone_lib="CHORI-211"
     misc_feature    1. .4825
                     /note="assembly_fragment:00162
                     fragment_chain:1
                     clone_end:SP6
                     vector_side:left"
     misc_feature    4926. .14966
                     /note="assembly_fragment:00231
                     fragment_chain:1"
     misc_feature    15067. .17900
                     /note="assembly_fragment:00038
                     fragment_chain:1"
     misc_feature    18001. .22764
                     /note="assembly_fragment:00112
                     fragment_chain:1"
     misc_feature    22865. .40985
                     /note="assembly_fragment:00460
                     fragment_chain:2"
     misc_feature    41086. .65693
                     /note="assembly_fragment:00820
                     fragment_chain:2"
     misc_feature    65794. .118569
                     /note="assembly_fragment:01583
                     fragment_chain:2"
     misc_feature    118670. .131148
                     /note="assembly_fragment:00325
                     fragment_chain:2"
     misc_feature    131249. .172432
                     /note="assembly_fragment:01060
                     fragment_chain:2"
     misc_feature    172533. .178335
                     /note="assembly_fragment:00062
                     fragment_chain:2"
     misc_feature    178436. .195311
                     /note="assembly_fragment:00627
                     fragment_chain:2"
ORIGIN

  Query Match             7.6%;  Score 349.2;  DB 2;  Length 195311;
  Best Local Similarity   63.3%;  Pred. No. 1.7e-59;
  Matches  475;  Conservative    0;  Mismatches  273;  Indels    2;  Gaps    2;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||||| | | | | | |||||||||| | |||||||
Db     178944 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 178885

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db     178884 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 178825

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db     178824 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 178765

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db     178764 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 178705

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db     178704 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 178645

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db     178644 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 178585

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db     178584 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 178525

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db     178524 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 178465

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| |||||||||||          |  ||
Db     178464 TATATATATATATATATATATATATATATNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 178405

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825

Db     178404 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 178345

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTA 884
                    | ||| ||| || | ||   | |  ||| | || |  ||  ||   |  ||
Db     178344 NNNNNNNNATATATATATATATATATATATATATATATATATATATATATATATATATA 178285
```

```
Qy        885 ACACTATATATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAA 944
              ||||||||||||||   | |    | || |  ||      | |   ||    ||
Db     178284 TATATATATATATATATATATATATATATATATATATATTCATAGATCTGTGGCGAGAAT 178225

Qy        945 TGTTATAGAAGAATGGCTAAGATGCCAAAG 974
              ||  | | |||   |  |||   |  ||||
Db     178224 TGCAAGAAAAGTTAGAATAACTGGTTAAAG 178195


RESULT 79
CR762470
LOCUS       CR762470              199230 bp    DNA     linear   HTG 16-SEP-2004
DEFINITION  Danio rerio clone DKEY-12N16, *** SEQUENCING IN PROGRESS ***, 15
            unordered pieces.
ACCESSION   CR762470
VERSION     CR762470.2  GI:52213994
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 199230)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 16, 2004 this sequence version replaced gi:52000316.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK12N16
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 192957 bases at least Q40
            Consensus quality: 193980 bases at least Q30
            Consensus quality: 194701 bases at least Q20
            Insert size: 197830; sum-of-contigs
            Insert size: 204603; 6.7% error; agarose-fp
            Quality coverage: 9.51x in Q20 bases; sum-of-contigs Quality
            coverage: 9.54x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 15 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    22618: contig of 22618 bp in length
            *    22619    22718: gap of 100 bp
            *    22719    28510: contig of 5792 bp in length
            *    28511    28610: gap of 100 bp
            *    28611    61395: contig of 32785 bp in length
            *    61396    61495: gap of 100 bp
            *    61496    65559: contig of 4064 bp in length
            *    65560    65659: gap of 100 bp
            *    65660    72529: contig of 6870 bp in length
            *    72530    72629: gap of 100 bp
            *    72630    92239: contig of 19610 bp in length
            *    92240    92339: gap of 100 bp
            *    92340    98985: contig of 6646 bp in length
            *    98986    99085: gap of 100 bp
            *    99086   146980: contig of 47895 bp in length
            *   146981   147080: gap of 100 bp
            *   147081   152421: contig of 5341 bp in length
            *   152422   152521: gap of 100 bp
            *   152522   157138: contig of 4617 bp in length
            *   157139   157238: gap of 100 bp
            *   157239   167850: contig of 10612 bp in length
            *   167851   167950: gap of 100 bp
            *   167951   170933: contig of 2983 bp in length
            *   170934   171033: gap of 100 bp
            *   171034   186729: contig of 15696 bp in length
            *   186730   186829: gap of 100 bp
            *   186830   195213: contig of 8384 bp in length
            *   195214   195313: gap of 100 bp
            *   195314   199230: contig of 3917 bp in length.
FEATURES             Location/Qualifiers
     source          1. .199230
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-12N16"
                     /clone_lib="DanioKey"
     misc_feature    1. .22618
                     /note="assembly_fragment:01219
                     fragment_chain:1"
     misc_feature    22719. .28510
                     /note="assembly_fragment:00267
                     fragment_chain:1"
     misc_feature    28611. .61395
                     /note="assembly_fragment:01636
                     fragment_chain:1"
     misc_feature    61496. .65559
                     /note="assembly_fragment:00116
                     fragment_chain:1"
     misc_feature    65660. .72529
                     /note="assembly_fragment:00202
                     fragment_chain:1"
     misc_feature    72630. .92239
                     /note="assembly_fragment:00921
                     fragment_chain:2"
     misc_feature    92340. .98985
                     /note="assembly_fragment:00329
                     fragment_chain:2"
     misc_feature    99086. .146980
                     /note="assembly_fragment:02193
                     fragment_chain:2"
     misc_feature    147081. .152421
                     /note="assembly_fragment:00157
                     fragment_chain:3"
     misc_feature    152522. .157138
                     /note="assembly_fragment:00407
                     fragment_chain:3"
     misc_feature    157239. .167850
                     /note="assembly_fragment:00593
                     fragment_chain:4"
     misc_feature    167951. .170933
                     /note="assembly_fragment:00070
                     fragment_chain:4"
     misc_feature    171034. .186729
                     /note="assembly_fragment:00727
                     fragment_chain:5"
     misc_feature    186830. .195213
                     /note="assembly_fragment:00491
                     fragment_chain:5"
     misc_feature    195314. .199230
                     /note="assembly_fragment:00081.0"
ORIGIN

  Query Match             7.6%;  Score 349.2;  DB 2;  Length 199230;
  Best Local Similarity   68.5%;  Pred. No. 1.7e-59;
  Matches  511;  Conservative    0;  Mismatches  233;  Indels    2;  Gaps    2;

Qy        181 CCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATA 240
              |||| ||| || |   |||  | |  | |  ||| ||| ||   | |  ||||||||||
Db      26837 CCAGACAAAAAAAAAAAAAAATTAAATAAATAAATAAATAAATAAATAAATATATATATA 26896
```

```
Qy        241 TATATATATATACACACACACACATATATATATACACATATATATACACATATATACA 300
              ||||||||||||| | | | | | |||||||||||| | ||||||||| | ||||||| |
Db      26897 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 26956

Qy        301 TATATACACATATATACACATATATATACACATACATATGTACACATATATATACACATA 360
              |||||| | ||||||| | ||||||||| | ||| |||| || | ||||||||| | |||
Db      26957 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 27016

Qy        361 TGTATACACATATATACACATATATACACACATATATACACATATATACACACATATATA 420
              | |||| | ||||||| | ||||||| | | ||||||| | ||||||| | | |||||||
Db      27017 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 27076

Qy        421 CACATATATACACATATATACACATATACACATATATACACATATATACATATATACACA 480
              | |||||||| | ||||||| | ||||| | ||||||| | ||||||| ||||||| | |
Db      27077 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 27136

Qy        481 TATATATA-ATATACACACATATATATACACATATATACACACATATATACACATATATA 539
              |||||||| ||||| | | ||||||||| | ||||||| | | ||||||| | |||||||
Db      27137 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 27196

Qy        540 CACATATATATACACATATATACACATATATACATATATACACATATATATACATATATA 599
              | ||||||||| | ||||||| | ||||||| ||||||| | ||||||||| |||||||
Db      27197 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 27256

Qy        600 CACATATATACATATATACACATATATACATATATACACACATATATACACATACATATA 659
              | ||||||| ||||||| | ||||||| ||||||| | | ||||||| | ||| |||||
Db      27257 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 27316

Qy        660 CACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATA 719
              | | | ||| ||||| ||||||||| | ||||||||| |||||||||| ||||||||||
Db      27317 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 27376

Qy        720 TGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGT 779
              |          | |||          | |   ||     |    ||| |   | | | |
Db      27377 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 27436

Qy        780 TTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATA 839
              |  | |    | |||    | |   | || |     |     | |||       | |||
Db      27437 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 27496

Qy        840 CATAAATGTGTAATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTAACACTATATATATA 898
              ||| || | ||   | |  ||| | || |  ||  ||   |  ||    ||||||||||
Db      27497 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 27556

Qy        899 TATAGTGAAAGCAGTGTAAACCTAAA 924
              ||||   | |    | || |  || |
Db      27557 TATATATATATATATATATATATA 27582
```

```
RESULT 80
AC098061/c
LOCUS       AC098061               247615 bp    DNA     linear   HTG 10-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-115G17, WORKING DRAFT SEQUENCE, 2
            unordered pieces.
ACCESSION   AC098061
VERSION     AC098061.7  GI:30522584
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 247615)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 247615)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-OCT-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 247615)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 10, 2003 this sequence version replaced gi:23096179.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GGBZ
```

```
            Center clone name: CH230-115G17
          ---------- Summary Statistics
            Assembly program: Atlas 3.0;
            Consensus quality: 219440 bases at least Q40
            Consensus quality: 221621 bases at least Q30
            Consensus quality: 223241 bases at least Q20
            Estimated insert size: 234299; sum-of-contigs estimation
            Quality coverage: 8x in Q20 bases; sum-of-contigs estimation
          ----------
          * NOTE: Estimated insert size may differ from sequence length
          *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 2 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *       1     9721: contig of 9721 bp in length
          *    9722     9821: gap of unknown length
          *    9822   247615: contig of 237794 bp in length.
FEATURES             Location/Qualifiers
     source          1. .247615
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-115G17"
ORIGIN

  Query Match             7.6%;  Score 349.2;  DB 2;  Length 247615;
  Best Local Similarity   67.1%;  Pred. No. 1.7e-59;
  Matches  523;  Conservative    0;  Mismatches  254;  Indels    2;  Gaps    2;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||||| | | | | | | |||||||||||| | |||||||
Db       9080 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 9021

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db       9020 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 8961

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db       8960 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 8901

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db       8900 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 8841

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db       8840 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 8781

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db       8780 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 8721

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db       8720 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 8661

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db       8660 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 8601

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||          | |||          | |   ||    |    |||
Db       8600 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 8541

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
              |  | | | |  | | |    | |||    | |   | || |      |      | |
Db       8540 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 8481

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTA 884
              ||       | ||| ||| || | ||   | |  ||| | || |  ||  ||    |  ||
Db       8480 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 8421

Qy        885 ACACTATATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAA 944
              |||||||||||||||  | ||   | || |   | | ||       ||  || || |
Db       8420 TATATATATATATATATAGANAGAGATATATATATAGATATGAATATATAAAAATGTAGA 8361

Qy        945 TGTTATAGAAGAATGGCTAAGATGCCAAAGCTCAGTGTATGTGGCAGAGGCATGGTGTA 1003
              | | | || |  || | | | |  || ||   |   |      || || | || | |
Db       8360 GATAAAAAAACAGAGGTTTAAAGGGGAAGGCATGGATCAATGACAAGGGGAAAGGGGGA 8302


RESULT 81
CR759820/c
LOCUS       CR759820              162435 bp    DNA     linear   HTG 10-SEP-2004
DEFINITION  Danio rerio clone DKEY-208C7, *** SEQUENCING IN PROGRESS ***, 7
            unordered pieces.
ACCESSION   CR759820
VERSION     CR759820.2  GI:51987786
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 162435)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 10, 2004 this sequence version replaced gi:51890335.
            ---------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            ---------- Project Information
            Center project name: zK208C7
            ---------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 159742 bases at least Q40
            Consensus quality: 160382 bases at least Q30
            Consensus quality: 160932 bases at least Q20
            Insert size: 161835; sum-of-contigs
            Insert size: 166033; 3.0% error; agarose-fp
            Quality coverage: 10.96x in Q20 bases; sum-of-contigs Quality
            coverage: 10.78x in Q20 bases; agarose-fp
            ----------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     9773: contig of 9773 bp in length
            *    9774     9873: gap of 100 bp
            *    9874    59608: contig of 49735 bp in length
            *   59609    59708: gap of 100 bp
            *   59709    64998: contig of 5290 bp in length
            *   64999    65098: gap of 100 bp
            *   65099    70537: contig of 5439 bp in length
            *   70538    70637: gap of 100 bp
            *   70638    90121: contig of 19484 bp in length
            *   90122    90221: gap of 100 bp
            *   90222    92310: contig of 2089 bp in length
            *   92311    92410: gap of 100 bp
            *   92411   162435: contig of 70025 bp in length.
```

```
FEATURES             Location/Qualifiers
     source          1. .162435
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-208C7"
                     /clone_lib="DanioKey"
     misc_feature    1. .9773
                     /note="assembly_fragment:00141
                     fragment_chain:1"
     misc_feature    9874. .59608
                     /note="assembly_fragment:00514
                     fragment_chain:1"
     misc_feature    59709. .64998
                     /note="assembly_fragment:00049
                     fragment_chain:1"
     misc_feature    65099. .70537
                     /note="assembly_fragment:00086
                     fragment_chain:2"
     misc_feature    70638. .90121
                     /note="assembly_fragment:00272
                     fragment_chain:2"
     misc_feature    90222. .92310
                     /note="assembly_fragment:00018.0"
     misc_feature    92411. .162435
                     /note="assembly_fragment:01362"
ORIGIN

  Query Match             7.6%;  Score 349;  DB 2;  Length 162435;
  Best Local Similarity   68.7%;  Pred. No. 2e-59;
  Matches  509;  Conservative    0;  Mismatches  230;  Indels    2;  Gaps    2;

Qy        209 AGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACATA 268
              |||||||||||    | | ||||||||||||||||||||||| | | | | | |||
Db        850 AAAAAAAAAAAAAAAAAATATATATATATATATATATATATATATATATATATATATA 791

Qy        269 TATATATACACATATATATACACATATATACATATATACACATATATACACATATATATA 328
              |||||||| | ||||||||| | ||||||| ||||||| | ||||||| | |||||||||
Db        790 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 731

Qy        329 CACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACA 388
              | ||| |||| || | ||||||||| | |||| |||| | ||||||| | ||||||| |
Db        730 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 671

Qy        389 CACATATATACACATATATACACACATATATACACATATATACACATATATACACATATA 448
              | ||||||| | ||||||| | | ||||||| | ||||||| | ||||||| | |||||
Db        670 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 611

Qy        449 CACATATATACACATATATACATATATACACATATATATA-ATATACACACATATATATA 507
              | ||||||| | ||||||| ||||||| | ||||||||| ||||| | | |||||||||
Db        610 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 551

Qy        508 CACATATATACACACATATATACACATATATACACATATATATACACATATATACACATA 567
              | ||||||| | | ||||||| | ||||||| | ||||||||| | ||||||| | |||
Db        550 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 491

Qy        568 TATACATATATACACATATATATACATATATACACATATATACATATATACACATATATA 627
              |||| ||||||| | ||||||||| ||||||| | ||||||| ||||||| | |||||||
Db        490 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 431

Qy        628 CATATATACACACATATATACACATACATATACACACACATAGATATACATATATATACA 687
              ||||||| | | ||||||| | ||| ||||| | | | ||| ||||| |||||||||
Db        430 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 371

Qy        688 CATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGT 747
              |||||||||  |||||||||  ||||||||||          |  |||          |
Db        370 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 311

Qy        748 TGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACA 807
              |   ||    |   ||| |  | | | |  | | |    | |||   | |   | |
Db        310 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 251

Qy        808 TTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA 867
              | |      |      | |||      | ||| ||| || | ||   | |  ||| | ||
Db        250 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 191

Qy        868 -AGGTAGTTAAGAAGGTAACACTATATATATATAGTGAAAGCAGTGTAAACCTAAAGG 926
               |  ||  ||   |  ||    |||||||||||||   | |    | || |  || |
Db        190 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 131

Qy        927 ATGGGCCAAGGATTTAAATGT 947
                |  |   ||||| || |
Db        130 TATAGAGATATATTTATATAT 110


RESULT 82
CR788258
LOCUS       CR788258              165992 bp    DNA     linear   HTG 25-SEP-2004
DEFINITION  Danio rerio clone DKEY-182D21, *** SEQUENCING IN PROGRESS ***, 10
            unordered pieces.
ACCESSION   CR788258
VERSION     CR788258.3  GI:52694485
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 165992)
  AUTHORS   McLay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 25, 2004 this sequence version replaced gi:52353217.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK182D21
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 162106 bases at least Q40
            Consensus quality: 163025 bases at least Q30
            Consensus quality: 163727 bases at least Q20
            Insert size: 165092; sum-of-contigs
            Insert size: 165748; 0.8% error; agarose-fp
            Quality coverage: 7.00x in Q20 bases; sum-of-contigs Quality
            coverage: 7.22x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    12528: contig of 12528 bp in length
            *    12529    12628: gap of 100 bp
            *    12629    31404: contig of 18776 bp in length
            *    31405    31504: gap of 100 bp
            *    31505    37666: contig of 6162 bp in length
            *    37667    37766: gap of 100 bp
            *    37767    62052: contig of 24286 bp in length
            *    62053    62152: gap of 100 bp
            *    62153    79002: contig of 16850 bp in length
            *    79003    79102: gap of 100 bp
            *    79103   113175: contig of 34073 bp in length
            *   113176   113275: gap of 100 bp
            *   113276   122334: contig of 9059 bp in length
            *   122335   122434: gap of 100 bp
            *   122435   137396: contig of 14962 bp in length
```

```
           *   137397   137496: gap of 100 bp
           *   137497   147411: contig of 9915 bp in length
           *   147412   147511: gap of 100 bp
           *   147512   165992: contig of 18481 bp in length.
FEATURES             Location/Qualifiers
     source          1. .165992
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-182D21"
                     /clone_lib="DanioKey"
     misc_feature    1. .12528
                     /note="assembly_fragment:00331
                     fragment_chain:1"
     misc_feature    12629. .31404
                     /note="assembly_fragment:00639
                     fragment_chain:1"
     misc_feature    31505. .37666
                     /note="assembly_fragment:00044
                     fragment_chain:1"
     misc_feature    37767. .62052
                     /note="assembly_fragment:01292
                     fragment_chain:2"
     misc_feature    62153. .79002
                     /note="assembly_fragment:00843
                     fragment_chain:2"
     misc_feature    79103. .113175
                     /note="assembly_fragment:01612
                     fragment_chain:2"
     misc_feature    113276. .122334
                     /note="assembly_fragment:00114
                     fragment_chain:3"
     misc_feature    122435. .137396
                     /note="assembly_fragment:00465
                     fragment_chain:3"
     misc_feature    137497. .147411
                     /note="assembly_fragment:00219"
     misc_feature    147512. .165992
                     /note="assembly_fragment:01055.0"
ORIGIN

  Query Match             7.6%;  Score 349;  DB 2;  Length 165992;
  Best Local Similarity   63.5%;  Pred. No. 2e-59;
  Matches  455;  Conservative    0;  Mismatches  261;  Indels    1;  Gaps    1;

Qy        209 AGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACATA 268
              |||||||||||      | ||||||||||||||||||||||| | | | | | |||
Db      11982 AAAAAAAAAAAAAAAATATATATATATATATATATATATATATATATATATATATA 12041

Qy        269 TATATATACACATATATATACACATATATACATATATACACATATATACACATATATATA 328
              |||||||| | ||||||||| | ||||||| ||||||| | ||||||| | |||||||||
Db      12042 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 12101

Qy        329 CACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACA 388
              | ||| |||| || | ||||||||| | |||| |||| | ||||||| | ||||||| |
Db      12102 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 12161

Qy        389 CACATATATACACATATATACACACATATATACACATATATACACATATATACACATATA 448
              | ||||||| | ||||||| | | ||||||| | ||||||| | ||||||| | |||||
Db      12162 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 12221

Qy        449 CACATATATACACATATATACATATATACACATATATATA-ATATACACACATATATATA 507
              | ||||||| | ||||||| ||||||| | ||||||||| ||||| | | |||||||||
Db      12222 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 12281

Qy        508 CACATATATACACACATATATACACATATATACACATATATATACACATATATACACATA 567
              | ||||||| | | ||||||| | ||||||| | ||||||||| | ||||||| | |||
Db      12282 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 12341

Qy        568 TATACATATATACACATATATATACATATATACACATATATACATATATACACATATATA 627
              |||| ||||||| | ||||||||| ||||||| | ||||||| ||||||| | |||||||
Db      12342 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 12401

Qy        628 CATATATACACACATATATACACATACATATACACACACATAGATATACATATATATACA 687
              ||||||| | | ||||||| | ||| ||||| | | | ||| ||||| |||||||||
Db      12402 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 12461

Qy        688 CATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGT 747
              |||||||||  |||||||| ||||||||||          | |||         |
Db      12462 TATATATATATATATATATATATATATATATATATATATATATATATATATATATA 12521

Qy        748 TGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACA 807
              |   ||
Db      12522 TATATATNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 12581

Qy        808 TTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA 867
                                                        |   |     ||
Db      12582 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTATATTTATATTA 12641

Qy        868 AGGTAGTTAAGAAGGTAACACTATATATATATAGTGAAAGCAGTGTAAACCTAAA 924
              |  ||  ||      |||   |||||||||||||   | |    | || |  || |
Db      12642 ATATAAATATTTTTTTAATTATATATATATATATATATATATATATATATATA 12698


RESULT 83
CR769782
LOCUS       CR769782              248767 bp    DNA     linear   HTG 27-SEP-2004
DEFINITION  Danio rerio clone DKEY-41C6, *** SEQUENCING IN PROGRESS ***, 16
            unordered pieces.
ACCESSION   CR769782
VERSION     CR769782.4  GI:52748693
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 248767)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 27, 2004 this sequence version replaced gi:52694483.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK41C6
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 243682 bases at least Q40
            Consensus quality: 244936 bases at least Q30
            Consensus quality: 245625 bases at least Q20
            Insert size: 247267; sum-of-contigs
            Insert size: 236837; 3.1% error; agarose-fp
            Quality coverage: 8.86x in Q20 bases; sum-of-contigs Quality
            coverage: 9.25x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 16 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    10635: contig of 10635 bp in length
            *    10636    10735: gap of 100 bp
            *    10736    54062: contig of 43327 bp in length
            *    54063    54162: gap of 100 bp
            *    54163    65940: contig of 11778 bp in length
            *    65941    66040: gap of 100 bp
```

```
            *    66041   115833: contig of 49793 bp in length
            *   115834   115933: gap of 100 bp
            *   115934   161746: contig of 45813 bp in length
            *   161747   161846: gap of 100 bp
            *   161847   167273: contig of 5427 bp in length
            *   167274   167373: gap of 100 bp
            *   167374   174818: contig of 7445 bp in length
            *   174819   174918: gap of 100 bp
            *   174919   184757: contig of 9839 bp in length
            *   184758   184857: gap of 100 bp
            *   184858   186977: contig of 2120 bp in length
            *   186978   187077: gap of 100 bp
            *   187078   202458: contig of 15381 bp in length
            *   202459   202558: gap of 100 bp
            *   202559   212782: contig of 10224 bp in length
            *   212783   212882: gap of 100 bp
            *   212883   215321: contig of 2439 bp in length
            *   215322   215421: gap of 100 bp
            *   215422   220884: contig of 5463 bp in length
            *   220885   220984: gap of 100 bp
            *   220985   237401: contig of 16417 bp in length
            *   237402   237501: gap of 100 bp
            *   237502   241277: contig of 3776 bp in length
            *   241278   241377: gap of 100 bp
            *   241378   248767: contig of 7390 bp in length.
FEATURES             Location/Qualifiers
     source          1. .248767
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-41C6"
                     /clone_lib="DanioKey"
     misc_feature    1. .10635
                     /note="assembly_fragment:00860
                     fragment_chain:1"
     misc_feature    10736. .54062
                     /note="assembly_fragment:02081
                     fragment_chain:1"
     misc_feature    54163. .65940
                     /note="assembly_fragment:00547
                     fragment_chain:1"
     misc_feature    66041. .115833
                     /note="assembly_fragment:02702
                     fragment_chain:1"
     misc_feature    115934. .161746
                     /note="assembly_fragment:01463
                     fragment_chain:1"
     misc_feature    161847. .167273
                     /note="assembly_fragment:00149
                     fragment_chain:1"
     misc_feature    167374. .174818
                     /note="assembly_fragment:00212
                     fragment_chain:1"
     misc_feature    174919. .184757
                     /note="assembly_fragment:00417
                     fragment_chain:2"
     misc_feature    184858. .186977
                     /note="assembly_fragment:00013
                     fragment_chain:2"
     misc_feature    187078. .202458
                     /note="assembly_fragment:01020
                     fragment_chain:2"
     misc_feature    202559. .212782
                     /note="assembly_fragment:00703
                     fragment_chain:2"
     misc_feature    212883. .215321
                     /note="assembly_fragment:00030
                     fragment_chain:2"
     misc_feature    215422. .220884
                     /note="assembly_fragment:00089
                     fragment_chain:3"
     misc_feature    220985. .237401
                     /note="assembly_fragment:01224
                     fragment_chain:3"
     misc_feature    237502. .241277
                     /note="assembly_fragment:00048.0"
     misc_feature    241378. .248767
                     /note="assembly_fragment:00309"
ORIGIN

  Query Match             7.6%;  Score 349;  DB 2;  Length 248767;
  Best Local Similarity   68.3%;  Pred. No. 1.8e-59;
  Matches  498;  Conservative    0;  Mismatches  230;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATACACACACACACATATATATATACACATATATA 286
              |||||||||||||||||||||||||| | | | | | |||||||||| | |||||||
Db      53249 TATATATATATATATATATATATATATATATATATATATATATATATATATATATA 53308

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db      53309 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 53368

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              ||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db      53369 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 53428

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db      53429 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 53488

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db      53489 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 53548

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db      53549 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 53608

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db      53609 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 53668

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db      53669 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 53728

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||||          | |||          |  |   ||    |   |||
Db      53729 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 53788

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
              |  | | | | |  | | |    | |||   | |   | || |     |     | | |
Db      53789 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 53848

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885
              ||       | ||| ||| ||  ||   | |  ||| | ||  ||  ||  |  ||
Db      53849 TATATATATATATATATATATACATATATATATATATATATAATAAATATATATATAT 53908

Qy        886 CACTATATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAAT 945
              ||||||||||||||   | |  | | |||   ||||  ||      |   | |||
Db      53909 ATATATATATATATATATATATATAAATATAATAATAAATAATATAAATATTTATATAAT 53968

Qy        946 GTTATAGAA 954
              |||  ||
Db      53969 ATTAATTAA 53977


RESULT 84
CR812467
LOCUS       CR812467               326441 bp    DNA     linear   HTG 27-SEP-2004
DEFINITION  Danio rerio clone CH211-134M19, *** SEQUENCING IN PROGRESS ***, 23
            unordered pieces.
ACCESSION   CR812467
VERSION     CR812467.2  GI:52748666
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
```

```
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 326441)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 27, 2004 this sequence version replaced gi:52694423.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zC134M19
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 317403 bases at least Q40
            Consensus quality: 319890 bases at least Q30
            Consensus quality: 321680 bases at least Q20
            Insert size: 324241; sum-of-contigs
            Insert size: 167531; 3.4% error; agarose-fp
            Quality coverage: 6.43x in Q20 bases; sum-of-contigs Quality
            coverage: 12.99x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 23 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     9379: contig of 9379 bp in length
            *    9380     9479: gap of 100 bp
            *    9480    17719: contig of 8240 bp in length
            *   17720    17819: gap of 100 bp
            *   17820    22219: contig of 4400 bp in length
            *   22220    22319: gap of 100 bp
            *   22320    33493: contig of 11174 bp in length
            *   33494    33593: gap of 100 bp
            *   33594    57232: contig of 23639 bp in length
            *   57233    57332: gap of 100 bp
            *   57333    63423: contig of 6091 bp in length
            *   63424    63523: gap of 100 bp
            *   63524    67014: contig of 3491 bp in length
            *   67015    67114: gap of 100 bp
            *   67115   106755: contig of 39641 bp in length
            *  106756   106855: gap of 100 bp
            *  106856   121891: contig of 15036 bp in length
            *  121892   121991: gap of 100 bp
            *  121992   135062: contig of 13071 bp in length
            *  135063   135162: gap of 100 bp
            *  135163   153032: contig of 17870 bp in length
            *  153033   153132: gap of 100 bp
            *  153133   190703: contig of 37571 bp in length
            *  190704   190803: gap of 100 bp
            *  190804   193044: contig of 2241 bp in length
            *  193045   193144: gap of 100 bp
            *  193145   198079: contig of 4935 bp in length
            *  198080   198179: gap of 100 bp
            *  198180   207026: contig of 8847 bp in length
            *  207027   207126: gap of 100 bp
            *  207127   214185: contig of 7059 bp in length
            *  214186   214285: gap of 100 bp
            *  214286   235774: contig of 21489 bp in length
            *  235775   235874: gap of 100 bp
            *  235875   268606: contig of 32732 bp in length
            *  268607   268706: gap of 100 bp
            *  268707   270870: contig of 2164 bp in length
            *  270871   270970: gap of 100 bp
            *  270971   278926: contig of 7956 bp in length
            *  278927   279026: gap of 100 bp
            *  279027   294890: contig of 15864 bp in length
            *  294891   294990: gap of 100 bp
            *  294991   302499: contig of 7509 bp in length
            *  302500   302599: gap of 100 bp
            *  302600   326441: contig of 23842 bp in length.
FEATURES             Location/Qualifiers
     source          1. .326441
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-134M19"
                     /clone_lib="CHORI-211"
     misc_feature    1. .9379
                     /note="assembly_fragment:00388
                     fragment_chain:1"
     misc_feature    9480. .17719
                     /note="assembly_fragment:00332
                     fragment_chain:1"
     misc_feature    17820. .22219
                     /note="assembly_fragment:00211
                     fragment_chain:1"
     misc_feature    22320. .33493
                     /note="assembly_fragment:00785
                     fragment_chain:1"
     misc_feature    33594. .57232
                     /note="assembly_fragment:01360
                     fragment_chain:1"
     misc_feature    57333. .63423
                     /note="assembly_fragment:00160
                     fragment_chain:1"
     misc_feature    63524. .67014
                     /note="assembly_fragment:00138
                     fragment_chain:1"
     misc_feature    67115. .106755
                     /note="assembly_fragment:02559
                     fragment_chain:2"
     misc_feature    106856. .121891
                     /note="assembly_fragment:00914
                     fragment_chain:2"
     misc_feature    121992. .135062
                     /note="assembly_fragment:00663
                     fragment_chain:2"
     misc_feature    135163. .153032
                     /note="assembly_fragment:00542
                     fragment_chain:2"
     misc_feature    153133. .190703
                     /note="assembly_fragment:01930
                     fragment_chain:3"
     misc_feature    190804. .193044
                     /note="assembly_fragment:00089
                     fragment_chain:3"
     misc_feature    193145. .198079
                     /note="assembly_fragment:00185
                     fragment_chain:3"
     misc_feature    198180. .207026
                     /note="assembly_fragment:00288
                     fragment_chain:4"
     misc_feature    207127. .214185
                     /note="assembly_fragment:00248
                     fragment_chain:4"
     misc_feature    214286. .235774
                     /note="assembly_fragment:01110
                     fragment_chain:5"
     misc_feature    235875. .268606
                     /note="assembly_fragment:02953
                     fragment_chain:5"
     misc_feature    268707. .270870
                     /note="assembly_fragment:00100"
     misc_feature    270971. .278926
```

```
                     /note="assembly_fragment:00463"
     misc_feature    279027. .294890
                     /note="assembly_fragment:01641.0"
     misc_feature    294991. .302499
                     /note="assembly_fragment:01641.1"
     misc_feature    302600. .326441
                     /note="assembly_fragment:02243"
ORIGIN

  Query Match             7.6%;  Score 349;  DB 2;  Length 326441;
  Best Local Similarity   64.1%;  Pred. No. 1.8e-59;
  Matches  467;  Conservative    0;  Mismatches  260;  Indels    2;  Gaps    2;

Qy        198 AATGTCATCTCAGAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACAC 257
              ||| | || |    | | | | |      | ||||||||||||||||||||||| | |
Db     235247 AATTTTATATATATATATATATATATATATATATATATATATATATATATATATATAT 235306

Qy        258 ACACACACATATATATATACACATATATATACACATATATACATATATACACATATATAC 317
              | | | | |||||||||| | ||||||||| | ||||||| ||||||| | |||||||
Db     235307 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 235366

Qy        318 ACATATATATACACATACATATGTACACATATATATACACATATGTATACACATATATAC 377
              | ||||||||| | ||| |||| || | ||||||||| | |||| |||| | |||||||
Db     235367 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 235426

Qy        378 ACATATATACACACATATATACACATATATACACACATATATACACATATATACACATAT 437
              | ||||||| | | ||||||| | ||||||| | | ||||||| | ||||||| | ||||
Db     235427 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 235486

Qy        438 ATACACATATACACATATATACACATATATACATATATACACATATATATA-ATATACAC 496
              ||| | ||||| | ||||||| | ||||||| ||||||| | ||||||||| ||||| |
Db     235487 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 235546

Qy        497 ACATATATATACACATATATACACACATATATACACATATATACACATATATATACACAT 556
              | ||||||||| | ||||||| | | ||||||| | ||||||| | ||||||||| | ||
Db     235547 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 235606

Qy        557 ATATACACATATATACATATATACACATATATATACATATATACACATATATACATATAT 616
              ||||| | ||||||| ||||||| | ||||||||| ||||||| | ||||||| ||||||
Db     235607 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 235666

Qy        617 ACACATATATACATATATACACACATATATACACATACATATACACACACATAGATATAC 676
              | | ||||||| ||||||| | | ||||||| | ||| ||||| | | | |||  |||| |
Db     235667 ATATATATATATATATATATATATATATATATATATATATATATATATATATATCTATAT 235726

Qy        677 ATATATATACACATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAA 736
              ||||||||| | ||||||||| |||||||||| ||||||||||
Db     235727 ATATATATATATATATATATATGTATATATATATATATATATATATNNNNNNNNNNNNN 235786

Qy        737 GAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTG 796

Db     235787 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 235846

Qy        797 TGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAA 856
                                          |||       | ||| ||| || | || | |
Db     235847 NNNNNNNNNNNNNNNNNNNNNNNNNNNNATATATATATATATATATATATATTTAT 235906

Qy        857 ACCTATCTGTA-AGGTAGTTAAGAAGGTAACACTATATATATATATAGTGAAAGCAGTGT 915
              |  ||| | || |  ||  ||   |  ||    |||||||||||||   | |    | |
Db     235907 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 235966

Qy        916 AAACCTAAA 924
              | |  || |
Db     235967 ATATATATA 235975


RESULT 85
CR627490
LOCUS       CR627490               216831 bp    DNA     linear   HTG 09-AUG-2004
DEFINITION  Danio rerio clone DKEY-52H23, WORKING DRAFT SEQUENCE, 15 unordered
            pieces.
ACCESSION   CR627490
VERSION     CR627490.3  GI:51090209
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 216831)
  AUTHORS   Burton,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-AUG-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 9, 2004 this sequence version replaced gi:51036472.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK52H23
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 210790 bases at least Q40
            Consensus quality: 211954 bases at least Q30
            Consensus quality: 213128 bases at least Q20
            Insert size: 215431; sum-of-contigs
            Insert size: 201083; 6.6% error; agarose-fp
            Quality coverage: 9.53x in Q20 bases; sum-of-contigs Quality
            coverage: 10.53x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 15 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    10834: contig of 10834 bp in length
            *    10835    10934: gap of 100 bp
            *    10935    15833: contig of 4899 bp in length
            *    15834    15933: gap of 100 bp
            *    15934    43416: contig of 27483 bp in length
            *    43417    43516: gap of 100 bp
            *    43517    50746: contig of 7230 bp in length
            *    50747    50846: gap of 100 bp
            *    50847    56683: contig of 5837 bp in length
            *    56684    56783: gap of 100 bp
            *    56784    61053: contig of 4270 bp in length
            *    61054    61153: gap of 100 bp
            *    61154    65291: contig of 4138 bp in length
            *    65292    65391: gap of 100 bp
            *    65392    93832: contig of 28441 bp in length
            *    93833    93932: gap of 100 bp
            *    93933   118737: contig of 24805 bp in length
            *   118738   118837: gap of 100 bp
            *   118838   122857: contig of 4020 bp in length
            *   122858   122957: gap of 100 bp
            *   122958   175495: contig of 52538 bp in length
            *   175496   175595: gap of 100 bp
            *   175596   203060: contig of 27465 bp in length
            *   203061   203160: gap of 100 bp
            *   203161   206216: contig of 3056 bp in length
            *   206217   206316: gap of 100 bp
            *   206317   208899: contig of 2583 bp in length
            *   208900   208999: gap of 100 bp
            *   209000   216831: contig of 7832 bp in length.
FEATURES             Location/Qualifiers
     source          1. .216831
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
```

```
                     /clone="DKEY-52H23"
                     /clone_lib="DanioKey"
     misc_feature    1. .10834
                     /note="assembly_fragment:00683
                     fragment_chain:1"
     misc_feature    10935. .15833
                     /note="assembly_fragment:00240
                     fragment_chain:1"
     misc_feature    15934. .43416
                     /note="assembly_fragment:01139
                     fragment_chain:1"
     misc_feature    43517. .50746
                     /note="assembly_fragment:00404
                     fragment_chain:1"
     misc_feature    50847. .56683
                     /note="assembly_fragment:00315
                     fragment_chain:1"
     misc_feature    56784. .61053
                     /note="assembly_fragment:00165
                     fragment_chain:1"
     misc_feature    61154. .65291
                     /note="assembly_fragment:00123
                     fragment_chain:1"
     misc_feature    65392. .93832
                     /note="assembly_fragment:02008
                     fragment_chain:1"
     misc_feature    93933. .118737
                     /note="assembly_fragment:00887
                     fragment_chain:1"
     misc_feature    118838. .122857
                     /note="assembly_fragment:00091
                     fragment_chain:1"
     misc_feature    122958. .175495
                     /note="assembly_fragment:02459
                     fragment_chain:2"
     misc_feature    175596. .203060
                     /note="assembly_fragment:01567
                     fragment_chain:2"
     misc_feature    203161. .206216
                     /note="assembly_fragment:00050"
     misc_feature    206317. .208899
                     /note="assembly_fragment:00069"
     misc_feature    209000. .216831
                     /note="assembly_fragment:00508.0"
ORIGIN

  Query Match             7.5%;  Score 348.8;  DB 2;  Length 216831;
  Best Local Similarity   64.1%;  Pred. No. 2.1e-59;
  Matches  448;  Conservative    0;  Mismatches  250;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db      93324 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 93383

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db      93384 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 93443

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db      93444 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 93503

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db      93504 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 93563

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| |||| | | ||||||||| | ||||||| | | |||
Db      93564 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 93623

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db      93624 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 93683

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db      93684 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 93743

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||| ||
Db      93744 TATATATATATATATATATATATATATATATATATATATATATANATATATATATATTTA 93803

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||          |  ||
Db      93804 TATATATATATATATATATATATATATATNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 93863

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825

Db      93864 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 93923

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885
                  | ||| | |||| | |    |  |||    |  ||  ||     ||
Db      93924 NNNNNNNNNATATATAAAAATATATTTATTTATATATAATATATATATATATTTTTAT 93983

Qy        886 CACTATATATATATATAGTGAAAGCAGTGTAAACCTAAA 924
                |||||||||||||| | | |    | || |  || |
Db      93984 ATATATATATATATATTTATATATATATANANATATA 94022


RESULT 86
CR450804/c
LOCUS       CR450804              220888 bp    DNA     linear   HTG 08-OCT-2004
DEFINITION  Danio rerio clone CH211-275D22, *** SEQUENCING IN PROGRESS ***, 28
            unordered pieces.
ACCESSION   CR450804
VERSION     CR450804.4  GI:54019855
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 220888)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 9, 2004 this sequence version replaced gi:53828284.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zC275D22
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 206726 bases at least Q40
            Consensus quality: 211339 bases at least Q30
            Consensus quality: 214628 bases at least Q20
            Insert size: 218188; sum-of-contigs
            Insert size: 163802; 4.5% error; agarose-fp
            Quality coverage: 7.02x in Q20 bases; sum-of-contigs Quality
            coverage: 11.27x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 28 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    25610: contig of 25610 bp in length
```

```
            *    25611    25710: gap of 100 bp
            *    25711    32841: contig of 7131 bp in length
            *    32842    32941: gap of 100 bp
            *    32942    47962: contig of 15021 bp in length
            *    47963    48062: gap of 100 bp
            *    48063    99124: contig of 51062 bp in length
            *    99125    99224: gap of 100 bp
            *    99225   126254: contig of 27030 bp in length
            *   126255   126354: gap of 100 bp
            *   126355   132148: contig of 5794 bp in length
            *   132149   132248: gap of 100 bp
            *   132249   135973: contig of 3725 bp in length
            *   135974   136073: gap of 100 bp
            *   136074   160340: contig of 24267 bp in length
            *   160341   160440: gap of 100 bp
            *   160441   163288: contig of 2848 bp in length
            *   163289   163388: gap of 100 bp
            *   163389   165705: contig of 2317 bp in length
            *   165706   165805: gap of 100 bp
            *   165806   168126: contig of 2321 bp in length
            *   168127   168226: gap of 100 bp
            *   168227   173446: contig of 5220 bp in length
            *   173447   173546: gap of 100 bp
            *   173547   175654: contig of 2108 bp in length
            *   175655   175754: gap of 100 bp
            *   175755   178489: contig of 2735 bp in length
            *   178490   178589: gap of 100 bp
            *   178590   181085: contig of 2496 bp in length
            *   181086   181185: gap of 100 bp
            *   181186   185198: contig of 4013 bp in length
            *   185199   185298: gap of 100 bp
            *   185299   188726: contig of 3428 bp in length
            *   188727   188826: gap of 100 bp
            *   188827   191247: contig of 2421 bp in length
            *   191248   191347: gap of 100 bp
            *   191348   194660: contig of 3313 bp in length
            *   194661   194760: gap of 100 bp
            *   194761   197060: contig of 2300 bp in length
            *   197061   197160: gap of 100 bp
            *   197161   199305: contig of 2145 bp in length
            *   199306   199405: gap of 100 bp
            *   199406   202170: contig of 2765 bp in length
            *   202171   202270: gap of 100 bp
            *   202271   204970: contig of 2700 bp in length
            *   204971   205070: gap of 100 bp
            *   205071   207333: contig of 2263 bp in length
            *   207334   207433: gap of 100 bp
            *   207434   210161: contig of 2728 bp in length
            *   210162   210261: gap of 100 bp
            *   210262   213219: contig of 2958 bp in length
            *   213220   213319: gap of 100 bp
            *   213320   217663: contig of 4344 bp in length
            *   217664   217763: gap of 100 bp
            *   217764   220888: contig of 3125 bp in length.
FEATURES             Location/Qualifiers
     source          1. .220888
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-275D22"
                     /clone_lib="CHORI-211"
     misc_feature    1. .25610
                     /note="assembly_fragment:01114
                     fragment_chain:1
                     clone_end:T7
                     vector_side:left"
     misc_feature    25711. .32841
                     /note="assembly_fragment:00426
                     fragment_chain:1"
     misc_feature    32942. .47962
                     /note="assembly_fragment:00502
                     fragment_chain:1"
     misc_feature    48063. .99124
                     /note="assembly_fragment:02105
                     fragment_chain:1"
     misc_feature    99225. .126254
                     /note="assembly_fragment:01573
                     fragment_chain:1"
     misc_feature    126355. .132148
                     /note="assembly_fragment:00373
                     fragment_chain:1"
     misc_feature    132249. .135973
                     /note="assembly_fragment:00352
                     fragment_chain:1"
     misc_feature    136074. .160340
                     /note="assembly_fragment:00794
                     fragment_chain:1
                     clone_end:SP6
                     vector_side:right"
     misc_feature    160441. .163288
                     /note="assembly_fragment:00075
                     fragment_chain:2"
     misc_feature    163389. .165705
                     /note="assembly_fragment:00116
                     fragment_chain:2"
     misc_feature    165806. .168126
                     /note="assembly_fragment:00155
                     fragment_chain:3"
     misc_feature    168227. .173446
                     /note="assembly_fragment:00298
                     fragment_chain:3"
     misc_feature    173547. .175654
                     /note="assembly_fragment:00176
                     fragment_chain:4"
     misc_feature    175755. .178489
                     /note="assembly_fragment:00200
                     fragment_chain:4"
     misc_feature    178590. .181085
                     /note="assembly_fragment:00192
                     fragment_chain:5"
     misc_feature    181186. .185198
                     /note="assembly_fragment:00276
                     fragment_chain:5"
     misc_feature    185299. .188726
                     /note="assembly_fragment:00244
                     fragment_chain:6"
     misc_feature    188827. .191247
                     /note="assembly_fragment:00254
                     fragment_chain:6"
     misc_feature    191348. .194660
                     /note="assembly_fragment:00311
                     fragment_chain:7"
     misc_feature    194761. .197060
                     /note="assembly_fragment:00225
                     fragment_chain:7"
     misc_feature    197161. .199305
                     /note="assembly_fragment:00131"
     misc_feature    199406. .202170
                     /note="assembly_fragment:00143"
     misc_feature    202271. .204970
                     /note="assembly_fragment:00162"
     misc_feature    205071. .207333
                     /note="assembly_fragment:00169"
     misc_feature    207434. .210161
                     /note="assembly_fragment:00184"
     misc_feature    210262. .213219
                     /note="assembly_fragment:00265"
     misc_feature    213320. .217663
                     /note="assembly_fragment:00324"
     misc_feature    217764. .220888
                     /note="assembly_fragment:00338"
ORIGIN

  Query Match             7.5%;  Score 348.8;  DB 2;  Length 220888;
  Best Local Similarity   65.1%;  Pred. No. 2.1e-59;
  Matches  456;  Conservative    0;  Mismatches  242;  Indels    2;  Gaps    2;
```

```
Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db      26206 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 26147

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |   *
Db      26146 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 26087

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db      26086 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 26027

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db      26026 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 25967

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db      25966 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 25907

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db      25906 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 25847

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db      25846 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 25787

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db      25786 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 25727

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||
Db      25726 TATATATATATATAAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 25667

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
                                                                       |
Db      25666 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNATTA 25607

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTA 884
              || ||    | ||| ||| || | ||   | |  ||| | || |  ||  |||  |  ||
Db      25606 TAATATATTATATATATATATTTATATATATATATATATATATATATATATATAAATATATA 25547

Qy        885 ACACTATATATATATATAGTGAAAGCAGTGTAAACCTAAA 924
                  |||||||||||||   | |    | || |  || |
Db      25546 TATATATATATATATATATATATATATATATATATA 25507


RESULT 87
CR812957/c
LOCUS       CR812957              224175 bp    DNA     linear   HTG 28-SEP-2004
DEFINITION  Danio rerio clone DKEY-41L5, *** SEQUENCING IN PROGRESS ***, 13
            unordered pieces.
ACCESSION   CR812957
VERSION     CR812957.2  GI:52839447
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 224175)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 28, 2004 this sequence version replaced gi:52749045.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK41L5
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 219065 bases at least Q40
            Consensus quality: 220415 bases at least Q30
            Consensus quality: 221446 bases at least Q20
            Insert size: 222975; sum-of-contigs
            Insert size: 218290; 5.1% error; agarose-fp
            Quality coverage: 6.26x in Q20 bases; sum-of-contigs Quality
            coverage: 6.49x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 13 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2025: contig of 2025 bp in length
            *     2026     2125: gap of 100 bp
            *     2126    41693: contig of 39568 bp in length
            *    41694    41793: gap of 100 bp
            *    41794    50195: contig of 8402 bp in length
            *    50196    50295: gap of 100 bp
            *    50296    54932: contig of 4637 bp in length
            *    54933    55032: gap of 100 bp
            *    55033    75269: contig of 20237 bp in length
            *    75270    75369: gap of 100 bp
            *    75370   105478: contig of 30109 bp in length
            *   105479   105578: gap of 100 bp
            *   105579   128261: contig of 22683 bp in length
            *   128262   128361: gap of 100 bp
            *   128362   130955: contig of 2594 bp in length
            *   130956   131055: gap of 100 bp
            *   131056   177561: contig of 46506 bp in length
            *   177562   177661: gap of 100 bp
            *   177662   185583: contig of 7922 bp in length
            *   185584   185683: gap of 100 bp
            *   185684   195396: contig of 9713 bp in length
            *   195397   195496: gap of 100 bp
            *   195497   208546: contig of 13050 bp in length
            *   208547   208646: gap of 100 bp
            *   208647   224175: contig of 15529 bp in length.
FEATURES             Location/Qualifiers
     source          1. .224175
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-41L5"
                     /clone_lib="DanioKey"
     misc_feature    1. .2025
                     /note="assembly_fragment:00011
                     fragment_chain:1"
     misc_feature    2126. .41693
                     /note="assembly_fragment:01456
                     fragment_chain:1"
     misc_feature    41794. .50195
                     /note="assembly_fragment:00090
                     fragment_chain:1"
     misc_feature    50296. .54932
                     /note="assembly_fragment:00059
                     fragment_chain:1"
     misc_feature    55033. .75269
                     /note="assembly_fragment:00674
                     fragment_chain:1"
     misc_feature    75370. .105478
                     /note="assembly_fragment:01132
                     fragment_chain:2"
```

```
     misc_feature    105579. .128261
                     /note="assembly_fragment:00900
                     fragment_chain:2"
     misc_feature    128362. .130955
                     /note="assembly_fragment:00024
                     fragment_chain:2"
     misc_feature    131056. .177561
                     /note="assembly_fragment:01832
                     fragment_chain:2"
     misc_feature    177662. .185583
                     /note="assembly_fragment:00258
                     fragment_chain:2"
     misc_feature    185684. .195396
                     /note="assembly_fragment:00172
                     fragment_chain:3"
     misc_feature    195497. .208546
                     /note="assembly_fragment:00383
                     fragment_chain:3"
     misc_feature    208647. .224175
                     /note="assembly_fragment:00528"
ORIGIN

  Query Match             7.5%;  Score 348.8;  DB 2;  Length 224175;
  Best Local Similarity   60.6%;  Pred. No. 2e-59;
  Matches  507;  Conservative    0;  Mismatches  327;  Indels    2;  Gaps    2;

Qy        191 AGAGCGAAATGTCATCTCAGAAAAAAAAAAAGGCATTTTATATATATATATATATATATA 250
              | | |  ||   |||  | ||||||||||    |  ||||||||||||||||||||
Db       2838 ATAACACAAAACCATAAAAAAAAAAAAAAAAAAAAAAAAAATATATATATATATATATATA 2779

Qy        251 TACACACACACACACATATATATATACACATATATATACACATATATACATATATACACA 310
              || | | | | | | |||||||||||| | ||||||||| | ||||||| ||||||| | |
Db       2778 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 2719

Qy        311 TATATACACATATATATACACATACATATGTACACATATATATACACATATGTATACACA 370
              |||||| | ||||||||| | ||| |||| || | ||||||||| | |||| |||| | |
Db       2718 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 2659

Qy        371 TATATACACATATATACACACATATATACACATATATACACACATATATACACATATATA 430
              |||||| | ||||||| | | ||||||| | ||||||| | | ||||||| | |||||||
Db       2658 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 2599

Qy        431 CACATATATACACATATACACATATATACACATATATACATATATACACATATATATAAT 490
              | ||||||| | ||||| | ||||||| | ||||||| ||||||| | |||||||| ||
Db       2598 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT-AT 2540

Qy        491 ATACACACATATATATACACATATATACACACATATATACACATATATACACATATATAT 550
              ||| | | ||||||||| | ||||||| | | ||||||| | ||||||| | ||||||||
Db       2539 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 2480

Qy        551 ACACATATATACACATATATACATATATACACATATATATACATATATACACATATATAC 610
              | | ||||||| | ||||||| ||||||| | ||||||||| ||||||| | ||||||||
Db       2479 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 2420

Qy        611 ATATATACACATATATACATATATACACACATATATACACATACATATACACACACATAG 670
              ||||||| | ||||||| ||||||| | | ||||||| | ||| ||||| | | | |||
Db       2419 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 2360

Qy        671 ATATACATATATATACACATATATATACGTATATATATGTATATATATATGCTCCAGAGT 730
              ||||| ||||||||| | |||||||||  ||||||||| ||||||||||          |
Db       2359 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 2300

Qy        731 TCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGT 790
                |||          |  |  ||    |    ||| |  | | | |  | |  | |    |
Db       2299 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 2240

Qy        791 GTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGT 850
              |||   | |   | || |     |     | |||       | ||| ||| || | |
Db       2239 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 2180

Qy        851 AATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTAACACTATATATATATATAGTGAAAG 909
              |   | |  ||| | || |  ||  ||    | ||    |||||||||||||||
Db       2179 ATATATATATATATATATATATATATATATATATATATATATATATATATATATNNNNNN 2120

Qy         910 CAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTATAGAAGAATGGCTAAGATGC 969

Db        2119 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2060

Qy         970 CAAAGCTCAGTGTATGTGGCAGAGGCATGGTGTAGGGTGTGTCCAGGTTCATATAT 1025
                                                    | | |  |  |  ||||||
Db        2059 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNATATATATATATATATATAT 2004


RESULT 88
AC094041/c
LOCUS       AC094041              236551 bp    DNA     linear   HTG 10-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-18C6, WORKING DRAFT SEQUENCE, 4
            unordered pieces.
ACCESSION   AC094041
VERSION     AC094041.8  GI:30520885
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 236551)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
```

```
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 236551)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 236551)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 10, 2003 this sequence version replaced gi:24942723.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GFNH
                Center clone name: CH230-18C6
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 222835 bases at least Q40
                Consensus quality: 225635 bases at least Q30
                Consensus quality: 228117 bases at least Q20
                Estimated insert size: 236184; sum-of-contigs estimation
                Quality coverage: 9x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   230804: contig of 230804 bp in length
            *  230805   230904: gap of unknown length
            *  230905   232373: contig of 1469 bp in length
            *  232374   232473: gap of unknown length
            *  232474   233531: contig of 1058 bp in length
            *  233532   233631: gap of unknown length
            *  233632   236551: contig of 2920 bp in length.
FEATURES             Location/Qualifiers
     source          1. .236551
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-18C6"
ORIGIN

  Query Match             7.5%;  Score 348.8;  DB 2;  Length 236551;
  Best Local Similarity   69.6%;  Pred. No. 2e-59;
  Matches  487;  Conservative    0;  Mismatches  212;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db     234988 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 234929

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db     234928 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 234869

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db     234868 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 234809

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db     234808 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 234749

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db     234748 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 234689

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db     234688 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 234629

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db     234628 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 234569

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db     234568 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 234509

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||||          | |||          |  |   ||    |   |||
Db     234508 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 234449

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
              |  | | | |  |  | |    | |||   | |   | || |     |      | |
Db     234448 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 234389

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885
              ||       | ||| ||| || | ||   | |  ||| | ||      |  | | | ||
Db     234388 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATGATAT 234329

Qy        886 CACTATATATATATAGTGAAAGCAGTGTAAACCTAAAG 925
                |||||||||||| | | |||  | | || |  || ||
Db     234328 ATATATATATATATAAAATTAAAAAATTATATATATATAG 234289


RESULT 89
BX322798/c
LOCUS       BX322798              109868 bp    DNA     linear   VRT 24-JUN-2004
DEFINITION  Zebrafish DNA sequence from clone CH211-231F6 in linkage group 24,
            complete sequence.
ACCESSION   BX322798
VERSION     BX322798.6  GI:49203424
KEYWORDS    HTG.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 109868)
  AUTHORS   Heath,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JUN-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jun 24, 2004 this sequence version replaced gi:49065270.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
```

```
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep Clone-derived
            Zebrafish pUC subclones occasionally display inconsistency over the
            length of mononucleotide A/T runs and conserved TA repeats. Where
            this is found the longest good quality representation will be
            submitted.
            Repeat names beginning 'Dr' were identified by the Recon repeat
            discovery system (Zhirong Bao and Sean Eddy, submitted), and those
            beginning 'drr' were identified by Rick Waterman (Stephen Johnson
            lab, WashU).  For further information see
            http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml CH211-231F6
            is from a CHORI-211 BAC library
            VECTOR: pTARBAC2.1.
FEATURES             Location/Qualifiers
     source          1. .109868
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-231F6"
                     /clone_lib="CHORI-211"
ORIGIN

  Query Match             7.5%;  Score 348.6;  DB 5;  Length 109868;
  Best Local Similarity   76.5%;  Pred. No. 2.5e-59;
  Matches  440;  Conservative    0;  Mismatches  134;  Indels    1;  Gaps    1;

Qy        147 TGCGGGGAGCAAGATCATGCCATTGCACTCCAGCCCAGGCAACAAGAGCGAAATGTCATC 206
              || | |||   ||  ||      ||||||  ||   |   ||   | |   || | ||
Db      78151 TGTGTGGAAACAGTCCAAAAACATGCACTATAGGTGAATGAATGGGTTCTTTATTTGACT 78092

Qy        207 TCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACACA 266
              | | | ||| | |   || ||||| ||| ||||| |||||||| | | | | | |||
Db      78091 ATATATATAAATATATATATATTATACATACATATACATATATATATATATATATATACA 78032

Qy        267 TATATATATACACATATATATACACATATATACATATATACACATATATACACATATATA 326
              ||||||||||||| ||||||||| | ||||| | ||||| | | |||||||||| |||||||
Db      78031 TATATATATACATATATATATATATATACATATATACATATATATATACATATATATA 77972

Qy        327 TACACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATA 386
              | |||||||||| |||| |||||||||| |||| |||||| ||||||| | |||||||
Db      77971 CATACATACATATATACATATATATATACATATATATATACATATATATATATATATA 77912

Qy        387 CACACATATATACACATATATACACACATATATACACATATATACACATATATACACATA 446
              || ||||||| | | ||||||||| | ||||||||| | ||||| | ||||| | | |||
Db      77911 CATACATATACATATATATATACATATATATATACATACATATATATATATACATATATA 77852

Qy        447 TACACATATATACACATATATACATATATACACATATATATA-ATATACACACATATATA 505
              || ||||| ||| ||||||||| ||| ||| | ||||| ||| ||||||| | |||||||
Db      77851 TATACATACATATACATATATATATACATATATATATACATACATATACATATATATATA 77792

Qy        506 TACACATATATACACACATATATACACATATATACACATATATATACACATATATACACA 565
              | | |||||||| ||||||| | | ||||||||| ||||||||| ||||| ||| |||
Db      77791 CATATATATATACATACATATACATATATATATACATATATATATATACATACATATACA 77732

Qy        566 TATATACATATATACACATATATATACATATATACACATATATACATATATACACATATA 625
              |||||| ||| ||||| ||||| ||||||||| | | ||||||| ||| ||||| |||||
Db      77731 TATATATATACATACATATATACATACATATACATATATATATATATACATACATATATA 77672

Qy        626 TACATATATACACACATATATACACATACATATACACACACATAGATATACATATATATA 685
              | ||||||| ||| | ||||| ||| |  |||| | | | ||| ||||| |||||||||
Db      77671 CATATATATATACATACATATATACACATGTATATATATATATATATATATATATATA 77612

Qy        686 CACATATATATACGTATATATATGTATATATATAT 720
              |||||| |||||  |||||||||| ||||||||||
Db      77611 CACATACATATATATATATATATATATATATAT 77577


RESULT 90
CR392338/c
LOCUS       CR392338              132545 bp    DNA     linear   HTG 26-AUG-2004
DEFINITION  Danio rerio clone DKEY-262M9, *** SEQUENCING IN PROGRESS ***, 9
            unordered pieces.
ACCESSION   CR392338
VERSION     CR392338.3  GI:51571696
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 132545)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-AUG-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 26, 2004 this sequence version replaced gi:51490965.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK262M9
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 130047 bases at least Q40
            Consensus quality: 130566 bases at least Q30
            Consensus quality: 130976 bases at least Q20
            Insert size: 131745; sum-of-contigs
            Insert size: 144213; 1.4% error; agarose-fp
            Quality coverage: 7.43x in Q20 bases; sum-of-contigs Quality
            coverage: 6.98x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2980: contig of 2980 bp in length
            *     2981     3080: gap of 100 bp
            *     3081     7869: contig of 4789 bp in length
            *     7870     7969: gap of 100 bp
            *     7970    25024: contig of 17055 bp in length
            *    25025    25124: gap of 100 bp
            *    25125    34501: contig of 9377 bp in length
            *    34502    34601: gap of 100 bp
            *    34602    42063: contig of 7462 bp in length
            *    42064    42163: gap of 100 bp
            *    42164    45417: contig of 3254 bp in length
            *    45418    45517: gap of 100 bp
            *    45518    62732: contig of 17215 bp in length
            *    62733    62832: gap of 100 bp
            *    62833    89992: contig of 27160 bp in length
```

```
          *    89993    90092: gap of 100 bp
          *    90093   132545: contig of 42453 bp in length.
FEATURES             Location/Qualifiers
     source          1. .132545
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-262M9"
                     /clone_lib="DanioKey"
     misc_feature    1. .2980
                     /note="assembly_fragment:00025
                     fragment_chain:1"
     misc_feature    3081. .7869
                     /note="assembly_fragment:00103
                     fragment_chain:1"
     misc_feature    7970. .25024
                     /note="assembly_fragment:00458
                     fragment_chain:1"
     misc_feature    25125. .34501
                     /note="assembly_fragment:00156
                     fragment_chain:1"
     misc_feature    34602. .42063
                     /note="assembly_fragment:00280
                     fragment_chain:1"
     misc_feature    42164. .45417
                     /note="assembly_fragment:00074
                     fragment_chain:2"
     misc_feature    45518. .62732
                     /note="assembly_fragment:00711
                     fragment_chain:2"
     misc_feature    62833. .89992
                     /note="assembly_fragment:00974
                     fragment_chain:2"
     misc_feature    90093. .132545
                     /note="assembly_fragment:01385
                     fragment_chain:2"
ORIGIN

  Query Match             7.5%;  Score 348.6;  DB 2;  Length 132545;
  Best Local Similarity   62.2%;  Pred. No. 2.4e-59;
  Matches  472;  Conservative    0;  Mismatches  284;  Indels    3;  Gaps    1;

Qy        169 TTGCACTCCAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTT 228
              ||  ||  |   |  |   ||  | |    |||||  |   ||| |||| || |    | ||
Db      63457 TTTTACACACCCTGAAAAAAGTAAAAACCAATGTAAATACAGCAAAACAATATTTTAATT 63398

Qy        229 TATATATATATATATATATATATACACACACACACATATATATATACACATATATATA 288
              ||||||||||||||||||||||| | | | | | ||||||||||| | |||||||||
Db      63397 AATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 63338

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |||
Db      63337 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 63278

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              |||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||||
Db      63277 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 63218

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||||
Db      63217 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 63158

Qy        469 CATATATACACATATATATA---ATATACACACATATATATACACATATATACACACATA 525
              ||||||| | |||||||||   ||||| | | ||||||||| | ||||||| | | |||
Db      63157 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 63098

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db      63097 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 63038

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db      63037 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 62978

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| |||||||||| | |||||||||  ||||||
Db      62977 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 62918

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||||          | |||          | |   ||    |   |||
Db      62917 TATATATATATATATATATATATATATATATATATATATATATATATATATATATAAATA 62858

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
              |  || | | |  | | |    |
Db      62857 AAATAATATATATATATATATATNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 62798

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885

Db      62797 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 62738

Qy        886 CACTATATATATATATAGTGAAAGCAGTGTAAACCTAAA 924
                ||||||||||||   | |    | || | ||| |
Db      62737 NNNNNTATATATATATATATATATATATATATCTATA 62699


RESULT 91
AC004966
LOCUS       AC004966               134084 bp    DNA     linear   PRI 03-OCT-2003
DEFINITION  Homo sapiens PAC clone RP5-1107K15 from 7, complete sequence.
ACCESSION   AC004966
VERSION     AC004966.2  GI:5306307
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 134084)
  AUTHORS   Sulston,J.E. and Wilson,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
   PUBMED   9847074
REFERENCE   2  (bases 1 to 134084)
  AUTHORS   Smith-Craig,Renee., Strowmatt,C. and Wohldmann,P.
  TITLE     The sequence of Homo sapiens PAC clone RP5-1107K15
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 134084)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-1998) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 134084)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JUL-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 134084)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-OCT-1999) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   6  (bases 1 to 134084)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-DEC-1999) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   7  (bases 1 to 134084)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-APR-2003) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   8  (bases 1 to 134084)
  AUTHORS   Wilson,R.
  TITLE     Direct Submission
```

JOURNAL Submitted (03-OCT-2003) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On Jul 1, 1999 this sequence version replaced gi:3907492.
---------- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: sapiens@watson.wustl.edu
---------- Summary Statistics
Center project name: H_DJ1107K15
----------.

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see http://www.nhgri.nih.gov/DIR/GTB/CHR7 , send mailto:egreen@nhgri.nih.gov , or see http://genome.wustl.edu

SOURCE INFORMATION:
This clone was derived from human PAC library RPCI-5, prepared by Pieter de Jong and coworkers at http://www.chori.org using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor.
The clone may be obtained either from Genome Systems, Inc. (http://www.genomesystems.com) or Research Genetics, Inc. (http://www.resgen.com); or from Pieter de Jong.
VECTOR: pCYPAC2

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP5-942I16, 200 bp overlap.
Actual start of this clone is at base position 197 of RP5-1107K15
actual end is at base position 134084 of RP5-1107K15.

FEATURES Location/Qualifiers
source 1. .134084
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="7"
/map="7"
/clone="RP5-1107K15"
/clone_lib="RPCI-5"
repeat_region 84. .925
/rpt_family="L1"
repeat_region 1199. .1495
/rpt_family="Alu"
repeat_region 1692. .2013
/rpt_family="Alu"
repeat_region 2023. .2340
/rpt_family="Alu"
repeat_region 3225. .3438
/rpt_family="CR1"
repeat_region 3459. .3768
/rpt_family="Alu"
repeat_region 4040. .4346
/rpt_family="Alu"
repeat_region 4474. .4594
/rpt_family="MIR"
repeat_region 4854. .5166
/rpt_family="Alu"
repeat_region 5172. .5337
/rpt_family="Alu"
repeat_region 5338. .5387
/rpt_family="(TTA)n"
repeat_region 5388. .5507
/rpt_family="Alu"
repeat_region 5523. .5800
/rpt_family="Alu"
repeat_region 6624. .6922
/rpt_family="Alu"
repeat_region 7057. .7353
/rpt_family="Alu"
repeat_region 7360. .7642
/rpt_family="Alu"
repeat_region 8314. .8428
/rpt_family="L2"
repeat_region 8464. .8762
/rpt_family="Alu"
repeat_region 8767. .9078
/rpt_family="Alu"
repeat_region 9642. .9950
/rpt_family="Alu"
repeat_region 9959. .10274
/rpt_family="Alu"
repeat_region 10789. .11096
/rpt_family="Alu"
repeat_region 11354. .11489
/rpt_family="Alu"
repeat_region 11526. .11806
/rpt_family="Alu"
repeat_region 11810. .12113
/rpt_family="Alu"
repeat_region 12388. .12679
/rpt_family="Alu"
repeat_region 13254. .13297
/rpt_family="(TG)n"
repeat_region 13298. .13445
/rpt_family="Alu"
repeat_region 13448. .13774
/rpt_family="Alu"
repeat_region 13776. .14066
/rpt_family="Alu"
repeat_region 14091. .14478
/rpt_family="ERVL"
repeat_region 14550. .14612
/rpt_family="Alu"
repeat_region 14618. .14897
/rpt_family="Alu"
repeat_region 14936. .15029
/rpt_family="MER1_type"
repeat_region 15078. .15244
/rpt_family="L2"
repeat_region 15248. .15469
/rpt_family="MaLR"
repeat_region 15485. .15561
/rpt_family="MaLR"
repeat_region 15700. .15955
/rpt_family="Alu"
repeat_region 15997. .16096
/rpt_family="L1"
repeat_region 16097. .16124
/rpt_family="(TTTA)n"
repeat_region 16135. .16422
/rpt_family="Alu"
repeat_region 16877. .17035
/rpt_family="Alu"
repeat_region 17036. .17345
/rpt_family="Alu"
repeat_region 17530. .17671
/rpt_family="MIR"

```
     repeat_region   18281. .18563
                     /rpt_family="Alu"
     repeat_region   18564. .18609
                     /rpt_family="(CA)n"
     repeat_region   18711. .18888

  Query Match             7.5%;  Score 348.6;  DB 9;  Length 134084;
  Best Local Similarity   83.3%;  Pred. No. 2.4e-59;
  Matches  409;  Conservative    0;  Mismatches   79;  Indels    3;  Gaps    1;

Qy        230 ATATATATATATATATATATACACACACACACATATATATATACACATATATATAC 289
              ||||||||| | |||||||||  | | | | ||| | |||||  |||||  | | |||||||||
Db     108308 ATATATATACACATATATATACATATATATACATATATATACATATATATACATATATAC 108367

Qy        290 ACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATAT 349
              ||||||||||| |||||  ||| |||||  ||| |||||||  |||||  ||||   | |  ||||
Db     108368 ACATATATACACATATATACACATATATACACATATATATACATATATATACATATATAT 108427

Qy        350 ATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATAC 409
              | |||  | |||   ||||  |||||  |||  |||||||||  |||  |||||||||  ||||||||
Db     108428 ACATATATATACATATATACATACATATACATATATATACATATATATACATATATATAC 108487

Qy        410 ACACATATATACACATATATACACATATATACACATATACACATATATACACATATATAC 469
              |||  |||||  |  |  |||||  |||  |||||  |||  |||||||||||||||||  ||||||||
Db     108488 ACATATATACATATATATACACATATATACACATATATACACATATATACATATATATAC 108547

Qy        470 ATATATACACATATATATA---ATATACACACATATATATACACATATATACACACATAT 526
              |||||||  ||||||||||    |||||||  |  |||  |  |||  |  |  |||||  |  |||  ||
Db     108548 ATATATATACATATATATACATATATACATATATACACATATATACATATATATACACAT 108607

Qy        527 ATACACATATATACACATATATATACACATATATACACATATATACATATATACACATAT 586
              |||  |||||||||  |||  |||||||||  |||||||||  |||||  ||||||
Db     108608 ATATACATATATATACACATATATACATATATATACATATATATACACATATATACATAT 108667

Qy        587 ATATACATATATACACATATATACATATATACACATATATACATATATACACACATATAT 646
              |||||||||||||||  ||||||||||||||||  ||||||||||||||||||  |||  ||||||
Db     108668 ATATACATATATACATATATATACATATATATACATATATACATATATATACATATATAT 108727

Qy        647 ACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATAT 706
              |||  |||  |  |||  |  |  |||||  ||||||||||||||||  |||||||  ||  |||||||
Db     108728 ACATATATACATATATATACATATATATACATATATATACATATATATACACATATATAT 108787

Qy        707 ATGTATATATA 717
              |   |||||||.
Db     108788 ACACATATATA 108798


RESULT 92
CR735106/c
LOCUS       CR735106              170355 bp    DNA     linear   HTG 19-AUG-2004
DEFINITION  Danio rerio clone CH211-232N15, WORKING DRAFT SEQUENCE, 13
            unordered pieces.
ACCESSION   CR735106
VERSION     CR735106.2  GI:51476923
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 170355)
  AUTHORS   Burton,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-AUG-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 20, 2004 this sequence version replaced gi:51241603.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zC232N15
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 167417 bases at least Q40
            Consensus quality: 168324 bases at least Q30
            Consensus quality: 168699 bases at least Q20
            Insert size: 169155; sum-of-contigs
            Insert size: 178984; 3.2% error; agarose-fp
            Quality coverage: 6.27x in Q20 bases; sum-of-contigs Quality
            coverage: 6.05x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 13 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    11635: contig of 11635 bp in length
            *    11636    11735: gap of 100 bp
            *    11736    22534: contig of 10799 bp in length
            *    22535    22634: gap of 100 bp
            *    22635    25737: contig of 3103 bp in length
            *    25738    25837: gap of 100 bp
            *    25838    30700: contig of 4863 bp in length
            *    30701    30800: gap of 100 bp
            *    30801    47115: contig of 16315 bp in length
            *    47116    47215: gap of 100 bp
            *    47216    67157: contig of 19942 bp in length
            *    67158    67257: gap of 100 bp
            *    67258    93368: contig of 26111 bp in length
            *    93369    93468: gap of 100 bp
            *    93469   103340: contig of 9872 bp in length
            *   103341   103440: gap of 100 bp
            *   103441   113531: contig of 10091 bp in length
            *   113532   113631: gap of 100 bp
            *   113632   125308: contig of 11677 bp in length
            *   125309   125408: gap of 100 bp
            *   125409   138161: contig of 12753 bp in length
            *   138162   138261: gap of 100 bp
            *   138262   143042: contig of 4781 bp in length
            *   143043   143142: gap of 100 bp
            *   143143   170355: contig of 27213 bp in length.
FEATURES             Location/Qualifiers
     source          1. .170355
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-232N15"
                     /clone_lib="CHORI-211"
     misc_feature    1. .11635
                     /note="assembly_fragment:00706
                     fragment_chain:1"
     misc_feature    11736. .22534
                     /note="assembly_fragment:00576
                     fragment_chain:1"
     misc_feature    22635. .25737
                     /note="assembly_fragment:00028
                     fragment_chain:1"
     misc_feature    25838. .30700
                     /note="assembly_fragment:00056
                     fragment_chain:1"
     misc_feature    30801. .47115
                     /note="assembly_fragment:00862
                     fragment_chain:1"
     misc_feature    47216. .67157
                     /note="assembly_fragment:01028
                     fragment_chain:1"
     misc_feature    67258. .93368
                     /note="assembly_fragment:01537
                     fragment_chain:1"
```

```
     misc_feature    93469. .103340
                     /note="assembly_fragment:00351
                     fragment_chain:2"
     misc_feature    103441. .113531
                     /note="assembly_fragment:00156
                     fragment_chain:2"
     misc_feature    113632. .125308
                     /note="assembly_fragment:00462
                     fragment_chain:2"
     misc_feature    125409. .138161
                     /note="assembly_fragment:00246
                     fragment_chain:2"
     misc_feature    138262. .143042
                     /note="assembly_fragment:00100.0"
     misc_feature    143143. .170355
                     /note="assembly_fragment:01234"
ORIGIN

  Query Match             7.5%;  Score 348.6;  DB 2;  Length 170355;
  Best Local Similarity   65.6%;  Pred. No. 2.3e-59;
  Matches  451;  Conservative    0;  Mismatches  234;  Indels    2;  Gaps    2;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db     114143 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 114084

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db     114083 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 114024

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db     114023 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 113964

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db     113963 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 113904

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db     113903 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 113844

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db     113843 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 113784

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db     113783 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 113724

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db     113723 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 113664

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||| |  ||      ||
Db     113663 TATATATATATATATATTGGAGTCAGATGAATNNNNNNNNNNNNNNNNNNNNNNNNNNN 113604

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825

Db     113603 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 113544

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTA 884
                      || ||| || | ||   | |  ||| | || |  ||  ||   |  ||
Db     113543 NNNNNNNNNNNTATATATATATATATATATATATATATATATATATATATATATATATA 113484

Qy        885 ACACTATATATATATATAGTGAAAGCA 911
              |||||||||||||| || | ||
Db     113483 TATATATATATATATATAGGGAGATCA 113457


RESULT 93
CR762394
```

```
LOCUS       CR762394             233491 bp    DNA     linear   HTG 13-SEP-2004
DEFINITION  Danio rerio clone DKEY-108C6, *** SEQUENCING IN PROGRESS ***, 11
            unordered pieces.
ACCESSION   CR762394
VERSION     CR762394.3  GI:52000387
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 233491)
  AUTHORS   McLay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 13, 2004 this sequence version replaced gi:51988247.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK108C6
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 229390 bases at least Q40
            Consensus quality: 230228 bases at least Q30
            Consensus quality: 230919 bases at least Q20
            Insert size: 232491; sum-of-contigs
            Insert size: 232003; 3.4% error; agarose-fp
            Quality coverage: 6.99x in Q20 bases; sum-of-contigs Quality
            coverage: 7.05x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2489: contig of 2489 bp in length
            *     2490     2589: gap of 100 bp
            *     2590    44861: contig of 42272 bp in length
            *    44862    44961: gap of 100 bp
            *    44962    83642: contig of 38681 bp in length
            *    83643    83742: gap of 100 bp
            *    83743   120587: contig of 36845 bp in length
            *   120588   120687: gap of 100 bp
            *   120688   129304: contig of 8617 bp in length
            *   129305   129404: gap of 100 bp
            *   129405   132761: contig of 3357 bp in length
            *   132762   132861: gap of 100 bp
            *   132862   147991: contig of 15130 bp in length
            *   147992   148091: gap of 100 bp
            *   148092   152438: contig of 4347 bp in length
            *   152439   152538: gap of 100 bp
            *   152539   199358: contig of 46820 bp in length
            *   199359   199458: gap of 100 bp
            *   199459   209632: contig of 10174 bp in length
            *   209633   209732: gap of 100 bp
            *   209733   233491: contig of 23759 bp in length.
FEATURES             Location/Qualifiers
     source          1. .233491
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-108C6"
                     /clone_lib="DanioKey"
     misc_feature    1. .2489
```

```
                     /note="assembly_fragment:00019
                     fragment_chain:1"
     misc_feature    2590. .44861
                     /note="assembly_fragment:01734
                     fragment_chain:1"
     misc_feature    44962. .83642
                     /note="assembly_fragment:01269
                     fragment_chain:1"
     misc_feature    83743. .120587
                     /note="assembly_fragment:00842
                     fragment_chain:1"
     misc_feature    120688. .129304
                     /note="assembly_fragment:00086
                     fragment_chain:1"
     misc_feature    129405. .132761
                     /note="assembly_fragment:00035
                     fragment_chain:1"
     misc_feature    132862. .147991
                     /note="assembly_fragment:00295
                     fragment_chain:1"
     misc_feature    148092. .152438
                     /note="assembly_fragment:00060
                     fragment_chain:2"
     misc_feature    152539. .199358
                     /note="assembly_fragment:02231
                     fragment_chain:2"
     misc_feature    199459. .209632
                     /note="assembly_fragment:00184
                     fragment_chain:2"
     misc_feature    209733. .233491
                     /note="assembly_fragment:00499.0"
ORIGIN

  Query Match             7.5%;  Score 348.6;  DB 2;  Length 233491;
  Best Local Similarity   64.2%;  Pred. No. 2.2e-59;
  Matches  449;  Conservative    0;  Mismatches  249;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db      83133 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 83192

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db      83193 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 83252

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db      83253 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 83312

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db      83313 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 83372

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || |||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db      83373 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 83432

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db      83433 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 83492

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db      83493 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 83552

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db      83553 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 83612

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||          |  |||
Db      83613 TATATATATATATATATATATATATATANNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 83672

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825

Db      83673 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 83732

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885
                      || ||| || |    || |   ||| | ||      || | |   ||
Db      83733 NNNNNNNNNNTTTATATATATATAATTTATATATTATATATATTTATATTTATATTATAT 83792

Qy        886 CACTATATATATATATAGTGAAAGCAGTGTAAACCTAAA 924
                | ||||||||||||   | |    | || |  || |
Db      83793 ATATTTATATATATATATATATATATATATATATATA 83831


RESULT 94
CR548629
LOCUS       CR548629              131550 bp    DNA     linear   HTG 06-JUL-2004
DEFINITION  Danio rerio clone DKEY-173G4, WORKING DRAFT SEQUENCE, 12 unordered
            pieces.
ACCESSION   CR548629
VERSION     CR548629.2  GI:50057711
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 131550)
  AUTHORS   Burton,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUL-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 8, 2004 this sequence version replaced gi:49658992.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK173G4
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 127208 bases at least Q40
            Consensus quality: 128066 bases at least Q30
            Consensus quality: 128642 bases at least Q20
            Insert size: 130450; sum-of-contigs
            Insert size: 145332; 1.1% error; agarose-fp
            Quality coverage: 7.95x in Q20 bases; sum-of-contigs Quality
            coverage: 7.21x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 12 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    18381: contig of 18381 bp in length
            *    18382    18481: gap of 100 bp
            *    18482    25435: contig of 6954 bp in length
            *    25436    25535: gap of 100 bp
            *    25536    29512: contig of 3977 bp in length
            *    29513    29612: gap of 100 bp
            *    29613    40983: contig of 11371 bp in length
            *    40984    41083: gap of 100 bp
            *    41084    56393: contig of 15310 bp in length
            *    56394    56493: gap of 100 bp
            *    56494    70939: contig of 14446 bp in length
            *    70940    71039: gap of 100 bp
            *    71040    80278: contig of 9239 bp in length
            *    80279    80378: gap of 100 bp
```

```
           *    80379    87641: contig of 7263 bp in length
           *    87642    87741: gap of 100 bp
           *    87742   103997: contig of 16256 bp in length
           *   103998   104097: gap of 100 bp
           *   104098   106761: contig of 2664 bp in length
           *   106762   106861: gap of 100 bp
           *   106862   110734: contig of 3873 bp in length
           *   110735   110834: gap of 100 bp
           *   110835   131550: contig of 20716 bp in length.
FEATURES             Location/Qualifiers
     source          1. .131550
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-173G4"
                     /clone_lib="DanioKey"
     misc_feature    1. .18381
                     /note="assembly_fragment:01412
                     fragment_chain:1"
     misc_feature    18482. .25435
                     /note="assembly_fragment:00193
                     fragment_chain:1"
     misc_feature    25536. .29512
                     /note="assembly_fragment:00013
                     fragment_chain:1"
     misc_feature    29613. .40983
                     /note="assembly_fragment:00264
                     fragment_chain:1"
     misc_feature    41084. .56393
                     /note="assembly_fragment:00930
                     fragment_chain:1"
     misc_feature    56494. .70939
                     /note="assembly_fragment:00480
                     fragment_chain:1"
     misc_feature    71040. .80278
                     /note="assembly_fragment:00367
                     fragment_chain:2"
     misc_feature    80379. .87641
                     /note="assembly_fragment:00063
                     fragment_chain:2"
     misc_feature    87742. .103997
                     /note="assembly_fragment:00696
                     fragment_chain:2"
     misc_feature    104098. .106761
                     /note="assembly_fragment:00032"
     misc_feature    106862. .110734
                     /note="assembly_fragment:00122.0"
     misc_feature    110835. .131550
                     /note="assembly_fragment:01164"
ORIGIN

  Query Match             7.5%;  Score 348.4;  DB 2;  Length 131550;
  Best Local Similarity   69.5%;  Pred. No. 2.7e-59;
  Matches  500;  Conservative    0;  Mismatches  217;  Indels    2;  Gaps    2;

Qy        208 CAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATATACACACACACACACAT 267
              || |||    || ||     ||||||||||||||||||||||||||||||||||||||||
Db      86675 CAAAAAGGCAATAAATACATTTATATATATATATATATATATATACACACACACACACAC 86734

Qy        268 ATATATATACACATATATATACACATATATACATATATACACATATATACACATATATAT 327
              | | | ||||| |||||||||| | ||||||| ||||||| | ||||||| | ||||||||
Db      86735 ACACACATACATATATATATATATATATATATATATATATATATATATATATATATATAT 86794

Qy        328 ACACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATAC 387
              | | ||| |||| || | ||||||||| | |||| |||| | ||||||| | |||||||
Db      86795 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 86854

Qy        388 ACACATATATACACATATATACACACATATATACACATATATACACATATATACACATAT 447
              | | ||||||| | ||||||| | | ||||||| | ||||||| | ||||||| | |||||
Db      86855 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 86914

Qy        448 ACACATATATACACATATATACATATATACACATATATATA-ATATACACACATATATAT 506
              | | ||||||| | ||||||| ||||||| | ||||||||| ||||| | | ||||||||
```

```
Db      86915 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 86974

Qy        507 ACACATATATACACACATATATACACATATATACACATATATATACACATATATACACAT 566
              | | ||||||| | | ||||||| | ||||||| | ||||||||| | ||||||| | ||
Db      86975 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 87034

Qy        567 ATATACATATATACACATATATATACATATATACACATATATACATATATACACATATAT 626
              ||||| ||||||| | ||||||||| ||||||| | ||||||| ||||||| | ||||||
Db      87035 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 87094

Qy        627 ACATATATACACACATATATACACATACATATACACACACATAGATATACATATATATAC 686
              | ||||||| | | ||||||| | ||| ||||| | | | ||| ||||| |||||||||
Db      87095 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 87154

Qy        687 ACATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAG 746
              | |||||||||  |||||||| |||||||||||          |  |||         |
Db      87155 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 87214

Qy        747 TTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGAC 806
              |   ||    |   ||| |  | | | |  | | |    | |||   | |  |
Db      87215 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 87274

Qy        807 ATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGT 866
              || |     |     | |||       | ||| ||| || | ||   | |  ||| | |
Db      87275 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 87334

Qy        867 A-AGGTAGTTAAGAAGGTAACACTATATATATATATAGTGAAAGCAGTGTAAACCTAAA 924
              | |  ||  ||   |  ||    ||||||||||||||   | |    | ||| |  ||| |
Db      87335 ATATATATATATATATATATATATATATATATATATATATATATATATATATNTATA 87393
```

```
RESULT 95
CR388149/c
LOCUS       CR388149              158085 bp    DNA     linear   HTG 19-AUG-2004
DEFINITION  Danio rerio clone DKEY-223J15, WORKING DRAFT SEQUENCE, 6 unordered
            pieces.
ACCESSION   CR388149
VERSION     CR388149.3  GI:51470665
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 158085)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-AUG-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 20, 2004 this sequence version replaced gi:46408652.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK223J15
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 156562 bases at least Q40
            Consensus quality: 156889 bases at least Q30
            Consensus quality: 157151 bases at least Q20
            Insert size: 157585; sum-of-contigs
            Insert size: 159679; 5.7% error; agarose-fp
            Quality coverage: 8.57x in Q20 bases; sum-of-contigs Quality
            coverage: 8.73x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
```

```
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    64210: contig of 64210 bp in length
            *    64211    64310: gap of 100 bp
            *    64311    82166: contig of 17856 bp in length
            *    82167    82266: gap of 100 bp
            *    82267    87201: contig of 4935 bp in length
            *    87202    87301: gap of 100 bp
            *    87302   120999: contig of 33698 bp in length
            *   121000   121099: gap of 100 bp
            *   121100   131190: contig of 10091 bp in length
            *   131191   131290: gap of 100 bp
            *   131291   158085: contig of 26795 bp in length.
FEATURES             Location/Qualifiers
     source          1. .158085
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-223J15"
                     /clone_lib="DanioKey"
     misc_feature    1. .64210
                     /note="assembly_fragment:01629
                     fragment_chain:1"
     misc_feature    64311. .82166
                     /note="assembly_fragment:00258
                     fragment_chain:1"
     misc_feature    82267. .87201
                     /note="assembly_fragment:00044
                     fragment_chain:1"
     misc_feature    87302. .120999
                     /note="assembly_fragment:00990
                     fragment_chain:1"
     misc_feature    121100. .131190
                     /note="assembly_fragment:00092
                     fragment_chain:1"
     misc_feature    131291. .158085
                     /note="assembly_fragment:00487.0"
ORIGIN

  Query Match             7.5%;  Score 348.4;  DB 2;  Length 158085;
  Best Local Similarity   67.9%;  Pred. No. 2.6e-59;
  Matches  515;  Conservative    0;  Mismatches  241;  Indels    2;  Gaps    2;

Qy        227 TTTATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||| | | | | | |||||||||| | |||||||
Db      65123 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 65064

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | |||||||| |||||||| | |||||||| | ||||||||| | ||| |||| || | |
Db      65063 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 65004

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db      65003 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 64944

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db      64943 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 64884

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db      64883 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 64824

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db      64823 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 64764

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| |||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db      64763 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 64704

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| |||||||||| | |||||||||  ||||||
Db      64703 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 64644

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||          | |||          |  |   ||    |    |||
Db      64643 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 64584

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
              |  |||| | |  |  | |    | |||    | |   | || |      |      | |
Db      64583 TATAAAAATATATATATATATATATATATATATATATATATATATATATATATATATATA 64524

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885
              ||       | ||| ||| || | ||   | |  | | | ||   ||  |||  |   ||
Db      64523 TATATATATATATATATATATATATATATATATATTTATATATAATAAATAAATAAAAAA 64464

Qy        886 CACTATATATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAAT 945
              |||||||||||||  |   |  | | || |  ||||  |      ||  || ||| |
Db      64463 ATATATATATATATATTATTTTATTA-TATATATATAAAAAAAAATAAAAATATATAATT 64405

Qy        946 GTTATAGAAGAATGGCTAAGATGCCAAAGCTCAGTGTA 983
              | | | |  |||   ||| |    | |  | | | ||
Db      64404 TTAAAATATAAATAATTAAAAAAATATATTTAATTATA 64367


RESULT 96
CR774185/c
LOCUS       CR774185             167023 bp    DNA     linear   HTG 01-OCT-2004
DEFINITION  Danio rerio clone CH211-237H18, WORKING DRAFT SEQUENCE, 11
            unordered pieces.
ACCESSION   CR774185
VERSION     CR774185.2  GI:52346261
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 167023)
  AUTHORS   McLaren,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 19, 2004 this sequence version replaced gi:52213970.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zC237H18
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 164787 bases at least Q40
            Consensus quality: 165392 bases at least Q30
            Consensus quality: 165717 bases at least Q20
            Insert size: 166023; sum-of-contigs
            Insert size: 188931; 10.3% error; agarose-fp
            Quality coverage: 9.96x in Q20 bases; sum-of-contigs Quality
            coverage: 8.81x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    21904: contig of 21904 bp in length
            *    21905    22004: gap of 100 bp
```

```
            *    22005    25287: contig of 3283 bp in length
            *    25288    25387: gap of 100 bp
            *    25388    69373: contig of 43986 bp in length
            *    69374    69473: gap of 100 bp
            *    69474    71725: contig of 2252 bp in length
            *    71726    71825: gap of 100 bp
            *    71826    74171: contig of 2346 bp in length
            *    74172    74271: gap of 100 bp
            *    74272    81210: contig of 6939 bp in length
            *    81211    81310: gap of 100 bp
            *    81311    91830: contig of 10520 bp in length
            *    91831    91930: gap of 100 bp
            *    91931    99897: contig of 7967 bp in length
            *    99898    99997: gap of 100 bp
            *    99998   111895: contig of 11898 bp in length
            *   111896   111995: gap of 100 bp
            *   111996   149605: contig of 37610 bp in length
            *   149606   149705: gap of 100 bp
            *   149706   167023: contig of 17318 bp in length.
FEATURES             Location/Qualifiers
     source          1. .167023
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-237H18"
                     /clone_lib="CHORI-211"
     misc_feature    1. .21904
                     /note="assembly_fragment:00892
                     clone_end:SP6
                     vector_side:left"
     misc_feature    22005. .25287
                     /note="assembly_fragment:00055
                     fragment_chain:1"
     misc_feature    25388. .69373
                     /note="assembly_fragment:01905
                     fragment_chain:1"
     misc_feature    69474. .71725
                     /note="assembly_fragment:00023"
     misc_feature    71826. .74171
                     /note="assembly_fragment:00031"
     misc_feature    74272. .81210
                     /note="assembly_fragment:00083
                     fragment_chain:2"
     misc_feature    81311. .91830
                     /note="assembly_fragment:00416
                     fragment_chain:2"
     misc_feature    91931. .99897
                     /note="assembly_fragment:00157
                     fragment_chain:2"
     misc_feature    99998. .111895
                     /note="assembly_fragment:00260
                     fragment_chain:2"
     misc_feature    111996. .149605
                     /note="assembly_fragment:01303
                     fragment_chain:2"
     misc_feature    149706. .167023
                     /note="assembly_fragment:00585
                     fragment_chain:2
                     clone_end:T7
                     vector_side:right"
ORIGIN

  Query Match             7.5%;  Score 348.4;  DB 2;  Length 167023;
  Best Local Similarity   64.6%;  Pred. No. 2.6e-59;
  Matches  461;  Conservative    0;  Mismatches  251;  Indels    2;  Gaps    2;

Qy        212 AAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACACATATAT 271
              || |  ||||   ||| ||||||||||||||||||||||| | | | | | ||||||
Db     100501 AATATTAAAAAAAATTATATATATATATATATATATATATATATATATATATATATAT 100442

Qy        272 ATATACACATATATATACACATATATACATATATACACATATATACACATATATATACAC 331
              |||||  | |||||||||  | |||||||  |||||||  | |||||||  | |||||||||  |
Db     100441 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 100382

Qy        332 ATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACACAC 391
              |||  ||||  ||  |  |||||||||  |  ||||  ||||  |  |||||||  |  |||||||  |  |
Db     100381 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 100322

Qy        392 ATATATACACATATATACACACATATATACACATATATACACATATATACACATATACAC 451
              |||||||  |  |||||||  |  |  |||||||  |  |||||||  |  |||||||  |  |||||  |
Db     100321 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 100262

Qy        452 ATATATACACATATATACATATATACACATATATATAATATACACACATATATATACACA 511
              |||||||  |  |||||||  |||||||  |  ||||||||  |||||  |  |  |||||||||  |  |
Db     100261 ATATATATATATATATATATATATATATATATATATAT-ATATATATATATATATATATATA 100203

Qy        512 TATATACACACATATATACACATATATACACATATATATACACATATATACACATATATA 571
              ||||||  |  |  |||||||  |  |||||||  |  |||||||||  |  |||||||  |  |||||||
Db     100202 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 100143

Qy        572 CATATATACACATATATATACATATATACACATATATACATATATACACATATATACATA 631
              |||||||  |  |||||||||  |||||||  |  |||||||  |||||||  |  |||||||  |||
Db     100142 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 100083

Qy        632 TATACACACATATATACACATACATATACACACACATAGATATACATATATATACACATA 691
              ||||  |  |  |||||||  |  |||  |||||  |  |  |  |||  |||||  |||||||||  |  |||
Db     100082 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 100023

Qy        692 TATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGAT 751
              ||||||  |||||||||  |||||||
Db     100022 TATATATATATATATATATATATATNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 99963

Qy        752 TACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTC 811

Db      99962 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 99903

Qy        812 AGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA-AGG 870
                      | |||       |||||  ||| ||  |  ||   |  |  |||  |  ||  |
Db      99902 NNNNNATATATATATATATATATAAATATATATATATATATATATATATATATATATA 99843

Qy        871 TAGTTAAGAAGGTAACACTATATATATATATAGTGAAAGCAGTGTAAACCTAAA 924
              ||  |||  |  ||    |||||||||||||   | |    | || |  || |
Db      99842 TATATAAATATATATATATATATATATATATATATATATATATATATATATATA 99789


RESULT 97
CR788285
LOCUS       CR788285              169053 bp    DNA     linear   HTG 03-OCT-2004
DEFINITION  Danio rerio clone DKEYP-99C4, *** SEQUENCING IN PROGRESS ***, 5
            unordered pieces.
ACCESSION   CR788285
VERSION     CR788285.4  GI:53755972
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 169053)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 3, 2004 this sequence version replaced gi:53145965.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zKp99C4
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 167822 bases at least Q40
```

```
           Consensus quality: 168197 bases at least Q30
           Consensus quality: 168423 bases at least Q20
           Insert size: 168653; sum-of-contigs
           Insert size: 171627; 2.8% error; agarose-fp
           Quality coverage: 12.96x in Q20 bases; sum-of-contigs Quality
           coverage: 12.74x in Q20 bases; agarose-fp
           --------------.
           * NOTE: This is a 'working draft' sequence. It currently
           * consists of 5 contigs. The true order of the pieces
           * is not known and their order in this sequence record is
           * arbitrary. Gaps between the contigs are represented as
           * runs of N, but the exact sizes of the gaps are unknown.
           * This record will be updated with the finished sequence
           * as soon as it is available and the accession number will
           * be preserved.
           *       1     5600: contig of 5600 bp in length
           *    5601     5700: gap of 100 bp
           *    5701    59937: contig of 54237 bp in length
           *   59938    60037: gap of 100 bp
           *   60038    78734: contig of 18697 bp in length
           *   78735    78834: gap of 100 bp
           *   78835   100059: contig of 21225 bp in length
           *  100060   100159: gap of 100 bp
           *  100160   169053: contig of 68894 bp in length.
FEATURES             Location/Qualifiers
     source          1. .169053
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEYP-99C4"
                     /clone_lib="DanioKeypilot"
     misc_feature    1. .5600
                     /note="assembly_fragment:00020
                     fragment_chain:1"
     misc_feature    5701. .59937
                     /note="assembly_fragment:00981
                     fragment_chain:1"
     misc_feature    60038. .78734
                     /note="assembly_fragment:00116
                     fragment_chain:2"
     misc_feature    78835. .100059
                     /note="assembly_fragment:00523
                     fragment_chain:2"
     misc_feature    100160. .169053
                     /note="assembly_fragment:02105.0"
ORIGIN

  Query Match             7.5%;  Score 348.4;  DB 2;  Length 169053;
  Best Local Similarity   64.5%;  Pred. No. 2.6e-59;
  Matches  463;  Conservative    0;  Mismatches  251;  Indels    4;  Gaps    2;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db       5108 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 5167

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || ||||||||| || |||| | ||||||| | ||||||||| | ||| |||| || | |
Db       5168 TATACATATATATATGTATATATATATATATATATATATATATATATATATATATATATA 5227

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db       5228 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 5287

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db       5288 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 5347

Qy        467 TACATATATACACATATATATAATATACACACATATATATACACATATATACACACATAT 526
              || ||||||| | |||||||| |||||  | | ||||||||| | ||||||| | | ||||
Db       5348 TATATATATATATATATATAT-ATATATATATATATATATATATATATATATATATATAT 5406

Qy        527 ATACACATATATACACATATATATACACATATATACACATATATACATATATACACATAT 586
              ||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | ||||

Db       5407 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 5466

Qy        587 ATATACATATATACACATATATACATATATACACATATATACATATATACACACATATAT 646
              ||||| ||||||||| ||||||| ||||||| | ||||||| ||||||| | | ||||||
Db       5467 ATATATATATATACATATATATATATATATATATATATATATATATATATATATATATAT 5526

Qy        647 ACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATAT 706
              | | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  |||||||
Db       5527 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 5586

Qy        707 ATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAG 766
              |||||||||||||||
Db       5587 ATGTATATATATATNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 5646

Qy        767 AGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAAT 826
                                                                      |||
Db       5647 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNAAAAAA 5706

Qy        827 AGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAAC 886
              |        | ||| ||| || | ||   | |  ||| | ||   ||  ||   |  ||
Db       5707 ATATATATATATATATATATATATATATATATATATA---TATATATATATATATA 5763

Qy        887 ACTATATATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAA 944
              |||||||||  | |   ||| || || | |     |||   |   |   ||||||
Db       5764 TATATATATATTAAAAAATAAAACATTGCAGTCTGTTGGGAGATGTGCATCGTTTAAA 5821


RESULT 98
CR749743/c
LOCUS       CR749743              171424 bp    DNA     linear   HTG 10-OCT-2004
DEFINITION  Danio rerio clone DKEY-97M3, WORKING DRAFT SEQUENCE, 9 unordered
            pieces.
ACCESSION   CR749743
VERSION     CR749743.4  GI:54021791
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 171424)
  AUTHORS   McLaren,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 10, 2004 this sequence version replaced gi:52213800.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK97M3
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 165755 bases at least Q40
            Consensus quality: 166566 bases at least Q30
            Consensus quality: 167672 bases at least Q20
            Insert size: 170624; sum-of-contigs
            Insert size: 189327; 3.5% error; agarose-fp
            Quality coverage: 8.39x in Q20 bases; sum-of-contigs Quality
            coverage: 7.61x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
```

```
            *        1    13261: contig of 13261 bp in length
            *    13262    13361: gap of 100 bp
            *    13362    19777: contig of 6416 bp in length
            *    19778    19877: gap of 100 bp
            *    19878    33915: contig of 14038 bp in length
            *    33916    34015: gap of 100 bp
            *    34016    61771: contig of 27756 bp in length
            *    61772    61871: gap of 100 bp
            *    61872    94572: contig of 32701 bp in length
            *    94573    94672: gap of 100 bp
            *    94673   131525: contig of 36853 bp in length
            *   131526   131625: gap of 100 bp
            *   131626   133888: contig of 2263 bp in length
            *   133889   133988: gap of 100 bp
            *   133989   157725: contig of 23737 bp in length
            *   157726   157825: gap of 100 bp
            *   157826   171424: contig of 13599 bp in length.
FEATURES             Location/Qualifiers
     source          1. .171424
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-97M3"
                     /clone_lib="DanioKey"
     misc_feature    1. .13261
                     /note="assembly_fragment:00444
                     fragment_chain:1"
     misc_feature    13362. .19777
                     /note="assembly_fragment:00033
                     fragment_chain:1"
     misc_feature    19878. .33915
                     /note="assembly_fragment:00098
                     fragment_chain:1"
     misc_feature    34016. .61771
                     /note="assembly_fragment:00628
                     fragment_chain:2"
     misc_feature    61872. .94572
                     /note="assembly_fragment:01012
                     fragment_chain:2"
     misc_feature    94673. .131525
                     /note="assembly_fragment:01459
                     fragment_chain:2"
     misc_feature    131626. .133888
                     /note="assembly_fragment:00008
                     fragment_chain:3"
     misc_feature    133989. .157725
                     /note="assembly_fragment:01969
                     fragment_chain:3"
     misc_feature    157826. .171424
                     /note="assembly_fragment:00264.0"
ORIGIN

  Query Match             7.5%;  Score 348.4;  DB 2;  Length 171424;
  Best Local Similarity   67.6%;  Pred. No. 2.6e-59;
  Matches  495;  Conservative    0;  Mismatches  236;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              |||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db     134802 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 134743

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db     134742 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 134683

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db     134682 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 134623

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db     134622 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 134563

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db     134562 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 134503

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db     134502 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 134443

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db     134442 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 134383

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db     134382 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 134323

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||||          | |||          |  |   ||    |    |||
Db     134322 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 134263

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
              |   | | | |  | | |    | |||    | |   | || |    | |    | |
Db     134262 TATATATATATATATATATATATATATATATATATATATATATATATATATAAATATATA 134203

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885
              ||  |    | ||| ||| || | ||   |      ||      |   |  ||  |  ||
Db     134202 TATAAATATATATATATATATTTATATATTTTAATAATATAANNAAAAAAATATATAN 134143

Qy        886 CACTATATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAAT 945
              |||||||||||||| | |||    | || | |      ||     ||  |   | |
Db     134142 ATATATATATATATATATTTAAATATATATATATTTTTTTNATNTTAAAANNANAAANAN 134083

Qy        946 GTTATAGAAGAA 957
                | | || ||
Db     134082 AAAAAATAATAA 134071


RESULT 99
CR759841
LOCUS       CR759841               178975 bp    DNA     linear   HTG 08-SEP-2004
DEFINITION  Danio rerio clone DKEY-192G7, *** SEQUENCING IN PROGRESS ***, 6
            unordered pieces.
ACCESSION   CR759841
VERSION     CR759841.2  GI:51965207
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 178975)
  AUTHORS   Burton,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 9, 2004 this sequence version replaced gi:51890356.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK192G7
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 176373 bases at least Q40
            Consensus quality: 176964 bases at least Q30
            Consensus quality: 177367 bases at least Q20
            Insert size: 178475; sum-of-contigs
            Insert size: 178296; 1.0% error; agarose-fp
            Quality coverage: 11.15x in Q20 bases; sum-of-contigs Quality
            coverage: 11.48x in Q20 bases; agarose-fp
```

```
             --------------.
             * NOTE: This is a 'working draft' sequence. It currently
             * consists of 6 contigs. The true order of the pieces
             * is not known and their order in this sequence record is
             * arbitrary. Gaps between the contigs are represented as
             * runs of N, but the exact sizes of the gaps are unknown.
             * This record will be updated with the finished sequence
             * as soon as it is available and the accession number will
             * be preserved.
             *       1     3686: contig of 3686 bp in length
             *    3687     3786: gap of 100 bp
             *    3787    62427: contig of 58641 bp in length
             *   62428    62527: gap of 100 bp
             *   62528    84899: contig of 22372 bp in length
             *   84900    84999: gap of 100 bp
             *   85000   136604: contig of 51605 bp in length
             *  136605   136704: gap of 100 bp
             *  136705   158073: contig of 21369 bp in length
             *  158074   158173: gap of 100 bp
             *  158174   178975: contig of 20802 bp in length.
FEATURES             Location/Qualifiers
     source          1. .178975
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-192G7"
                     /clone_lib="DanioKey"
     misc_feature    1. .3686
                     /note="assembly_fragment:00051
                     fragment_chain:1"
     misc_feature    3787. .62427
                     /note="assembly_fragment:01908
                     fragment_chain:1"
     misc_feature    62528. .84899
                     /note="assembly_fragment:00400
                     fragment_chain:1"
     misc_feature    85000. .136604
                     /note="assembly_fragment:01093
                     fragment_chain:2"
     misc_feature    136705. .158073
                     /note="assembly_fragment:00722
                     fragment_chain:2"
     misc_feature    158174. .178975
                     /note="assembly_fragment:00096
                     fragment_chain:2"
ORIGIN

  Query Match             7.5%;  Score 348.4;  DB 2;  Length 178975;
  Best Local Similarity   60.4%;  Pred. No. 2.5e-59;
  Matches  492;  Conservative    0;  Mismatches  321;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              |||||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db     135912 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 135971

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db     135972 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 136031

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db     136032 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 136091

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db     136092 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 136151

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db     136152 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 136211

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||

Db     136212 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 136271

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db     136272 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 136331

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db     136332 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 136391

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||||          |  |||          |  |   ||    |   |||
Db     136392 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 136451

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
              |   | | | |  |  | |   | |||   | |   | || |     |   |  | |
Db     136452 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 136511

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885
              ||       | ||| ||| || | ||   | |  ||| | ||      |  |  |  ||
Db     136512 TATATATATATATATATATATATATATATATATATATATATATATATATATAATATATAT 136571

Qy        886 CACTATATATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAAT 945
                |||||||||||||   | |    | || |
Db     136572 ATATATATATATATATATATATATATATATANNNNNNNNNNNNNNNNNNNNNNNNNN 136631

Qy        946 GTTATAGAAGAATGGCTAAGATGCCAAAGCTCAGTGTATGTGGCAGAGGCATGGTGTAGG 1005

Db     136632 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 136691

Qy       1006 GTGTGTCCAGGTTCATATATTGCATTAAGTGTGA 1039
                         | | ||  ||||| | | |
Db     136692 NNNNNNNNNNNNTTTTTTTTTTATTAATTATTA 136725


RESULT 100
CR792414
LOCUS       CR792414              179213 bp    DNA     linear   HTG 25-SEP-2004
DEFINITION  Danio rerio clone CH211-95O10, *** SEQUENCING IN PROGRESS ***, 9
            unordered pieces.
ACCESSION   CR792414
VERSION     CR792414.2  GI:52694443
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 179213)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 25, 2004 this sequence version replaced gi:52626621.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zC95O10
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 176437 bases at least Q40
            Consensus quality: 176944 bases at least Q30
            Consensus quality: 177295 bases at least Q20
            Insert size: 178413; sum-of-contigs
            Insert size: 195379; 9.0% error; agarose-fp
            Quality coverage: 8.87x in Q20 bases; sum-of-contigs Quality
            coverage: 8.31x in Q20 bases; agarose-fp
            --------------.
```

```
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     4542: contig of 4542 bp in length
            *    4543     4642: gap of 100 bp
            *    4643    41427: contig of 36785 bp in length
            *   41428    41527: gap of 100 bp
            *   41528    50062: contig of 8535 bp in length
            *   50063    50162: gap of 100 bp
            *   50163    59923: contig of 9761 bp in length
            *   59924    60023: gap of 100 bp
            *   60024   103384: contig of 43361 bp in length
            *  103385   103484: gap of 100 bp
            *  103485   121057: contig of 17573 bp in length
            *  121058   121157: gap of 100 bp
            *  121158   148007: contig of 26850 bp in length
            *  148008   148107: gap of 100 bp
            *  148108   169047: contig of 20940 bp in length
            *  169048   169147: gap of 100 bp
            *  169148   179213: contig of 10066 bp in length.
FEATURES             Location/Qualifiers
     source          1. .179213
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-95O10"
                     /clone_lib="CHORI-211"
     misc_feature    1. .4542
                     /note="assembly_fragment:00045
                     fragment_chain:1
                     clone_end:T7
                     vector_side:left"
     misc_feature    4643. .41427
                     /note="assembly_fragment:01381
                     fragment_chain:1"
     misc_feature    41528. .50062
                     /note="assembly_fragment:00108
                     fragment_chain:1"
     misc_feature    50163. .59923
                     /note="assembly_fragment:00205
                     fragment_chain:1"
     misc_feature    60024. .103384
                     /note="assembly_fragment:01959"
     misc_feature    103485. .121057
                     /note="assembly_fragment:00463
                     fragment_chain:2"
     misc_feature    121158. .148007
                     /note="assembly_fragment:00984
                     fragment_chain:2"
     misc_feature    148108. .169047
                     /note="assembly_fragment:00719
                     fragment_chain:2"
     misc_feature    169148. .179213
                     /note="assembly_fragment:00316
                     fragment_chain:2
                     clone_end:SP6
                     vector_side:right"
ORIGIN

  Query Match          7.5%;  Score 348.4;  DB 2;  Length 179213;
  Best Local Similarity   66.8%;  Pred. No. 2.5e-59;
  Matches  524;  Conservative    0;  Mismatches  258;  Indels    2;  Gaps    2;

Qy        191 AGAGCGAAATGTCATCTCAGAAAAAAAAAAAGGCATTTTATATATATATATATATATATA 250
              |||  |  || ||  |   ||  ||||||  |   |  ||||||||||||||||||||||
Db     120224 AGACAGGCATTTCGACAGTGATTAAAAAACAACAACATTATATATATATATATATATATA 120283

Qy        251 TACACACACACACACATATATATATACACATATATATACACATATATACATATATACACA 310
              ||| | | | | | | |||||||||| | ||||||||| | ||||||| ||||||| | |
Db     120284 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 120343

Qy        311 TATATACACATATATATACACATACATATGTACACATATATATACACATATGTATACACA 370
              |||||| | ||||||||| | ||| |||| || | ||||||||| | |||| |||| | |
Db     120344 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 120403

Qy        371 TATATACACATATATACACACATATATACACATATATACACACATATATACACATATATA 430
              |||||| | ||||||| | | ||||||| | ||||||| | | ||||||| | |||||||
Db     120404 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 120463

Qy        431 CACATATATACACATATACACATATATACACATATATACATATATACACATATATATAAT 490
              | ||||||| | ||||| | ||||||| | ||||||| ||||||| | |||||||| ||
Db     120464 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT-AT 120522

Qy        491 ATACACACATATATATACACATATATACACACATATATACACATATATACACATATATAT 550
              ||| | | |||||||| | ||||||| | | ||||||| | ||||||| | ||||||||
Db     120523 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 120582

Qy        551 ACACATATATACACATATATACATATATACACATATATATACATATATACACATATATAC 610
              | | ||||||| | ||||||| ||||||| | ||||||||| ||||||| | |||||||
Db     120583 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 120642

Qy        611 ATATATACACATATATACATATATACACACATATATACACATACATATACACACACATAG 670
              ||||||| | ||||||| ||||||| | | ||||||| | ||| ||||| | | | |||
Db     120643 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 120702

Qy        671 ATATACATATATATACACATATATATACGTATATATATGTATATATATATGCTCCAGAGT 730
              ||||| ||||||||| | |||||||||  |||||||| ||||||||||          |
Db     120703 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 120762

Qy        731 TCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGT 790
              |||          |  |   ||    |    ||| |   | | | |  | | |   |
Db     120763 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 120822

Qy        791 GTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGT 850
              |||    | |   | || |      |       | |||       | ||| ||| || | |
Db     120823 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 120882

Qy        851 AATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTAACACTATATATATATATAGTGAAAG 909
              |   | | ||| | || |  ||  ||   |  ||    ||||||||||||||   |||
Db     120883 ATATATATATATATATATATATATATATATATATATATATATATATATATATATAAAT 120942

Qy        910 CAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTATAGAAGAATGGCTAAGATGC 969
              | || |  || |   |     |  || || || |    || | |   ||  |
Db     120943 ATATATATATATATATATTTATATATATATATATATATATATAAAAANTAAATATAAATA 121002

Qy        970 CAAA 973
              |||
Db     121003 NAAA 121006


Search completed: April 21, 2005, 17:02:39
Job time : 12969.1 secs
```

This Page Blank (uspto)

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 20, 2005, 22:49:10 ; Search time 11529.3 Seconds
(without alignments)
17319.707 Million cell updates/sec

Title: US-09-939-209A-3_COPY_1_4121
Perfect score: 4121
Sequence: 1 agttcaagaccagcctgagc.....tgtgagggcaaattgaagtc 4121

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	4121	100.0	20300	6	AX451337	AX451337 Sequence
	2	4107.4	99.7	165329	9	AL583850	AL583850 Human DNA
c	3	3720	90.3	191699	2	AC031977	AC031977 Homo sapi
	4	411.2	10.0	203216	2	AC097328	AC097328 Pan trogl
	5	405.4	9.8	104321	2	CR759943	CR759943 Homo sapi
c	6	405.2	9.8	181946	2	AC097330	AC097330 Pan trogl
	7	402.2	9.8	109081	9	AL358074	AL358074 Human DNA
	8	396.8	9.6	164844	9	AC009058	AC009058 Homo sapi
	9	395.2	9.6	130349	9	AC011593	AC011593 Homo sapi
c	10	394.8	9.6	196413	2	AC097269	AC097269 Pan trogl
	11	390.8	9.5	189662	9	AC015468	AC015468 Homo sapi
c	12	387.4	9.4	125685	9	AC073310	AC073310 Homo sapi
	13	385.8	9.4	139118	9	AL133477	AL133477 Human DNA
	14	382.6	9.3	216021	9	HUAC004787	AC004787 Homo sapi
	15	379.8	9.2	92564	9	AY007685	AY007685 Homo sapi
c	16	379.4	9.2	161252	2	AC114955	AC114955 Homo sapi
c	17	378	9.2	202305	9	AC114291	AC114291 Homo sapi
c	18	373.6	9.1	192139	9	AC021955	AC021955 Homo sapi
c	19	372	9.0	99596	9	AC106873	AC106873 Homo sapi
	20	371.4	9.0	38096	9	BX901949	BX901949 Human DNA
c	21	371.4	9.0	110000	2	BX324168_0	BX324168 Homo sapi
	22	370.4	9.0	252512	2	AC097326	AC097326 Pan trogl
c	23	368.2	8.9	170646	2	AC123545	AC123545 Pan trogl
	24	367	8.9	112762	9	HS323B6	Z83841 Human DNA s
c	25	365.4	8.9	163976	9	AC108483	AC108483 Homo sapi
	26	365.4	8.9	163998	9	AC022166	AC022166 Homo sapi
c	27	364.2	8.8	153201	2	AC092863	AC092863 Homo sapi
c	28	363.6	8.8	151549	9	AC092694	AC092694 Homo sapi
	29	357.4	8.7	217985	9	AC021193	AC021193 Homo sapi
c	30	357.2	8.7	151424	2	CR812471	CR812471 Danio rer
c	31	356.6	8.7	121600	9	HS141H5	AL049176 Human DNA
c	32	356.6	8.7	127172	2	CR788239	CR788239 Danio rer
	33	355.4	8.6	282240	2	CR753865	CR753865 Danio rer
c	34	355.2	8.6	149969	2	AC024180	AC024180 Homo sapi
c	35	355.2	8.6	151047	2	CR759776	CR759776 Danio rer
c	36	355.2	8.6	162996	9	AC006441	AC006441 Homo sapi
c	37	354.8	8.6	182430	9	AL590381	AL590381 Human DNA
c	38	354	8.6	252450	2	CR753861	CR753861 Danio rer
	39	353.6	8.6	163651	2	CR790381	CR790381 Danio rer
c	40	353.6	8.6	164994	2	AC141288	AC141288 Homo sapi
c	41	353	8.6	165979	2	CR376741	CR376741 Danio rer
c	42	353	8.6	248167	2	CR627491	CR627491 Danio rer
	43	352.4	8.6	178254	9	AL731541	AL731541 Human DNA
	44	352.4	8.6	204486	2	CR752655	CR752655 Danio rer
c	45	352.2	8.5	178427	2	CR392006	CR392006 Danio rer

ALIGNMENTS

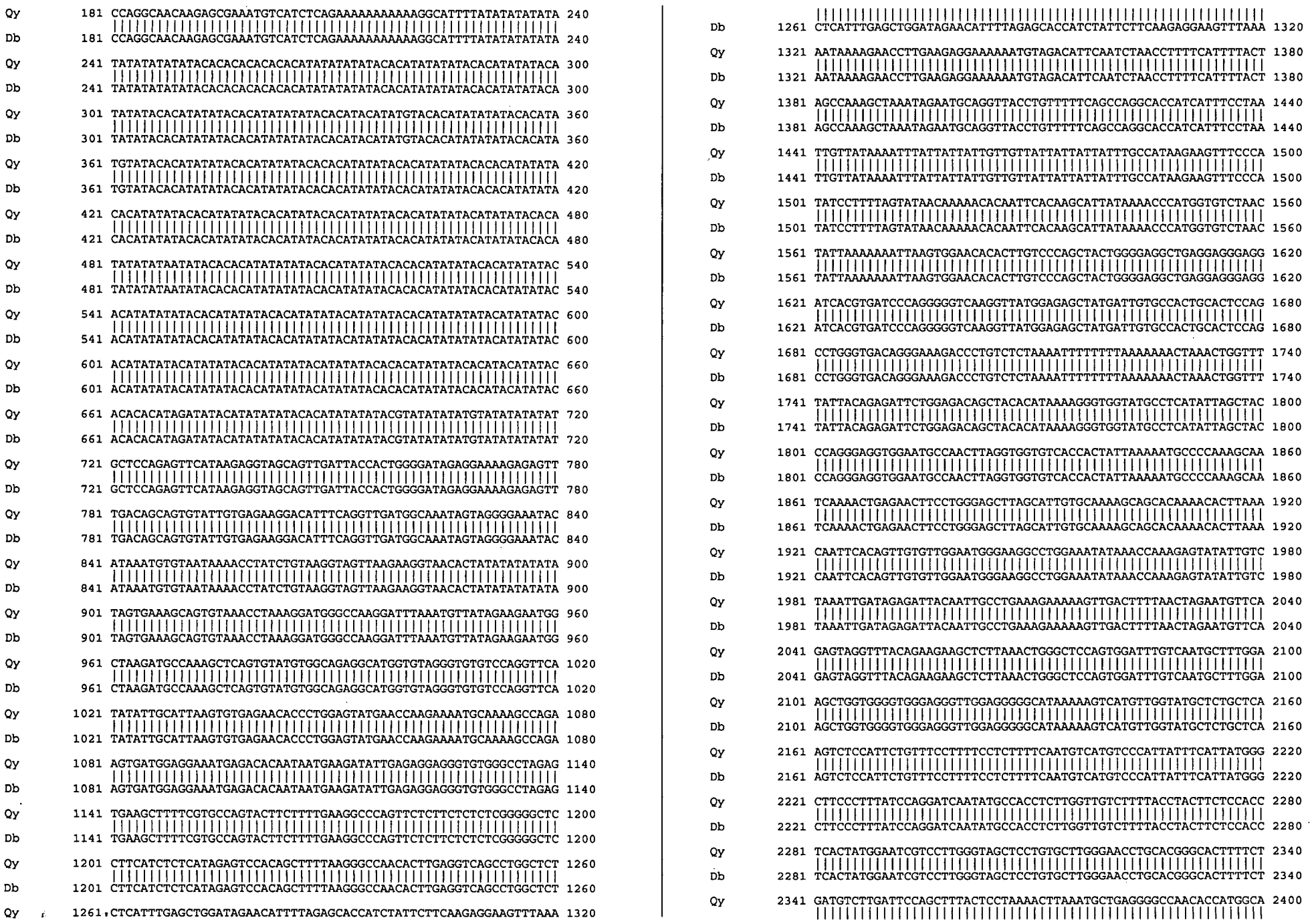
RESULT 1
AX451337
LOCUS AX451337 20300 bp DNA linear PAT 03-JUL-2002
DEFINITION Sequence 3 from Patent WO0216653.
ACCESSION AX451337
VERSION AX451337.1 GI:21698388
KEYWORDS .
SOURCE synthetic construct
ORGANISM synthetic construct
other sequences; artificial sequences.
REFERENCE 1
AUTHORS Levitt,P.R., Mirnics,K., Kodavali,V.C. and Nimgaonkar,V.L.
TITLE Methods and systems for facilitating the diagnosis and treatment of schizophrenia
JOURNAL Patent: WO 0216653-A 3 28-FEB-2002;
University of Pittsburgh (US)
FEATURES Location/Qualifiers
source 1. .20300
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="A genomic sequence containing RGS4 nucleic acid sequence and sequ ences upstream and downstream to the RGS4 nucleic acid sequence"
ORIGIN

Query Match 100.0%; Score 4121; DB 6; Length 20300;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4121; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAATTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAATTA 60

Qy         61 GACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAAT 120

Qy        121 CACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAGC 180
```


Qy 181 CCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATA 240
Db 181 CCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAGGCATTTTATATATATATA 240
Qy 241 TATATATATATACACACACACACACATATATATATACACATATATATACACATATATACA 300
Db 241 TATATATATATACACACACACACACATATATATATACACATATATATACACATATATACA 300
Qy 301 TATATACACATATATACACATATATATACACATACATATGTACACATATATATACACATA 360
Db 301 TATATACACATATATACACATATATATACACATACATATGTACACATATATATACACATA 360
Qy 361 TGTATACACATATATACACATATATACACACATATATACACATATATACACACATATATA 420
Db 361 TGTATACACATATATACACATATATACACACATATATACACATATATACACACATATATA 420
Qy 421 CACATATATACACATATATACACATATACACATATATACACATATATACATATATACACA 480
Db 421 CACATATATACACATATATACACATATACACATATATACACATATATACATATATACACA 480
Qy 481 TATATATAATATACACACATATATATACACATATATACACACATATATACACATATATAC 540
Db 481 TATATATAATATACACACATATATATACACATATATACACACATATATACACATATATAC 540
Qy 541 ACATATATATACACATATATACACATATATACATATATACACATATATATACATATATAC 600
Db 541 ACATATATATACACATATATACACATATATACATATATACACATATATATACATATATAC 600
Qy 601 ACATATATACATATATACACATATATACATATATACACACATATATACACATACATATAC 660
Db 601 ACATATATACATATATACACATATATACATATATACACACATATATACACATACATATAC 660
Qy 661 ACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATAT 720
Db 661 ACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATAT 720
Qy 721 GCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTT 780
Db 721 GCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTT 780
Qy 781 TGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATAC 840
Db 781 TGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATAC 840
Qy 841 ATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATATATATA 900
Db 841 ATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATATATATA 900
Qy 901 TAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTATAGAAGAATGG 960
Db 901 TAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTATAGAAGAATGG 960
Qy 961 CTAAGATGCCAAAGCTCAGTGTATGTGGCAGAGGCATGGTGTAGGGTGTGTCCAGGTTCA 1020
Db 961 CTAAGATGCCAAAGCTCAGTGTATGTGGCAGAGGCATGGTGTAGGGTGTGTCCAGGTTCA 1020
Qy 1021 TATATTGCATTAAGTGTGAGAACACCCTGGAGTATGAACCAAGAAAATGCAAAAGCCAGA 1080
Db 1021 TATATTGCATTAAGTGTGAGAACACCCTGGAGTATGAACCAAGAAAATGCAAAAGCCAGA 1080
Qy 1081 AGTGATGGAGGAAATGAGACACAATAATGAAGATATTGAGAGGAGGGTGTGGGCCTAGAG 1140
Db 1081 AGTGATGGAGGAAATGAGACACAATAATGAAGATATTGAGAGGAGGGTGTGGGCCTAGAG 1140
Qy 1141 TGAAGCTTTTCGTGCCAGTACTTCTTTTGAAGGCCCAGTTCTCTTCTCTCTCGGGGGCTC 1200
Db 1141 TGAAGCTTTTCGTGCCAGTACTTCTTTTGAAGGCCCAGTTCTCTTCTCTCTCGGGGGCTC 1200
Qy 1201 CTTCATCTCTCATAGAGTCCACAGCTTTTAAGGGCCAACACTTGAGGTCAGCCTGGCTCT 1260
Db 1201 CTTCATCTCTCATAGAGTCCACAGCTTTTAAGGGCCAACACTTGAGGTCAGCCTGGCTCT 1260
Qy 1261 CTCATTTGAGCTGGATAGAACATTTTAGAGCACCATCTATTCTTCAAGAGGAAGTTTAAA 1320
Db 1261 CTCATTTGAGCTGGATAGAACATTTTAGAGCACCATCTATTCTTCAAGAGGAAGTTTAAA 1320
Qy 1321 AATAAAAGAACCTTGAAGAGGAAAAAATGTAGACATTCAATCTAACCTTTTCATTTTACT 1380
Db 1321 AATAAAAGAACCTTGAAGAGGAAAAAATGTAGACATTCAATCTAACCTTTTCATTTTACT 1380
Qy 1381 AGCCAAAGCTAAATAGAATGCAGGTTACCTGTTTTTCAGCCAGGCACCATCATTTCCTAA 1440
Db 1381 AGCCAAAGCTAAATAGAATGCAGGTTACCTGTTTTTCAGCCAGGCACCATCATTTCCTAA 1440
Qy 1441 TTGTTATAAAATTTATTATTATTGTTGTTATTATTATTATTTGCCATAAGAAGTTTCCCA 1500
Db 1441 TTGTTATAAAATTTATTATTATTGTTGTTATTATTATTATTTGCCATAAGAAGTTTCCCA 1500
Qy 1501 TATCCTTTTAGTATAACAAAAACACAATTCACAAGCATTATAAAACCCATGGTGTCTAAC 1560
Db 1501 TATCCTTTTAGTATAACAAAAACACAATTCACAAGCATTATAAAACCCATGGTGTCTAAC 1560
Qy 1561 TATTAAAAAAAATTAAGTGGAACACACTTGTCCCAGCTACTGGGGAGGCTGAGGAGGGAGG 1620
Db 1561 TATTAAAAAAAATTAAGTGGAACACACTTGTCCCAGCTACTGGGGAGGCTGAGGAGGGAGG 1620
Qy 1621 ATCACGTGATCCCAGGGGGTCAAGGTTATGGAGAGCTATGATTGTGCCACTGCACTCCAG 1680
Db 1621 ATCACGTGATCCCAGGGGGTCAAGGTTATGGAGAGCTATGATTGTGCCACTGCACTCCAG 1680
Qy 1681 CCTGGGTGACAGGGAAAGACCCTGTCTCTAAAATTTTTTTTAAAAAAACTAAACTGGTTT 1740
Db 1681 CCTGGGTGACAGGGAAAGACCCTGTCTCTAAAATTTTTTTTAAAAAAACTAAACTGGTTT 1740
Qy 1741 TATTACAGAGATTCTGGAGACAGCTACACATAAAAGGGTGGTATGCCTCATATTAGCTAC 1800
Db 1741 TATTACAGAGATTCTGGAGACAGCTACACATAAAAGGGTGGTATGCCTCATATTAGCTAC 1800
Qy 1801 CCAGGGAGGTGGAATGCCAACTTAGGTGGTGTCACCACTATTAAAAATGCCCCAAAGCAA 1860
Db 1801 CCAGGGAGGTGGAATGCCAACTTAGGTGGTGTCACCACTATTAAAAATGCCCCAAAGCAA 1860
Qy 1861 TCAAAACTGAGAACTTCCTGGGAGCTTAGCATTGTGCAAAAGCAGCACAAAACACTTAAA 1920
Db 1861 TCAAAACTGAGAACTTCCTGGGAGCTTAGCATTGTGCAAAAGCAGCACAAAACACTTAAA 1920
Qy 1921 CAATTCACAGTTGTGTTGGAATGGGAAGGCCTGGAAATATAAACCAAAGAGTATATTGTC 1980
Db 1921 CAATTCACAGTTGTGTTGGAATGGGAAGGCCTGGAAATATAAACCAAAGAGTATATTGTC 1980
Qy 1981 TAAATTGATAGAGATTACAATTGCCTGAAAGAAAAAGTTGACTTTTAACTAGAATGTTCA 2040
Db 1981 TAAATTGATAGAGATTACAATTGCCTGAAAGAAAAAGTTGACTTTTAACTAGAATGTTCA 2040
Qy 2041 GAGTAGGTTTACAGAAGAAGCTCTTAAACTGGGCTCCAGTGGATTTGTCAATGCTTTGGA 2100
Db 2041 GAGTAGGTTTACAGAAGAAGCTCTTAAACTGGGCTCCAGTGGATTTGTCAATGCTTTGGA 2100
Qy 2101 AGCTGGTGGGGTGGGAGGGTTGGAGGGGGCATAAAAAGTCATGTTGGTATGCTCTGCTCA 2160
Db 2101 AGCTGGTGGGGTGGGAGGGTTGGAGGGGGCATAAAAAGTCATGTTGGTATGCTCTGCTCA 2160
Qy 2161 AGTCTCCATTCTGTTTCCTTTTCCTCTTTTCAATGTCATGTCCCATTATTTCATTATGGG 2220
Db 2161 AGTCTCCATTCTGTTTCCTTTTCCTCTTTTCAATGTCATGTCCCATTATTTCATTATGGG 2220
Qy 2221 CTTCCCTTTATCCAGGATCAATATGCCACCTCTTGGTTGTCTTTTACCTACTTCTCCACC 2280
Db 2221 CTTCCCTTTATCCAGGATCAATATGCCACCTCTTGGTTGTCTTTTACCTACTTCTCCACC 2280
Qy 2281 TCACTATGGAATCGTCCTTGGGTAGCTCCTGTGCTTGGGAACCTGCACGGGCACTTTTCT 2340
Db 2281 TCACTATGGAATCGTCCTTGGGTAGCTCCTGTGCTTGGGAACCTGCACGGGCACTTTTCT 2340
Qy 2341 GATGTCTTGATTCCAGCTTTACTCCTAAAACTTAAATGCTGAGGGGCCAACACCATGGCA 2400

```
Db      2341 GATGTCTTGATTCCAGCTTTACTCCTAAAACTTAAATGCTGAGGGGCCAACACCATGGCA 2400
Qy      2401 GTGGTAGGGATGGGAATGGGGGTCTTGTAACACACTACATAAACTACACGAAATAAACTA 2460
Db      2401 GTGGTAGGGATGGGAATGGGGGTCTTGTAACACACTACATAAACTACACGAAATAAACTA 2460
Qy      2461 CATGAAACTCAACATGTTTGCAAGACTCAGTTCACATCCATGAGGAGCTCATGCTTCTCC 2520
Db      2461 CATGAAACTCAACATGTTTGCAAGACTCAGTTCACATCCATGAGGAGCTCATGCTTCTCC 2520
Qy      2521 CTCCTGCTCCCCTAGCACACATGATTATCTCTATTTGGAAATGTTTGGCATTTTTGGTGA 2580
Db      2521 CTCCTGCTCCCCTAGCACACATGATTATCTCTATTTGGAAATGTTTGGCATTTTTGGTGA 2580
Qy      2581 AGTGAATGGTTCAATAACTTTCTCCACCATCAGAACAAAAGCTCTTTAAGGTTAGGGATG 2640
Db      2581 AGTGAATGGTTCAATAACTTTCTCCACCATCAGAACAAAAGCTCTTTAAGGTTAGGGATG 2640
Qy      2641 GGATCATACACACTTCCCTTGTCCAAGTCCCCATCACCCCTTATCTAGACAATTGCTACA 2700
Db      2641 GGATCATACACACTTCCCTTGTCCAAGTCCCCATCACCCCTTATCTAGACAATTGCTACA 2700
Qy      2701 GTTTCCTACACACTCTTCTAACCTCTTGCAGTCTATTTTCATAAAACAGCTAGAGAACTT 2760
Db      2701 GTTTCCTACACACTCTTCTAACCTCTTGCAGTCTATTTTCATAAAACAGCTAGAGAACTT 2760
Qy      2761 TGAGATGTAAGTCAAAAAATAGAACATGTCGCTCTTTCCCATTGTTTTTGAAATAAAGTT 2820
Db      2761 TGAGATGTAAGTCAAAAAATAGAACATGTCGCTCTTTCCCATTGTTTTTGAAATAAAGTT 2820
Qy      2821 CAACCCCCTTACCAGGGTCAACAAGGCCCTGCAATGATTTGGTCCTGTTAAAAATTCTTT 2880
Db      2821 CAACCCCCTTACCAGGGTCAACAAGGCCCTGCAATGATTTGGTCCTGTTAAAAATTCTTT 2880
Qy      2881 AGCCTTAACTCATGCTGTTCTTCCTTACACTCACTGCATTCTAGCCATTGAGGTTTCTAT 2940
Db      2881 AGCCTTAACTCATGCTGTTCTTCCTTACACTCACTGCATTCTAGCCATTGAGGTTTCTAT 2940
Qy      2941 GCATCAAACTTTTTTTGGTCCCAGCACTGTGCACATCCTTCTGGGTAGAATGCCCCTTGA 3000
Db      2941 GCATCAAACTTTTTTTGGTCCCAGCACTGTGCACATCCTTCTGGGTAGAATGCCCCTTGA 3000
Qy      3001 TTTGTATAATTAGCACCTCCTTCATCATTTAGGTCTTAGTATAACTACTACCTTCTTAGA 3060
Db      3001 TTTGTATAATTAGCACCTCCTTCATCATTTAGGTCTTAGTATAACTACTACCTTCTTAGA 3060
Qy      3061 GAAGCTCTGCTTCTTCATCCTATAAAAAAGTAAAATTCCTTACCCTGTTATTTTTTAAGT 3120
Db      3061 GAAGCTCTGCTTCTTCATCCTATAAAAAAGTAAAATTCCTTACCCTGTTATTTTTTAAGT 3120
Qy      3121 CATCCGTGTTTCATTCTGTTAAAGTTCTTATCACAATTTATCATTATTTTATTTACAGTC 3180
Db      3121 CATCCGTGTTTCATTCTGTTAAAGTTCTTATCACAATTTATCATTATTTTATTTACAGTC 3180
Qy      3181 ATGTGCCACATAACAATGTTTCAGTCAGGGATAGAACACAAATGTATCTGGCCCCATAAT 3240
Db      3181 ATGTGCCACATAACAATGTTTCAGTCAGGGATAGAACACAAATGTATCTGGCCCCATAAT 3240
Qy      3241 ATTATAAGCTGAGAAATTTCTATTAACTAGTGATATCGCAGCCATCATAAGTGTAATGCA 3300
Db      3241 ATTATAAGCTGAGAAATTTCTATTAACTAGTGATATCGCAGCCATCATAAGTGTAATGCA 3300
Qy      3301 GGACATTACCTTTTCTATGTTTAGATATGTTAGATACACAAATATATTTCATTGTGTTAT 3360
Db      3301 GGACATTACCTTTTCTATGTTTAGATATGTTAGATACACAAATATATTTCATTGTGTTAT 3360
Qy      3361 AATTTCCTACAGTATTCAGTACAGTAACATGCTGTACAGGTTTGTAACCTAGGAGTAATA 3420
Db      3361 AATTTCCTACAGTATTCAGTACAGTAACATGCTGTACAGGTTTGTAACCTAGGAGTAATA 3420
Qy      3421 GGCTATACCATACAGCTTAGGTGTGTAGTAGGCTATAACCATCTAGGTTTGTGTAAGTAC 3480
Db      3421 GGCTATACCATACAGCTTAGGTGTGTAGTAGGCTATAACCATCTAGGTTTGTGTAAGTAC 3480
Qy      3481 ATTCTATGATATTCCCACAATGATGAAATCACCTAACTACACATTTCTCAGAATGTTTCA 3540
Db      3481 ATTCTATGATATTCCCACAATGATGAAATCACCTAACTACACATTTCTCAGAATGTTTCA 3540
Qy      3541 CTGTTGTGAAGTGACCCATGACTATATTTTCCTATATACTTGATATTTTTGTGCATCTGC 3600
Db      3541 CTGTTGTGAAGTGACCCATGACTATATTTTCCTATATACTTGATATTTTTGTGCATCTGC 3600
Qy      3601 CCATGAGAATGTAGTGTAAGATCAAAGGATGCAAGAATGGGTTCTATCCAGTATAGTACC 3660
Db      3601 CCATGAGAATGTAGTGTAAGATCAAAGGATGCAAGAATGGGTTCTATCCAGTATAGTACC 3660
Qy      3661 CACTACACTGGTGGATGTCAATATGTATTTGTTAGATTAATATCTCAAGAATGAGCACCT 3720
Db      3661 CACTACACTGGTGGATGTCAATATGTATTTGTTAGATTAATATCTCAAGAATGAGCACCT 3720
Qy      3721 TTCTCAGACACATAAAAGATGCTCAATATAAAAGTTTGTTGAACTGAACGTTATTGGCAA 3780
Db      3721 TTCTCAGACACATAAAAGATGCTCAATATAAAAGTTTGTTGAACTGAACGTTATTGGCAA 3780
Qy      3781 ATGTAACATGATCGGATTTAAAGAGGAGCGAAACAGAGGTCTGGCTCAAACACCATACTT 3840
Db      3781 ATGTAACATGATCGGATTTAAAGAGGAGCGAAACAGAGGTCTGGCTCAAACACCATACTT 3840
Qy      3841 CTAGAGTGCATAAGAGGTAGCAGTTGATTACCACTGGCGACAGGAGAAAAAAGAGCTTGA 3900
Db      3841 CTAGAGTGCATAAGAGGTAGCAGTTGATTACCACTGGCGACAGGAGAAAAAAGAGCTTGA 3900
Qy      3901 CCGCAGGGTACTGTGAAGACATTTCAGGTTGATGGCACAGAACAGGGGAAATACATAAAT 3960
Db      3901 CCGCAGGGTACTGTGAAGACATTTCAGGTTGATGGCACAGAACAGGGGAAATACATAAAT 3960
Qy      3961 GTGTGGGAATATTCAGTGGTCTGGGATGACTACATAGTAGAATATAATGAAGAAAAGAGT 4020
Db      3961 GTGTGGGAATATTCAGTGGTCTGGGATGACTACATAGTAGAATATAATGAAGAAAAGAGT 4020
Qy      4021 GGAAGGGAAAGATGAAAAGTTGGAATGGGGATGAATTATGAAAGTACCAGAATGTTATGC 4080
Db      4021 GGAAGGGAAAGATGAAAAGTTGGAATGGGGATGAATTATGAAAGTACCAGAATGTTATGC 4080
Qy      4081 TAAGGAATCTAGATTTTAAAATGTGAGGGCAAATTGAAGTC 4121
Db      4081 TAAGGAATCTAGATTTTAAAATGTGAGGGCAAATTGAAGTC 4121
```

RESULT 2
AL583850

```
LOCUS       AL583850              165329 bp    DNA     linear   PRI 15-NOV-2001
DEFINITION  Human DNA sequence from clone RP11-430G6 on chromosome 1, complete
            sequence.
ACCESSION   AL583850
VERSION     AL583850.5  GI:16973044
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 165329)
  AUTHORS   Tracey,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Nov 16, 2001 this sequence version replaced gi:15020514.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
```

chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr1
RP11-430G6 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
IMPORTANT: This sequence is not the entire insert of clone RP11-430G6 It may be shorter because we sequence overlapping sections only once, except for a short overlap.
The true right end of clone RP11-430G6 is at 165329 in this sequence. The true right end of clone RP11-331H2 is at 2000 in this sequence.

FEATURES Location/Qualifiers
source 1. .165329
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-430G6"
/clone_lib="RPCI-11.2"

ORIGIN

Query Match 99.7%; Score 4107.4; DB 9; Length 165329;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4119; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

```
Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAATTA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71683 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAATTA 71742

Qy         61 GACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAAT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71743 GACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAAT 71802

Qy        121 CACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71803 CACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAGC 71862

Qy        181 CCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71863 CCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATA 71922

Qy        241 TATATATATATACACACACACACATATATATATACACATATATATACACATATATACA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71923 TATATATATATACACACACACACATATATATATACACATATATATACACATATATACA 71982

Qy        301 TATATACACATATATACACATATATATACACATACATATGTACACATATATATACACATA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71983 TATATACACATATATACACATATATATACACATACATATGTACACATATATATACACATA 72042

Qy        361 TGTATACACATATATACACATATATACACACATATATACACATATATACACACATATATA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72043 TGTATACACATATATACACATATATACACACATATATACACATATATACACACATATATA 72102

Qy        421 CACATATATACACATATATACACATATACACATATATACACATATATACATATATACACA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72103 CACATATATACACATATATACACATATACACATATATACACATATATACATATATACACA 72162

Qy        481 TATATATAATATACACACATATATATACACATATATACACACATATATACACATATATAC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72163 TATATATAATATACACACATATATATACACATATATACACACATATATACACATATATAC 72222

Qy        541 ACATATATATACACATATATACACATATATACATATATACACATATATATACATATATAC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72223 ACATATATATACACATATATACACATATATACATATATACACATATATATACATATATAC 72282

Qy        601 ACATATATACATATATACACATATATACATATATACACACATATATACACATACATATAC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72283 ACATATATACATATATACACATATATACATATATACACACATATATACACATACATATAC 72342

Qy        661 ACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATAT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72343 ACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATAT 72402

Qy        721 GCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72403 GCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTT 72462

Qy        781 TGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72463 TGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATAC 72522

Qy        841 ATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATATATATA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72523 ATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATATATATA 72582

Qy        901 TAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTATAGAAGAATGG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72583 TAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTATAGAAGAATGG 72642

Qy        961 CTAAGATGCCAAAGCTCAGTGTATGTGGCAGAGGCATGGTGTAGGGTGTGTCCAGGTTCA 1020
              ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db      72643 CTAAGATGCCAAAGCTCAGTGTATGTGGCAGAGGCATGGTGGAGGGTGTGTCCAGGTTCA 72702

Qy       1021 TATATTGCATTAAGTGTGAGAACACCCTGGAGTATGAACCAAGAAAATGCAAAAGCCAGA 1080
              |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db      72703 TATATTGCATTAAGTGTGAGAA-ACCCTGGAGTATGAACCAAGAAAATGCAAAAGCCAGA 72761

Qy       1081 AGTGATGGAGGAAATGAGACACAATAATGAAGATATTGAGAGGAGGGTGTGGGCCTAGAG 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72762 AGTGATGGAGGAAATGAGACACAATAATGAAGATATTGAGAGGAGGGTGTGGGCCTAGAG 72821

Qy       1141 TGAAGCTTTTCGTGCCAGTACTTCTTTTGAAGGCCCAGTTCTCTTCTCTCTCGGGGGCTC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72822 TGAAGCTTTTCGTGCCAGTACTTCTTTTGAAGGCCCAGTTCTCTTCTCTCTCGGGGGCTC 72881

Qy       1201 CTTCATCTCTCATAGAGTCCACAGCTTTTAAGGGCCAACACTTGAGGTCAGCCTGGCTCT 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72882 CTTCATCTCTCATAGAGTCCACAGCTTTTAAGGGCCAACACTTGAGGTCAGCCTGGCTCT 72941

Qy       1261 CTCATTTGAGCTGGATAGAACATTTTAGAGCACCATCTATTCTTCAAGAGGAAGTTTAAA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72942 CTCATTTGAGCTGGATAGAACATTTTAGAGCACCATCTATTCTTCAAGAGGAAGTTTAAA 73001

Qy       1321 AATAAAAGAACCTTGAAGAGGAAAAAATGTAGACATTCAATCTAACCTTTTCATTTTACT 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73002 AATAAAAGAACCTTGAAGAGGAAAAAATGTAGACATTCAATCTAACCTTTTCATTTTACT 73061

Qy       1381 AGCCAAAGCTAAATAGAATGCAGGTTACCTGTTTTTCAGCCAGGCACCATCATTTCCTAA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73062 AGCCAAAGCTAAATAGAATGCAGGTTACCTGTTTTTCAGCCAGGCACCATCATTTCCTAA 73121

Qy       1441 TTGTTATAAAATTTATTATTATTGTTGTTATTATTATTATTTGCCATAAGAAGTTTCCCA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73122 TTGTTATAAAATTTATTATTATTGTTGTTATTATTATTATTTGCCATAAGAAGTTTCCCA 73181

Qy       1501 TATCCTTTTAGTATAACAAAAACACAATTCACAAGCATTATAAAACCCATGGTGTCTAAC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73182 TATCCTTTTAGTATAACAAAAACACAATTCACAAGCATTATAAAACCCATGGTGTCTAAC 73241

Qy       1561 TATTAAAAAAATTAAGTGGAACACACTTGTCCCAGCTACTGGGGAGGCTGAGGAGGGAGG 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73242 TATTAAAAAAATTAAGTGGAACACACTTGTCCCAGCTACTGGGGAGGCTGAGGAGGGAGG 73301
```

```
Qy       1621 ATCACGTGATCCCAGGGGGTCAAGGTTATGGAGAGCTATGATTGTGCCACTGCACTCCAG 1680
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73302 ATCACGTGATCCCAGGGGGTCAAGGTTATGGAGAGCTATGATTGTGCCACTGCACTCCAG 73361

Qy       1681 CCTGGGTGACAGGGAAAGACCCTGTCTCTAAAATTTTTTTTTAAAAAAACTAAACTGGTTT 1740
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73362 CCTGGGTGACAGGGAAAGACCCTGTCTCTAAAATTTTTTTTTAAAAAAACTAAACTGGTTT 73421

Qy       1741 TATTACAGAGATTCTGGAGACAGCTACACATAAAAGGGTGGTATGCCTCATATTAGCTAC 1800
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73422 TATTACAGAGATTCTGGAGACAGCTACACATAAAAGGGTGGTATGCCTCATATTAGCTAC 73481

Qy       1801 CCAGGGAGGTGGAATGCCAACTTAGGTGGTGTCACCACTATTAAAAATGCCCCAAAGCAA 1860
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73482 CCAGGGAGGTGGAATGCCAACTTAGGTGGTGTCACCACTATTAAAAATGCCCCAAAGCAA 73541

Qy       1861 TCAAAACTGAGAACTTCCTGGGAGCTTAGCATTGTGCAAAAGCAGCACAAAACACTTAAA 1920
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73542 TCAAAACTGAGAACTTCCTGGGAGCTTAGCATTGTGCAAAAGCAGCACAAAACACTTAAA 73601

Qy       1921 CAATTCACAGTTGTGTTGGAATGGGAAGGCCTGGAAATATAAACCAAAGAGTATATTGTC 1980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73602 CAATTCACAGTTGTGTTGGAATGGGAAGGCCTGGAAATATAAACCAAAGAGTATATTGTC 73661

Qy       1981 TAAATTGATAGAGATTACAATTGCCTGAAAGAAAAAGTTGACTTTTAACTAGAATGTTCA 2040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73662 TAAATTGATAGAGATTACAATTGCCTGAAAGAAAAAGTTGACTTTTAACTAGAATGTTCA 73721

Qy       2041 GAGTAGGTTTACAGAAGAAGCTCTTAAACTGGGCTCCAGTGGATTTGTCAATGCTTTGGA 2100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73722 GAGTAGGTTTACAGAAGAAGCTCTTAAACTGGGCTCCAGTGGATTTGTCAATGCTTTGGA 73781

Qy       2101 AGCTGGTGGGGTGGGAGGGTTGGAGGGGGCATAAAAAGTCATGTTGGTATGCTCTGCTCA 2160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73782 AGCTGGTGGGGTGGGAGGGTTGGAGGGGGCATAAAAAGTCATGTTGGTATGCTCTGCTCA 73841

Qy       2161 AGTCTCCATTCTGTTTCCTTTTCCTCTTTTCAATGTCATGTCCCATTATTTCATTATGGG 2220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73842 AGTCTCCATTCTGTTTCCTTTTCCTCTTTTCAATGTCATGTCCCATTATTTCATTATGGG 73901

Qy       2221 CTTCCCTTTATCCAGGATCAATATGCCACCTCTTGGTTGTCTTTTACCTACTTCTCCACC 2280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73902 CTTCCCTTTATCCAGGATCAATATGCCACCTCTTGGTTGTCTTTTACCTACTTCTCCACC 73961

Qy       2281 TCACTATGGAATCGTCCTTGGGTAGCTCCTGTGCTTGGGAACCTGCACGGGCACTTTTCT 2340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      73962 TCACTATGGAATCGTCCTTGGGTAGCTCCTGTGCTTGGGAACCTGCACGGGCACTTTTCT 74021

Qy       2341 GATGTCTTGATTCCAGCTTTACTCCTAAAACTTAAATGCTGAGGGGCCAACACCATGGCA 2400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74022 GATGTCTTGATTCCAGCTTTACTCCTAAAACTTAAATGCTGAGGGGCCAACACCATGGCA 74081

Qy       2401 GTGGTAGGGATGGGAATGGGGGTCTTGTAACACACTACATAAACTACACGAAATAAACTA 2460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74082 GTGGTAGGGATGGGAATGGGGGTCTTGTAACACACTACATAAACTACACGAAATAAACTA 74141

Qy       2461 CATGAAACTCAACATGTTTGCAAGACTCAGTTCACATCCATGAGGAGCTCATGCTTCTCC 2520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74142 CATGAAACTCAACATGTTTGCAAGACTCAGTTCACATCCATGAGGAGCTCATGCTTCTCC 74201

Qy       2521 CTCCTGCTCCCCTAGCACACATGATTATCTCTATTTGGAAATGTTTGGCATTTTTGGTGA 2580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74202 CTCCTGCTCCCCTAGCACACATGATTATCTCTATTTGGAAATGTTTGGCATTTTTGGTGA 74261

Qy       2581 AGTGAATGGTTCAATAACTTTCTCCACCATCAGAACAAAAGCTCTTTAAGGTTAGGGATG 2640
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74262 AGTGAATGGTTCAATAACTTTCTCCACCATCAGAACAAAAGCTCTTTAAGGTTAGGGATG 74321

Qy       2641 GGATCATACACACTTCCCTTGTCCAAGTCCCCATCACCCCTTATCTAGACAATTGCTACA 2700
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74322 GGATCATACACACTTCCCTTGTCCAAGTCCCCATCACCCCTTATCTAGACAATTGCTACA 74381

Qy       2701 GTTTCCTACACACTCTTCTAACCTCTTGCAGTCTATTTTCATAAAACAGCTAGAGAACTT 2760
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74382 GTTTCCTACACACTCTTCTAACCTCTTGCAGTCTATTTTCATAAAACAGCTAGAGAACTT 74441

Qy       2761 TGAGATGTAAGTCAAAAAATAGAACATGTCGCTCTTTCCCATTGTTTTTGAAATAAAGTT 2820
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74442 TGAGATGTAAGTCAAAAAATAGAACATGTCGCTCTTTCCCATTGTTTTTGAAATAAAGTT 74501

Qy       2821 CAACCCCCTTACCAGGGTCAACAAGGCCCTGCAATGATTTGGTCCTGTTAAAAATTCTTT 2880
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74502 CAACCCCCTTACCAGGGTCAACAAGGCCCTGCAATGATTTGGTCCTGTTAAAAATTCTTT 74561

Qy       2881 AGCCTTAACTCATGCTGTTCTTCCTTACACTCACTGCATTCTAGCCATTGAGGTTTCTAT 2940
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74562 AGCCTTAACTCATGCTGTTCTTCCTTACACTCACTGCATTCTAGCCATTGAGGTTTCTAT 74621

Qy       2941 GCATCAAACTTTTTTTGGTCCCAGCACTGTGCACATCCTTCTGGGTAGAATGCCCCTTGA 3000
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74622 GCATCAAACTTTTTTTGGTCCCAGCACTGTGCACATCCTTCTGGGTAGAATGCCCCTTGA 74681

Qy       3001 TTTGTATAATTAGCACCTCCTTCATCATTTAGGTCTTAGTATAACTACTACCTTCTTAGA 3060
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74682 TTTGTATAATTAGCACCTCCTTCATCATTTAGGTCTTAGTATAACTACTACCTTCTTAGA 74741

Qy       3061 GAAGCTCTGCTTCTTCATCCTATAAAAAAGTAAAATTCCTTACCCTGTTATTTTTTAAGT 3120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74742 GAAGCTCTGCTTCTTCATCCTATAAAAAAGTAAAATTCCTTACCCTGTTATTTTTTAAGT 74801

Qy       3121 CATCCGTGTTTCATTCTGTTAAAGTTCTTATCACAATTTATCATTATTTTATTTACAGTC 3180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74802 CATCCGTGTTTCATTCTGTTAAAGTTCTTATCACAATTTATCATTATTTTATTTACAGTC 74861

Qy       3181 ATGTGCCACATAACAATGTTTCAGTCAGGGATAGAACACAAATGTATCTGGCCCCATAAT 3240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74862 ATGTGCCACATAACAATGTTTCAGTCAGGGATAGAACACAAATGTATCTGGCCCCATAAT 74921

Qy       3241 ATTATAAGCTGAGAAATTTCTATTAACTAGTGATATCGCAGCCATCATAAGTGTAATGCA 3300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74922 ATTATAAGCTGAGAAATTTCTATTAACTAGTGATATCGCAGCCATCATAAGTGTAATGCA 74981

Qy       3301 GGACATTACCTTTTCTATGTTTAGATATGTTAGATACACAAATATATTTCATTGTGTTAT 3360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      74982 GGACATTACCTTTTCTATGTTTAGATATGTTAGATACACAAATATATTTCATTGTGTTAT 75041

Qy       3361 AATTTCCTACAGTATTCAGTACAGTAACATGCTGTACAGGTTTGTAACCTAGGAGTAATA 3420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75042 AATTTCCTACAGTATTCAGTACAGTAACATGCTGTACAGGTTTGTAACCTAGGAGTAATA 75101

Qy       3421 GGCTATACCATACAGCTTAGGTGTGTAGTAGGCTATAACCATCTAGGTTTGTGTAAGTAC 3480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75102 GGCTATACCATACAGCTTAGGTGTGTAGTAGGCTATAACCATCTAGGTTTGTGTAAGTAC 75161

Qy       3481 ATTCTATGATATTCCCACAATGATGAAATCACCTAACTACACATTTCTCAGAATGTTTCA 3540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75162 ATTCTATGATATTCCCACAATGATGAAATCACCTAACTACACATTTCTCAGAATGTTTCA 75221

Qy       3541 CTGTTGTGAAGTGACCCATGACTATATTTTCCTATATACTTGATATTTTTGTGCATCTGC 3600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75222 CTGTTGTGAAGTGACCCATGACTATATTTTCCTATATACTTGATATTTTTGTGCATCTGC 75281

Qy       3601 CCATGAGAATGTAGTGTAAGATCAAAGGATGCAAGAATGGGTTCTATCCAGTATAGTACC 3660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75282 CCATGAGAATGTAGTGTAAGATCAAAGGATGCAAGAATGGGTTCTATCCAGTATAGTACC 75341

Qy       3661 CACTACACTGGTGGATGTCAATATGTATTTGTTAGATTAATATCTCAAGAATGAGCACCT 3720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75342 CACTACACTGGTGGATGTCAATATGTATTTGTTAGATTAATATCTCAAGAATGAGCACCT 75401

Qy       3721 TTCTCAGACACATAAAAGATGCTCAATATAAAAGTTTGTTGAACTGAACGTTATTGGCAA 3780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75402 TTCTCAGACACATAAAAGATGCTCAATATAAAAGTTTGTTGAACTGAACGTTATTGGCAA 75461

Qy       3781 ATGTAACATGATCGGATTTAAAGAGGAGCGAAACAGAGGTCTGGCTCAAACACCATACTT 3840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      75462 ATGTAACATGATCGGATTTAAAGAGGAGCGAAACAGAGGTCTGGCTCAAACACCATACTT 75521

Qy       3841 CTAGAGTGCATAAGAGGTAGCAGTTGATTACCACTGGCGACAGGAGAAAAAAGAGCTTGA 3900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75522 CTAGAGTGCATAAGAGGTAGCAGTTGATTACCACTGGCGACAGGAGAAAAAAGAGCTTGA 75581

Qy       3901 CCGCAGGGTACTGTGAAGACATTTCAGGTTGATGGCACAGAACAGGGGAAATACATAAAT 3960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75582 CCGCAGGGTACTGTGAAGACATTTCAGGTTGATGGCACAGAACAGGGGAAATACATAAAT 75641

Qy       3961 GTGTGGGAATATTCAGTGGTCTGGGATGACTACATAGTAGAATATAATGAAGAAAAGAGT 4020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75642 GTGTGGGAATATTCAGTGGTCTGGGATGACTACATAGTAGAATATAATGAAGAAAAGAGT 75701

Qy       4021 GGAAGGGAAAGATGAAAAGTTGGAATGGGGATGAATTATGAAAGTACCAGAATGTTATGC 4080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      75702 GGAAGGGAAAGATGAAAAGTTGGAATGGGGATGAATTATGAAAGTACCAGAATGTTATGC 75761

Qy       4081 TAAGGAATCTAGATTTTAAAATGTGAGGGCAAATTGAAGTC 4121
              |||||||||||||||||||||||||||||||||||||||||
Db      75762 TAAGGAATCTAGATTTTAAAATGTGAGGGCAAATTGAAGTC 75802


RESULT 3
AC031977/c
LOCUS       AC031977              191699 bp    DNA     linear   HTG 12-APR-2001
DEFINITION  Homo sapiens chromosome 1 clone RP11-288O18, WORKING DRAFT
            SEQUENCE, 3 unordered pieces.
ACCESSION   AC031977
VERSION     AC031977.7  GI:13194952
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 191699)
  AUTHORS   Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
            Federspiel,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,
            Mao,J., Komp,C., Kottler,S., Lam,B., Marathe,R., Miranda,M.,
            Morehouse,A.J., Nguyen,M., Oefner,P., Palm,C.J., Ramirez,D.,
            Southwick,A.M., Webb,C., Wilhelmy,J., Yu,S. and Davis,R.W.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 191699)
  AUTHORS   Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
            Federspiel,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,
            Mao,J., Marathe,R., Morehouse,A.J., Oefner,P., Palm,C.J.,
            Ramirez,D., Wilhelmy,J., Yu,S. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-APR-2000) DNA Sequencing and Technology Center,
            Stanford University, 855 California Avenue, Palo Alto, CA 94304,
            USA
COMMENT     On Mar 4, 2001 this sequence version replaced gi:9665085.
            -------------- Genome Center
                 Center: Stanford DNA Sequencing and Technology Development
            Center
                 Center code: SDSTDC
                 Web site: http://sequence-www.stanford.edu/group/human/
                 Contact: hum-info@sequence.stanford.edu
            -------------- Project Information
                 Center project name: 880
                 Center clone name: RP11-288O18
            -------------- Summary Statistics
                 Sequencing Vector: M13mp18; X02513; 100% of reads
                 Sequencing Vector: plasmid; plasmid_accession; 0% of reads
                 Chemistry: Dye-primer; 1% of reads
                 Chemistry: Dye-terminator Big Dye; 99% of reads
                 Assembly program: Phrap; version 0.990319
                 Consensus quality: 190680 bases at least Q40
                 Consensus quality: 191287 bases at least Q30
                 Consensus quality: 191336 bases at least Q20
                 Insert size: 195548; agarose-fp
                 Insert size: 191499; sum-of-contigs
                 Quality coverage:  7.9x in Q20 bases; agarose-fp
                 Quality coverage:  8.1x in Q20 bases; sum-of-contigs.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    12646: contig of 12646 bp in length
            *    12647    12746: gap of unknown length
            *    12747    94961: contig of 82215 bp in length
            *    94962    95061: gap of unknown length
            *    95062   191699: contig of 96638 bp in length.
FEATURES             Location/Qualifiers
     source          1. .191699
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-288O18"
                     /clone_lib="RPCI human BAC library 11"
     misc_feature    1. .12646
                     /note="assembly_name:Contig5"
     misc_feature    12747. .94961
                     /note="assembly_name:Contig6
                     clone_end:T7"
     misc_feature    95062. .191699
                     /note="assembly_name:Contig7
                     clone_end:SP6"
ORIGIN

  Query Match             90.3%;  Score 3720;  DB 2;  Length 191699;
  Best Local Similarity   96.3%;  Pred. No. 0;
  Matches 3793;  Conservative    0;  Mismatches  130;  Indels   15;  Gaps    4;

Qy        194 GCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATATAC 253
              |||
Db      12749 GCGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 12690

Qy        254 ACACACACACACATATATATATACACATATATATACACATATATACATATATACACATAT 313
                                                        |||| ||||| | | |
Db      12689 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTACACATATATATACAC 12630

Qy        314 ATACACATATATATACACATACATATGTACACATATATATACACATATGTATACACATAT 373
              |||   ||| | ||| | | |||||| |||||| |||||||||||||| || ||| ||||
Db      12629 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACATATAT 12570

Qy        374 ATACACATATATACACACATATATACACATATATACACACATATATACACATATATACAC 433
              | ||||||||||||||||||||||||||||||    ||||||||||||||||||||||||
Db      12569 ACACACATATATACACACATATATACACATAT----ACACATATATACACATATATACAC 12514

Qy        434 ATATATACACATATACACATATATACACATATATACATATATACACATATATATAATATA 493
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db      12513 ATATATACACATATATACATATATACACATATATACATATATACACATATATATAATATA 12454

Qy        494 CACACATATATATACACATATATACACACATATATACACATATATACACATAT------A 547
              ||||||||||||||||||||||||||||||||||||||||||||||||||||      |
Db      12453 CACACATATATATACACATATATACACACATATATACACATATATACACATATATACACA 12394

Qy        548 TATACACATATATACACATATATACATATATACACATATATATACATATATACACATATA 607
              |||| |||||||||||||||||| | ||||| ||| ||||||||||||||||||||||||
Db      12393 TATATACATATATACACATATATATACATATATACACATATATACATATATACACATATA 12334

Qy        608 TACATATATACACATATATACATATATACACACATATATACACATACATATACACACACA 667
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12333 TACATATATACACATATATACATATATACACACATATATACACATACATATACACACACA 12274

Qy        668 TAGATATACATATATATACACATATATATACGTATATATATGTATATATATATGCTCCAG 727
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db      12273 TAGATATACATATATATACACATATATATGTATATATATGTATATATATATGCTCCAG 12214

Qy        728 AGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGC 787
```

```
Db      12213 AGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGC 12154
Qy        788 AGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATG 847
Db      12153 AGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATG 12094
Qy        848 TGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACAC----TATATATATATATAG 903
Db      12093 TGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATATATATATATATAG 12034
Qy        904 TGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTATAGAAGAATGGCTA 963
Db      12033 TGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTATAGAAGAATGGCTA 11974
Qy        964 AGATGCCAAAGCTCAGTGTATGTGGCAGAGGCATGGTGTAGGGTGTGTCCAGGTTCATAT 1023
Db      11973 AGATGCCAAAGCTCAGTGTATGTGGCAGAGGCATGGTGGAGGGTGTGTCCAGGTTCATAT 11914
Qy       1024 ATTGCATTAAGTGTGAGAACACCCTGGAGTATGAACCAAGAAAATGCAAAAGCCAGAAGT 1083
Db      11913 ATTGCATTAAGTGTGAGAA-ACCCTGGAGTATGAACCAAGAAAATGCAAAAGCCAGAAGT 11855
Qy       1084 GATGGAGGAAATGAGACACAATAATGAAGATATTGAGAGGAGGGTGTGGGCCTAGAGTGA 1143
Db      11854 GATGGAGGAAATGAGACACAATAATGAAGATATTGAGAGGAGGGTGTGGGCCTAGAGTGA 11795
Qy       1144 AGCTTTTCGTGCCAGTACTTCTTTTGAAGGCCCAGTTCTCTTCTCTCTCGGGGGCTCCTT 1203
Db      11794 AGCTTTTCGTGCCAGTACTTCTTTTGAAGGCCCAGTTCTCTTCTCTCTCGGGGGCTCCTT 11735
Qy       1204 CATCTCTCATAGAGTCCACAGCTTTTAAGGGCCAACACTTGAGGTCAGCCTGGCTCTCTC 1263
Db      11734 CATCTCTCATAGAGTCCACAGCTTTTAAGGGCCAACACTTGAGGTCAGCCTGGCTCTCTC 11675
Qy       1264 ATTTGAGCTGGATAGAACATTTTAGAGCACCATCTATTCTTCAAGAGGAAGTTTAAAAAT 1323
Db      11674 ATTTGAGCTGGATAGAACATTTTAGAGCACCATCTATTCTTCAAGAGGAAGTTTAAAAAT 11615
Qy       1324 AAAAGAACCTTGAAGAGGAAAAAATGTAGACATTCAATCTAACCTTTTCATTTTACTAGC 1383
Db      11614 AAAAGAACCTTGAAGAGGAAAAAATGTAGACATTCAATCTAACCTTTTCATTTTACTAGC 11555
Qy       1384 CAAAGCTAAATAGAATGCAGGTTACCTGTTTTTCAGCCAGGCACCATCATTTCCTAATTG 1443
Db      11554 CAAAGCTAAATAGAATGCAGGTTACCTGTTTTTCAGCCAGGCACCATCATTTCCTAATTG 11495
Qy       1444 TTATAAAATTTATTATTATTGTTGTTATTATTATTATTTGCCATAAGAAGTTTCCCATAT 1503
Db      11494 TTATAAAATTTATTATTATTGTTGTTATTATTATTATTTGCCATAAGAAGTTTCCCATAT 11435
Qy       1504 CCTTTTAGTATAACAAAAACACAATTCACAAGCATTATAAAACCCATGGTGTCTAACTAT 1563
Db      11434 CCTTTTAGTATAACAAAAACACAATTCACAAGCATTATAAAACCCATGGTGTCTAACTAT 11375
Qy       1564 TAAAAAAATTAAGTGGAACACACTTGTCCCAGCTACTGGGGAGGCTGAGGAGGGAGGATC 1623
Db      11374 TAAAAAAATTAAGTGGAACACACTTGTCCCAGCTACTGGGGAGGCCGAGGAGGGAGGATC 11315
Qy       1624 ACGTGATCCCAGGGGGTCAAGGTTATGGAGAGCTATGATTGTGCCACTGCACTCCAGCCT 1683
Db      11314 ACGTGATCCCAGGGGGTCAAGGTTATGGAGAGCTATGATTGTGCCACTGCACTCCAGCCT 11255
Qy       1684 GGGTGACAGGGAAAGACCCTGTCTCTAAAATTTTTTTTAAAAAAACTAAACTGGTTTTAT 1743
Db      11254 GGGTGACAGGGAAAGACCCTGTCTCTAAAATTTTTTTTAAAAAAACTAAACTGGTTTTAT 11195
Qy       1744 TACAGAGATTCTGGAGACAGCTACACATAAAAGGGTGGTATGCCTCATATTAGCTACCCA 1803
Db      11194 TACAGAGATTCTGGAGACAGCTACACATAAAAGGGTGGTATGCCTCATATTAGCTACCCA 11135
Qy       1804 GGGAGGTGGAATGCCAACTTAGGTGGTGTCACCACTATTAAAAATGCCCCAAAGCAATCA 1863
Db      11134 GGGAGGTGGAATGCCAACTTAGGTGGTGTCACCACTATTAAAAATGCCCCAAAGCAATCA 11075
Qy       1864 AAACTGAGAACTTCCTGGGAGCTTAGCATTGTGCAAAAGCAGCACAAAACACTTAAACAA 1923
Db      11074 AAACTGAGAACTTCCTGGGAGCTTAGCATTGTGCAAAAGCAGCACAAAACACTTAAACAA 11015
Qy       1924 TTCACAGTTGTGTTGGAATGGGAAGGCCTGGAAATATAAACCAAAGAGTATATTGTCTAA 1983
Db      11014 TTCACAGTTGTGTTGGAATGGGAAGGCCTGGAAATATAAACCAAAGAGTATATTGTCTAA 10955
Qy       1984 ATTGATAGAGATTACAATTGCCTGAAAGAAAAAGTTGACTTTTAACTAGAATGTTCAGAG 2043
Db      10954 ATTGATAGAGATTACAATTGCCTGAAAGAAAAAGTTGACTTTTAACTAGAATGTTCAGAG 10895
Qy       2044 TAGGTTTACAGAAGAAGCTCTTAAACTGGGCTCCAGTGGATTTGTCAATGCTTTGGAAGC 2103
Db      10894 TAGGTTTACAGAAGAAGCTCTTAAACTGGGCTCCAGTGGATTTGTCAATGCTTTGGAAGC 10835
Qy       2104 TGGTGGGGTGGGAGGGTTGGAGGGGGCATAAAAAGTCATGTTGGTATGCTCTGCTCAAGT 2163
Db      10834 TGGTGGGGTGGGAGGGTTGGAGGGGGCATAAAAAGTCATGTTGGTATGCTCTGCTCAAGT 10775
Qy       2164 CTCCATTCTGTTTCCTTTTCCTCTTTTCAATGTCATGTCCCATTATTTCATTATGGGCTT 2223
Db      10774 CTCCATTCTGTTTCCTTTTCCTCTTTTCAATGTCATGTCCCATTATTTCATTATGGGCTT 10715
Qy       2224 CCCTTTATCCAGGATCAATATGCCACCTCTTGGTTGTCTTTTACCTACTTCTCCACCTCA 2283
Db      10714 CCCTTTATCCAGGATCAATATGCCACCTCTTGGTTGTCTTTTACCTACTTCTCCACCTCA 10655
Qy       2284 CTATGGAATCGTCCTTGGGTAGCTCCTGTGCTTGGGAACCTGCACGGGCACTTTTCTGAT 2343
Db      10654 CTATGGAATCGTCCTTGGGTAGCTCCTGTGCTTGGGAACCTGCACGGGCACTTTTCTGAT 10595
Qy       2344 GTCTTGATTCCAGCTTTACTCCTAAAACTTAAATGCTGAGGGGCCAACACCATGGCAGTG 2403
Db      10594 GTCTTGATTCCAGCTTTACTCCTAAAACTTAAATGCTGAGGGGCCAACACCATGGCAGTG 10535
Qy       2404 GTAGGGATGGGAATGGGGGTCTTGTAACACACTACATAAACTACACGAAATAAACTACAT 2463
Db      10534 GTAGGGATGGGAATGGGGGTCTTGTAACACACTACATAAACTACACGAAATAAACTACAT 10475
Qy       2464 GAAACTCAACATGTTTGCAAGACTCAGTTCACATCCATGAGGAGCTCATGCTTCTCCCTC 2523
Db      10474 GAAACTCAACATGTTTGCAAGACTCAGTTCACATCCATGAGGAGCTCATGCTTCTCCCTC 10415
Qy       2524 CTGCTCCCCTAGCACACATGATTATCTCTATTTGGAAATGTTTGGCATTTTTGGTGAAGT 2583
Db      10414 CTGCTCCCCTAGCACACATGATTATCTCTATTTGGAAATGTTTGGCATTTTTGGTGAAGT 10355
Qy       2584 GAATGGTTCAATAACTTTCTCCACCATCAGAACAAAAGCTCTTTAAGGTTAGGGATGGGA 2643
Db      10354 GAATGGTTCAATAACTTTCTCCACCATCAGAACAAAAGCTCTTTAAGGTTAGGGATGGGA 10295
Qy       2644 TCATACACACTTCCCTTGTCCAAGTCCCCATCACCCCTTATCTAGACAATTGCTACAGTT 2703
Db      10294 TCATACACACTTCCCTTGTCCAAGTCCCCATCACCCCTTATCTAGACAATTGCTACAGTT 10235
Qy       2704 TCCTACACACTCTTCTAACCTCTTGCAGTCTATTTTCATAAAACAGCTAGAGAACTTTGA 2763
Db      10234 TCCTACACACTCTTCTAACCTCTTGCAGTCTATTTTCATAAAACAGCTAGAGAACTTTGA 10175
Qy       2764 GATGTAAGTCAAAAAATAGAACATGTCGCTCTTTCCCATTGTTTTTGAAATAAAGTTCAA 2823
Db      10174 GATGTAAGTCAAAAAATAGAACATGTCGCTCTTTCCCATTGTTTTTGAAATAAAGTTCAA 10115
Qy       2824 CCCCCTTACCAGGGTCAACAAGGCCCTGCAATGATTTGGTCCTGTTAAAAATTCTTTAGC 2883
Db      10114 CCCCCTTACCAGGGTCAACAAGGCCCTGCAATGATCTGGTCCTGTTAAAAATTCTTTAGC 10055
Qy       2884 CTTAACTCATGCTGTTCTTCCTTACACTCACTGCATTCTAGCCATTGAGGTTTCTATGCA 2943
Db      10054 CTTAACTCATGCTGTTTTTCCTTACACTCACTGCATTCTAGCCATTGAGGTTTCTATGCA 9995
```

```
Qy       2944 TCAAACTTTTTTTGGTCCCAGCACTGTGCACATCCTTCTGGGTAGAATGCCCCTTGATTT 3003
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9994 TCAAACTTTTTTTGGTCCCAGCACTGTGCACATCCTTCTGGGTAGAATGCCCCTTGATTT 9935

Qy       3004 GTATAATTAGCACCTCCTTCATCATTTAGGTCTTAGTATAACTACTACCTTCTTAGAGAA 3063
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9934 GTATAATTAGCACCTCCTTCATCATTTAGGTCTTAGTATAACTACTACCTTCTTAGAGAA 9875

Qy       3064 GCTCTGCTTCTTCATCCTATAAAAAAGTAAAATTCCTTACCCTGTTATTTTTTAAGTCAT 3123
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9874 GCTCTGCTTCTTCATCCTATAAAAAAGTAAAATTCCTTACCCTGTTATTTTTTAAGTCAT 9815

Qy       3124 CCGTGTTTCATTCTGTTAAAGTTCTTATCACAATTTATCATTATTTTATTTACAGTCATG 3183
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9814 CCGTGTTTCATTCTGTTAAAGTTCTTATCACAATTTATCATTATTTTATTTACAGTCATG 9755

Qy       3184 TGCCACATAACAATGTTTCAGTCAGGGATAGAACACAAATGTATCTGGCCCCATAATATT 3243
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9754 TGCCACATAACAATGTTTCAGTCAGGGATAGAACACAAATGTATCTGGCCCCATAATATT 9695

Qy       3244 ATAAGCTGAGAAATTTCTATTAACTAGTGATATCGCAGCCATCATAAGTGTAATGCAGGA 3303
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9694 ATAAGCTGAGAAATTTCTATTAACTAGTGATATCGCAGCCATCATAAGTGTAATGCAGGA 9635

Qy       3304 CATTACCTTTTCTATGTTTAGATATGTTAGATACACAAATATATTTCATTGTGTTATAAT 3363
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9634 CATTACCTTTTCTATGTTTAGATATGTTAGATACACAAATATATTTCATTGTGTTATAAT 9575

Qy       3364 TTCCTACAGTATTCAGTACAGTAACATGCTGTACAGGTTTGTAACCTAGGAGTAATAGGC 3423
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9574 TTCCTACAGTATTCAGTACAGTAACATGCTGTACAGGTTTGTAACCTAGGAGTAATAGGC 9515

Qy       3424 TATACCATACAGCTTAGGTGTGTAGTAGGCTATAACCATCTAGGTTTGTGTAAGTACATT 3483
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9514 TATACCATACAGCTTAGGTGTGTAGTAGGCTATAACCATCTAGGTTTGTGTAAGTACATT 9455

Qy       3484 CTATGATATTCCCACAATGATGAAATCACCTAACTACACATTTCTCAGAATGTTTCACTG 3543
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9454 CTATGATATTCCCACAATGATGAAATCACCTAACTACACATTTCTCAGAATGTTTCACTG 9395

Qy       3544 TTGTGAAGTGACCCATGACTATATTTTCCTATATACTTGATATTTTTGTGCATCTGCCCA 3603
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9394 TTGTGAAGTGACCCATGACTATATTTTCCTATATACTTGATATTTTTGTGCATCTGCCCA 9335

Qy       3604 TGAGAATGTAGTGTAAGATCAAAGGATGCAAGAATGGGTTCTATCCAGTATAGTACCCAC 3663
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9334 TGAGAATGTAGTGTAAGATCAAAGGATGCAAGAATGGGTTCTATCCAGTATAGTACCCAC 9275

Qy       3664 TACACTGGTGGATGTCAATATGTATTTGTTAGATTAATATCTCAAGAATGAGCACCTTTC 3723
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9274 TACACTGGTGGATGTCAATATGTATTTGTTAGATTAATATCTCAAGAATGAGCACCTTTC 9215

Qy       3724 TCAGACACATAAAAGATGCTCAATATAAAAGTTTGTTGAACTGAACGTTATTGGCAAATG 3783
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9214 TCAGACACATAAAAGATGCTCAATATAAAAGTTTGTTGAACTGAACGTTATTGGCAAATG 9155

Qy       3784 TAACATGATCGGATTTAAAGAGGAGCGAAACAGAGGTCTGGCTCAAACACCATACTTCTA 3843
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9154 TAACATGATCGGATTTAAAGAGGAGCGAAACAGAGGTCTGGCTCAAACACCATACTTCTA 9095

Qy       3844 GAGTGCATAAGAGGTAGCAGTTGATTACCACTGGCGACAGGAGAAAAAAGAGCTTGACCG 3903
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9094 GAGTGCATAAGAGGTAGCAGTTGATTACCACTGGCGACAGGAGAAAAAAGAGCTTGACCG 9035

Qy       3904 CAGGGTACTGTGAAGACATTTCAGGTTGATGGCACAGAACAGGGGAAATACATAAATGTG 3963
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9034 CAGGGTACTGTGAAGACATTTCAGGTTGATGGCACAGAACAGGGGAAATACATAAATGTG 8975

Qy       3964 TGGGAATATTCAGTGGTCTGGGATGACTACATAGTAGAATATAATGAAGAAAAGAGTGGA 4023
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8974 TGGGAATATTCAGTGGTCTGGGATGACTACATAGTAGAATATAATGAAGAAAAGAGTGGA 8915

Qy       4024 AGGGAAAGATGAAAAGTTGGAATGGGGATGAATTATGAAAGTACCAGAATGTTATGCTAA 4083
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8914 AGGGAAAGATGAAAAGTTGGAATGGGGATGAATTATGAAAGTACCAGAATGTTATGCTAA 8855

Qy       4084 GGAATCTAGATTTTAAAATGTGAGGGCAAATTGAAGTC 4121
             |||||||||||||||||||||||||||||||||||||||
Db       8854 GGAATCTAGATTTTAAAATGTGAGGGCAAATTGAAGTC 8817
```

RESULT 4
AC097328
LOCUS AC097328 203216 bp DNA linear HTG 20-OCT-2001
DEFINITION Pan troglodytes clone RP43-174K14, WORKING DRAFT SEQUENCE, 21 unordered pieces.
ACCESSION AC097328
VERSION AC097328.1 GI:16118151
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE 1 (bases 1 to 203216)
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G. and Gibbs,R.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 203216)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (14-OCT-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT ---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/

```
    Contact: hgsc-help@bcm.tmc.edu
-------------- Project Information
    Center project name: ZUAM
    Center clone name: RP43-174K14
-------------- Summary Statistics
    Sequencing vector: Plasmid; M77789
    Chemistry: Dye-terminator Big Dye: 100% of reads
    Assembly program: Phrap; version 0.990329
    Consensus quality: 196812 bases at least Q40
    Consensus quality: 208422 bases at least Q30
    Consensus quality: 218080 bases at least Q20
    Estimated insert size: 228347; sum-of-contigs estimation
    Quality coverage: 0x in Q20 bases; agarose-fp estimation
    Quality coverage: 7.1x in Q20 bases; sum-of-contigs estimation
--------------
* NOTE: Estimated insert size may differ from sequence length
*   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*       1    30770: contig of 30770 bp in length
*   30771    30870: gap of unknown length
*   30871    52356: contig of 21486 bp in length
*   52357    52456: gap of unknown length
*   52457    72697: contig of 20241 bp in length
*   72698    72797: gap of unknown length
*   72798    90396: contig of 17599 bp in length
*   90397    90496: gap of unknown length
*   90497    96134: contig of 5638 bp in length
*   96135    96234: gap of unknown length
*   96235   114433: contig of 18199 bp in length
*  114434   114533: gap of unknown length
*  114534   125184: contig of 10651 bp in length
*  125185   125284: gap of unknown length
*  125285   135194: contig of 9910 bp in length
*  135195   135294: gap of unknown length
*  135295   140273: contig of 4979 bp in length
*  140274   140373: gap of unknown length
*  140374   148490: contig of 8117 bp in length
*  148491   148590: gap of unknown length
*  148591   153435: contig of 4845 bp in length
*  153436   153535: gap of unknown length
*  153536   162229: contig of 8694 bp in length
*  162230   162329: gap of unknown length
*  162330   167843: contig of 5514 bp in length
*  167844   167943: gap of unknown length
*  167944   175297: contig of 7354 bp in length
*  175298   175397: gap of unknown length
*  175398   179409: contig of 4012 bp in length
*  179410   179509: gap of unknown length
*  179510   185931: contig of 6422 bp in length
*  185932   186031: gap of unknown length
*  186032   190690: contig of 4659 bp in length
*  190691   190790: gap of unknown length
*  190791   194808: contig of 4018 bp in length
*  194809   194908: gap of unknown length
*  194909   198788: contig of 3880 bp in length
*  198789   198888: gap of unknown length
*  198889   201055: contig of 2167 bp in length
*  201056   201155: gap of unknown length
*  201156   203216: contig of 2061 bp in length.
FEATURES             Location/Qualifiers
     source          1. .203216
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="RP43-174K14"
ORIGIN
```

```
  Query Match           10.0%;  Score 411.2;  DB 2;  Length 203216;
  Best Local Similarity  76.7%;  Pred. No. 5e-71;
  Matches  554;  Conservative    0;  Mismatches  163;  Indels    5;  Gaps    4;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCA-TCTCTACTAAAAATACAAAATT 59
              ||||||| |||||||||| |||||||||||||||||  |||||||||||| |||||||||
Db     114617 AGTTCAAAACCAGCCTGACCAACATGGTGAAACCCCCGTCTCTACTAAAAACACAAAATT 114676

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   |||||||| | || ||||||| ||||||||||   ||| ||  ||||||||||
Db     114677 AGCTGGGCATGGTGGTGCATGCCTGTAGTCCCAGCTACTCAGGAAGCTAAGGCAGGAGAA 114736

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGC-AAGATCATGCCATTGCACTCCA 178
              ||||||||||  | || || |||||||| | || |  ||||||||||| |||||||||
Db     114737 TCACTTGAACC--CAGGAGGCGGAGGTTGCAGTGAACTGAGATCATGCCACTGCACTCCA 114794

Qy        179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATA 238
              |||   | ||||| |||         |  |    | | |    |       | | ||||||||
Db     114795 GCCTGAGTAACAAAAGCATCTCAAAAAATATATATATATGTATATATGTGTGTATATATA 114854

Qy        239 TATATATATATATACACACACACACATATATATATACACATATATATACACATATATA 298
              | | |||||||||      | | | |  | ||||||||    | | |||| |  | ||||
Db     114855 TGTGTATATATATGTGTGTATATATATGTGTATATATATGTGTGTGTATATATGTGTATA 114914

Qy        299 CATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACA 358
              |||||||  | |||||||  | |||||||||  | |||  ||||  ||  | |||||||||  | |
Db     114915 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 114974

Qy        359 TATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATA 418
              |||  |||| | |||||||  | |||||||  | | |||||||  | |||||||  | |  |||||
Db     114975 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 115034

Qy        419 TACACATATATACACATATATACACATATACACATATATACACATATATACATATATACA 478
              ||  | |||||||  | |||||||  | |||||  | |||||||  | |||||||  |||||||  |
Db     115035 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 115094

Qy        479 CATATATATAATATACACACATATATATACACATATATACACACATATATACACATATAT 538
              ||||||||  |||||  | |  |||||||||  | |||||||  | |  |||||||  | ||||||
Db     115095 TATATATAT-ATATATATATATATATATATATATATATATATATATATATATATATAT 115153

Qy        539 ACACATATATATACACATATATACACATATATACATATATACACATATATATACATATAT 598
              |  | |||||||||  | |||||||  |||||||  |||||||  | |||||||||  ||||||
Db     115154 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 115213

Qy        599 ACACATATATACATATATACACATATATACATATATACACACATATATACACATACATAT 658
              | | |||||||  |||||||  | |||||||  |||||||  | |  |||||||  | |||  ||||
Db     115214 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 115273

Qy        659 ACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATAT 718
              | | |   |||  |||||  |||||||||  | |||||||||   |||||||||  |||||||||
Db     115274 ATATATCTATATATATATATATATATATATATATATATATATATATATATATATATAT 115333

Qy        719 AT 720
              ||
Db     115334 AT 115335
```

```
RESULT 5
CR759943
LOCUS       CR759943              104321 bp    DNA     linear   HTG 01-OCT-2004
DEFINITION  Homo sapiens chromosome 6 clone DAAP-21A6, WORKING DRAFT SEQUENCE.
ACCESSION   CR759943
VERSION     CR759943.2  GI:53748680
KEYWORDS    HTG; HTGS_PHASE2; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 104321)
  AUTHORS   Phillimore,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
```

```
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 1, 2004 this sequence version replaced gi:51965768.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bAP21A6
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 103081 bases at least Q40
            Consensus quality: 103313 bases at least Q30
            Consensus quality: 103455 bases at least Q20
            Insert size: 104321; sum-of-contigs
            Insert size: 110232; 0.4% error; agarose-fp
            Quality coverage: 7.35x in Q20 bases; sum-of-contigs Quality
            coverage: 7.39x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   104321: contig of 104321 bp in length.
FEATURES            Location/Qualifiers
     source         1. .104321
                    /organism="Homo sapiens"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:9606"
                    /chromosome="6"
                    /clone="DAAP-21A6"
                    /clone_lib="APD"
     misc_feature   1. .104321
                    /note="assembly_fragment:01274"
ORIGIN

  Query Match             9.8%;  Score 405.4;  DB 2;  Length 104321;
  Best Local Similarity   75.3%;  Pred. No. 7.5e-70;
  Matches  558;  Conservative    0;  Mismatches  176;  Indels    7;  Gaps    4;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC---AAAA 57
              ||||| ||||||||||||  || ||||||||||||||||||||||||||||||   ||||
Db      96311 AGTTCGAGACCAGCCTGGCCAGCATGGTGAAACCCCATCTCTACTAAAAATACAAAAAAA 96370

Qy         58 TTAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAG 117
              |||| ||||  ||||     |  | || || ||||| |||| | |||||| | ||||||| |
Db      96371 TTAGCCAGGTGTGGTAGCGCGTGTCTCTAGTCCCACCTACCTGGGAGGCTGTGGCAGGCG 96430

Qy        118 AATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGC-AAGATCATGCCATTGCACTC 176
              |||  ||| ||||  |  | || |||||||||| | ||||  |||||| |||| |||||||
Db      96431 AATTGCTTCAACC--CGTGAGGCGGAGGTTGCAGTGAGCTGAGATCACGCCACTGCACTC 96488

Qy        177 CAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATA 236
              |||||  |||||| ||||| | |     ||||| |||||||| |       | ||||||||
Db      96489 CAGCCTGGGCAAC-AGAGCAAGAATCAGTCTCAAAAAAAAAATATATATATATATATATA 96547

Qy        237 TATATATATATATATACACACACACACACATATATATATACACATATATATACACATATA 296
              |||||||||||||||| | | | | | | |||||||||||| | |||||||| | |||||
Db      96548 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 96607

Qy        297 TACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACA 356
              || ||||||| | ||||||| | ||||||||| | ||| |||| || | |||||||| |
Db      96608 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 96667

Qy        357 CATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATA 416
              |||| |||| | ||||||| | ||||||| | | ||||||| | ||||||| | | |||

Db      96668 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 96727

Qy        417 TATACACATATATACACATATATACACATATACACATATATACACATATATACATATATA 476
              |||| | |||||||| | ||||||| | ||||| | ||||||| | ||||||| ||| |||
Db      96728 TATATATATATATATATATATATATATATATATATATATATATATATATATATAAATA 96787

Qy        477 CACATATATATAATATACACACATATATATACACATATATACACACATATATACACATAT 536
              | ||||| |||  |              ||| | ||||||| | | ||||||| | ||||
Db      96788 TAAATATAAATATAAATATATATATATAATATATATATATATATATATATATATATAT 96847

Qy        537 ATACACATATATATACACATATATACACATATATACATATATACACATATATATACATAT 596
              ||| | |||||||||| | ||||||| | ||||||| ||||||| | |||||||||| ||||
Db      96848 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 96907

Qy        597 ATACACATATATACATATATACACATATATACATATATACACACATATATACACATACAT 656
              ||| | ||||||| ||||||| | ||||||| ||||||| | | ||||||| | ||| ||
Db      96908 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 96967

Qy        657 ATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATAT 716
              ||| | | | ||| ||||| |||| |||| | |||| | ||  ||  | ||||| |||||
Db      96968 ATATATATATATATATATATATATGTATATATATATGTGTATATACGTGTATGTGTATAT 97027

Qy        717 ATATGCTCCAGAGTTCATAAG 737
              |||       | ||  ||| |
Db      97028 ATACATGTATGTGTATATATG 97048


RESULT 6
AC097330/c
LOCUS       AC097330              181946 bp    DNA     linear   HTG 20-OCT-2001
DEFINITION  Pan troglodytes clone RP43-54O15, WORKING DRAFT SEQUENCE, 16
            unordered pieces.
ACCESSION   AC097330
VERSION     AC097330.1  GI:16118153
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 181946)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
```

```
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 181946)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-OCT-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: ZUAP
                Center clone name: RP43-54O15
            -------------- Summary Statistics
                Sequencing vector: Plasmid; M77789
                Sequencing vector: M13; L08821
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 169928 bases at least Q40
                Consensus quality: 176100 bases at least Q30
                Consensus quality: 180968 bases at least Q20
                Estimated insert size: 183838; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 6.9x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 16 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    17715: contig of 17715 bp in length
            *   17716    17815: gap of unknown length
            *   17816    44117: contig of 26302 bp in length
            *   44118    44217: gap of unknown length
            *   44218    67185: contig of 22968 bp in length
            *   67186    67285: gap of unknown length
            *   67286    84740: contig of 17455 bp in length
            *   84741    84840: gap of unknown length
            *   84841   100199: contig of 15359 bp in length
            *  100200   100299: gap of unknown length
            *  100300   112190: contig of 11891 bp in length
            *  112191   112290: gap of unknown length
            *  112291   124548: contig of 12258 bp in length
            *  124549   124648: gap of unknown length
            *  124649   134134: contig of 9486 bp in length
            *  134135   134234: gap of unknown length
            *  134235   143821: contig of 9587 bp in length
            *  143822   143921: gap of unknown length
            *  143922   152987: contig of 9066 bp in length
            *  152988   153087: gap of unknown length
            *  153088   159063: contig of 5976 bp in length
            *  159064   159163: gap of unknown length
            *  159164   166547: contig of 7384 bp in length
            *  166548   166647: gap of unknown length
            *  166648   171663: contig of 5016 bp in length
            *  171664   171763: gap of unknown length
            *  171764   177471: contig of 5708 bp in length
            *  177472   177571: gap of unknown length
            *  177572   179822: contig of 2251 bp in length
            *  179823   179922: gap of unknown length
            *  179923   181946: contig of 2024 bp in length.
FEATURES             Location/Qualifiers
     source          1. .181946
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="RP43-54O15"
ORIGIN

  Query Match             9.8%;  Score 405.2;  DB 2;  Length 181946;
  Best Local Similarity   68.3%;  Pred. No. 7.6e-70;
  Matches  634;  Conservative    0;  Mismatches  288;  Indels    6;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||| |||||| ||||  | | ||||||||||||||| |||||||| |||||| ||||||
Db     155131 AGTTCGAGACCAACCTGGACTATATGGTGAAACCCCGTCTCTACTAGAAATACAAAAATT 155072

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   |||||||||  ||||||||||||||||| |||||| ||||||  ||||||||||
Db     155071 AGCTGGGCATGGTGGCACACGCCTGTAATCCCAACTACTTGGGAGGCTAAGGCAGGAGAA 155012

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCCA 178
              ||||||||||    | || ||  |||||||| | ||| | |||||   ||| ||||||||||
Db     155011 TCACTTGAAC--TCGGGAGGCAGAGGTTGCAGTGAGCCCAGATCGCACCACTGCACTCCA 154954

Qy        179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATA 238
              ||| |||| |      |   || |||     | |||| |||| |     |      |  | ||
Db     154953 GCCTAGGCGAGTGAGACTCCATCTCAAAAAAAAAAAGAAAAGAAAATTGAGGGCTGGACA 154894

Qy        239 TATATATATATATACACACACACACATATATATATACACATATATATACACATATATA 298
              |     || |     | | ||    | |  ||||| | ||||||||| | |||||||
Db     154893 CAGTGGCTTACACCTGTAATCCCAGCACTTTGGATATATATATATATATATATATATA 154834

Qy        299 CATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACA 358
              ||||||| | ||||||| | ||||||||| | ||| |||| || | ||||||||| | |
Db     154833 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 154774

Qy        359 TATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATA 418
              ||| |||| | ||||||| | ||||||| | | ||||||| | ||||||| | | |||||
Db     154773 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 154714

Qy        419 TACACATATATACACATATATACACATATACACATATATACACATATATACATATATACA 478
              || | ||||||| | ||||||| | ||||| | ||||||| | ||||||| ||||||| |
Db     154713 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 154654

Qy        479 CATATATATA-ATATACACACATATATATACACATATATACACACATATATACACATATA 537
              |||||||||| ||||| | | ||||||||| | ||||||| | | ||||||| | |||||
Db     154653 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 154594

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATACATATA 597
              || | |||||||||| | ||||||| | ||||||| ||||||| | |||||||| |||||
Db     154593 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 154534

Qy        598 TACACATATATACATATATACACATATATACATATATACACACATATATACACATACATA 657
              || | ||||||| ||||||| | ||||||| ||||||| | | ||||||| | ||| |||
Db     154533 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 154474

Qy        658 TACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATA 717
              || | | | ||| ||||| |||||||||  | |||||||||  ||||||||| ||||||||
Db     154473 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 154414

Qy        718 TATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGA 777
              |||          | |||          |  |   ||      |    ||| |   | | | |
Db     154413 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 154354

Qy        778 GTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAA 837
              |  | |   | |||    | |   | || |      |     | |||      | |
Db     154353 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 154294

Qy        838 TACATAAATGTGTAATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTAACACTATATATA 896
              || ||| || | ||   | |  ||| | || |  ||  ||    |  ||    ||||||||
Db     154293 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 154234
```

```
Qy        897 TATATAGTGAAAGCAGTGTAAACCTAAA 924
              ||||||   | |    | || |  || |
Db     154233 TATATATATATATATATATATATATATA 154206


RESULT 7
AL358074
LOCUS       AL358074              109081 bp    DNA     linear   PRI 21-NOV-2001
DEFINITION  Human DNA sequence from clone RP11-423C15 on chromosome 9, complete
            sequence.
ACCESSION   AL358074
VERSION     AL358074.22  GI:17048358
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 109081)
  AUTHORS   Hammond,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Nov 22, 2001 this sequence version replaced gi:17043779.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr9
            RP11-423C15 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-423C15 It may be shorter because we sequence overlapping
            sections only once, except for a short overlap.
            The true right end of clone RP11-423C15 is at 109081 in this
            sequence. The true left end of clone RP11-770B5 is at 78953 in this
            sequence. The true right end of clone RP11-12A16 is at 2000 in this
            sequence.
FEATURES             Location/Qualifiers
     source          1. .109081
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /clone="RP11-423C15"
                     /clone_lib="RPCI-11.2"
     misc_feature    complement(65197. .65223)
                     /note="Single clone region. Sequence from reads from a
                     short insert library derived from a single pUC clone.
                     Restriction digest data confirm the assembly."
     misc_feature    65533. .65757
                     /note="Single clone region. Sequence from reads from a
                     short insert library derived from a single pUC clone.
                     Restriction digest data confirm the assembly."
     misc_feature    98714. .98750
                     /note="Single clone region. Sequence from reads from a
                     short insert library derived from a single pUC clone.
                     Restriction digest data confirm the assembly."
     misc_feature    98834. .98842
                     /note="Single clone region. Sequence from reads from a
                     short insert library derived from a single pUC clone.
                     Restriction digest data confirm the assembly."
     misc_feature    99111. .99161
                     /note="Single clone region. Sequence from reads from a
                     short insert library derived from a single pUC clone.
                     Restriction digest data confirm the assembly"
     misc_feature    99163. .99191
                     /note="Single clone region. Sequence from reads from a
                     short insert library derived from a single pUC clone.
                     Restriction digest data confirm the assembly"
ORIGIN

  Query Match             9.8%;  Score 402.2;  DB 9;  Length 109081;
  Best Local Similarity   78.2%;  Pred. No. 3.2e-69;
  Matches  564;  Conservative    0;  Mismatches  138;  Indels   19;  Gaps    6;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC--AAAAT 58
              ||||  ||||||||||  ||||||||||||||||| ||||||||||||||  ||||
Db      11620 AGTTTGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAAAAAC 11679

Qy         59 TAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGA 118
              ||| || || |||||  || | ||||||||||||||||||  ||  || ||||||||||
Db      11680 TAGCCAAGCGTGGTGGCACGCACCTGTAATCCCAGCTACTCGGGCAGCTGAGGCAGGAGA 11739

Qy        119 ATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCC 177
              ||| ||||||||  |  | ||  ||||||||| | ||| ||||||||||||||||||||
Db      11740 ATCGCTTGAACC--CAAGAGGCAGAGGTTGCAGTGAGCCAAGATCATGCCATTGCACTCC 11797

Qy        178 AGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATAT 237
              ||| || | ||||||||||||||  |  |||| ||||||||||          ||||||||
Db      11798 AGCTCAAGTAACAAGAGCGAAATTCCGCCTCAAAAAAAAAAAA---------ATATATAT 11848

Qy        238 ATATATATATATATACACACACACACATATATATATACACATATATATACACATATAT 297
              |||||||| | | |||||||| |||  |||||||| ||||| ||||| ||| |||||
Db      11849 ATATATATACACACACACACACATACATGTATATATACACACACATATACACATATATAC 11908

Qy        298 ACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACAC 357
              ||| ||||| ||| ||| ||||| |||||||||||||| | |   ||| ||||| ||| ||
Db      11909 ACACATATATACACATACACACACATATATACACATATACACACACATATATACATATAC 11968

Qy        358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417
              | |   ||| | | ||| ||||| ||||| |||| |||| ||||| | ||||||| ||||
Db      11969 ACACACATATATACATACACACACATATATACACGTATACACACACACATACACATATAT 12028

Qy        418 ATACACATATATACACATATATACACATATACACATATATACACATATATACATATATAC 477
              | |||||||||||||||||||| ||||| ||| |||  |||||| | | | ||||||| ||
Db      12029 ACACACATATATACACATATACACACACATATACACGTATACATACACACACATATACAC 12088

Qy        478 ACATATATATA-ATATACACACATATATATACACATATATACACACATATATACACATAT 536
                |||||||  |  |||| | |||  |||||| ||| ||||| |||| ||||||||||||||
Db      12089 GTATATATACACGTATATATACACGTATATATACACATATATACACGTATATACACATAT 12148

Qy        537 ATACACATATATATACACATATATACACATATATACATATATACAC----ATATATATAC 592
              |||||| |||||||||||||||||||| |||||||| ||||| ||    ||||||||||
Db      12149 ATACACGTATATATACACATATATACACGTATATACACATATATACACGTATATATATAC 12208

Qy        593 ATATATACACATATATACATATATACACATATATACATATATACACACATATATACACAT 652
              | | ||| | ||| | ||| |||||||||  | | ||||||||| ||||| ||| |||||
Db      12209 ACACATATATATACACACACATATACACACACACATATATATACACACATACACACAC 12268

Qy        653 ACATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTAT 712
              ||| | |||||||| | ||||| |||||  |   |     |||   |   | |  ||||
Db      12269 ACACACACACACACACACATATATATATAGTTGGCCGGGCATAGGGGCTCACACCTGTAA 12328

Qy        713 A 713
              |
```

```
Db      12329 A 12329


RESULT 8
AC009058
LOCUS       AC009058              164844 bp    DNA     linear   PRI 13-MAR-2003
DEFINITION  Homo sapiens chromosome 16 clone RP11-292B23, complete sequence.
ACCESSION   AC009058
VERSION     AC009058.10  GI:28933531
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 164844)
  AUTHORS   DOE Joint Genome Institute, Stanford Human Genome Center and Los
            Alamos National Laboratory.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 164844)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 164844)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-DEC-2001) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   4  (bases 1 to 164844)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-2001) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   5  (bases 1 to 164844)
  AUTHORS   DOE Joint Genome Institute, Stanford Human Genome Center and Los
            Alamos National Laboratory.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAR-2003) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Mar 13, 2003 this sequence version replaced gi:17432456.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center and Los Alamos
            National Laboratory
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 100% of Sequence;
            Estimated Total Number of Errors is 0.
FEATURES             Location/Qualifiers
     source          1. .164844
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone="RP11-292B23"
ORIGIN

  Query Match             9.6%;  Score 396.8;  DB 9;  Length 164844;
  Best Local Similarity   75.4%;  Pred. No. 3.4e-68;
  Matches  559;  Conservative    0;  Mismatches  173;  Indels    9;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAATTA 60
              ||||| | |||| ||||  |||||| ||||||||||||||||||||||||    |||||||
Db     140343 AGTTCGAAACCACCCTGGCCAACATAGTGAAACCCCATCTCTACTAAAATACAAAATTA 140402

Qy         61 GACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAAT 120
              ||||| |||||  | ||||||||||||||||||||  |||||| |||||||||| ||
Db     140403 GCCAGGCGTGGTGGCAGACGCCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGAGGAT 140462

Qy        121 CACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCCAG 179
               ||||||||  | || ||  |||||||| | ||| | ||||  ||||| || |||||||
Db     140463 TGCTTGAACC--CAGGAGGCAGAGGTTGCAGTGAGCCCAGATTGTGCCACTGGACTCCAG 140520

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  ||  |||      | |       |  | | | | |        ||||||||
Db     140521 CCTGGGAGACAGAGTAAGACTCCGTCTCAAATACATATATACATACACACATATATATAC 140580

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACAC----ATAT 295
              | ||||| ||| ||||| | | | ||| ||||| ||||||| ||| |||||||    ||||
Db     140581 ACATATACATACACACATATATATACACATATACATACACACATACATACACACATATAT 140640

Qy        296 ATACATATATACACATATATACACATATATATACACATACATATGTACACATATATAT-A 354
              | ||| ||||||| ||||||||||||||||||||| ||||| || || ||| |||||| |
Db     140641 ACACACATATACATATATATACACATATATATACATATACACATATATACACATATATAA 140700

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              | | |   ||| | ||| ||| | | | ||| ||||||||||||| ||||| ||||| |
Db     140701 TATACACACATATATATACATATATACACATATACACATATATACATATATACACACATA 140760

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              |||| |||||||||||||||||||||| ||| |||||||||||||| ||| | |||||||
Db     140761 TATACACACATATATACACATATATACATATACACACATATATACATATACACACATATA 140820

Qy        475 TACACATATATATAATATACACACATATATATACACATATATACACACATATATACACAT 534
              ||||||| | |  | |||||||||| | ||| ||||||| | ||||| | | | ||||||
Db     140821 CACACATACACA-CACATACACACACACATACACACATACACACACATACACACACACAT 140879

Qy        535 ATATACACATATATATACACATATATACACATATATACATATATACACATATATATACAT 594
              | | | ||| | ||| ||||| |||||||| ||| |  || | | ||| | |||||||
Db     140880 ACACATACACACATACACACACATATACACACATACATGTACACATACACACATATACAC 140939

Qy        595 ATATACACATATATACATATATACACATATATACATATATACACACATATATACACATAC 654
              |||||||| |||||||||||||||| |||||||||||||||||||||||||| |||||
Db     140940 ATATACACACATATACATATATACACACATATACATATATACACACATATATATACATAT 140999

Qy        655 ATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATAT 714
              ||| ||| | | | | ||||||| |||| |||| ||||| |||  || | | |  |||||
Db     141000 ATACACATATATACATATATACACATATGTACATATATACATATATACACACACATATAT 141059

Qy        715 ATATATGCTCCAGAGTTCATA 735
              | |||  |    | ||||
Db     141060 GTGTATATACACACATACATA 141080


RESULT 9
AC011593
LOCUS       AC011593              130349 bp    DNA     linear   PRI 07-NOV-2001
DEFINITION  Homo sapiens BAC clone RP11-437E9 from 2, complete sequence.
ACCESSION   AC011593
VERSION     AC011593.8  GI:14190772
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 130349)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
   PUBMED   9847074
REFERENCE   2  (bases 1 to 130349)
  AUTHORS   Pearman,C. and Cotton,M.
  TITLE     The sequence of Homo sapiens BAC clone RP11-437E9
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 130349)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 130349)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-MAY-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
```

```
            MO 63108, USA
REFERENCE   5  (bases 1 to 130349)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-NOV-2001) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On May 23, 2001 this sequence version replaced gi:13518261.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0437E09
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. John D.
            McPherson, Department of Genetics, Washington University, St. Louis
            MO.  For additional information about the map position of this
            sequence, see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries.  Genomics 51:1-8.  The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at the Roswell Park Cancer Institute
            (http://bacpac.med.buffalo.edu)
            VECTOR:  pBACe3.6
            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is RP11-289J14, 200 bp overlap; the
            clone sequenced to the right is AC061984.  Actual start of this
            clone is at base position 150577 of RP11-289J14; actual end is at
            base position 130349 of RP11-437E9.
FEATURES                Location/Qualifiers
     source             1. .130349
                        /organism="Homo sapiens"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9606"
                        /chromosome="2"
                        /map="2"
                        /clone="RP11-437E9"
                        /clone_lib="RPCI-11"
     repeat_region      1813. .2297
                        /rpt_family="MaLR"
     repeat_region      2618. .3136
                        /rpt_family="MaLR"
     repeat_region      3200. .3370
                        /rpt_family="MaLR"
     repeat_region      3522. .3856
                        /rpt_family="Alu"
     repeat_region      3866. .4290
                        /rpt_family="MaLR"
     repeat_region      4325. .4805
                        /rpt_family="ERV1"
     repeat_region      4910. .5014
                        /rpt_family="MER1_type"
     repeat_region      5258. .5600
                        /rpt_family="L1"
     repeat_region      5601. .5772
                        /rpt_family="Alu"
     repeat_region      5773. .6644
                        /rpt_family="L1"
     repeat_region      6773. .7356
                        /rpt_family="L1"
     repeat_region      7389. .7516
                        /rpt_family="L1"
     repeat_region      7529. .8188
                        /rpt_family="L1"
     repeat_region      8189. .8490
                        /rpt_family="Alu"
     repeat_region      8491. .8590
                        /rpt_family="L1"
     repeat_region      8605. .8943
                        /rpt_family="L1"
     repeat_region      9017. .9042
                        /rpt_family="(CA)n"
     repeat_region      9062. .9104
                        /rpt_family="MER53"
     repeat_region      9105. .9394
                        /rpt_family="Alu"
     repeat_region      9395. .9508
                        /rpt_family="MER53"
     repeat_region      9786. .9966
                        /rpt_family="C-rich"
     repeat_region      9946. .10036
                        /rpt_family="(TCC)n"
     repeat_region      10080. .10374
                        /rpt_family="Alu"
     repeat_region      12491. .12521
                        /rpt_family="(CA)n"
     repeat_region      13138. .13321
                        /rpt_family="MIR"
     repeat_region      13414. .13442
                        /rpt_family="(TAA)n"
     repeat_region      13615. .13925
                        /rpt_family="Alu"
     repeat_region      16183. .16266
                        /rpt_family="MIR"
     repeat_region      16629. .16997
                        /rpt_family="MaLR"
     repeat_region      17643. .17821
                        /rpt_family="Tc2"
     repeat_region      18157. .18200
                        /rpt_family="(TG)n"
     repeat_region      19591. .19619
                        /rpt_family="AT_rich"
     repeat_region      21204. .21375
                        /rpt_family="MaLR"
     repeat_region      21376. .21940
                        /rpt_family="ERVL"
     repeat_region      21962. .22901
                        /rpt_family="L1"
     repeat_region      22909. .23443
                        /rpt_family="L1"
     repeat_region      23492. .23987
                        /rpt_family="L1"
     repeat_region      23988. .24016
                        /rpt_family="(TTTTG)n"
     repeat_region      24017. .24634
                        /rpt_family="L1"
     repeat_region      24636. .25726
                        /rpt_family="L1"
     repeat_region      25727. .26040
                        /rpt_family="Alu"
     repeat_region      26041. .27537
                        /rpt_family="L1"
     repeat_region      27607. .28123
                        /rpt_family="ERVL"
     repeat_region      28282. .28808
```

```
                     /rpt_family="ERVL"
     repeat_region   28809. .29192
                     /rpt_family="MaLR"
     repeat_region   29193. .29577
                     /rpt_family="ERVL"
     repeat_region   29628. .29895
                     /rpt_family="Alu"
     repeat_region   29960. .30264
                     /rpt_family="ERVL"
     repeat_region   30265. .30573
                     /rpt_family="Alu"
     repeat_region   30574. .30753
                     /rpt_family="ERVL"
     repeat_region   30774. .32056
                     /rpt_family="ERVL"
     repeat_region   32100. .32242
                     /rpt_family="ERVL"
     repeat_region   32477. .32750
                     /rpt_family="MaLR"
     repeat_region   32751. .32847
                     /rpt_family="L1"
     repeat_region   32848. .33164
                     /rpt_family="Alu"
     repeat_region   33165. .33302

  Query Match             9.6%;  Score 395.2;  DB 9;  Length 130349;
  Best Local Similarity   67.8%;  Pred. No. 7.2e-68;
  Matches  645;  Conservative    0;  Mismatches  293;  Indels   14;  Gaps    6;

Qy          4 TCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC--AAAATTAG 61
              ||||||||| ||||   |||||||||| |||||||||||||||||||||  ||||||||
Db      39934 TCAAGACCATCCTGGCTAACATGGTGAATCCCCATCTCTACTAAAAATACAAAAAATTAG 39993

Qy         62 ACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAATC 121
              ||||| |||||      ||||||| ||||||||||  |||||| ||||||||||||||
Db      39994 CCAGGCGTGGTGGCGGGTGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATG 40053

Qy        122 ACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAGCC 181
              | |||||||  | || |||||||  ||||   |||| |||||   ||||||||||||||||
Db      40054 GCGTGAACC--CGGGAGGTGGAGCTTGCAATGAGCCAGATCGCACCATTGCACTCCAGCC 40111

Qy        182 CAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTT---TATATATATA 238
              | ||||  |||| | |    |||||| | ||| | | | |    || |   || |||||||
Db      40112 TGGACAAC-GGAGCAAGACTCCATCTAAAAAATATATATAAATATATACATAAATATATA 40170

Qy        239 TATATATATATATACACACACACACATATATATATACACATATATATACACATATATA 298
              |||||| ||||| ||| | | | | | ||||| ||||| ||||| ||||| |   ||| |
Db      40171 TATATAAATATAGACATAAATATATATATATAAATATAGACATAAATATATATGCATAAA 40230

Qy        299 CATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACA 358
              |||||||  |||| ||| | ||| | ||| | ||| |||| |||| | ||||||| | |
Db      40231 TATATATATGCATAAATATATATAAAAATATATATAAATATATACATAAATATATATAAA 40290

Qy        359 TATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATA 418
              ||| || | | ||||||| | ||||||| | | ||||||| | |||||||||| |||||
Db      40291 TATATACAAAAATATATATAAATATATAAAAAAATATATAAATATATATACACATATATA 40350

Qy        419 TACACATATATACACATATATACACATATACACATATATACACATATATACATATATACA 478
              | | ||||| | ||||||||| ||||||| ||||| ||| | || |||| ||||||| |
Db      40351 AATATATATACATACATATATAAACATATATACATAAATATATATGTATAAATATATATA 40410

Qy        479 CATATATATA---ATATACACACATATATATACACATATATACACACATATATACACATA 535
              || ||| |||    |||  | | ||||| ||| | ||||||  | | ||||||| |||||
Db      40411 CACATAAATATATGTATGAATATATATACATAAATATATATGTATAAATATATATACATA 40470

Qy        536 TATACACATATATATACACATATATACACATATATACATATATACACATATATATACATA 595
              ||| | | | ||||| ||||| ||| | ||| ||| ||||| | | | |||||||||| |||
Db      40471 AATATATAAAGATATATACATAAATATATATAAATATATATACATAAATATATATAAATA 40530

Qy        596 TATACACATATATACATATATACACATATATACATATATACACACATATATACACATACA 655
              |||| | ||| ||| ||| ||| | ||||| | ||||| | | | ||||| | | ||| |
Db      40531 TATATAAATAGATATATAAATATATATATAAATATATAAATATATATATAAATATATAAA 40590
```

```
Qy        656 TATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATA 715
              |||| | | | | | ||||| ||||||||| | ||||||| |  ||||||||  ||||||
Db      40591 TATATAAAAATAGATATATAAATATATATATAAATATATAAATATATATATAAATATATA 40650

Qy        716 TATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGA 775
              || ||     |   |  ||| |.      |     ||  | |    ||| |   | | |
Db      40651 TAAATATATAAATATATATATAAATATATATAAATATATAAATATATATAAATATATAAA 40710

Qy        776 GAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGA 835
              |  |  | |    | |||    | |   | || |      |      |||||       |
Db      40711 TATATATATAAATATATATAAATATATAAATATATATAAATATATATAAATATATAAATA 40770

Qy        836 AATACATAAATGTGTAATAAAACCTATCTGTA---AGGTAGTTAAGAAGGTAACACTATA 892
              ||| |||||| | ||  || |  ||  | ||   |  |  ||  ||  ||  | ||||
Db      40771 TATATATAAATATATATAAATATATAAATATATATATAAATATATAAATATATAAATATA 40830

Qy        893 TATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAA 944
              |||||| |||   |||    | ||||  || |  |      ||  || || |
Db      40831 TATATAAATATATAAATATATATAAATATATATAAATATATAAATATATATA 40882
```

```
RESULT 10
AC097269/c
LOCUS       AC097269               196413 bp    DNA     linear   HTG 20-OCT-2001
DEFINITION  Pan troglodytes clone RP43-11J15, WORKING DRAFT SEQUENCE, 18
            unordered pieces.
ACCESSION   AC097269
VERSION     AC097269.1  GI:16117532
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 196413)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
```

```
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 196413)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-OCT-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     -------------- Genome Center
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
            Center project name: ZUAW
            Center clone name: RP43-11J15
            -------------- Summary Statistics
            Sequencing vector: Plasmid; M77789
            Chemistry: Dye-terminator Big Dye: 100% of reads
            Assembly program: Phrap; version 0.990329
            Consensus quality: 190674 bases at least Q40
            Consensus quality: 194102 bases at least Q30
            Consensus quality: 196834 bases at least Q20
            Estimated insert size: 194298; sum-of-contigs estimation
            Quality coverage: 0x in Q20 bases; agarose-fp estimation
            Quality coverage: 8.1x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 18 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    17120: contig of 17120 bp in length
            *    17121    17220: gap of unknown length
            *    17221    36306: contig of 19086 bp in length
            *    36307    36406: gap of unknown length
            *    36407    58958: contig of 22552 bp in length
            *    58959    59058: gap of unknown length
            *    59059    82373: contig of 23315 bp in length
            *    82374    82473: gap of unknown length
            *    82474    96867: contig of 14394 bp in length
            *    96868    96967: gap of unknown length
            *    96968   109418: contig of 12451 bp in length
            *   109419   109518: gap of unknown length
            *   109519   120755: contig of 11237 bp in length
            *   120756   120855: gap of unknown length
            *   120856   134377: contig of 13522 bp in length
            *   134378   134477: gap of unknown length
            *   134478   144432: contig of 9955 bp in length
            *   144433   144532: gap of unknown length
            *   144533   155013: contig of 10481 bp in length
            *   155014   155113: gap of unknown length
            *   155114   161933: contig of 6820 bp in length
            *   161934   162033: gap of unknown length
            *   162034   168622: contig of 6589 bp in length
            *   168623   168722: gap of unknown length
            *   168723   175724: contig of 7002 bp in length
            *   175725   175824: gap of unknown length
            *   175825   180213: contig of 4389 bp in length
            *   180214   180313: gap of unknown length
            *   180314   186640: contig of 6327 bp in length
            *   186641   186740: gap of unknown length
            *   186741   190642: contig of 3902 bp in length
            *   190643   190742: gap of unknown length
            *   190743   194248: contig of 3506 bp in length
            *   194249   194348: gap of unknown length
            *   194349   196413: contig of 2065 bp in length.
FEATURES            Location/Qualifiers
     source         1. .196413
                    /organism="Pan troglodytes"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:9598"
                    /clone="RP43-11J15"
ORIGIN

  Query Match              9.6%;  Score 394.8;  DB 2;  Length 196413;
  Best Local Similarity   75.8%;  Pred. No. 8.1e-68;
  Matches  547;  Conservative    0;  Mismatches  157;  Indels   18;  Gaps    4;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC------- 53
              |||||||||||||||||| ||||||||| |||||||||||| |||||| |
Db     196321 AGTTCAAGACCAGCCTGACCAACATGGTGGAACCCCATCTCTAGTAAAAAAAAAAAAAAA 196262

Qy         54 -------AAAATTAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGC 106
                    |||||||| |||| ||||| || ||||| ||||| | ||||||| ||||||
Db     196261 AAAAAAAAAAATTAGCCAGGTGTGGTGGGGCATGCCTGAAATCCTAACTACTTGGGAGGC 196202

Qy        107 CGAGGCAGGAGAATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGC-AAGATCATG 165
              ||||||||| |||  |||||| |  |  | ||||||||||||    ||||  | |||| |
Db     196201 TGAGGCAGGAAAATTGCTTGAAGC--CAAGAGGTGGAGGTTGCAATGAGCTGATATCACG 196144

Qy        166 CCATTGCACTCCAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAGGCA 225
              || ||||||||||||  |||||||| |  ||||   |||||| |||||| |    | ||
Db     196143 CCGTTGCACTCCAGCCTGGGCAACAAAAATGAAACTCCATCTCAAAAAAAAGAGGGGCCG 196084

Qy        226 TTTTATATATATATATATATATATACACACACACACACATATATATATACACATATAT 285
                |  |  ||  |  || | || || | | | | | | ||||||||||| | ||||||
Db     196083 GGAGAGGTGGATCAAGCTAGAAATCTATATATATATATATATATATATATATATATAT 196024

Qy        286 ATACACATATATACATATATACACATATATACACATATATATACACATACATATGTACAC 345
              ||| | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db     196023 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 195964

Qy        346 ATATATATACACATATGTATACACATATATACACATATATACACACATATATACACATAT 405
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | ||||
Db     195963 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 195904

Qy        406 ATACACACATATATACACATATATACACATATATACACATATACACATATATACACATAT 465
              ||| | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | ||||
Db     195903 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 195844

Qy        466 ATACATATATACACATATATATAATATACACACATATATATACACATATATACACACATA 525
              ||| ||||||| | |||||||| ||||| | | ||||||||| | |||||||  | | |||
Db     195843 ATATATATATATATATATATAT-ATATATATATATATATATATATATATATATATATA 195785

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | |||||||| | ||||||| | ||||||| ||||||| | |||
Db     195784 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 195725

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db     195724 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 195665

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| |||||||| | ||||||||   |||  |
Db     195664 TATATATATATATATATATATATATATATATATATATATATATATATATATATCTATATCCA 195605

Qy        706 TA 707
                |
Db     195604 CA 195603


RESULT 11
AC015468
LOCUS       AC015468              189662 bp    DNA     linear   PRI 01-MAY-2001
DEFINITION  Homo sapiens chromosome 8, clone RP11-369E15, complete sequence.
ACCESSION   AC015468
VERSION     AC015468.5  GI:13899433
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
```

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 189662)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 8, clone RP11-369E15
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 189662)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 189662)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M.,
            Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K.,
            Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W.,
            Iliev,I., Johnson,R., Jones,C., Karatas,A., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Sougnez,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
            Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAY-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 1, 2001 this sequence version replaced gi:12313808.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L2466
                Center clone name: 369_E_15
            --------------.
FEATURES             Location/Qualifiers
     source          1. .189662
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="RP11-369E15"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   1. .100
                     /rpt_family="MLT1A1"
     repeat_region   1594. .1741
                     /rpt_family="GA-rich"
     repeat_region   complement(2782. .2957)
                     /rpt_family="MIR"
     repeat_region   3311. .3471
                     /rpt_family="MIR"
     repeat_region   3964. .4067
                     /rpt_family="MIR"
     repeat_region   complement(4528. .4771)
                     /rpt_family="L2"
     repeat_region   4776. .4871
                     /rpt_family="MER4A"
     repeat_region   4869. .5223
                     /rpt_family="MER4A"
     repeat_region   complement(5255. .5765)
                     /rpt_family="L2"
     repeat_region   complement(5805. .6003)
                     /rpt_family="THE1A"
     repeat_region   6004. .6035
                     /rpt_family="(TTA)n"
     repeat_region   complement(6151. .6319)
                     /rpt_family="SVA"
     repeat_region   complement(6320. .6468)
                     /rpt_family="THE1A"
     repeat_region   8722. .8751
                     /rpt_family="(TTTTA)n"
     repeat_region   9155. .9553
                     /rpt_family="MSTB"
     repeat_region   9599. .9982
                     /rpt_family="THE1C"
     repeat_region   complement(10113. .10320)
                     /rpt_family="AluY"
     repeat_region   10822. .10970
                     /rpt_family="MIR3"
     unsure          11210. .11215
                     /note="<30 qual SNGL region"
     repeat_region   11794. .12080
                     /rpt_family="AluSx"
     repeat_region   12974. .13010
                     /rpt_family="(TG)n"
     repeat_region   complement(13091. .13304)
                     /rpt_family="MER106A"
     repeat_region   13484. .13858
                     /rpt_family="MLT1C"
     repeat_region   14025. .14260
                     /rpt_family="L2"
     repeat_region   14584. .14679
                     /rpt_family="MIR"
     repeat_region   14722. .14825
                     /rpt_family="L2"
     repeat_region   15167. .15332
                     /rpt_family="MER33"
     repeat_region   15436. .15788
                     /rpt_family="L2"
     repeat_region   complement(16078. .16165)
                     /rpt_family="AluJ/FRAM"
     repeat_region   16532. .17094
                     /rpt_family="MLT1F1"
     repeat_region   17653. .17674
                     /rpt_family="(TAA)n"
     repeat_region   complement(17675. .17892)
                     /rpt_family="MER7C"
     repeat_region   complement(17897. .18650)
                     /rpt_family="LTR17"
     repeat_region   complement(18651. .18996)
                     /rpt_family="MER7C"
     repeat_region   18999. .19103
                     /rpt_family="(TAA)n"
     repeat_region   complement(19203. .19283)
                     /rpt_family="MIR"
     repeat_region   20464. .20523
                     /rpt_family="A-rich"
     repeat_region   complement(21279. .21902)
```

```
                     /rpt_family="MER41B"
     repeat_region   21922. .22092
                     /rpt_family="MIR"
     repeat_region   22249. .22373
                     /rpt_family="MIR"
     repeat_region   23956. .24058
                     /rpt_family="L2"
     repeat_region   complement(24069. .24198)
                     /rpt_family="MIR"
     repeat_region   25149. .25213
                     /rpt_family="L2"
     repeat_region   complement(25783. .26130)
                     /rpt_family="MLT1I"
     repeat_region   complement(26174. .26539)
                     /rpt_family="MLT1A1"
     repeat_region   complement(28187. .28528)
                     /rpt_family="AluY"
     repeat_region   complement(28792. .29263)
                     /rpt_family="Charlie8"
     repeat_region   complement(30652. .30730)
                     /rpt_family="MER53"
     repeat_region   complement(30986. .31117)
                     /rpt_family="L1ME1"
     repeat_region   31113. .31556
                     /rpt_family="L1ME2"
     repeat_region   31590. .32174
                     /rpt_family="L1ME2"
     repeat_region   32162. .32588
                     /rpt_family="L1ME3"
     repeat_region   33242. .33490
                     /rpt_family="MIR"
     repeat_region   complement(34416. .34803)
                     /rpt_family="MSTB"
     repeat_region   complement(35044. .35350)
                     /rpt_family="AluSg"
     repeat_region   36407. .36695
                     /rpt_family="AluSx"
     repeat_region   complement(37670. .37790)
                     /rpt_family="FLAM_A"
     repeat_region   complement(38065. .38132)
                     /rpt_family="L2"
     repeat_region   complement(38504. .38590)
                     /rpt_family="MIR"
     repeat_region   39044. .39228
                     /rpt_family="MIR3"
     repeat_region   complement(40317. .40473)
                     /rpt_family="MIR"

  Query Match             9.5%;  Score 390.8;  DB 9;  Length 189662;
  Best Local Similarity   65.9%;  Pred. No. 4.9e-67;
  Matches  643;  Conservative    0;  Mismatches  327;  Indels    6;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAAT-ACAAAATT 59
              ||||| |||||  ||||  ||||||| ||||||||||||||||||||||| | ||||||
Db      65793 AGTTCGAGACCCACCTGGCCAACATGATGAAACCCCATCTCTACTAAAAATGAAAAAATT 65852

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGA-GGCCGAGGCAGGAGA 118
              ||   ||  |||||  |   || | |||||||||||||  ||| ||| ||||||| |||
Db      65853 AGCTGGGTGTGGTGGCAGGTGCATCTAATCCCAGCTACTCTGGAGGGCTGAGGCAGAAGA 65912

Qy        119 ATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCC 177
              ||  ||||||||  | || |||||| ||||| | ||| | |||||  ||||||||||||
Db      65913 ATTGCTTGAACC--CAGGAGGTGGACGTTGCAGTGAGCCGAGATCGCGCCATTGCACTCC 65970

Qy        178 AGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATAT 237
              ||||  |||||||||||| |||     ||||| |||||||||          |||||||||
Db      65971 AGCCTGGGCAACAAGAGCAAAACTCTGTCTCAAAAAAAAAAATTTTATAAATATATATAT 66030

Qy        238 ATATATATATATATACACACACACACACATATATATATACACATATATATACACATATAT 297
              | |||||||||| | | | | | | | ||||||| ||| | ||| ||||| | | ||||
Db      66031 AAATATATATATAAATATATATAAATATATATATAAATATATATAAATATATATAAATAT 66090

Qy    ,   298 ACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACAC 357
```

```
              | ||||| | | ||||| | | ||||| ||| | ||| |||| || | ||||||||| |
Db      66091 ATATATAAATATATATAAATATATATAAATATATATAAATATATATAAATATATATATAA 66150

Qy        358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417
              |||| |||| | ||||| | | ||||||| | | |||||||| | ||||||| | | ||||
Db      66151 ATATATATAAATATATAAAAATATATATATAAATATATATAAATATATATAAATATATAT 66210

Qy        418 ATACACATATATACACATATATACACATATACACATATATACACATATATACATATATAC 477
              | | | ||||||| | ||||||| | ||||| | ||||||| | | | ||| |||||||
Db      66211 AAATATATATATAAATATATATAAATATATATAAATATATATATAAAAATATATATATAA 66270

Qy        478 ACATATATATAATATACACACATATATATACACATATATACACACATATATACACATATA 537
              | ||||||| |||||| | | | ||||||| | |||||||  | | |||||||  | || ||
Db      66271 ATATATATA-AATATATATATAAATATATATAAATATATATATAAATATATAAATATGTA 66329

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATACATATA 597
              || | ||||||| | |  |||||| | || |||| | |||||  | |||| |||| | |||
Db      66330 TATAAATATATAAATATGTATATAAATATGTATATAAATATATAAATATGTATATAAATA 66389

Qy        598 TACACATATATACATATATACACATATATACATATATACACACATATATACACATACATA 657
              || | ||| ||| ||| |||    ||||| | ||||||| | | ||||||| | ||  |||
Db      66390 TATATATAAATATATAAATATGTATATAAATATATATATAAATATATATAAATATGTATA 66449

Qy        658 TACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATA 717
              || | | | ||| ||||  ||||| ||| | ||||||| |  ||||||  |  ||||||||
Db      66450 TAAATATATATAAATATGTATATAAATATATATATATAAATATATATAAAAATATATATA 66509

Qy        718 TATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGA 777
              ||         | |  |||           | | | ||    |    ||| |   ||| | |
Db      66510 TAAATATATAAATATATATATAAATATATATAAATATATATAAATATATATATAAATATA 66569

Qy        778 GTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAA 837
                | |    | ||     | |   | || |        |     |||||        | |
Db      66570 TAAATATATATATATAAATATATATAAATATATATAAAAATATAAATATATATAAATA 66629

Qy        838 TACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATATAT 897
              || ||| |  | ||  | |  |     |  |  ||  |  |   |||| ||||
Db      66630 TATATATAAATATATATATAAAAAATATATAAAATATATATATAAATATATATAAATAT 66689

Qy        898 ATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTATAGAAGAA 957
              |||||   | |      || |  || || |  |   ||  | | |||| || |     ||
Db      66690 ATATAAATATATATAAATATATATATAGTAGTGAAGAATCACTAAAATATTTTTATGAAA 66749

Qy        958 TGGCTAAGATGCCAAA 973
                || ||  | | |
Db      66750 AAAGTATGAATCTAGA 66765
```

```
RESULT 12
AC073310/c
LOCUS       AC073310              125685 bp    DNA     linear   PRI 08-OCT-2003
DEFINITION  Homo sapiens BAC clone RP11-49G5 from 7, complete sequence.
ACCESSION   AC073310
VERSION     AC073310.7  GI:13176604
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 125685)
  AUTHORS   Sulston,J.E. and Wilson,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
   PUBMED   9847074
REFERENCE   2  (bases 1 to 125685)
  AUTHORS   Haakenson,W., Hannah,C. and Kang,K.
  TITLE     The sequence of Homo sapiens BAC clone RP11-49G5
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 125685)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUN-2000) Genome Sequencing Center, Washington
```

```
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 125685)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 125685)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2001) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   6  (bases 1 to 125685)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2001) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   7  (bases 1 to 125685)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-APR-2003) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   8  (bases 1 to 125685)
  AUTHORS   Wilson,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-OCT-2003) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Mar 1, 2001 this sequence version replaced gi:11024926.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0049G05
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            The sequence of this clone was established as part of a mapping and
            sequencing collaboration between the NHGRI Chromosome 7 Mapping
            Project (Eric D. Green, Director), John D. McPherson in the
            Department of Genetics (Washington University), and the Washington
            University Genome Sequencing Center. For additional information
            about the map position of this sequence, see
            http://www.nhgri.nih.gov/DIR/GTB/CHR7 , send
            mailto:egreen@nhgri.nih.gov , or see http://genome.wustl.edu

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries. Genomics 51:1-8. The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at http://www.chori.org
            VECTOR:  pBACe3.6

            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is RP11-409M7, 200 bp overlap the
            clone sequenced to the right is RP11-374N8.  Actual start of this
            clone is at base position 14572 of RP11-409M7 actual end is at base
            position 125685 of RP11-49G5.
FEATURES             Location/Qualifiers
     source          1. .125685
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7"
                     /clone="RP11-49G5"
                     /clone_lib="RPCI-11"
     repeat_region   1013. .1151
                     /rpt_family="CR1"
     repeat_region   2238. .2394
                     /rpt_family="MIR"
     repeat_region   2548. .2668
                     /rpt_family="L2"
     repeat_region   3206. .5231
                     /rpt_family="L1"
     repeat_region   5694. .5715
                     /rpt_family="AT_rich"
     repeat_region   7142. .7353
                     /rpt_family="MER1_type"
     repeat_region   7490. .7666
                     /rpt_family="ERVL"
     repeat_region   7671. .7771
                     /rpt_family="L1"
     repeat_region   7951. .7985
                     /rpt_family="AT_rich"
     repeat_region   9572. .9809
                     /rpt_family="L1"
     repeat_region   9811. .9831
                     /rpt_family="AT_rich"
     repeat_region   9895. .10017
                     /rpt_family="MaLR"
     repeat_region   10320. .10375
                     /rpt_family="MIR"
     repeat_region   10395. .10732
                     /rpt_family="L1"
     repeat_region   10911. .11118
                     /rpt_family="(TA)n"
     unsure          10914. .11684
                     /note="unresolved simple sequence repeat"
     repeat_region   11130. .11196
                     /rpt_family="L1"
     repeat_region   11217. .11663
                     /rpt_family="(TA)n"
     repeat_region   11667. .11918
                     /rpt_family="Alu"
     repeat_region   11919. .11970
                     /rpt_family="L1"
     repeat_region   12356. .12508
                     /rpt_family="L1"
     repeat_region   12646. .12813
                     /rpt_family="L1"
     repeat_region   12863. .13111
                     /rpt_family="L1"
     repeat_region   13280. .13592
                     /rpt_family="L1"
     repeat_region   13768. .14032
                     /rpt_family="L1"
     repeat_region   14048. .14489
                     /rpt_family="L1"
     repeat_region   15108. .15378
                     /rpt_family="Alu"
     repeat_region   15563. .15584
                     /rpt_family="AT_rich"
     repeat_region   17280. .17442
                     /rpt_family="MIR"
     repeat_region   18313. .18467
                     /rpt_family="L1"
     repeat_region   18468. .18676
```

```
                    /rpt_family="L1"
     repeat_region  18677. .18701
                    /rpt_family="AT_rich"
     repeat_region  18797. .19155
                    /rpt_family="MaLR"
     repeat_region  19594. .19911
                    /rpt_family="Alu"
     repeat_region  21451. .21639
                    /rpt_family="MER53"
     repeat_region  22809. .22986
                    /rpt_family="L2"
     repeat_region  23094. .23371
                    /rpt_family="L1"
     repeat_region  23372. .23678
                    /rpt_family="Alu"
     repeat_region  23679. .23806
                    /rpt_family="L1"
     repeat_region  23820. .23999
                    /rpt_family="L1"
     repeat_region  24445. .24470
                    /rpt_family="AT_rich"
     repeat_region  24596. .24649
                    /rpt_family="(CA)n"
     repeat_region  26292. .26765
                    /rpt_family="L1"
     repeat_region  26766. .27123
                    /rpt_family="MaLR"
     repeat_region  27124. .27202
                    /rpt_family="L1"
     repeat_region  27203. .27556
                    /rpt_family="MaLR"

  Query Match             9.4%;  Score 387.4;  DB 9;  Length 125685;
  Best Local Similarity   67.7%;  Pred. No. 2.4e-66;
  Matches  647;  Conservative    0;  Mismatches  296;  Indels   12;  Gaps    7;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              |||||||||||||||   ||||||||||||||||||| | | ||||||||||| ||||||
Db      11886 AGTTCAAGACCAGCCTAGCCAACATGGTGAAACCCCGTTTTTACTAAAAATACAAAAATT 11827

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              || ||||| ||||| |    |||||||||||||||| ||   ||||||  ||||||||||
Db      11826 AGCCAGGCGTGGTGGT---TGCCTGTAATCCCAGCCACACGGGAGGCTAAGGCAGGAGAA 11770

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGC-AAGATCATGCCATTGCACTCCA 178
              |  |||||| |||  || |  |||||||| | ||||  |||||| |||| |||||||||
Db      11769 TTGCTTGAATCTG--GGAGACGGAGGTTGCAGTGAGCTGAGATCACGCCACTGCACTCCA 11712

Qy        179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAGGCATTTTATATATATA 238
              |||  ||  || |||| || |   ||||||| ||||||||||     |  ||||||||||
Db      11711 GCCTGGGTGAC-AGAGTGAGACTCCATCTCAAAAAAAAAAAAAAAAAGTCATATATATATA 11653

Qy        239 TATATATATATATACACACACACACATATATATATACACATATATATACACATATATA 298
              |||||||||||||| | | | | | | | | ||||||| | | ||||||| | | |||||
Db      11652 TATATATATATATATATATATATATAAAAATATATATATAAATATATATATAAATATA 11593

Qy        299 CATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACA 358
              |||||||  | |||| || | ||||||||| | ||| || | || | |||||||  | | |
Db      11592 TATATATAAATATATTTATAAATATATATATAAATATATTTATAAATATATATAAATATA 11533

Qy        359 TATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATA 418
              ||| |||| | ||||||| | ||||||   | | | ||||| | ||| ||| | | |||||
Db      11532 TATATATAAATATATATATAAATATATGTATAAAAATATATATATAAATATAAATATATA 11473

Qy        419 TACACATATATACACATATATACACATATACACATATATACACATATATACATATATACA 478
              || | |||||||  | | | ||| | ||||| | ||| ||| | ||||| | |||||||  |
Db      11472 TATAAATATATATATAAAAATATATATATAAATATAAATATATATATAAATATATATATA 11413

Qy        479 CATATATATA---ATATACACACATATATATACACATATATACACACATATATACACATA 535
              |||||||||   |||||  | | |||||  ||| | |||||||  | | |||||  | |   ||
Db      11412 AATATATATATATATATATAAATATATATAAATATATATATATATATATATATAAATATCTA 11353

Qy        536 TATACACATATATATACACATATATACACATATATACATATATACACATATATATACATA 595
```

```
              |||| |||||||||||  | |||||||  | | | |||  |||||||  | |||||||||  | |
Db      11352 TATAAACATATATATATAAATATATATATAAACATATATATATAAATATATATATAAACA 11293

Qy        596 TATACACATATATACATATATACACATATATACATATATACACACATATATACACATACA 655
              ||||  | |||  |||  |||||||  |||||||||  |||  |||  |  |  |||||||  | |||  |
Db      11292 TATATATATAAATATATATATAAACATATATATATAAATATATATATATATATATATA 11233

Qy        656 TATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATA 715
              ||||  |  |  |  |||  |  |  |  |||||||| ||  |||||||||   |  ||||||  ||||  |
Db      11232 TATATATATATATATAAACATATATATATTCAAATATATATATAAACATATATATATAAA 11173

Qy        716 TATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGA 775
              |||||      |   |  |||          |  |    ||   |         |   | | |
Db      11172 TATATATATAAACATATATATATAAATATATATATAAACATATATAATATAACATATATA 11113

Qy        776 GAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGA 835
              |  |  | |    | ||     | |   | || |      | |    | |||      |
Db      11112 TATATATATATATATATAAATATATATATAAATATATATATAAATATATATATATAAATA 11053

Qy        836 AATACATAAATGTGTAATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTAACACTATATA 894
              |||  ||||||  | ||   | |  |||  | ||  |  ||  ||   |   |  ||||||
Db      11052 TATATATAAATATATATATATATATATATATATAAATATATATATATAAATATATATATA 10993

Qy        895 TATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTA 949
              || |||||   | |    | || |  ||||          ||  || || || | |
Db      10992 TAAATATATATATATATATATATATATAAATATATATATAAATATATATATATAA 10938


RESULT 13
AL133477
LOCUS       AL133477              139118 bp    DNA     linear   PRI 02-FEB-2001
DEFINITION  Human DNA sequence from clone RP11-172F4 on chromosome 9q22.2-31.1,
            complete sequence.
ACCESSION   AL133477
VERSION     AL133477.16  GI:12666198
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 139118)
  AUTHORS   Lloyd,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Feb 5, 2001 this sequence version replaced gi:12584335.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence has been finished according to sequence map criteria
            as follows.  An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated repeat sequence elements.  Where the sequence is
            ambiguous, there is an annotation using the 'unsure' feature key.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr9
            RP11-172F4 is from the library RPCI-11.1 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-172F4 It may be shorter because we sequence overlapping
```

```
          sections only once, except for a 100 base overlap.
          The true left end of clone RP11-172F4 is at 1 in this sequence. The
          true left end of clone RP11-240L7 is at 139019 in this sequence.
          The true right end of clone RP11-392G7 is at 43511 in this
          sequence.
FEATURES             Location/Qualifiers
     source          1. .139118
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /map="q22.2-31.1"
                     /clone="RP11-172F4"
                     /clone_lib="RPCI-11.1"
     repeat_region   45. .78
                     /note="L1MB2 repeat: matches 6125. .6158 of consensus"
     repeat_region   195. .361
                     /note="MER20 repeat: matches 39. .218 of consensus"
     repeat_region   488. .688
                     /note="MLT1B repeat: matches 205. .390 of consensus"
     repeat_region   689. .986
                     /note="AluY repeat: matches 1. .297 of consensus"
     repeat_region   987. .1144
                     /note="MLT1B repeat: matches 5. .205 of consensus"
     repeat_region   1149. .1500
                     /note="MST-INTERNAL repeat: matches 1289. .1651 of
                     consensus"
     repeat_region   1812. .2130
                     /note="MLT1-INTERNAL repeat: matches 404. .728 of
                     consensus"
     repeat_region   2175. .2472
                     /note="MLT2FB repeat: matches 101. .414 of consensus"
     repeat_region   2729. .2832
                     /note="MER5B repeat: matches 2. .80 of consensus"
     repeat_region   2833. .3146
                     /note="AluSx repeat: matches 1. .308 of consensus"
     repeat_region   3147. .3236
                     /note="MER5B repeat: matches 80. .177 of consensus"
     repeat_region   5694. .6008
                     /note="AluSx repeat: matches 1. .306 of consensus"
     repeat_region   6183. .6318
                     /note="MER5A repeat: matches 10. .189 of consensus"
     repeat_region   6556. .6717
                     /note="FRAM repeat: matches -5. .161 of consensus"
     repeat_region   7313. .7619
                     /note="AluSx repeat: matches 1. .311 of consensus"
     repeat_region   7907. .8217
                     /note="AluY repeat: matches 1. .311 of consensus"
     repeat_region   8545. .8837
                     /note="AluSg repeat: matches 1. .293 of consensus"
     repeat_region   9101. .9389
                     /note="AluJo repeat: matches 1. .282 of consensus"
     repeat_region   9399. .9701
                     /note="AluY repeat: matches 1. .311 of consensus"
     repeat_region   9702. .10014
                     /note="AluSx repeat: matches 1. .311 of consensus"
     repeat_region   10050. .10292
                     /note="AluSp repeat: matches 66. .310 of consensus"
     repeat_region   10300. .10615
                     /note="AluSx repeat: matches 4. .312 of consensus"
     repeat_region   14460. .14754
                     /note="AluJo repeat: matches 13. .308 of consensus"
     repeat_region   14846. .15042
                     /note="AluJb repeat: matches 96. .296 of consensus"
     repeat_region   15166. .15281
                     /note="AluJb repeat: matches 9. .96 of consensus"
     repeat_region   15506. .15781
                     /note="AluSg repeat: matches 1. .276 of consensus"
     repeat_region   16165. .16466
                     /note="AluSx repeat: matches 1. .300 of consensus"
     repeat_region   18757. .19057
                     /note="AluSx repeat: matches 1. .301 of consensus"
     repeat_region   19097. .19144
                     /note="12 copies 4 mer tata 77% conserved"
     repeat_region   19359. .19662
                     /note="AluSg repeat: matches 1. .304 of consensus"
     repeat_region   19743. .20060
                     /note="AluSx repeat: matches 1. .305 of consensus"
     repeat_region   21872. .22179
                     /note="AluSx repeat: matches 1. .299 of consensus"
     repeat_region   27049. .27427
                     /note="MLT1B repeat: matches 7. .390 of consensus"
     repeat_region   29141. .29384
                     /note="MIR repeat: matches 2. .260 of consensus"
     misc_feature    29833. .30559
                     /note="CpG island"
                     /evidence=not_experimental
     repeat_region   31569. .31653
                     /note="MER91A repeat: matches 3. .63 of consensus"
     repeat_region   31939. .32227
                     /note="AluY repeat: matches 1. .291 of consensus"
     repeat_region   32228. .32323
                     /note="48 copies 2 mer gg 67% conserved"
     repeat_region   32248. .32297
                     /note="10 copies 5 mer gacgg 88% conserved"
     repeat_region   32324. .32378
                     /note="11 copies 5 mer aagaa 78% conserved"
     repeat_region   33239. .33518
                     /note="AluJo repeat: matches 1. .289 of consensus"
     repeat_region   34313. .34610
                     /note="AluSq repeat: matches 1. .298 of consensus"
     repeat_region   35793. .35899
                     /note="MIR repeat: matches 1. .116 of consensus"
     repeat_region   37385. .37604
                     /note="L2 repeat: matches 2436. .2659 of consensus"
     repeat_region   38070. .38181
                     /note="L2 repeat: matches 2563. .2729 of consensus"
     repeat_region   40862. .41123
                     /note="AluSq repeat: matches 39. .302 of consensus"
     repeat_region   41722. .42016
                     /note="AluSp repeat: matches 1. .297 of consensus"
     repeat_region   43048. .43521
                     /note="L1MB5 repeat: matches 4783. .5239 of consensus"
     repeat_region   43522. .43695
                     /note="FAM repeat: matches 1. .174 of consensus"
     repeat_region   43696. .44552
                     /note="L1MB5 repeat: matches 5239. .6138 of consensus"
     repeat_region   44598. .44834
                     /note="L1MB4 repeat: matches 5814. .6065 of consensus"
     repeat_region   44835. .45133
                     /note="AluSg repeat: matches 1. .295 of consensus"
     repeat_region   45134. .45257
                     /note="L1MB4 repeat: matches 6065. .6184 of consensus"
     repeat_region   45312. .45611
                     /note="AluJo repeat: matches 1. .293 of consensus"
     repeat_region   45681. .45989
                     /note="AluSg repeat: matches 1. .300 of consensus"
     repeat_region   46204. .46347
                     /note="L1ME2 repeat: matches 6007. .6153 of consensus"
     repeat_region   46876. .47188
                     /note="AluJo repeat: matches 1. .304 of consensus"
     repeat_region   47262. .47560
                     /note="AluSq repeat: matches 1. .301 of consensus"
     repeat_region   47923. .48041
                     /note="L1MC4 repeat: matches 6562. .6679 of consensus"
     repeat_region   48060. .48148
                     /note="L1ME1 repeat: matches 5347. .5428 of consensus"
     repeat_region   48149. .48413
                     /note="AluJo repeat: matches 2. .302 of consensus"
     repeat_region   48414. .48641
                     /note="L1ME1 repeat: matches 5428. .5637 of consensus"
     repeat_region   48642. .48935
                     /note="AluSg repeat: matches 1. .294 of consensus"
     repeat_region   48936. .49372
                     /note="L1ME1 repeat: matches 5637. .6160 of consensus"
     repeat_region   50733. .50947
```

```
                     /note="MIR repeat: matches 2. .257 of consensus"
     repeat_region   51470. .51541
                     /note="MIR repeat: matches 44. .124 of consensus"
     repeat_region   51667. .51977
                     /note="AluY repeat: matches 1. .311 of consensus"
     repeat_region   52966. .53214
                     /note="MIR repeat: matches 12. .262 of consensus"
     repeat_region   53551. .54965
                     /note="L2 repeat: matches 505. .2101 of consensus"
     repeat_region   54984. .55272
                     /note="AluSp repeat: matches 1. .291 of consensus"

  Query Match             9.4%;  Score 385.8;  DB 9;  Length 139118;
  Best Local Similarity   77.4%;  Pred. No. 4.9e-66;
  Matches  583;  Conservative    0;  Mismatches  132;  Indels   38;  Gaps    8;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC---AAAA 57
              ||||| ||||||||||  ||||||||||||||||||||||||||||||||    ||||
Db      97498 AGTTCGAGACCAGCCTGGCCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAAAAA 97557

Qy         58 TTAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAG 117
              ||||   ||| |||||  ||| ||||||| || ||||||| | |||||| ||||||||||
Db      97558 TTAGCTGGGCGTGGTGGCACATGCCTGTAGTCTCAGCTACCTGGGAGGCTGAGGCAGGAG 97617

Qy        118 AATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTC 176
              |||| ||||||||  | || |||||||||||| | ||| | |||||    ||| |||||||
Db      97618 AATCGCTTGAACC--CAGGAGGTGGAGGTTGCAGTGAGCCGAGATCGCTCCACTGCACTC 97675

Qy        177 CAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCAT---------- 226
              | ||   ||| || |||||   |   | ||||| ||||||||||||   ||
Db      97676 CGGCTTGGGCGAC-AGAGCATGACTCCGTCTCAAAAAAAAAAAAAATATATATATATATAT 97734

Qy        227 --------------TTTATATATATATATATATATATACACACACACACACATATAT 271
                           | |||||||||| | ||||||| ||| | ||| ||| | ||| ||
Db      97735 ATGTGTGTGTGTGTGTATATATATACACATATATACATATATACATACATATATACAT 97794

Qy        272 ATATACACATATATATACACATATATACATATATACACATATATACACATATATATACAC 331
              ||||||| | |||||||||||||||||||| ||| |||||||||||| ||||| | ||
Db      97795 ATATACATACATATATACACATATATACATACATATACATATATACACACATATACATAC 97854

Qy        332 ATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACACAC 391
              ||| | || || |||||||| ||||| ||  ||| | | ||||| ||||||||||||||
Db      97855 ATATACATATATACATATATACACACACATACATATATACATATATACATATATACACAT 97914

Qy        392 ATATATACACATATATACACACATATATACACATATATACACATATATACACATATACAC 451
              ||||| | | ||| ||| | ||| ||||| ||| | ||| ||||| ||| ||||||| ||
Db      97915 ATATACATATATACATATATACACATATATACACACATATACATACATATACATATATAC 97974

Qy        452 ATATATACACATATATACATATATACACATATATATA-ATATACACACATATATATACAC 510
              |||||||||||||||||||||||||||||| |||||  || |||||||| | ||| ||||| ||
Db      97975 GTATATACACATATATACATATATACATATATACGTATATATACATATATACATATATAC 98034

Qy        511 ATATATACACACATATATACACATATATACACATATATATACACATATATACACATATAT 570
              |||||| ||| ||||| | | |||  || | ||| ||||| |||||||| ||||||||
Db      98035 GTATATATACATATATACATATATACGTATATATACATATATACATATATATACATATAT 98094

Qy        571 ACATATATACACATAT----ATATACATATATACACATATATACATATATACACATATAT 626
              |||||||| ||||||    ||||| | ||||| || |||| |||||||| ||||||||
Db      98095 ACATATATATACATATATACATATATACATATATACGTATATGCATATATATACATATAT 98154

Qy        627 ACATATATACACACATATATACACATACATATACACACACATAGATATACATATATATAC 686
              ||||||||| ||| |||||||| ||| | | | | ||| | ||||||||| ||||||
Db      98155 ACATATATATACATATATATACATATATACACATATATACACA-ATATACATACATATAC 98213

Qy        687 ACATATATATACGTATATATATGTATATATATA 719
              | ||||| | |  || |||||| || |||||||
Db      98214 ATATATACACATATACATATATATACATATATA 98246


RESULT 14
HUAC004787
LOCUS       HUAC004787            216021 bp    DNA     linear   PRI 30-OCT-2002
DEFINITION  Homo sapiens Chromosome 16 BAC clone CIT987SK-A-952F10, complete
            sequence.
ACCESSION   AC004787
VERSION     AC004787.1  GI:3337381
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 216021)
  AUTHORS   Loftus,B.J., Kim,U.J., Sneddon,V.P., Kalush,F., Brandon,R.,
            Fuhrmann,J., Mason,T., Crosby,M.L., Barnstead,M., Cronin,L.,
            Deslattes Mays,A., Cao,Y., Xu,R.X., Kang,H.L., Mitchell,S.,
            Eichler,E.E., Harris,P.C., Venter,J.C. and Adams,M.D.
  TITLE     Genome duplications and other features in 12 Mb of DNA sequence
            from human chromosome 16p and 16q
  JOURNAL   Genomics 60 (3), 295-308 (1999)
  MEDLINE   99425270
   PUBMED   10493829
REFERENCE   2  (bases 1 to 216021)
  AUTHORS   Adams,M.D., Loftus,B.J., Zhou,L., Crosby,M., Fuhrmann,J.,
            Mason,T.M., Brandon,R., Kim,U.J., Kerlavage,A.R. and Venter,J.C.
  TITLE     Homo sapiens Chromosome 16 BAC clone CIT987SK-A-952F10
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 216021)
  AUTHORS   Adams,M.D. and Loftus,B.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1998) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, Email:
            bjloftus@tigr.org
REFERENCE   4  (bases 1 to 216021)
  AUTHORS   Adams,M.D. and Loftus,B.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JUL-1998) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
COMMENT     On Jul 24, 1998 this sequence version replaced gi:3241936.
            Address all correspondence to: Mark Adams The Institute for Genomic
            Research 9712 Medical Center Dr, Rockville, MD 20850, USA e-mail
            address: humgen@tigr.org.   The orientation of the sequence is from
            SP6 end to T7 end. Genes were identified by a combination of five
            methods including: XGRAIL (available by anonymous ftp from
            arthur.epm.ornl.gov), Genefinder (Phil Green, University of
            Washington), Genscan  (Chris Burge,
            http://gnomic.stanford.edu/~chris/GENSCANW.html)searches of the
            complete sequence against a peptide database,and the Human gene
            Index database at TIGR (http://www.tigr.org/tdb/hgi/hgi.html).
            Genes without pepetide homolgy having spliced EST hits are termed
            'Unknown gene product'. Genes encoding tRNAs are predicted by
            tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
FEATURES             Location/Qualifiers
     source          1. .216021
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /map="#16q21-22"
                     /clone="A-952F10"
     STS             27765. .27872
                     /note="7766, STS1-cSRL-27g3-uA/cSRL-27g3-uZ, Chr. -, Homo
                     sapiens"
                     /db_xref="dbSTS:G02280"
     STS             73826. .73943
                     /note="7608, STS1-cSRL-24g1-uA/cSRL-24g1-uZ, Chr. -, Homo
                     sapiens"
                     /db_xref="dbSTS:G02122"
     STS             175801. .175945
                     /note="16084, CHLC.GCT10B02, Chr. -, Homo sapiens"
                     /db_xref="dbSTS:G09703"
     STS             175810. .175945
                     /note="16316, CHLC.GCT15C04, Chr. -, Homo sapiens"
                     /db_xref="dbSTS:G09935"
     STS             199463. .199572
                     /note="9824, WI-3555, Chr. 16, Homo sapiens"
                     /db_xref="dbSTS:G04338"
```

ORIGIN

Query Match 9.3%; Score 382.6; DB 9; Length 216021;
Best Local Similarity 76.2%; Pred. No. 1.9e-65;
Matches 551; Conservative 0; Mismatches 159; Indels 13; Gaps 6;

```
Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAATTA 60
              ||||| | |||| ||||  |||||| ||||||||||||||||||||||||    |||||||
Db      80976 AGTTCGAAACCACCCTGGCCAACATAGTGAAACCCCATCTCTACTAAAATACAAAATTA 81035

Qy         61 GACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAAT 120
              | ||||| |||||  | |||||||||||||||||||||  |||||| |||||||||| ||
Db      81036 GCCAGGCGTGGTGGCAGACGCCTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGAGGAT 81095

Qy        121 CACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCCAG 179
              ||||||||  | || ||  |||||||| | ||| | |||| |||||  || ||||||||
Db      81096 TGCTTGAACC--CAGGAGGCAGAGGTTGCAGTGAGCCCAGATTGTGCCACTGGACTCCAG 81153

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  ||  || |||| | |   | ||||| | | | | | |        || | | ||||
Db      81154 CCTGGGAGAC-AGAGTAAGACTCCGTCTCAAATACATATATA-----CATACACACATAT 81207

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              ||||| | ||||||| ||||||| | ||||| | ||| ||||| | | | ||||| | ||
Db      81208 ATATACACATATACATACACACATATATATACACATATACATACACACATACATACACAC 81267

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACAT 359
              |||||||||| ||||||| ||||||| ||| ||| |||  || ||| ||||||||||||
Db      81268 ATATATACACACATATACATATATATACACATATATATACATATACACATATATACACAT 81327

Qy        360 ATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATAT 419
              || || | | | | | | | ||||| | | | | | ||| ||| ||||| ||| |||||
Db      81328 ATATA-ATATACACACATATATATACATATATACACATATACACATATATACATATATAC 81386

Qy        420 ACACATATATACACATATATACACATATACACATATATACACATATATACATATATACAC 479
              |||||||||||||||||||||||| ||| ||||||||||||| ||| | ||||||| ||||
Db      81387 ACACATATATACACATATATACATATACACACATATATACATATACACACATATACACAC 81446

Qy        480 ATATATATA---ATATACACACATATATATACACATATATACACACATATATACACATAT 536
              ||| | | |   | | |||||||||| | | | ||| | | | ||||| | | | ||| |
Db      81447 ATACACACACATACACACACACATACACACATACACACACATACACACACACATACACAC 81506

Qy        537 ATACACATATATATACACATATATACACATATATACATATATACACATATATATACATAT 596
              | | ||| | ||| ||||| ||||||||| ||| |  || | | ||| | ||||||| ||
Db      81507 ACATACACACATACACACACATATACACACATACATGTACACATACACACATATACACAT 81566

Qy        597 ATACACATATATACATATATACACATATATACATATATACACACATATATACACATACAT 656
              ||||||| |||||||||||||||| |||||||||||||||||||||||| ||||| ||
Db      81567 ATACACACATATACATATATACACACATATACATATATACACACATATATATACATATAT 81626

Qy        657 ATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATAT 716
              | ||| | | | | ||||||| |||||||| |||| | |  |||| |||| || | ||
Db      81627 ACACATATATACATATATACACATATATACATATATGCACATATATACATATATACACAT 81686

Qy        717 ATA 719
              |||
Db      81687 GTA 81689
```

RESULT 15
AY007685
LOCUS AY007685 92564 bp DNA linear PRI 22-FEB-2002
DEFINITION Homo sapiens telomerase catalytic subunit (TERT) and sodium channel-like protein genes, complete cds.
ACCESSION AY007685
VERSION AY007685.2 GI:15991796
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 92564)
AUTHORS Leem,S.H., Londono-Vallejo,J.A., Kim,J.H., Bui,H., Tubacher,E., Solomon,G., Park,J.E., Horikawa,I., Kouprina,N., Barrett,J.C. and Larionov,V.
TITLE The human telomerase gene: complete genomic sequence and analysis of tandem repeat polymorphisms in intronic regions
JOURNAL Oncogene 21 (5), 769-777 (2002)
MEDLINE 21839826
PUBMED 11850805
REFERENCE 2 (bases 1 to 92564)
AUTHORS Londono-Vallejo,J.A.
TITLE Direct Submission
JOURNAL Submitted (06-SEP-2000) Centre d'Etudes du Polymorphisme Humain, 27 rue Juliette Dodu, Paris 75010, France
REFERENCE 3 (bases 1 to 92564)
AUTHORS Londono-Vallejo,J.A.
TITLE Direct Submission
JOURNAL Submitted (10-OCT-2001) Centre d'Etudes du Polymorphisme Humain, 27 rue Juliette Dodu, Paris 75010, France
REMARK Sequence update by submitter
COMMENT On Oct 10, 2001 this sequence version replaced gi:12642956.
FEATURES Location/Qualifiers

```
     source          1. .92564
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /map="5p13"
     repeat_region   complement(2853. .3034)
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   complement(6174. .6429)
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   complement(7445. .7837)
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   complement(8199. .8341)
                     /rpt_family="MLT1"
                     /rpt_type=dispersed
     repeat_region   complement(10300. .10916)
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   12028. .12301
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   13088. .13597
                     /rpt_family="L1"
                     /rpt_type=dispersed
     repeat_region   13716. .13988
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   14271. .14556
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   14599. .17711
                     /rpt_family="L1"
                     /rpt_type=dispersed
     repeat_region   18109. .18398
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   complement(20818. .21123)
                     /rpt_family="LTR1"
                     /rpt_type=dispersed
     repeat_region   21650. .21912
                     /rpt_family="MLT1"
                     /rpt_type=dispersed
     repeat_region   complement(22236. .22484)
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     repeat_region   complement(22986. .23283)
                     /rpt_family="Alu"
                     /rpt_type=dispersed
     gene            <24442. .64466
                     /gene="TERT"
```

```
mRNA            join(<24442. .24660,24765. .26118,35473. .35668,37759. .37939,
                38627. .38806,39301. .39456,45360. .45455,46436. .46521,
                49007. .49120,51105. .51176,53048. .53236,57040. .57166,
                58981. .59042,62229. .62353,63135. .63272,63809. .>63912)
                /gene="TERT"
                /product="telomerase catalytic subunit"
CDS             join(24442. .24660,24765. .26118,35473. .35668,37759. .37939,
                38627. .38806,39301. .39456,45360. .45455,46436. .46521,
                49007. .49120,51105. .51176,53048. .53236,57040. .57166,
                58981. .59042,62229. .62353,63135. .63272,63809. .63912)
                /gene="TERT"
                /codon_start=1
                /product="telomerase catalytic subunit"
                /protein_id="AAG23289.1"
                /db_xref="GI:12642957"
                /translation="MPRAPRCRAVRSLLRSHYREVLPLATFVRRLGPQGWRLVQRGDP
                AAFRALVAQCLVCVPWDARPPPAAPSFRQVSCLKELVARVLQRLCERGAKNVLAFGFA
                LLDGARGGPPEAFTTSVRSYLPNTVTDALRGSGAWGLLLRRVGDDVLVHLLARCALFV
                LVAPSCAYQVCGPPLYQLGAATQARPPPHASGPRRRLGCERAWNHSVREAGVPLGLPA
                PGARRRGGSASRSLPLPKRPRRGAAPEPERTPVGQGSWAHPGRTRGPSDRGFCVVSPA
                RPAEEATSLEGALSGTRHSHPSVGRQHHAGPPSTSRPPRPWDTPCPPVYAETKHFLYS
                SGDKEQLRPSFLLSSLRPSLTGARRLVETIFLGSRPWMPGTPRRLPRLPQRYWQMRPL
                FLELLGNHAQCPYGVLLKTHCPLRAAVTPAAGVCAREKPQGSVAAPEEEDTDPRRLVQ
                LLRQHSSPWQVYGFVRACLRRLVPPGLWGSRHNERRFLRNTKKFISLGKHAKLSLQEL
                TWKMSVRDCAWLRRSPGVGCVPAAEHRLREEILAKFLHWLMSVYVVELLRSFFYVTET
                TFQKNRLFFYRKSVWSKLQSIGIRQHLKRVQLRELSEAEVRQHREARPALLTSRLRFI
                PKPDGLRPIVNMDYVVGARTFRREKRAERLTSRVKALFSVLNYERARRPGLLGASVLG
                LDDIHRAWRTFVLRVRAQDPPPELYFVKVDVTGAYDTIPQDRLTEVIASIIKPQNTYC
                VRRYAVVQKAAHGHVRKAFKSHVSTLTDLQPYMRQFVAHLQETSPLRDAVVIEQSSSL
                NEASSGLFDVFLRFMCHHAVRIRGKSYVQCQGIPQGSILSTLLCSLCYGDMENKLFAG
                IRRDGLLLRLVDDFLLVTPHLTHAKTFLRTLVRGVPEYGCVVNLRKTVVNFPVEDEAL
                GGTAFVQMPAHGLFPWCGLLLDTRTLEVQSDYSSYARTSIRASLTFNRGFKAGRNMRR
                KLFGVLRLKCHSLFLDLQVNSLQTVCTNIYKILLLQAYRFHACVLQLPFHQQVWKNPT
                FFLRVISDTASLCYSILKAKNAGMSLGAKGAAGPLPSEAVQWLCHQAFLLKLTRHRVT
                YVPLLGSLRTAQTQLSRKLPGTTLTALEAAANPALPSDFKTILD"
repeat_region   27341. .27628
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   27650. .29650
                /note="VNTR"
                /rpt_type=tandem
                /rpt_unit="gagtgaggcgccgtccccgggtgtccctgtcacgtgcagggt"
repeat_region   complement(31926. .32203)
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   complement(32555. .32835)
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   33569. .33861
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   complement(33873. .33967)
                /rpt_family="MER33"
                /rpt_type=dispersed
repeat_region   34111. .35453
                /note="VNTR"
                /rpt_type=tandem
repeat_region   complement(35993. .36281)
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   41485. .42324
                /note="VNTR"
                /rpt_type=tandem
                /rpt_unit="gtgggattggttttcatgtgtggggtaggtggggatct"
repeat_region   complement(42656. .42954)
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   43312. .43618
                /rpt_family="MER1"
                /rpt_type=dispersed
repeat_region   44225. .45115
                /note="VNTR"
                /rpt_type=tandem
                /rpt_unit="ggggtctgatgtgtggtgactgtggatggcggtcgt"
repeat_region   complement(48015. .48300)
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   complement(49647. .50276)
                /rpt_family="L1"
                /rpt_type=dispersed
repeat_region   53945. .54234
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   57630. .58735
                /note="VNTR"
                /rpt_type=tandem
                /rpt_unit="tgggcatctgcgtccactccccctctcctg"
repeat_region   61332. .61597
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   complement(66201. .66323)
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   66389. .66661
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   66661. .66736
                /note="microsatellite"
                /rpt_type=tandem
                /rpt_unit="aag"
repeat_region   67140. .67583
                /rpt_family="MLT1"
                /rpt_type=dispersed
mRNA            complement(join(<73264. .73377,73925. .74084,76991. .77080,
                78049. .78146,79551. .79661,81950. .82131,84724. .84861,
                85253. .85393,91909. .>92068))
                /product="sodium channel-like protein"
CDS             complement(join(73264. .73377,73925. .74084,76991. .77080,
                78049. .78146,79551. .79661,81950. .82131,84724. .84861,
                85253. .85393,91909. .92068))
                /codon_start=1
                /evidence=not_experimental
                /product="sodium channel-like protein"
                /protein_id="AAG23290.1"
                /db_xref="GI:12642958"
```

Query Match 9.2%; Score 379.8; DB 9; Length 92564;
Best Local Similarity 75.0%; Pred. No. 7.8e-65;
Matches 542; Conservative 0; Mismatches 172; Indels 9; Gaps 5;

```
Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||| |||||||||||  ||| || | ||||||||||| |||||||| || ||||||
Db      11380 AGTTCGAGACCAGCCTGGCCAATATAGCGAAACCCCATCTGTACTAAAAACACAAAAATT 11439

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   ||  |||||| ||| |||||||||||||| ||||  ||||||  ||||||||||
Db      11440 AGCTGGGTGTGGTGACACATGCCTGTAATCCCAGGTACTCAGGAGGCTAAGGCAGGAGAA 11499

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              | |||||||  |  ||  | || |||     |   ||||||    ||| ||||||||||
Db      11500 TTGCTTGAACTGGGAGGCAGAGGTGGT-----GAGCCAAGATTGCACCAGTGCACTCCAG 11554

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  ||  || |||| ||||    ||||||| ||| ||| |||   | |  ||||| |||
Db      11555 CCTTGGTGAC-AGAGTGAAACTCCATCTCAAAAACAAACAAACAAAATACATATACATAA 11613

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              ||||||||  | | | | | ||| | ||| ||||||| ||| ||||||| |  |||||||
Db      11614 ATATATATGCACATATATATACATATATAAATATATATACACATATATAAATCTATATAC 11673

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACATATAT-ATACACA 358
              |||||||||  ||||| ||| ||||| ||  | |||||||| || |||||||| ||| |
Db      11674 ATATATACATATATACACATATATAAATCTATATACATATATATACATATATAATATATT 11733

Qy        359 TATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATA 418
              ||  |||| | |||||||||| ||||| | | |||||||||| | | ||||| | | ||| |
```

```
Db      11734 TACATATATAAATATATACATATATAAATATACATATATAAATACATATATAAATATACA 11793

Qy        419 TACACATATATACACATATATACACATATACACATATATACACATATATACATATATACA 478
              || | | ||||| ||||||||| | ||| | | ||| ||||| ||||| | ||| ||| |
Db      11794 TATATAAATATATACATATATAAATATACATATATAAATACATATATAAATATACATATA 11853

Qy        479 CATATATATA-ATATACACACATATATATACACATATATACACACATATATACACATATA 537
                | ||||||| ||||| | | ||| ||||| | ||||||||| | ||| ||| |||||||
Db      11854 TAAATATATACATATATAAATATACATATATAAATATATACATATATAAATATACATATA 11913

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATACATATA 597
              |||| ||||| ||| ||||||| | | ||| ||| ||||| ||| ||||| ||| | |||
Db      11914 TACATATATAAATATACATATACATATATAAATATATATACACATATATAAATATACATA 11973

Qy        598 TACACATATATACATATATACACATATATACATATATACACACATATATACACATACATA 657
              || | ||||| | ||||| | | ||| ||| ||| ||| ||| ||||| | | ||| | |
Db      11974 TATAAATATACATATATAAATATATACATATATAAATATACATATATAAATATATATACA 12033

Qy        658 TACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATA 717
              || | | | ||| | ||| | |  |||| ||| |||||||| |||||| |||||||||
Db      12034 TATATAAATATATAAATATACAAGTATATACAAATATATACATATATAAATGTATATACG 12093

Qy        718 TAT 720
              |||
Db      12094 TAT 12096


RESULT 16
AC114955/c
LOCUS       AC114955               161252 bp    DNA     linear   HTG 14-MAR-2002
DEFINITION  Homo sapiens chromosome 5 clone RP11-117B23, WORKING DRAFT
            SEQUENCE, 9 unordered pieces.
ACCESSION   AC114955
VERSION     AC114955.1  GI:19424416
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 161252)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 161252)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-MAR-2002) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 451340
            Center clone name: RPCI-11_117B23
            -------
            Summary Statistics
            Consensus quality: 148939 bases at least Q40
            Consensus quality: 156608 bases at least Q30
            Consensus quality: 158826 bases at least Q20
            Estimated insert size: 175000; agarose-fp estimation
            Estimated insert size: 160452; sum-of-contigs estimation
            Quality coverage:  7.26 in Q20 bases; agarose-fp estimation
            Quality coverage:  7.92 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     3008: contig of 3008 bp in length
            *     3009     3108: gap of unknown length
            *     3109     9209: contig of 6101 bp in length
            *     9210     9309: gap of unknown length
            *     9310    15613: contig of 6304 bp in length
            *    15614    15713: gap of unknown length
            *    15714    23243: contig of 7530 bp in length
            *    23244    23343: gap of unknown length
            *    23344    39674: contig of 16331 bp in length
            *    39675    39774: gap of unknown length
            *    39775    60001: contig of 20227 bp in length
            *    60002    60101: gap of unknown length
            *    60102    85303: contig of 25202 bp in length
            *    85304    85403: gap of unknown length
            *    85404   117460: contig of 32057 bp in length
            *   117461   117560: gap of unknown length
            *   117561   161252: contig of 43692 bp in length.
FEATURES             Location/Qualifiers
     source          1. .161252
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="RP11-117B23"
                     /clone_lib="RPCI human BAC library 11"
ORIGIN

  Query Match             9.2%;  Score 379.4;  DB 2;  Length 161252;
  Best Local Similarity   68.7%;  Pred. No. 8.6e-65;
  Matches  600;  Conservative    0;  Mismatches  256;  Indels   17;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||| |||||||||||  ||| || | ||||||||||||| |||||||||  || ||||||
Db     149753 AGTTCGAGACCAGCCTGGCCAATATAGCGAAACCCCATCTGTACTAAAAACACAAAAATT 149694

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   ||  |||||| ||| |||||||||||||| ||||  ||||||  ||||||||||
Db     149693 AGCTGGGTGTGGTGACACATGCCTGTAATCCCAGGTACTCAGGAGGCTAAGGCAGGAGAA 149634

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              |  |||||||  |  ||  | ||  |||     |   ||||||    ||| ||||||||||
Db     149633 TTGCTTGAACTGGGAGGCAGAGGTGGT-----GAGCCAAGATTGCACCAGTGCACTCCAG 149579

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||   ||  || |||| ||||    ||||||| ||| ||| |||    | |  ||||| |||
Db     149578 CCTTGGTGAC-AGAGTGAAACTCCATCTCAAAAACAAACAAACAAAATACATATACATAA 149520

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              ||||||||  | | | | ||| | ||| |||||||| ||| |||||||  |  |||||||
Db     149519 ATATATATGCACATATATATACATATATAAATATATATACACATATATAAATCTATATAC 149460

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACATATATATA----- 354
              |||||||| ||||| ||| ||||| ||  | |||||||| || | | |||||||
Db     149459 ATATATACATATATACACATATATAAATCTATATACATATATATATACATATATAATATA 149400

Qy        355 --CACATATGTATACACATATATACACATATATACACACATATATACACACATATATACACA 412
                |||||| || | | ||| ||| | | ||||||| | ||| ||||| ||||| | | |
Db     149399 TTTACATATATAAATATATACATATATAAATATACATATATAAATACATATATAAATATA 149340

Qy        413 CATATATACACATATATACACATATATACACATATACACATATATACACATATATACATA 472
              |||||||| | ||||| | | ||| ||| ||||||| | ||| ||| | ||| |||||||
Db     149339 CATATATAAATATATACATATATAAATATACATATATAAATACATATATATAAATACATA 149280

Qy        473 TATACACATATATAT---AATATACACACATATATATACACATATATACACACATATATA 529
              |||| | ||| ||||   |||||| ||| ||||| ||| |||||||||| | | ||| |||
Db     149279 TATAAATATACATATATAAATATATACATATATAAATATACATATATAAATATATACATA 149220

Qy        530 CACATATATACACATATATATACACATATATACACATATATACATATATACACATATATA 589
              | | ||||||| ||||| ||| | | ||||||| ||| ||| ||| ||| | ||| | |
Db     149219 TATAAATATACATATATACATATATAAATATACATATACATATATAAATATATATACACA 149160

Qy        590 TACATATATACACATATATACATATATACACATATATACATATATACACACATATATACA 649
              || ||| ||| ||||||||| ||||| | | ||| ||| ||| ||| | | ||||| | |
```

```
Db     149159 TATATAAATATACATATATAAATATACATATATAAATATATACATATATAAATATACATA 149100

Qy        650 CATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATATG 709

Db     149099 TATAAATATATATACATATATAAATATATAAATATACAAGTATATACAAATATATACATA 149040

Qy        710 TATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGG 769

Db     149039 TATAAACATATATACGTATATACGTATATACATGTATATACGTATATACGTATATACATG 148980

Qy        770 AAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGT 829

Db     148979 TATATACGTATATACGTATATACATGTATATACGTATATATATAAATGTATATACGTATA 148920

Qy        830 AGGGGAAATACATAAATGTGTAATAAAACCTAT 862

Db     148919 TACATATATATATAAATATATAAAAAAACTTTT 148887


RESULT 17
AC114291/c
LOCUS       AC114291               202305 bp    DNA     linear   PRI 15-MAR-2003
DEFINITION  Homo sapiens chromosome 5 clone CTD-3080P12, complete sequence.
ACCESSION   AC114291
VERSION     AC114291.2  GI:28973811
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 202305)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 202305)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAR-2002) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 202305)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-MAR-2003) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Mar 15, 2003 this sequence version replaced gi:19224940.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 99.9% of Sequence;
            Estimated Total Number of Errors is 0.2.
            NOTE: Transposon sequencing failed to verify number of repeat
            copies 173000-175780. Unsure number of repeat copies 173000-175780.
            BAC and subclones unstable 76080. Force join at 76080.
FEATURES             Location/Qualifiers
     source          1. .202305
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTD-3080P12"
     misc_feature    76080
                     /note="NOTE: BAC and subclones unstable 76080. Force join
                     at 76080."
     misc_feature    173000. .175780
                     /note="NOTE: Transposon sequencing failed to verify number
                     of repeat copies 173000-175780. Unsure number of repeat
                     copies 173000-175780"
ORIGIN

  Query Match             9.2%;  Score 378;  DB 9;  Length 202305;
  Best Local Similarity   66.3%;  Pred. No. 1.6e-64;
  Matches  640;  Conservative    0;  Mismatches  310;  Indels   16;  Gaps    6;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59

Db     192244 AGTTCGAGACCAGCCTGGCCAATATAGCGAAACCCCATCTGTACTAAAAACACAAAAATT 192185

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119

Db     192184 AGCTGGGTGTGGTGACACATGCCTGTAATCCCAGGTACTCAGGAGGCTAAGGCAGGAGAA 192125

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179

Db     192124 TTGCTTGAACTGGGAGGCAGAGGTGGT-----GAGCCAAGATTGCACCAGTGCACTCCAG 192070

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239

Db     192069 CCTTGGTGAC-AGAGTGAAACTCCATCTCAAAAACAAACAAACAAAATACATATACATAA 192011

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299

Db     192010 ATATATATGCACATATATATACATATATAAATATATATACACATATATAAATCTATATAC 191951

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACAT-----ATATATA 354

Db     191950 ATATATACATATATACACATATATAAATCTATATACATATATATACATATATAATATATT 191891

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414

Db     191890 TACATATATAAATATATACATATATAAATATACATATATAAATACATATATAAATATACA 191831

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474

Db     191830 TATATAAATATATACATATATAAATATACATATATAAATACATATATAAATATACATATA 191771

Qy        475 TACACATATATATA---ATATACACACATATATATACACATATATACACACATATATACA 531

Db     191770 TAAATATATACATATATAAATATACATATATAAATATATACATATATAAATATACATATA 191711

Qy        532 CATATATACACATATATATACACATATATACACATATATACATATATACACATATATATA 591

Db     191710 TACATATATAAATATACATATACATATATAAATATATATACACATATATAAATATACATA 191651

Qy        592 CATATATACACATATATACATATATACACATATATACATATATACACACATATATACACA 651

Db     191650 TATAAATATACATATATAAATATATACATATATAAATATACATATATAAATATATACATA 191591

Qy        652 TACATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTA 711

Db     191590 TATATAAATATATAAATATACAAGTATATACAAATATATACATATATAAACATATATACG 191531

Qy        712 TATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAA 771

Db     191530 TATATACATGTATATACGTATATACGTATATACATGTATATACGTACGTATATACATGTA 191471

Qy        772 AAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAG 831

Db     191470 TATACGTATATACGTATATACATGTATATACGTATATACGTATATACATGTATATACGTA 191411

Qy        832 GGGAAATACATAAATGTGTAATAAAACCTATCT-GTAAGGTAGTTAAGAAGGTAACACTA 890

Db     191410 TATACGTATATATATAAATGTATATACGTATATACGTATATATATAAATGTATATACGTA 191351

Qy        891 TATATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTAT 950

Db     191350 TATACATATATATATAAATATATAAAAAAACTTTTGGCTGGGCACCTTTCCAAATCTCAT 191291

Qy        951 AGAAGA 956

Db     191290 GGCACA 191285


RESULT 18
AC021955/c
LOCUS       AC021955               192139 bp    DNA     linear   PRI 22-MAR-2002
```

```
DEFINITION  Homo sapiens chromosome 8, clone RP11-238G17, complete sequence.
ACCESSION   AC021955
VERSION     AC021955.11  GI:19683766
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 192139)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 8, clone RP11-238G17
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 192139)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 192139)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C.,
            Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-FEB-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 192139)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 22, 2002 this sequence version replaced gi:18875268.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L5949
                Center clone name: 238_G_17
            --------------.
FEATURES             Location/Qualifiers
     source          1. .192139
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="RP11-238G17"
                     /clone_lib="RPCI-11 Human Male BAC"
     unsure          441. .445
                     /note="<30 qual SNGL region"
     unsure          455. .461
                     /note="<30 qual SNGL region"
     repeat_region   complement(1301. .1824)
                     /rpt_family="MLT1F"
     repeat_region   complement(1885. .2017)
                     /rpt_family="L2"
     repeat_region   2167. .2264
                     /rpt_family="GA-rich"
     repeat_region   2333. .2853
                     /rpt_family="(TA)n"
     repeat_region   complement(2862. .3156)
                     /rpt_family="AluSq"
     repeat_region   3916. .3974
                     /rpt_family="MLT1F"
     repeat_region   3977. .4182
                     /rpt_family="AluJo"
     repeat_region   complement(4553. .4782)
                     /rpt_family="MIR"
     repeat_region   5326. .5402
                     /rpt_family="LTR33"
     repeat_region   5570. .5756
                     /rpt_family="LTR33"
     repeat_region   complement(6278. .7151)
                     /rpt_family="MER4D"
     repeat_region   7424. .7644
                     /rpt_family="MER58A"
     repeat_region   7701. .7999
                     /rpt_family="AluY"
     repeat_region   9337. .9635
                     /rpt_family="AluSg"
     repeat_region   9723. .9878
                     /rpt_family="L1M4c"
     repeat_region   9877. .10902
                     /rpt_family="L1M4c"
     repeat_region   10903. .11141
                     /rpt_family="AluSc"
     repeat_region   11142. .11261
```

```
                     /rpt_family="L1M4c"
     repeat_region   11261. .11381
                     /rpt_family="L1MB4"
     repeat_region   11383. .11839
                     /rpt_family="MER51A"
     repeat_region   11956. .12100
                     /rpt_family="L1MB4"
     repeat_region   12101. .12132
                     /rpt_family="(CA)n"
     repeat_region   12133. .12496
                     /rpt_family="L1MB4"
     repeat_region   12994. .13074
                     /rpt_family="L2"
     repeat_region   13900. .13930
                     /rpt_family="AT_rich"
     repeat_region   14087. .14128
                     /rpt_family="(TG)n"
     repeat_region   complement(15008. .15212)
                     /rpt_family="MIR"
     repeat_region   15423. .15924
                     /rpt_family="MER87"
     repeat_region   16247. .16353
                     /rpt_family="MLT1I"
     repeat_region   17718. .17749
                     /rpt_family="(TTA)n"
     repeat_region   complement(18035. .18137)
                     /rpt_family="MIR3"
     repeat_region   18182. .18418
                     /rpt_family="MIR"
     repeat_region   18433. .18573
                     /rpt_family="L3"
     repeat_region   complement(18574. .18883)
                     /rpt_family="AluSg"
     repeat_region   18884. .18963
                     /rpt_family="L3"
     repeat_region   19497. .19592
                     /rpt_family="MIR"
     repeat_region   19909. .20082
                     /rpt_family="L3"
     repeat_region   complement(20319. .20627)
                     /rpt_family="AluSx"
     repeat_region   20785. .20932
                     /rpt_family="MLT1A1"
     unsure          20910. .20953
                     /note="<30 qual SNGL region"
     repeat_region   20933. .21252
                     /rpt_family="AluSg"
     unsure          20944
                     /note="probably G, possibly C"
     unsure          20965. .21026
                     /note="<30 qual SNGL region"
     repeat_region   21253. .21466
                     /rpt_family="MLT1A1"

  Query Match           9.1%;  Score 373.6;  DB 9;  Length 192139;
  Best Local Similarity   74.2%;  Pred. No. 1.1e-63;
  Matches  539;  Conservative    0;  Mismatches  179;  Indels    8;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAAT-ACAAAATT 59
              ||||| |||||  ||||  ||||||| ||||||||||||||||||||||| | ||||||
Db       3081 AGTTCGAGACCCACCTGGCCAACATGATGAAACCCCATCTCTACTAAAAATGAAAAAATT 3022

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGA-GGCCGAGGCAGGAGA 118
              ||   ||  |||||  |   || | ||||||||||||||  ||| ||| ||||||| |||
Db       3021 AGCTGGGTGTGGTGGCAGGTGCATCTAATCCCAGCTACTCTGGAGGGCTGAGGCAGAAGA 2962

Qy        119 ATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCC 177
              ||  ||||||||  | || |||||| ||||| | ||| | |||||  ||||||||||||
Db       2961 ATTGCTTGAACC--CAGGAGGTGGACGTTGCAGTGAGCCGAGATCGCGCCATTGCACTCC 2904

Qy        178 AGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATAT 237
              ||||  |||||||||||| |||     ||||  |||||||||          |||||||||
Db   .    2903 AGCCTGGGCAACAAGAGCAAAACTCTGTCTCGAAAAAAAAAATTTTATAAATATATATAT 2844

Qy        238 ATATATATATATATACACACACACACATATATATATACACATATATATACACATATAT 297
              | |||||||||||| | | | | | | ||||| ||||| | | |||||||| | ||||||
Db       2843 AAATATATATATAAATATATATAAATATATATAAATATATATAAATATATATAAATATAT 2784

Qy        298 ACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACAC 357
              | ||| ||| | ||| ||| | ||||| ||| | ||| |||| || | | ||||| | |
Db       2783 ATATAAATATATATAAATATATATATAAATATATATAAATATATATATAAATATAAATAA 2724

Qy        358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417
              |||| |||| | ||| ||| | ||||| | |    ||||| | | ||| ||| | | ||||
Db       2723 ATATATATAAATATAAATAAATATATAAATATGTATATAAATATATAAATATATAAATAT 2664

Qy        418 ATACACATATATACACATATATACACATATACACATATATACACATATATACATATATAC 477
              || | | ||||| | ||| ||| | | |||   ||||| | | ||||| | |||||||
Db       2663 GTATATAAATATATATATAAATATATAAATATGTATATAAATATATATAAATATATATAA 2604

Qy        478 ACATATATATA---ATATACACACATATATATACACATATATACACACATATATACACAT 534
              | || ||||||   ||||| | | || |||||| | ||||||| | | ||||||| | |
Db       2603 ATATGTATATAAATATATATAAATATGTATATAAATATATATATAAAAATATATATAAAA 2544

Qy        535 ATATACACATATATATACACATATATACACATATATACATATATACACATATATATACAT 594
              ||||| | | ||||||| | ||||||| | ||||||| | ||||| | |||||||||| |
Db       2543 ATATATATAAATATATAAATATATATATAAATATATATAAATATATAAATATATATATAA 2484

Qy        595 ATATACACATATATACATATATACACATATATACATATATACACACATATATACACATAC 654
              ||||| | | ||||| ||||| | | ||||| | ||||||| | | ||||||| | |||
Db       2483 ATATATATAAATATATATATAAATATATATAAATATATATAAATATATATATAAATATAT 2424

Qy        655 ATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATAT 714
              ||||| | | | ||| | ||| ||||| ||| | | |||||||  |||||||| || ||
Db       2423 ATATAAATATATATAAAAATATATATAAATATATAAATATATAAATATATATATATAAAT 2364

Qy        715 ATATAT 720
              ||||||
Db       2363 ATATAT 2358
```

```
RESULT 19
AC106873/c
LOCUS       AC106873               99596 bp    DNA     linear   PRI 31-JAN-2004
DEFINITION  Homo sapiens BAC clone RP11-374M7 from 7, complete sequence.
ACCESSION   AC106873
VERSION     AC106873.3  GI:18698877
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 99596)
  AUTHORS   Hillier,L.W., Fulton,R.S., Fulton,L.A., Graves,T.A., Pepin,K.H.,
            Wagner-McPherson,C., Layman,D., Maas,J., Jaeger,S., Walker,R.,
            Wylie,K., Sekhon,M., Becker,M.C., O'Laughlin,M.D., Schaller,M.E.,
            Fewell,G.A., Delehaunty,K.D., Miner,T.L., Nash,W.E., Cordes,M.,
            Du,H., Sun,H., Edwards,J., Bradshaw-Cordum,H., Ali,J., Andrews,S.,
            Isak,A., Vanbrunt,A., Nguyen,C., Du,F., Lamar,B., Courtney,L.,
            Kalicki,J., Ozersky,P., Bielicki,L., Scott,K., Holmes,A.,
            Harkins,R., Harris,A., Strong,C.M., Hou,S., Tomlinson,C.,
            Dauphin-Kohlberg,S., Kozlowicz-Reilly,A., Leonard,S., Rohlfing,T.,
            Rock,S.M., Tin-Wollam,A.M., Abbott,A., Minx,P., Maupin,R.,
            Strowmatt,C., Latreille,P., Miller,N., Johnson,D., Murray,J.,
            Woessner,J.P., Wendl,M.C., Yang,S.P., Schultz,B.R., Wallis,J.W.,
            Spieth,J., Bieri,T.A., Nelson,J.O., Berkowicz,N., Wohldmann,P.E.,
            Cook,L.L., Hickenbotham,M.T., Eldred,J., Williams,D., Bedell,J.A.,
            Mardis,E.R., Clifton,S.W., Chissoe,S.L., Marra,M.A., Raymond,C.,
            Haugen,E., Gillett,W., Zhou,Y., James,R., Phelps,K., Iadanoto,S.,
            Bubb,K., Simms,E., Levy,R., Clendenning,J., Kaul,R., Kent,W.J.,
            Furey,T.S., Baertsch,R.A., Brent,M.R., Keibler,E., Flicek,P.,
            Bork,P., Suyama,M., Bailey,J.A., Portnoy,M.E., Torrents,D.,
            Chinwalla,A.T., Gish,W.R., Eddy,S.R., McPherson,J.D., Olson,M.V.,
            Eichler,E.E., Green,E.D., Waterston,R.H. and Wilson,R.K.
  TITLE     The DNA sequence of human chromosome 7
  JOURNAL   Nature 424 (6945), 157-164 (2003)
```

```
  MEDLINE   22737999
   PUBMED   12853948
REFERENCE   2  (bases 1 to 99596)
  AUTHORS   Cotton,M. and Grewal,N.
  TITLE     The sequence of Homo sapiens BAC clone RP11-374M7
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 99596)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JAN-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 99596)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-FEB-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 99596)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-FEB-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   6  (bases 1 to 99596)
  AUTHORS   Wilson,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-JAN-2004) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Feb 17, 2002 this sequence version replaced gi:18158390.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0374M07
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            The sequence of this clone was established as part of a mapping and
            sequencing collaboration between the NHGRI Chromosome 7 Mapping
            Project (Eric D. Green, Director), John D. McPherson in the
            Department of Genetics (Washington University), and the Washington
            University Genome Sequencing Center. For additional information
            about the map position of this sequence, see
            http://www.nhgri.nih.gov/DIR/GTB/CHR7 , send
            mailto:egreen@nhgri.nih.gov , or see http://genome.wustl.edu

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries. Genomics 51:1-8. The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at http://www.chori.org
            VECTOR:  pBACe3.6.
FEATURES                Location/Qualifiers
     source             1. .99596
                        /organism="Homo sapiens"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9606"
                        /chromosome="7"
                        /map="7"
                        /clone="RP11-374M7"
                        /clone_lib="RPCI-11"
     repeat_region      5. .138
                        /rpt_family="Alu"
     repeat_region      399. .604
                        /rpt_family="L1"
     repeat_region      605. .919
                        /rpt_family="Alu"
     repeat_region      920. .1039
                        /rpt_family="L1"
     repeat_region      1045. .1325
                        /rpt_family="Alu"
     repeat_region      1330. .1444
                        /rpt_family="(GAAA)n"
     repeat_region      1552. .1572
                        /rpt_family="AT_rich"
     repeat_region      1699. .2061
                        /rpt_family="MER2_type"
     repeat_region      2062. .2284
                        /rpt_family="Alu"
     repeat_region      2285. .2351
                        /rpt_family="MER2_type"
     repeat_region      2949. .3225
                        /rpt_family="Alu"
     repeat_region      3237. .3413
                        /rpt_family="Alu"
     repeat_region      3438. .3548
                        /rpt_family="L1"
     repeat_region      3549. .3859
                        /rpt_family="Alu"
     repeat_region      3860. .3891
                        /rpt_family="L1"
     repeat_region      3936. .4009
                        /rpt_family="CR1"
     repeat_region      5320. .5362
                        /rpt_family="CR1"
     repeat_region      5365. .5446
                        /rpt_family="AcHobo"
     repeat_region      5491. .5521
                        /rpt_family="MIR"
     repeat_region      5526. .5571
                        /rpt_family="CR1"
     rRNA               complement(5771. .5860)
                        /product="rRNA-5S_rRNA-related"
     repeat_region      5771. .5860
                        /rpt_family="5S"
     repeat_region      6245. .6555
                        /rpt_family="Alu"
     repeat_region      6785. .6812
                        /rpt_family="AT_rich"
     repeat_region      7501. .7800
                        /rpt_family="Alu"
     repeat_region      7864. .7884
                        /rpt_family="AT_rich"
     repeat_region      8393. .8707
                        /rpt_family="Alu"
     repeat_region      9733. .10028
                        /rpt_family="Alu"
     repeat_region      10324. .10521
                        /rpt_family="MIR"
     repeat_region      11074. .11279
                        /rpt_family="L2"
     repeat_region      11338. .11642
                        /rpt_family="Alu"
     repeat_region      11690. .12111
                        /rpt_family="CT-rich"
     repeat_region      12118. .12392
                        /rpt_family="Alu"
     repeat_region      12924. .13569
```

```
                     /rpt_family="ERV1"
     repeat_region   14427. .14642
                     /rpt_family="MIR"
     misc_feature    14589. .15717
                     /note="CpG_island (%GC=70.4, o/e=0.91, #CpGs=143)"
     gene            14788. .15137
                     /gene="ST7"
     mRNA            14788. .15137
                     /gene="ST7"
     repeat_region   14814. .14913
                     /rpt_family="GC_rich"
     CDS             14987. .>15137
                     /gene="ST7"
                     /note="Homo sapiens suppression of tumorigenicity 7 (ST7),
                     transcript variant a, mRNA.; H_NH0374M07.1
                     This gene was based on gi(11761623 11761625)
                     Continues as H_RG114A06.2.1"
                     /codon_start=1
                     /product="unknown"
                     /protein_id="AAS07529.1"
                     /db_xref="GI:41473937"

  Query Match             9.0%;  Score 372;  DB 9;  Length 99596;
  Best Local Similarity   72.5%;  Pred. No. 2.6e-63;
  Matches  522;  Conservative    0;  Mismatches  195;  Indels    3;  Gaps    3;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              | ||| ||||||||||  |||||||  ||||| ||||||||||||||||||| ||||||
Db      93552 AATTCGAGACCAGCCTGGTCAACATGGCTAAACCTCATCTCTACTAAAAATACAAAAATT 93493

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              || ||  | |||||  |   |||||||||||||||||| |||||| ||||||||||||
Db      93492 AGCCACACGTGGTGGCATTTGCCTGTAATCCCAGCTACTCAGGAGGCTGAGGCAGGAGAA 93433

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              |  ||||||||||  |  |  | || |   | |   ||| ||||  ||| ||||||||||
Db      93432 TTGCTTGAACCTG-GGAAGCAGAAGTTGCAGTGAGCCAAAATCACACCACTGCACTCCAG 93374

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  ||  ||||      | |     |  || ||||||||||      | |||||||| ||
Db      93373 CCTGGGTGACAAAGCAAGACTCCATTGCCAAAAAAAAAAAAAAAAATATATATATAGAT 93314

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              | ||| ||| ||| | | | | | | ||||| ||| | ||||| ||| |   |||||
Db      93313 AGATAGATAGATATAGATATAGATATAGATATAGATATAGATATAGATATATGCATATAT 93254

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACAT 359
              ||| | ||| ||||  | | ||||| | | | || ||||| || ||| |||||| || ||
Db      93253 ATACACACATATATGCATATATATACACATATATGCATATATATACACATATATGCATAT 93194

Qy        360 ATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATAT 419
              || || ||| ||||  | | ||||  | | | ||  ||| | | | ||| | |||||| |
Db      93193 ATATACACATATATGCATATATATGCATATATATGCATATATACACATATATACATATGT 93134

Qy        420 ACACATATATACACATATATACACATATACACATATATACACATATATACATATATACAC 479
              |||||| | | ||||  ||| ||| |||||||||||||| |||||| || ||||||||
Db      93133 GTGTATATACATATATATGCACATATACACACATATATACATATATATGCACATATACAC 93074

Qy        480 ATATATATAATATACACACATATATATACACATATATACACACATATATACACATATATA 539
              | ||||||   ||| | |   | ||||||||| |||||  || |||||| ||||||||
Db      93073 ACATATAT-GCATATATATGCACATATACACACATATATGCATATATATGCACATATACA 93015

Qy        540 CACATATATATACACATATATACACATATATACATATATACACATATATATACATATATA 599
              |||||||||  | | ||||  ||| ||| | |||||||| || ||||||  ||||||| |
Db      93014 CACATATATGCATATATATGCACATATACACACATATATGCATATATATGCACATATACA 92955

Qy        600 CACATATATACATATATACACATATATACATATATACACACATATATACACATACATATA 659
              ||||||||| |||||||  || |||||||| | ||| |  ||||||||  || | ||| |
Db      92954 CACATATATGCATATATATGCACATATACACACATATATGCATATATATGCACATATACA 92895

Qy        660 CACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATA 719
              |||| | ||  ||||| ||  | ||| ||| | ||||| |  |||||||| | |||||||
Db      92894 CACATATATGCATATATATGCACATATACACACATATATGCGTATATATGCACATATATA 92835
```

```
RESULT 20
BX901949
LOCUS       BX901949               38096 bp    DNA     linear   PRI 24-MAY-2004
DEFINITION  Human DNA sequence from clone WI2-86554F10 on chromosome X,
            complete sequence.
ACCESSION   BX901949
VERSION     BX901949.9  GI:47601709
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 38096)
  AUTHORS   Whitehead,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAY-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On May 24, 2004 this sequence version replaced gi:46559263.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome X, constructed by the Sanger Centre Chromosome X Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/ChrX
            WI2-86554F10 is from a Whitehead human fosmid library VECTOR:
            pEpiFos-5_Eco72I.
FEATURES             Location/Qualifiers
     source          1. .38096
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /clone="WI2-86554F10"
                     /clone_lib="Whitehead_fosmid"
ORIGIN

  Query Match             9.0%;  Score 371.4;  DB 9;  Length 38096;
  Best Local Similarity   77.9%;  Pred. No. 3.9e-63;
  Matches  540;  Conservative    0;  Mismatches  136;  Indels   17;  Gaps    7;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              |||||||||||||||||  |||  |||||||||||||||| ||||||||||| |||||
Db      35292 AGTTCAAGACCAGCCTGGCCAATGTGGTGAAACCCCATCTCAACTAAAAATACAAAAATC 35351

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   ||| |||||  |   |||||||||||||||||||| |||||| ||||||||||||
Db      35352 AGCTGGGCGTGGTGGCAGGTGCCTGTAATCCCAGCTACTTGGGAGGCTGAGGCAGGAGAA 35411
```

```
Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGC-AAGATCATGCCATTGCACTCCA 178
              |  ||||||||||  || ||| |||||||| | ||||  |||    |||||||| || |
Db      35412 TTGCTTGAACCTG--GGAGGTAGAGGTTGCAGTGAGCGGAGATTGCACCATTGCATTCTA 35469

Qy        179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCAT---TTTATATAT 235
              |||  || ||| | || |||||   |||| | ||| ||| | |   ||    | |||||||
Db      35470 GCCTGGGTAACGAAAGTGAAATTCTATCTAAAAAACAAATATATATATACATATATATAT 35529

Qy        236 ATATATATATATATATACACACACACACAT---ATATATATACACATATATATACACA 292
              ||| |  | | | | |       | | ||| ||   ||||||| ||| ||||| ||| | |
Db      35530 ATACACGTGTGTGTGTGTGTATATACATATACCATATATACACATATATACATATATA 35589

Qy        293 TATATACATATATACACATATATACACATATATATACACATACATATGTACACATATATA 352
               |||||||||||||||||||||||||||||| ||| ||||| | || || ||| ||| |
Db      35590 CATATACATATATACACATATATACACATATACATATACATATACATATATACACATACA 35649

Qy        353 TACACATATGTATACACATATATACACATATATACACACATATATACACATATATACACA 412
              ||||||||| || | | ||| ||| ||| ||||| |||||| || ||| ||||       |
Db      35650 TACACATATATACATATATACATATACACATATATACACATGTACACATATAT------A 35703

Qy        413 CATATATACACATATATACACATATATACACATATACACATATATACACATATATACATA 472
              |||||||||||||||||||| ||||| ||||||| ||| |||||||||| ||||| |
Db      35704 CATATATACACATATATACACACATATATACATATATACACATATACACATAGATACACA 35763

Qy        473 TATACACATATATATAATATACACACATATATATACACATATATACACACATATATACAC 532
              |||||| |||||||  | ||| | ||||||||| ||| ||| ||| ||| ||||| |||
Db      35764 TATACAGATATATA-CACATATATACATATATACACATATACATATACATATATACACAT 35822

Qy        533 ATATATACACATATATATACACATATATACACATATATACATATATACACATATATATAC 592
              | ||| ||| ||||| ||| | | |||||||||||||||| ||||||||||||||| |
Db      35823 ACATACACATATATACATATATACATATACACATATATACACATATACACATATATACAT 35882

Qy        593 ATATATACACATATATACATATATACACATATATACATATATACACACATATATACACAT 652
              ||||| |||||||||||||||||||||||| ||||||| ||| | ||| | | | ||
Db      35883 ATATACACACATATATACATATATACACATATACACATATACACATAGATACACATATAT 35942

Qy        653 ACATATACACACACATAGATATACATATATATA 685
              ||||||| ||||| ||| |||     | | |||
Db      35943 ACATATATACACATATATATAATTTAAAAAATA 35975


RESULT 21
BX324168_0/c
WPCOMMENT
  Sequence split into 6 fragments  LOCUS BX324168 Accession BX324168
     Fragment Name      Begin     End
     BX324168_0             1     110000
     BX324168_1        100001     210000
     BX324168_2        200001     310000
     BX324168_3        300001     410000
     BX324168_4        400001     510000
     BX324168_5        500001     600124
LOCUS       BX324168               600124 bp    DNA     linear   HTG 13-JUN-2004
DEFINITION  Homo sapiens chromosome X clone XXyac-89H7, WORKING DRAFT SEQUENCE,
            31 unordered pieces.
ACCESSION   BX324168
VERSION     BX324168.8  GI:37777001
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 600124)
  AUTHORS   McLay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 20, 2003 this sequence version replaced gi:37650784.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: yX89H7
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 591433 bases at least Q40
            Consensus quality: 594175 bases at least Q30
            Consensus quality: 595562 bases at least Q20
            Insert size: 597124; sum-of-contigs
            Quality coverage: 6.16x in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 31 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    20165: contig of 20165 bp in length
            *    20166    20265: gap of 100 bp
            *    20266    44493: contig of 24228 bp in length
            *    44494    44593: gap of 100 bp
            *    44594    61605: contig of 17012 bp in length
            *    61606    61705: gap of 100 bp
            *    61706    67907: contig of 6202 bp in length
            *    67908    68007: gap of 100 bp
            *    68008    74311: contig of 6304 bp in length
            *    74312    74411: gap of 100 bp
            *    74412    79880: contig of 5469 bp in length
            *    79881    79980: gap of 100 bp
            *    79981    88557: contig of 8577 bp in length
            *    88558    88657: gap of 100 bp
            *    88658   112320: contig of 23663 bp in length
            *   112321   112420: gap of 100 bp
            *   112421   166393: contig of 53973 bp in length
            *   166394   166493: gap of 100 bp
            *   166494   227100: contig of 60607 bp in length
            *   227101   227200: gap of 100 bp
            *   227201   239356: contig of 12156 bp in length
            *   239357   239456: gap of 100 bp
            *   239457   242526: contig of 3070 bp in length
            *   242527   242626: gap of 100 bp
            *   242627   319013: contig of 76387 bp in length
            *   319014   319113: gap of 100 bp
            *   319114   327413: contig of 8300 bp in length
            *   327414   327513: gap of 100 bp
            *   327514   343545: contig of 16032 bp in length
            *   343546   343645: gap of 100 bp
            *   343646   391808: contig of 48163 bp in length
            *   391809   391908: gap of 100 bp
            *   391909   426746: contig of 34838 bp in length
            *   426747   426846: gap of 100 bp
            *   426847   440337: contig of 13491 bp in length
            *   440338   440437: gap of 100 bp
            *   440438   460971: contig of 20534 bp in length
            *   460972   461071: gap of 100 bp
            *   461072   465425: contig of 4354 bp in length
            *   465426   465525: gap of 100 bp
            *   465526   488190: contig of 22665 bp in length
            *   488191   488290: gap of 100 bp
            *   488291   497966: contig of 9676 bp in length
            *   497967   498066: gap of 100 bp
            *   498067   517262: contig of 19196 bp in length
            *   517263   517362: gap of 100 bp
            *   517363   567911: contig of 50549 bp in length
            *   567912   568011: gap of 100 bp
            *   568012   582171: contig of 14160 bp in length
            *   582172   582271: gap of 100 bp
            *   582272   584652: contig of 2381 bp in length
            *   584653   584752: gap of 100 bp
```

```
          *   584753   586940: contig of 2188 bp in length
          *   586941   587040: gap of 100 bp
          *   587041   589956: contig of 2916 bp in length
          *   589957   590056: gap of 100 bp
          *   590057   592460: contig of 2404 bp in length
          *   592461   592560: gap of 100 bp
          *   592561   597736: contig of 5176 bp in length
          *   597737   597836: gap of 100 bp
          *   597837   600124: contig of 2288 bp in length.
FEATURES             Location/Qualifiers
     source          1. .600124
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /clone="XXyac-89H7"
                     /clone_lib="Human-ICRF-YAC"
     misc_feature    1. .20165
                     /note="assembly_fragment:00356
                     fragment_chain:1"
     misc_feature    20266. .44493
                     /note="assembly_fragment:00589
                     fragment_chain:1"
     misc_feature    44594. .61605
                     /note="assembly_fragment:02258
                     fragment_chain:1"
     misc_feature    61706. .67907
                     /note="assembly_fragment:05479
                     fragment_chain:1"
     misc_feature    68008. .74311
                     /note="assembly_fragment:00274
                     fragment_chain:1"
     misc_feature    74412. .79880
                     /note="assembly_fragment:06012
                     fragment_chain:1"
     misc_feature    79981. .88557
                     /note="assembly_fragment:06208
                     fragment_chain:1"
     misc_feature    88658. .112320
                     /note="assembly_fragment:06075
                     fragment_chain:2"
     misc_feature    112421. .166393
                     /note="assembly_fragment:02111
                     fragment_chain:2"
     misc_feature    166494. .227100
                     /note="assembly_fragment:00544
                     fragment_chain:2"
     misc_feature    227201. .239356
                     /note="assembly_fragment:03674
                     fragment_chain:2"
     misc_feature    239457. .242526
                     /note="assembly_fragment:04283
                     fragment_chain:2"
     misc_feature    242627. .319013
                     /note="assembly_fragment:01306
                     fragment_chain:2"
     misc_feature    319114. .327413
                     /note="assembly_fragment:02688
                     fragment_chain:2"
     misc_feature    327514. .343545
                     /note="assembly_fragment:00658
                     fragment_chain:3"
     misc_feature    343646. .391808
                     /note="assembly_fragment:04025
                     fragment_chain:3"
     misc_feature    391909. .426746
                     /note="assembly_fragment:01126
                     fragment_chain:3"
     misc_feature    426847. .440337
                     /note="assembly_fragment:06618
                     fragment_chain:3"
     misc_feature    440438. .460971
                     /note="assembly_fragment:03859
                     fragment_chain:3"
     misc_feature    461072. .465425
                     /note="assembly_fragment:03896
                     fragment_chain:3"
     misc_feature    465526. .488190
                     /note="assembly_fragment:06799
                     fragment_chain:4"
     misc_feature    488291. .497966
                     /note="assembly_fragment:02743
                     fragment_chain:4"
     misc_feature    498067. .517262
                     /note="assembly_fragment:06602
                     fragment_chain:4"
     misc_feature    517363. .567911
                     /note="assembly_fragment:02958
                     fragment_chain:5"
     misc_feature    568012. .582171
                     /note="assembly_fragment:05773
                     fragment_chain:5"
     misc_feature    582272. .584652
                     /note="assembly_fragment:00466"
     misc_feature    584753. .586940
                     /note="assembly_fragment:02122"
     misc_feature    587041. .589956
                     /note="assembly_fragment:04085"

  Query Match             9.0%;  Score 371.4;  DB 2;  Length 110000;
  Best Local Similarity   77.9%;  Pred. No. 3.3e-63;
  Matches  540;  Conservative    0;  Mismatches  136;  Indels   17;  Gaps    7;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              |||||||||||||||||  |||  ||||||||||||||||| |||||||||| |||||
Db      82809 AGTTCAAGACCAGCCTGGCCAATGTGGTGAAACCCCATCTCAACTAAAAATACAAAAATC 82750

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   ||| |||||   |   |||||||||||||||||||| |||||| ||||||||||||
Db      82749 AGCTGGGCGTGGTGGCAGGTGCCTGTAATCCCAGCTACTTGGGAGGCTGAGGCAGGAGAA 82690

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGC-AAGATCATGCCATTGCACTCCA 178
              | |||||||||  || ||| ||||||| | ||||  ||||    |||||||| || |
Db      82689 TTGCTTGAACCTG--GGAGGTAGAGGTTGCAGTGAGCGGAGATTGCACCATTGCATTCTA 82632

Qy        179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAGGCAT---TTTATATAT 235
              |||  || ||| | || |||||   |||| | ||| ||| | |    ||   | |||||||
Db      82631 GCCTGGGTAACGAAAGTGAAATTCTATCTAAAAAACAAATATATATATACATATATATAT 82572

Qy        236 ATATATATATATATATACACACACACACAT---ATATATATACACATATATATACACA 292
              ||| |  | | | | |       | | ||| ||   ||||||| ||| ||||| ||| | |
Db      82571 ATACACGTGTGTGTGTGTGTATATACATATACCATATATACACATATATACATATATA 82512

Qy        293 TATATACATATATACACATATATACACATATATATACACATACATATGTACACATATATA 352
              |||||||||||||||||||||||||||||| ||| ||||| | || || ||| ||| |
Db      82511 CATATACATATATACACATATATACACATATACATATACATATATACACATACA 82452

Qy        353 TACACATATGTATACACATATATACACATATATACACACATATATACACATATATACACA 412
              ||||||||| || | | ||| ||| ||| ||||| |||||| || ||| ||||      |
Db      82451 TACACATATATACATATATACATATACACATATATACACATGTACACATATAT------A 82398

Qy        413 CATATATACACATATATACACATATATACACATATACACATATATACACATATATACATA 472
              |||||||||||||||||||| ||||| ||||||| ||| |||||||||| ||||| |
Db      82397 CATATATACACATATATACACACATATATACATATATACACATATACACATAGATACACA 82338

Qy        473 TATACACATATATATAATATACACACATATATATACACATATATACACACATATATACAC 532
              |||||| |||||||  | ||| | ||||||||| ||| ||| ||| ||| ||||| |||
Db      82337 TATACAGATATATA-CACATATATACATATATACACATATACATATATATACACAT 82279

Qy        533 ATATATACACATATATATACACATATATACACATATATACATATATACACATATATATAC 592
              | ||| ||| ||||| ||| | | |||||||||||||||| ||||||||||||||| |
Db      82278 ACATACACATATATACATATATACATATACACATATATACACATATACACATATATACAT 82219

Qy        593 ATATATACACATATATACATATATACACATATATACATATATACACACATATATACACAT 652
              ||||| ||||||||||||||||||||||||| ||||||| ||| | ||| | | | ||
Db      82218 ATATACACACATATATACATATATACACATATACACATATACACATAGATACACATATAT 82159
```

```
Qy        653 ACATATACACACACATAGATATACATATATATA 685
              ||||||| ||||| ||| |||     | | |||
Db      82158 ACATATATACACATATATATAATTTAAAAAATA 82126


RESULT 22
AC097326
LOCUS       AC097326               252512 bp    DNA     linear   HTG 15-JUN-2002
DEFINITION  Pan troglodytes clone RP43-10G9, WORKING DRAFT SEQUENCE, 14
            unordered pieces.
ACCESSION   AC097326
VERSION     AC097326.3  GI:21397202
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 252512)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 252512)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-OCT-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 252512)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JUN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jun 13, 2002 this sequence version replaced gi:21392442.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: ZUAK
                Center clone name: RP43-10G9
            -------------- Summary Statistics
                Sequencing vector: Plasmid; M77789
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 241130 bases at least Q40
                Consensus quality: 243666 bases at least Q30
                Consensus quality: 245673 bases at least Q20
                Estimated insert size: 259407; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 7.4x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
            * NOTE: This sequence may represent more than one clone.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 14 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    55034: contig of 55034 bp in length
            *    55035    55134: gap of unknown length
            *    55135    92628: contig of 37494 bp in length
            *    92629    92728: gap of unknown length
            *    92729   119376: contig of 26648 bp in length
            *   119377   119476: gap of unknown length
            *   119477   142089: contig of 22613 bp in length
            *   142090   142189: gap of unknown length
            *   142190   165691: contig of 23502 bp in length
            *   165692   165791: gap of unknown length
            *   165792   184091: contig of 18300 bp in length
            *   184092   184191: gap of unknown length
            *   184192   198445: contig of 14254 bp in length
            *   198446   198545: gap of unknown length
            *   198546   211127: contig of 12582 bp in length
            *   211128   211227: gap of unknown length
            *   211228   223003: contig of 11776 bp in length
            *   223004   223103: gap of unknown length
            *   223104   234877: contig of 11774 bp in length
            *   234878   234977: gap of unknown length
            *   234978   239452: contig of 4475 bp in length
            *   239453   239552: gap of unknown length
            *   239553   246173: contig of 6621 bp in length
            *   246174   246273: gap of unknown length
            *   246274   250219: contig of 3946 bp in length
            *   250220   250319: gap of unknown length
            *   250320   252512: contig of 2193 bp in length.
FEATURES             Location/Qualifiers
     source          1. .252512
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="RP43-10G9"
ORIGIN

  Query Match             9.0%;  Score 370.4;  DB 2;  Length 252512;
  Best Local Similarity   72.5%;  Pred. No. 4.6e-63;
  Matches  535;  Conservative    0;  Mismatches  196;  Indels    7;  Gaps    4;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAATTA 60
              ||||  |||||||||||| |||||||||||||||| |||||||||| |||||||||
Db     249133 AGTTTGAGACCAGCCTGACCAACATGGTGAAACCCCGTCTCTACTAAAACTACAAAATTA 249192
```

```
Qy         61 GACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAAT 120
              | ||||| ||||| | ||||||||||||  ||||||||| ||| || | |||| ||||||
Db     249193 GCCAGGCGTGGTGGTGCACGCCTGTAATATCAGCTACTTGGGAAGCTGGGGCAAGAGAAT 249252

Qy        121 CACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCCAG 179
               ||||||| |  | || |||||||| ||| | ||| |||||||   ||| ||||||||||
Db     249253 AACTTGAAGC--CGGGAGGTGGAGGCTGCAGTGAGCCAAGATCGCACCACTGCACTCCAG 249310

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              ||  | |||||||||| |||   ||||||  | ||  ||         |    ||||  |
Db     249311 CCTGGACAACAAGAGCAAAACTCCATCTCGAATAATTAATTTATTTATAATACTATAATT 249370

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATAC 299
              |       ||  |  |     |   |||   |  |||         |  | | | |||
Db     249371 AACCGGGCATGGAGGCGTGTGCCTGTATAGCCAGCTACTTGGGAGGCTGAAGAAAGATAA 249430

Qy        300 ATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACAT 359
                |  |  || | |||| | ||| ||||| | ||| |||| ||   |||||||| | ||
Db     249431 TCACTTGAACCTGTATATATATAGATATATAGATATATATATATCTATATATATATATAT 249490

Qy        360 ATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATAT 419
              || || | | ||||||| | ||||||| | | |||||||| | |||||||| | | ||||||
Db     249491 ATATAGATATATATATATATATATATATATATATATATATATATATATATATATATAT 249550

Qy        420 ACACATATATACACATATATACACATATACACATATATACACATATATACATATATACAC 479
              | | ||||||| |  |||||| | ||||| | ||||||| |  |||||| |||||||  |
Db     249551 ATATATATATATATCTATATATATATATATATATATATATATGTATATATATATATATAT 249610

Qy        480 ATATATATA---ATATACACACATATATATACACATATATACACACATATATACACATAT 536
              |||||||||   |||||  | | |||||||||  | |||||||  | | |||||||  | ||||
Db     249611 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 249670

Qy        537 ATACACATATATATACACATATATACACATATATACATATATACACA-TATATATACATA 595
              ||| | |||||||||  | |||||||  | |||||||  |||||||  | |  ||||||||  |||
Db     249671 ATATATATATATATATATATATATATATATATATATATATATATATAGTATATATATATA 249730

Qy        596 TATACACATATATACATATATACACATATATACATATATACACACATATATACACATACA 655
              ||||  | |||||||  |||||||  | |||||||  |||||||  |  | |||||||  | |||  |
Db     249731 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 249790

Qy        656 TATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATA 715
              ||||   | | |||  ||||||||||||||||  | |||||||||  |  ||||||||||||
Db     249791 TATATTTATATATATATATATACATATATATATATATATATATATTTTTATATATGTATATC 249850

Qy        716 TATATGCTCCAGAGTTCA 733
              | | |  | |   ||||
Db     249851 TTTCTCTTTCCAGATTCA 249868
```

```
RESULT 23
AC123545/c
LOCUS       AC123545               170646 bp    DNA     linear   HTG 25-JUL-2003
DEFINITION  Pan troglodytes clone RP43-165N9, WORKING DRAFT SEQUENCE, 10
            ordered pieces.
ACCESSION   AC123545
VERSION     AC123545.3  GI:33235839
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 170646)
  AUTHORS   Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
            Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S., Cariaga,K.,
            Chu,G., Coleman,B., Coleman,H., Engle,J., Granite,S., Guan,X.,
            Gupta,J., Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Hu,P.,
            Hurle,B., Idol,J.R., Karlins,E., Kwong,P., Laric,P., Lee-Lin,S.-Q.,
            Legaspi,R., Maduro,Q.L., Maduro,V.B., Margulies,E.H., Masiello,C.,
            Maskeri,B., McDowell,J., Paguirigan,C., Pearson,R., Portnoy,M.E.,
            Prasad,A., Reddix-Dugue,N., Schandler,K., Schueler,M.G., Shah,K.,
            Sison,C., Stantripop,S., Thomas,J.W., Thomas,P.J., Tsipouri,V.,
            Vogt,J.L., Wetherby,K.D., Young,A. and Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 170646)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAY-2002) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE   3  (bases 1 to 170646)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUL-2003) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     On Jul 25, 2003 this sequence version replaced gi:24414279.
            ---------- Genome Center
                 Center: NIH Intramural Sequencing Center
                 Center code: NISC
                 Web site: http://www.nisc.nih.gov
                 Contact: nisc_zoo@nhgri.nih.gov
            ---------- Project Information
                 Center project name: cmz
                 Center clone name: 165N09

            The sequence data in this record represents an 'enhanced'
            version of a Phase 2 submission. Specifically, the indicated
            order and orientation of each sequence contig has been
            established using one or more of the following: read-pair
            data from individual subclones, overlaps with neighboring
            clones, alignment with available reference sequence (e.g.,
            human), and/or confirmation by PCR testing. In addition,
            the sequence assembly is based on at least 8X average
            coverage in Q20 bases and has been reviewed to rule out
            gross misassemblies, the low-quality ends of sequence
            contigs have been trimmed away, and each base is associated
            with a Phrap-derived quality score.
            ---------- Summary Statistics
                 Sequencing vector: plasmid; n/a; 100% of reads
                 Chemistry: Dye-terminator Big Dye; 100% of reads
                 Assembly program: Phrap; version 0.990319
                 Consensus quality: 167110 bases at least Q40
                 Consensus quality: 168559 bases at least Q30
                 Consensus quality: 169363 bases at least Q20
                 Insert size: 152000; agarose-fp
                 Insert size: 169746; sum-of-contigs
                 Quality coverage: 11.31x in Q20 bases; agarose-fp
                 Quality coverage: 10.13x in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1     6384: contig of 6384 bp in length
            *    6385     6484: gap of unknown length
            *    6485    37411: contig of 30927 bp in length
            *   37412    37511: gap of unknown length
            *   37512    46351: contig of 8840 bp in length
            *   46352    46451: gap of unknown length
            *   46452    51816: contig of 5365 bp in length
            *   51817    51916: gap of unknown length
            *   51917    67976: contig of 16060 bp in length
            *   67977    68076: gap of unknown length
            *   68077   114680: contig of 46604 bp in length
            *  114681   114780: gap of unknown length
            *  114781   121651: contig of 6871 bp in length
            *  121652   121751: gap of unknown length
            *  121752   125931: contig of 4180 bp in length
            *  125932   126031: gap of unknown length
            *  126032   128356: contig of 2325 bp in length
            *  128357   128456: gap of unknown length
```

```
          *   128457   170646: contig of 42190 bp in length.
FEATURES             Location/Qualifiers
     source          1. .170646
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="RP43-165N9"
                     /clone_lib="RP43"
     misc_feature    1. .6384
                     /note="assembly_fragment
                     missing T7 clone end on 5' end of insert"
     misc_feature    6485. .37411
                     /note="assembly_fragment"
     misc_feature    37512. .46351
                     /note="assembly_fragment"
     misc_feature    46452. .51816
                     /note="assembly_fragment"
     misc_feature    51917. .67976
                     /note="assembly_fragment"
     misc_feature    68077. .114680
                     /note="assembly_fragment"
     misc_feature    114781. .121651
                     /note="assembly_fragment"
     misc_feature    121752. .125931
                     /note="assembly_fragment"
     misc_feature    126032. .128356
                     /note="assembly_fragment"
     misc_feature    128457. .170646
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
ORIGIN

  Query Match             8.9%;  Score 368.2;  DB 2;  Length 170646;
  Best Local Similarity   74.9%;  Pred. No. 1.3e-62;
  Matches  543;  Conservative    0;  Mismatches  168;  Indels   14;  Gaps    6;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||| |||||||||||  ||| || | ||||||||||||| ||| ||||| || ||||||
Db     103362 AGTTCGAGACCAGCCTGGCCAATATAGCGAAACCCCATCTGTACCAAAAACACAAAAATT 103303

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              ||   ||  |||||| ||| ||||||||||||| ||||  ||||||  ||||||||||
Db     103302 AGCTGGGTGTGGTGACACATGCCTGTAATCCCAGGTACTCAGGAGGCTAAGGCAGGAGAA 103243

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCAG 179
              | |||||||  |  ||  | || |||     |   ||||||    ||| || |||||||
Db     103242 TTGCTTGAACTGGGAGGCAGAGGTGGT-----GAGCCAAGATTGCACCAGTGTACTCCAG 103188

Qy        180 CCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATAT 239
              |  ||  || |||| ||||   ||||||| ||| ||| |||   | |  ||||| |||
Db     103187 GCTTGGTGAC-AGAGTGAAACTCCATCTCAAAAACAAACAAACAAAATACATATACATAA 103129

Qy        240 ATATATATATATACACACACACACATATATATATACACATATATATACACATATATA- 298
              ||||||||| | | | | ||| | ||| ||||||| || ||||||||||  ||||| |
Db     103128 ATATATATACACATATATATACATATATAAATATATATACCTATATATACATATATACAT 103069

Qy        299 ----CATATATACACATATATACACATATATATACACATACATATGTACACATATATATA 354
                  ||||||| |||||||||| |||||||||   ||| |||  || | | ||||| |
Db     103068 ATATAATATATATACATATATACATATATATATA-TTATATATACATATATAAATATACA 103010

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              | |||  |||||| ||| | | | ||| ||| | | |||||||||  |||| | | |||
Db     103009 TATATAAATATACATATAAAAATATATACATATAAAAATATATACATGTATAAATATACA 102950

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              | |||| | ||| ||| | | ||||| ||||||| ||||||||| | ||| ||| ||| |
Db     102949 TGTATAAATATACATATATAAATATATACATATATACATATATAAATATACATATATAAA 102890

Qy        475 TACACATATATATAATATACACACATATATATACACATATATACACACATATATACACAT 534
              || | | |||||||| ||| |||  |||| ||| ||||||||| | | ||| ||| | |
Db     102889 TATATACATATATAA-ATATACATGTATAAATATACATATATAAATATATACATATATAA 102831
```

```
Qy        535 ATATACACATATATATACACATATATACACATATATACATATATACACATATATATACAT 594
              ||||||| ||||| ||| | | ||||| | |||| |||| ||||| ||||||||| | ||
Db     102830 ATATACATATATAAATATATAAATATATAAATATGTACAAATATATACATATATAAATAT 102771

Qy        595 ATATACACATATATACATATATACACATATATACATATATACACACATATATACACATAC 654
              |||||||||||||||| ||||||| | |||||||||||||||||| | ||| ||| | ||||
Db     102770 ATATACACATATATAGATATATAAATATATATACATATATACATATATAAATATATATAC 102711

Qy        655 ATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATAT 714
              ||||| | | | ||| | ||| ||| ||||||| ||||| |||  |||| |||| || ||
Db     102710 ATATATAAATATATATACATATATAAATATACATATATAAATATATATACATATATAAAT 102651

Qy        715 ATATA 719
              || ||
Db     102650 ATGTA 102646
```

```
RESULT 24
HS323B6
LOCUS       HS323B6               112762 bp    DNA     linear   PRI 05-MAR-2003
DEFINITION  Human DNA sequence from clone RP3-323B6 on chromosome Xq11-12,
            complete sequence.
ACCESSION   Z83841
VERSION     Z83841.1  GI:1871528
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 112762)
  AUTHORS   McDougall,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAR-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Mar 10, 1997 this sequence version replaced gi:1772950.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome X, constructed by the Sanger Centre Chromosome X Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/ChrX
            RP3-323B6 is from the library RPCI-3 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pCYPAC2.
FEATURES             Location/Qualifiers
     source          1. .112762
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
```

```
                     /db_xref="RZPD:RPCIP704B06323"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="q11-12"
                     /clone="RP3-323B6"
                     /clone_lib="RPCI-3"
ORIGIN

  Query Match             8.9%;  Score 367;  DB 9;  Length 112762;
  Best Local Similarity   67.1%;  Pred. No. 2.4e-62;
  Matches  599;  Conservative    0;  Mismatches  280;  Indels   14;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC---AAAA 57
              |||||||||||||| ||   |||| ||||||||||||||||||||| ||||||   ||||
Db      39093 AGTTCAAGACCAGCTTGGCTAACACGGTGAAACCCCATCTCTACTATAAATACAAAAAAA 39152

Qy         58 TTAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAG 117
              |||| |||||||||||  |  |||||||||| |||| |  ||  ||||| ||||||| ||
Db      39153 TTAGCCAGGCATGGTGGCATGCGCCTGTAATTCCAGTTGTTTGCGAGGCTGAGGCAGAAG 39212

Qy        118 AATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGA-GCAAGATCATGCCATTGCACTC 176
              |||| ||||||||  | || ||||||| |||| | || |  |||||   ||| ||||||
Db      39213 AATCGCTTGAACC--CAGGAGGTGGAGCTTGCAGTGAGGTGAGATCGCACCACTGCACTT 39270

Qy        177 CAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATA 236
              |||||  |   | | |   | |  | |  | | | |                | ||||| ||
Db      39271 CAGCCTGGATGAACAAA---AGAATTGAATTAATACTATTGCTTTTCACATATATATGTA 39327

Qy        237 TATATATATATATATACACACACACACACATATATATATACACATATATATACACATATA 296
              | | ||||||||||||               |||||| |||| | ||| |||||  ||| |||
Db      39328 TGTGTATATATATATATGTGTGTGTGTGTATATATGTATATATATACATATATACACATA 39387

Qy        297 TACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACA 356
              |  ||||||| |||||||||||||||| |||| | |||||||| ||||||||| |||| |
Db      39388 TGTATATATATACATATATACACATATGTATATATATACATATATACACATATGTATATA 39447

Qy        357 CATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATA 416
              |||  |||| ||| ||||  | ||||||||| | ||| | | |  |||| | | |||||
Db      39448 TATACATATATACACATATGTATATATATACATATATACACATATGTATACATATACATA 39507

Qy        417 TATACACATATATACACATATATACACATATACACATATATACACATATATACATATATA 476
              |||||||||| || | ||||||||| |||||||||||| || ||||||| | ||||| |
Db      39508 TATACACATATGTATATATATATACATATATACACATATGTATACATATACATATATACA 39567

Qy        477 CACATATATATAATATACACACATATATATACACATATATACACACATATATACACACATAT 536
              || || |   ||| | | ||||||||| |||  |||| | | ||||||||||||||||
Db      39568 CATATGT-----ATACATATACATATATACACATGTATACATATACATATATACACATAT 39622

Qy        537 ATACACATATATATACACATATATACACATATATACATATATACACATATATATACATAT 596
              || ||||||| ||| | | | |||   ||| ||| | ||||| ||| |||| || | ||
Db      39623 GTATACATATACATATATACACATATGTATACATATACATATATACACATATGTATACAT 39682

Qy        597 ATACACATATATACATATATACACATATATACATATATACACACATATATACACATACAT 656
              |||||.|||| |||||| || ||||||| | ||||| ||| |  |||| | | | | ||
Db      39683 ATACATATATACACATATGTATACATATACATATATACACATATGTATACATATACATAT 39742

Qy        657 ATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATAT 716
              ||||||| |  || | ||| | |||||||||||||| |||| |||| |||| || | ||
Db      39743 ATACACATATGTATACATATACATATATACACATATGTATACATATACATATATACACAT 39802

Qy        717 ATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAG 776
              ||||  |     | |||          ||  | | ||    |  | ||| |   | | |
Db      39803 ATATATACATATATGTATATATATACACATATAAATATATATATGTATATACACACATAT 39862

Qy        777 AGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAA 836
              |  |  | |     |  |     | |   | || |     |     || |||        |
Db      39863 ATATACATATATATACACATATAAATATATATATGTATATACACACATATATATACACAT 39922

Qy        837 ATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACT 889
              ||| ||| || | ||   | |  ||| | ||    |  ||  ||    ||  ||
Db      39923 ATATATACATATATATGTATATATATATATTTTTTTTTTGAGATGAAGTCT 39975
```

```
RESULT 25
AC108483/c
LOCUS       AC108483              163976 bp    DNA     linear   PRI 29-MAR-2003
DEFINITION  Homo sapiens chromosome 16 clone RP11-170P24, complete sequence.
ACCESSION   AC108483
VERSION     AC108483.3  GI:29366946
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 163976)
  AUTHORS   DOE Joint Genome Institute, Stanford Human Genome Center and Los
            Alamos National Laboratory.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 163976)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JAN-2002) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 163976)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-2002) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   4  (bases 1 to 163976)
  AUTHORS   DOE Joint Genome Institute, Stanford Human Genome Center and Los
            Alamos National Laboratory.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAR-2003) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Mar 29, 2003 this sequence version replaced gi:19073802.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center and Los Alamos
            National Laboratory
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 100% of Sequence;
            Estimated Total Number of Errors is 0.
FEATURES             Location/Qualifiers
     source          1. .163976
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone="RP11-170P24"
ORIGIN

  Query Match             8.9%;  Score 365.4;  DB 9;  Length 163976;
  Best Local Similarity   72.7%;  Pred. No. 4.7e-62;
  Matches  527;  Conservative    0;  Mismatches  191;  Indels    7;  Gaps    4;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC--AAAAT 58
              || |  |||||| ||||   |||||||||||||  || ||||| ||||||||||  |||||
Db     146516 AGATTGAGACCATCCTGGCTAACATGGTGAAAGTCCGTCTCTCCTAAAAATACAAAAAAT 146457

Qy         59 TAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGA 118
              |||  ||   ||||| ||  ||||||| ||||||||||||   ||||| ||||||||||
Db     146456 TAGCTAGTTGTGGTGGCACGTGCCTGTAGTCCCAGCTACTCGAGAGGCTGAGGCAGGAGA 146397

Qy        119 ATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCA 178
              ||  | ||||||    | ||  | || |   | |   ||||||  |||||  ||||||||
Db     146396 ATAGCCTGAACC-AAGGAGGCGGAAGTTGCAGTGAGCCAAGATTGTGCCACCGCACTCCA 146338

Qy        179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATA 238
              | || |||||||| | | | |          | || || || ||      | |||| | | |
Db     146337 G-CCTGGCAACAACAACAACAACAACAACAACAACAACAACAAATATATATATACACACA 146279

Qy        239 TATATATATATATACACACACACACATATATATATACACATATATATACACATATATA 298
              | | | | ||||| | | |||||||| |||||||| | ||||||||||||||||| | | |
Db     146278 CACACACACATATATATATACACACATATATATACACACACATATATATACACACACACA 146219
```

```
Qy        299 CATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACA 358
              ||||||| ||| | | | | ||||| | ||||| | |||| |||||| | | | ||| |
Db     146218 TATATATATACACACACATATATATACACACACACATATATATACACACACACACACATA 146159

Qy        359 TATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATA 418
              ||| || ||||| | | | | |||||||||||||||||||| ||| | | ||| | |||||
Db     146158 TATATACACACACACACATATATATATACACACATATATATACACACACACATATATATA 146099

Qy        419 TACACATATATACACATATATACACATATACACATATATACACATATATACATATATACA 478
              ||||| | | | | |||||||||||| ||| | ||| | ||||| ||||| ||| | |||
Db     146098 CACACACACACATATATATATACACACATATATATACACACACACATATATATACACACA 146039

Qy        479 CATATATATA---ATATACACACATATATATACACATATATACACACATATATACACATA 535
              || | | |||   |||||||||||||||||||||||| | | ||| | ||||||||||| |
Db     146038 CACACACATATATATATACACACATATATATACACACACACACATATATATATACACACA 145979

Qy        536 TATACACATATATATACACATATATACACATATATACATATATACACATATATATACATA 595
              |||| | | | | | ||| ||||||||||| | | ||||||||| | | | | | |||||
Db     145978 TATATATACACACACACATATATATACACACACACACATATATATATACACACACACATA 145919

Qy        596 TATACACATATATATACATATATACACATATATACATATATACACACATATATACACATACA 655
              |||| | | | | | ||||||||||| | ||| ||||| ||||||| ||||| | | |||
Db     145918 TATATATACACACATATATATACACACACATATATATACACACACACATATATATACACA 145859

Qy        656 TATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATA 715
              | ||| | | ||| | | |||||||||||||||| | ||||| |||| | |   | | |
Db     145858 CACACATATATATATACACACATATATATACACACACATATATATATACACACACACACA 145799

Qy        716 TATAT 720
              |||||
Db     145798 TATAT 145794


RESULT 26
AC022166
LOCUS       AC022166               163998 bp    DNA     linear   PRI 21-JUL-2001
DEFINITION  Homo sapiens chromosome 16 clone RP11-525K10, complete sequence.
ACCESSION   AC022166
VERSION     AC022166.7  GI:14993688
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 163998)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 163998)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JAN-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 163998)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
REFERENCE   4  (bases 1 to 163998)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JUL-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Jul 21, 2001 this sequence version replaced gi:14327778.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 99.8% of Sequence;
            Estimated Total Number of Errors is 0.2.
            STS Content:
            WI-15858 G24147
            SHGC-79420 G52213.

FEATURES             Location/Qualifiers
     source          1. .163998
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone="RP11-525K10"
ORIGIN

  Query Match             8.9%;  Score 365.4;  DB 9;  Length 163998;
  Best Local Similarity   72.7%;  Pred. No. 4.7e-62;
  Matches  527;  Conservative    0;  Mismatches  191;  Indels    7;  Gaps    4;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC--AAAAT 58
              || |  |||||| ||||   |||||||||||  || ||||| ||||||||||  |||||
Db     130747 AGATTGAGACCATCCTGGCTAACATGGTGAAAGTCCGTCTCTCCTAAAAATACAAAAAAT 130806

Qy         59 TAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGA 118
              |||  ||   |||||  ||  ||||||| ||||||||||   ||||| ||||||||||
Db     130807 TAGCTAGTTGTGGTGGCACGTGCCTGTAGTCCCAGCTACTCGAGAGGCTGAGGCAGGAGA 130866

Qy        119 ATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCAAGATCATGCCATTGCACTCCA 178
              ||  | ||||||    | ||  | || |   | |   ||||||  |||||  ||||||||
Db     130867 ATAGCCTGAACC-AAGGAGGCGGAAGTTGCAGTGAGCCAAGATTGTGCCACCGCACTCCA 130925

Qy        179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATA 238
              | || |||||||| | | | |          | || || || ||     | |||| | | |
Db     130926 G-CCTGGCAACAACAACAACAACAACAACAACAACAACAACAAATATATATATACACACA 130984

Qy        239 TATATATATATATACACACACACACATATATATATACACATATATATACACATATATA 298
              | | | | ||||| | | ||||||| |||||||| | ||||||||||||||||| | | |
Db     130985 CACACACACATATATATATACACACATATATATACACACACATATATATACACACACACA 131044

Qy        299 CATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACA 358
              ||||||| ||| | | | | ||||| | ||||| | |||| |||||| | | | ||| |
Db     131045 TATATATATACACACACATATATATACACACACACATATATATACACACACACACACATA 131104

Qy        359 TATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATA 418
              ||| || ||||| | | | | |||||||||||||||||||| ||| | | ||| | |||||
Db     131105 TATATACACACACACACATATATATATACACACATATATATACACACACACATATATATA 131164

Qy        419 TACACATATATACACATATATACACATATACACATATATACACATATATACATATATACA 478
              |||||  | | | |||||||||||| ||| | ||| | ||||| ||||| ||| | |||
Db     131165 CACACACACACATATATATATACACACATATATATACACACACACATATATATACACACA 131224

Qy        479 CATATATATA---ATATACACACATATATATACACATATATACACACATATATACACATA 535
              || | | |||   |||||||||||||||||||||||| | | ||| | ||||||||||| |
Db     131225 CACACACATATATATATACACACATATATATACACACACACACATATATATATACACACA 131284

Qy        536 TATACACATATATATACACATATATACACATATATACATATATACACATATATATACATA 595
              |||| | | | | | ||| |||||||||||| | | |||||||||| | | | | | |||||
Db     131285 TATATATACACACACACATATATATACACACACACACATATATATATACACACACACATA 131344

Qy        596 TATACACATATATACATATATACACATATATACATATATACACACATATATACACATACA 655
              |||| | | | | | |||||||||||| | ||| ||||| ||||||| ||||| | | |||
Db     131345 TATATATACACACATATATATACACACACATATATATACACACACACATATATATACACA 131404

Qy        656 TATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATA 715
              | ||| | | ||| | | |||||||||||||||| | |||||  |||| | |   | | |
Db     131405 CACACATATATATATACACACATATATATACACACACATATATATATACACACACACACA 131464

Qy        716 TATAT 720
              |||||
Db     131465 TATAT 131469


RESULT 27
AC092863/c
LOCUS       AC092863               153201 bp    DNA     linear   HTG 11-DEC-2002
DEFINITION  Homo sapiens chromosome 12 clone RP11-355N1, WORKING DRAFT
            SEQUENCE, 5 unordered pieces.
ACCESSION   AC092863
VERSION     AC092863.10  GI:26449042
```

```
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 153201)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 153201)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-AUG-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 153201)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-DEC-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Dec 11, 2002 this sequence version replaced gi:19172705.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Drafting Center Code: BCM
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: HDJU
                Center clone name: RP11-355N1
            -------------- Summary Statistics
                Sequencing vector: Plasmid;
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 150350 bases at least Q40
                Consensus quality: 150698 bases at least Q30
                Consensus quality: 151119 bases at least Q20
                Estimated insert size: 144092; sum-of-contigs estimation
                Quality coverage: 14x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2895: contig of 2895 bp in length
            *     2896     2995: gap of unknown length
            *     2996    16175: contig of 13180 bp in length
            *    16176    16275: gap of unknown length
            *    16276    56442: contig of 40167 bp in length
            *    56443    56542: gap of unknown length
            *    56543    99843: contig of 43301 bp in length
            *    99844    99943: gap of unknown length
            *    99944   153201: contig of 53258 bp in length.
FEATURES             Location/Qualifiers
     source          1. .153201
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="12"
                     /clone="RP11-355N1"
ORIGIN

  Query Match             8.8%;  Score 364.2;  DB 2;  Length 153201;
  Best Local Similarity   74.5%;  Pred. No. 8.1e-62;
  Matches  543;  Conservative    0;  Mismatches  168;  Indels   18;  Gaps    6;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAATTA 60
              |||| |||||| |||| ||||||||  |||||||||| |||||       ||||||
Db       4255 AGTTTGAGACCATCCTGGCCAACATGGCAAAACCCCATCTTTACTA-------AAAATTG 4203

Qy         61 GACAGGCATGGTGATACACGCCTG-----TAATCCCAGCTACTTCGGAGGCCGAGGCAGG 115
              | | ||| |||||| ||| |||||      ||||||||||||||  |||| |  ||||| |
Db       4202 GTCGGGCGTGGTGACACATGCCTGTAATCTAATCCCAGCTACTCAGGAGCCTAAGGCATG 4143

Qy        116 AGAATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGCA-AGATCATGCCATTGCAC 174
              |||||| ||||||||||  || ||||||||||| | ||||| | |    |||| || ||
Db       4142 AGAATCGCTTGAACCTG--GGAGGTGGAGGTTGCAGTGAGCAGATAATGGGCCACTGTAC 4085

Qy        175 TCCAGCCCAGGCAAC--AAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATA 232
              |||||||  ||  ||  | |       |||||     |  ||| |||| | ||| || ||
Db       4084 TCCAGCCTGGGTGACAGAGCAAGACTCTGTCAAAAAAACAAACAAAACATGCAGTTCATC 4025

Qy        233 TATATATATATATATATATACACACACACACATATATATATACACATATATATACACA 292
              | |   |   || ||||  | | |      | | |   | | ||| | || | | || | |
Db       4024 TTTCCCTCAGTAGATATCGAAAGAGTGGGATAAAAGGAAAAATATATATGTGTGTATATA 3965

Qy        293 TATATACATATATACACATATATACACATATATATACACATACATATGTACACATATATA 352
              |||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |||||||
Db       3964 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 3905

Qy        353 TACACATATGTATACACATATATACACATATATACACACATATATACACATATATACACA 412
              || | |||| |||| | ||||||| | ||||||| | | ||||||| | ||||||| | |
Db       3904 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 3845

Qy        413 CATATATACACATATATACACATATATACACATATACACATATATACACATATATACATA 472
              ||||||| | ||||||| | ||||||| | ||||| | ||||||| | ||||||| |||
Db       3844 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 3785

Qy        473 TATACACATATATATA-ATATACACACATATATATACACATATATACACACATATATACA 531
              |||| | ||||||||| ||||| | | ||||||||| | ||||||| | | ||||||| |
```

```
Db       3784 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 3725

Qy        532 CATATATACACATATATATACACATATATACACATATATACATATATACACATATATATA 591
              ||||||| | |||||||||| | ||||||| | ||||||| ||||||| | |||||||||
Db       3724 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 3665

Qy        592 CATATATACACATATATACATATATACACATATATACATATATACACACATATATACACA 651
              ||||||| | ||||||| ||||||| | ||||||| ||||||| | | ||||||| | |
Db       3664 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 3605

Qy        652 TACATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTA 711
              || ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||||| ||
Db       3604 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 3545

Qy        712 TATATATAT 720
              |||||||||
Db       3544 TATATATAT 3536
```

```
RESULT 28
AC092694/c
LOCUS       AC092694               151549 bp    DNA     linear   PRI 07-MAR-2002
DEFINITION  Homo sapiens 3q BAC RP11-172A10 (Roswell Park Cancer Institute
            Human BAC Library) complete sequence.
ACCESSION   AC092694
VERSION     AC092694.6  GI:18449730
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 151549)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., He,X.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M., Neal,D.,
            Nelson,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N.,
            Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N.,
            Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L.,
            Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M.,
            Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S.,
            Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E.,
            Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A.,
            Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C.,
            Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L.,
            Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S.,
            Warren,R., Washington,C., Watlington,S., Williams,G.,
            Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y.,
            Wu,Y.F., Zhou,J., Zorrilla,S., Naylor,S.L., Weinstock,G. and
            Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 151549)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUL-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 151549)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-JAN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   4  (bases 1 to 151549)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   5  (bases 1 to 151549)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jan 31, 2002 this sequence version replaced gi:17977520.
            INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
            gc-help@bcm.tmc.edu

            CLONE LENGTH: This sequence does not necessarily represent the
            entire insert of this clone. Overlapping regions of clones are only
            sequenced and submitted once, so the sequence for the remainder of
            the insert may be found in the record for the adjacent clones.
            Overlapping clones are noted at the beginning and end of the
            Features listing.

            ANNOTATION OF FEATURES:
               STSs are identified using ePCR (Genome Res. 7:541-550) searches
            of a local database that includes entries from dbSTS, GDB, and
            local mapping efforts.
               Repeats are identified using RepeatMasker (A. Smit and P. Green,
            unpublished.) for Human and Mouse sequences.
               Genes and Region of sequence similarity are identified by BLAST
            (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
            EST and cDNA sequences. Genes demonstrate at least two exons
            flanked by consensus splice sites that maintained sequence
            continuity across the splice junctions. Sequences that are not
            identical matches are annotated as similar.

            SEQUENCING READ COVERAGE:Sequencing is completed to a minimum
            standard of double strand coverage with a minimum of 2 clones and 2
            reads with no ambiguities or 2 chemistries with a minimum of 2
            clones and 3 reads with no ambiguities. If the sequence quality for
            a region does not meet this standard, it will be indicated in the
            annotation as Low Coverage.

            QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality
            standards - estimated error rate less than 1 per 10,000 bases.
            Reports of lowest quality individual bases and measures of base
            quality are listed below. Description of the metrics can be found
            at URL:
            http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

            QUALSTAT-REPORT.
FEATURES                Location/Qualifiers
     source             1. .151549
                        /organism="Homo sapiens"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9606"
                        /chromosome="3"
                        /clone="RP11-172A10"
     repeat_region      137. .646
```

```
                     /rpt_family="L1M4c"
repeat_region        696. .1026
                     /rpt_family="L1M4c"
repeat_region        complement(1065. .1364)
                     /rpt_family="AluSx"
repeat_region        1365. .1966
                     /rpt_family="L1M1"
repeat_region        1967. .2309
                     /rpt_family="MLT1A1"
repeat_region        2310. .3436
                     /rpt_family="L1M1"
repeat_region        3437. .3473
                     /rpt_family="(TTTA)n"
repeat_region        complement(3474. .3802)
                     /rpt_family="LTR17"
repeat_region        complement(3803. .4094)
                     /rpt_family="HERV17"
repeat_region        complement(8054. .9082)
                     /rpt_family="HERV17"
repeat_region        complement(8719. .9163)
                     /rpt_family="Harlequin"
repeat_region        9166. .9611
                     /rpt_family="L1M1"
repeat_region        complement(9605. .9798)
                     /rpt_family="L1PA16"
repeat_region        complement(9805. .9857)
                     /rpt_family="L1"
repeat_region        9880. .10350
                     /rpt_family="L1M1"
repeat_region        10786. .11108
                     /rpt_family="L1MA2"
repeat_region        11109. .11492
                     /rpt_family="L1MA2"
repeat_region        11612. .12041
                     /rpt_family="L1M1"
repeat_region        12053. .12130
                     /rpt_family="(TA)n"
misc_feature         12121. .13100
                     /function="unresolved tandem repeat"
repeat_region        12131. .12462
                     /rpt_family="(CATATA)n"
repeat_region        12466. .13020
                     /rpt_family="(TATATG)n"
repeat_region        13040. .13100
                     /rpt_family="(TATATG)n"
repeat_region        complement(13101. .13387)
                     /rpt_family="AluY"
repeat_region        13388. .13466
                     /rpt_family="(TATATG)n"
repeat_region        13469. .13599
                     /rpt_family="L1M1"
repeat_region        13611. .14584
                     /rpt_family="L1M1"
repeat_region        14669. .15274
                     /rpt_family="L1M1"
repeat_region        15275. .17378
                     /rpt_family="L1M1"
repeat_region        17379. .17684
                     /rpt_family="AluSc"
repeat_region        17685. .18217
                     /rpt_family="L1M1"
repeat_region        18200. .18827
                     /rpt_family="L1"
repeat_region        18829. .19120
                     /rpt_family="AluJb"
repeat_region        19123. .19483
                     /rpt_family="L1"
repeat_region        complement(19478. .19591)
                     /rpt_family="L1"
repeat_region        20015. .20050
                     /rpt_family="(TA)n"
repeat_region        20531. .20559
                     /rpt_family="AT_rich"
repeat_region        20570. .20647
                     /rpt_family="AluSx"

  Query Match             8.8%;  Score 363.6;  DB 9;  Length 151549;
  Best Local Similarity   68.3%;  Pred. No. 1.1e-61;
  Matches  623;  Conservative    0;  Mismatches  274;  Indels   15;  Gaps    8;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              || ||||||||| ||||    |||| |||||||||||||||||||||||||| ||||||
Db      13316 AGATCAAGACCATCCTGGCTAACACGGTGAAACCCCATCTCTACTAAAAATACAAAAATT 13257

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              || ||||||||||     | |||||| |||||||||||| |||||| ||||||||||||
Db      13256 AGCCAGGCATGGTGGCGGGCACCTGTAGTCCCAGCTACTTGGGAGGCTGAGGCAGGAGAA 13197

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGC-AAGATCATGCCATTGCACTCCA 178
              |  | ||||||||  || |||||||| |||| | |||| |||||   |||| ||||||||||
Db      13196 TGGCCTGAACCTG--GGAGGTGGAGCTTGCAGTGAGCTAAGATGGCGCCACTGCACTCCA 13139

Qy        179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATA 238
              |||  |||||| |||||  | | | |  |      |||||| | |      |  |||| ||||
Db      13138 GCCTGGGCAAC-AGAGCAAGAAGACTCCATCTCAAAAAATATATTATGTACATATGTATA 13080

Qy        239 TATATATATATATACACACACACACATATATATATACACATATATATACACATATATA 298
              | | | |||| |||   | | |||||||| |||| ||| | |  ||| |||| |||| ||
Db      13079 TGTGTCTATACATATGTATATACACACATGTATACATATATACTTATGTACATATATGTA 13020

Qy        299 CATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACA 358
              |||  ||| |||||| || | | ||||| | |   || |||  || ||||||  |||| |||
Db      13019 TATACATATACATATGTATATACATATACATATGTATATACATATACATATGTATATACA 12960

Qy        359 TATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATA 418
              ||||||||     ||||| || ||| |||| | || ||||||||||||| | |    |||||
Db      12959 CATGTATAC----ATATATACTTATGTACATATATGTATACACATATACATATGTATATA 12904

Qy        419 TACACATATATACACATATATACACATATACACATATATACACATATATACATATATACA 478
              ||||||  ||||||||||||    |  || |||||| ||||||||||| | || |||| |
Db      12903 CACACATGTATACACATATACTTATGTACACACATGTATACACATATACATATGTATATA 12844

Qy        479 CATATATATA---ATATACACACATATATATACACATATATACACACATATATACACATA 535
              || | || ||    | ||| ||| || |||||||||| || || ||||||||| | |  ||
Db      12843 CACACATGTATACACATATACATATGTATATACACACATGTATACACATATACATATGTA 12784

Qy        536 TATACACATATATATACACATATATACACATATATACATATATACACATATATATACATA 595
              |||||||| || |||||||||||| | |  |||||||| | ||  | | | |||||||||
Db      12783 TATACACACATGTATACACATATACATATGTATATACACACATGTATACACATATACATA 12724

Qy        596 TATACACATATATACATATATACACATATATACATATATACACACATATATACACATACA 655
              | || | | | |  |||| ||| ||| |||  ||||||||||||| ||||||||||| |
Db      12723 TGTATATACACACATGTATACACATATACATATGTATATACACACATGTATACACATATA 12664

Qy        656 TATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATA 715
              |||   | | | | | ||  ||| | ||| |||||| |||  || | | |||||||| |
Db      12663 CATATGTATATACACACATGTATACACATATACATATGTATA--TACACACATGTATACA 12606

Qy        716 TATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGA 775
              |||| |    | | | |           |      |||   ||   ||| | | | | |
Db      12605 CATATACATATGTATATACACACATGTATACACATATACATATGTACATATATGTATATA 12546

Qy        776 GAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGA 835
              |  |  |||   ||  ||      |  |||| |    ||  .  || |||      |
Db      12545 CACATATACATATGTACATATATGTATATACATATATACTTATGTACATATATGTATATA 12486

Qy        836 AATACATAAATGTGTAATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTAACACTATATA 894
              |||  | | |||| ||   |  | | | |  | |  || ||| | |  ||    ||||||
Db      12485 CATATACATATGTATATACACCCATGTATACATATATACTTATGTACATATATGTATATA 12426

Qy        895 TATATATAGTGA 906
              |||||||  | |
Db      12425 CATATATACTTA 12414

RESULT 29
```

```
AC021193
LOCUS       AC021193               217985 bp    DNA     linear   PRI 21-FEB-2002
DEFINITION  Homo sapiens BAC clone RP11-774G5 from 4, complete sequence.
ACCESSION   AC021193
VERSION     AC021193.7  GI:15321574
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 217985)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
   PUBMED   9847074
REFERENCE   2  (bases 1 to 217985)
  AUTHORS   Paulson,E., Meyer,R., Elliott,G., Boyer,E. and Dixon,R.
  TITLE     The sequence of Homo sapiens BAC clone RP11-774G5
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 217985)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JAN-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 217985)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-AUG-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 217985)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JAN-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   6  (bases 1 to 217985)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-FEB-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Aug 28, 2001 this sequence version replaced gi:15022788.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0774G05
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. John D.
            McPherson, Department of Genetics, Washington University, St. Louis
            MO. For additional information about the map position of this
            sequence, see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries. Genomics 51:1-8. The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at http://www.chori.org
            VECTOR:  pBACe3.6

            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is AC079084.  Actual start of this
            clone is at base position 1 of RP11-774G5; actual end is at base
            position 217985 of RP11-774G5.
FEATURES             Location/Qualifiers
     source          1. .217985
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /map="4"
                     /clone="RP11-774G5"
                     /clone_lib="RPCI-11"
     repeat_region   1077. .1483
                     /rpt_family="MER2_type"
     repeat_region   1511. .1555
                     /rpt_family="AT_rich"
     repeat_region   1669. .1888
                     /rpt_family="MaLR"
     repeat_region   1889. .2198
                     /rpt_family="Alu"
     repeat_region   2172. .2191
                     /rpt_family="(CAAA)n"
     repeat_region   2192. .2219
                     /rpt_family="(CAAAA)n"
     repeat_region   2199. .2388
                     /rpt_family="MaLR"
     repeat_region   2668. .2947
                     /rpt_family="Alu"
     repeat_region   3099. .3125
                     /rpt_family="AT_rich"
     repeat_region   3100. .3462
                     /rpt_family="L1"
     repeat_region   3730. .3751
                     /rpt_family="AT_rich"
     repeat_region   3999. .4659
                     /rpt_family="L1"
     repeat_region   4460. .4551
                     /rpt_family="(TA)n"
     repeat_region   4698. .4897
                     /rpt_family="L1"
     repeat_region   4898. .5162
                     /rpt_family="L1"
     repeat_region   5377. .5405
                     /rpt_family="AT_rich"
     repeat_region   6918. .7257
                     /rpt_family="MaLR"
     repeat_region   8680. .8986
                     /rpt_family="MaLR"
     repeat_region   8989. .10092
                     /rpt_family="MER2_type"
     repeat_region   10216. .10313
                     /rpt_family="L1"
     repeat_region   11999. .12023
                     /rpt_family="AT_rich"
     repeat_region   12490. .12719
                     /rpt_family="MIR"
     repeat_region   13417. .13960
                     /rpt_family="L1"
     repeat_region   13945. .13969
                     /rpt_family="(A)n"
     repeat_region   14697. .15010
                     /rpt_family="Alu"
     repeat_region   14989. .15023
                     /rpt_family="AT_rich"
```

```
     repeat_region   16415. .16850
                     /rpt_family="MaLR"
     repeat_region   18007. .18299
                     /rpt_family="Alu"
     repeat_region   19150. .19172
                     /rpt_family="AT_rich"
     repeat_region   19177. .19375
                     /rpt_family="ERV1"
     repeat_region   19490. .19640
                     /rpt_family="ERV1"
     repeat_region   19679. .20025
                     /rpt_family="L1"
     repeat_region   20302. .20588
                     /rpt_family="ERV1"
     repeat_region   21525. .21574
                     /rpt_family="Mariner"
     repeat_region   22586. .22642
                     /rpt_family="(TG)n"
     repeat_region   22960. .22980
                     /rpt_family="AT_rich"
     repeat_region   23682. .23709
                     /rpt_family="(A)n"
     repeat_region   23892. .24010
                     /rpt_family="ERV1"
     repeat_region   24012. .24039
                     /rpt_family="AT_rich"
     repeat_region   24187. .24257
                     /rpt_family="(TTATA)n"
     repeat_region   24472. .24493
                     /rpt_family="(T)n"
     repeat_region   25473. .25523
                     /rpt_family="AT_rich"
     repeat_region   25899. .25993
                     /rpt_family="MIR"
     repeat_region   25994. .27011
                     /rpt_family="L1"
     repeat_region   26007. .26053
                     /rpt_family="AT_rich"
     repeat_region   26269. .26421
                     /rpt_family="T-rich"
     repeat_region   26617. .26647
                     /rpt_family="AT_rich"
     repeat_region   27102. .27425
                     /rpt_family="L1"
     repeat_region   27217. .27237
                     /rpt_family="AT_rich"
     repeat_region   27512. .27828
                     /rpt_family="L1"
     repeat_region   27595. .27630
                     /rpt_family="AT_rich"
     repeat_region   27923. .28308
                     /rpt_family="L1"
     repeat_region   28362. .29158
                     /rpt_family="L1"
     repeat_region   29254. .29527

  Query Match             8.7%;  Score 357.4;  DB 9;  Length 217985;
  Best Local Similarity   78.0%;  Pred. No. 1.6e-60;
  Matches  495;  Conservative    0;  Mismatches  131;  Indels    9;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC--AAAAT 58
              || || ||||||  |||   ||||||||||||||| ||||||||||||||  |  |||||
Db     146265 AGATCGAGACCATTCTGGCTAACATGGTGAAACCCCGTCTCTACTAAAAATTCCAAAAAT 146324

Qy         59 TAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGA 118
              ||| ||| | |||||     |||||||| ||||||| |||    ||||| ||||||||||
Db     146325 TAGCCAGACGTGGTGGCGGGCGCCTGTAGTCCCAGCCACTCGAGAGGCTGAGGCAGGAGA 146384

Qy        119 ATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGC-AAGATCATGCCATTGCACTCC 177
              ||  | ||||||  | || || ||||  ||| | ||||  |||||  |||| ||||||||
Db     146385 ATGGCGTGAACC--CGGGAGGCGGAGCCTGCAGTGAGCTGAGATCGCGCCACTGCACTCC 146442

Qy   .     178,AGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATAT 237
```

```
              ||||  ||| || ||||||| |   | ||||| |||||||||||     | |||||||||
Db     146443 AGCCTGGGCGAC-AGAGCGAGACTCCGTCTCAAAAAAAAAAAAAAATATATATATATAT 146501

Qy        238 ATATATATATATATACACACACACACATATATATATACACATATATATACACATATAT 297
              | | ||||||||||||| | | | ||| |||||||| ||| ||||||| ||| ||||
Db     146502 ACACATATATATATACATATATATATACACATATATATATACACATATATATACACATAT 146561

Qy        298 ACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACAC 357
              | ||| ||| | ||| | | | ||| ||| | | ||||| || || | | ||||||||||
Db     146562 ATATACATATATATACACATATATACATACATATATACACATATATATACATATATACAC 146621

Qy        358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417
              |||| |||| ||||||||| ||| ||| | | | | ||||| | | ||||| ||| ||||
Db     146622 ATATATATATACATATATATACACATACATATATACATATATATACATATATACATAT 146681

Qy        418 ATACACATATATACACATATATACACATATACACATATATACACATATATACATATATAC 477
              ||||| |||||||||| ||||| | | ||||||| ||||||||| ||||| ||||||||||
Db     146682 ATACATATATATACATATATACATATATATACATATATATACATATATACACATATATAC 146741

Qy        478 ACATATATATAATATACACACATATATATACACATATATACACACATATATACACATATA 537
              | |||||    ||||| ||| | ||||||||| ||||| ||| ||||||||||| |||||
Db     146742 ATATATA---CATATATACATACATATATACATATATACACATACATATATACATATATA 146798

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATACATATA 597
              |||  | ||||||| |||||||| ||| ||| | ||||||||| | ||| ||| |||||
Db     146799 TACGTACATATATATACATATATATACACATACATATATATACATACATACATATATATA 146858

Qy        598 TACACATATATACATATATACACATATATACATAT 632
              || | ||||||| ||||||| | ||||||| || |
Db     146859 TATATATATATATATATATATATATATATATGT 146893
```

```
RESULT 30
CR812471/c
LOCUS       CR812471              151424 bp    DNA     linear   HTG 27-SEP-2004
DEFINITION  Danio rerio clone CH211-237A4, *** SEQUENCING IN PROGRESS ***, 9
            unordered pieces.
ACCESSION   CR812471
VERSION     CR812471.2  GI:52748670
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 151424)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 27, 2004 this sequence version replaced gi:52694427.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zC237A4
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 147995 bases at least Q40
            Consensus quality: 148500 bases at least Q30
            Consensus quality: 148901 bases at least Q20
            Insert size: 150624; sum-of-contigs
            Insert size: 152063; 2.5% error; agarose-fp
            Quality coverage: 7.11x in Q20 bases; sum-of-contigs Quality
            coverage: 7.09x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
```

```
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    18105: contig of 18105 bp in length
            *   18106    18205: gap of 100 bp
            *   18206    20634: contig of 2429 bp in length
            *   20635    20734: gap of 100 bp
            *   20735    23916: contig of 3182 bp in length
            *   23917    24016: gap of 100 bp
            *   24017    35672: contig of 11656 bp in length
            *   35673    35772: gap of 100 bp
            *   35773    47596: contig of 11824 bp in length
            *   47597    47696: gap of 100 bp
            *   47697    63363: contig of 15667 bp in length
            *   63364    63463: gap of 100 bp
            *   63464   127690: contig of 64227 bp in length
            *  127691   127790: gap of 100 bp
            *  127791   133366: contig of 5576 bp in length
            *  133367   133466: gap of 100 bp
            *  133467   151424: contig of 17958 bp in length.
FEATURES             Location/Qualifiers
     source          1. .151424
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-237A4"
                     /clone_lib="CHORI-211"
     misc_feature    1. .18105
                     /note="assembly_fragment:00546
                     fragment_chain:1
                     clone_end:SP6
                     vector_side:left"
     misc_feature    18206. .20634
                     /note="assembly_fragment:00036
                     fragment_chain:1"
     misc_feature    20735. .23916
                     /note="assembly_fragment:00008
                     fragment_chain:1"
     misc_feature    24017. .35672
                     /note="assembly_fragment:00104
                     fragment_chain:2"
     misc_feature    35773. .47596
                     /note="assembly_fragment:00228
                     fragment_chain:2"
     misc_feature    47697. .63363
                     /note="assembly_fragment:00380
                     fragment_chain:2"
     misc_feature    63464. .127690
                     /note="assembly_fragment:01023
                     fragment_chain:2"
     misc_feature    127791. .133366
                     /note="assembly_fragment:00065"
     misc_feature    133467. .151424
                     /note="assembly_fragment:00782
                     clone_end:T7
                     vector_side:right"
ORIGIN

  Query Match             8.7%;  Score 357.2;  DB 2;  Length 151424;
  Best Local Similarity   65.2%;  Pred. No. 1.9e-60;
  Matches  468;  Conservative    0;  Mismatches  248;  Indels    2;  Gaps    2;

Qy        209 AGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACATA 268
              | ||||||||||||    || ||||||||||||||||||||||| | | | | | | |||
Db      21248 AAAAAAAAAAAAAGCATATTCATATATATATATATATATATATATATATATATATATA 21189

Qy        269 TATATATACACATATATATACACATATATACATATATACACATATATACACATATATATA 328
              |||||||| | ||||||||| | ||||||| ||||||| | ||||||| | ||||||||||
Db      21188 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 21129

Qy        329 CACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACA 388
              | ||| |||| || | ||||||||| | |||| |||| | ||||||| | ||||||| |
Db      21128 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 21069

Qy        389 CACATATATACACATATATACACACATATATACACATATATACACATATATACACATATA 448
              | ||||||| | ||||||| | | |||||| | ||||||| | ||||||| | |||||
Db      21068 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 21009

Qy        449 CACATATATACACATATATACATATATACACATATATATA-ATATACACACATATATATA 507
              | ||||||| | ||||||| ||||||| | ||||||||| ||||| | | |||||||||
Db      21008 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 20949

Qy        508 CACATATATACACACATATATACACATATATACACATATATATACACATATATACACATA 567
              | ||||||| | | ||||||| | ||||||| | ||||||||| | ||||||| | |||
Db      20948 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 20889

Qy        568 TATACATATATACACATATATATACATATATACACATATATACACATATATACACATATA 627
              |||| ||||||| | |||||||| ||||||| | ||||||| ||||||| | |||||||
Db      20888 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 20829

Qy        628 CATATATACACACATATATACACATACATATACACACACATAGATATACATATATATACA 687
              ||||||| | | ||||||| | ||| ||||| | | | ||| ||||| |||||||||
Db      20828 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 20769

Qy        688 CATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGT 747
              |||||||||  |||||||| ||||||||||
Db      20768 TATATATATATATATATATATATATATATATATANNNNNNNNNNNNNNNNNNNNNNNN 20709

Qy        748 TGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACA 807

Db      20708 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 20649

Qy        808 TTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA 867
                          | |||       | ||| ||| || | ||   | |  ||| | ||
Db      20648 NNNNNNNNNNNNNNTATATATATATATATATATATATATATATATATATATATATA 20589

Qy        868 -AGGTAGTTAAGAAGGTAACACTATATATATATATAGTGAAAGCAGTGTAAACCTAAA 924
               | ||  ||   |  ||    ||||||||||||||   | |    | || |  || |
Db      20588 TATATATATATATATATATATATATATATATATATATATATATATATATATATATA 20531


RESULT 31
HS141H5/c
LOCUS       HS141H5               121600 bp    DNA     linear   PRI 05-JUN-2003
DEFINITION  Human DNA sequence from clone RP6-141H5 on chromosome Xq22.1-23
            Contains the 3' end of the gene for a novel protein with von
            Willebrand factor type C domains, complete sequence.
ACCESSION   AL049176
VERSION     AL049176.3  GI:4808226
KEYWORDS    HTG; von Willebrand factor type C.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 121600)
  AUTHORS   Pearce,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JUN-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On May 12, 1999 this sequence version replaced gi:4495007.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
```

```
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome X, constructed by the Sanger Centre Chromosome X Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/ChrX
            RP6-141H5 is from the library RPCI-6 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pPAC4
            This sequence is the entire insert of clone RP6-141H5.
FEATURES             Location/Qualifiers
     source          1. .121600
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="RZPD:RPCIP709H05141"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="q22.1-23"
                     /clone="RP6-141H5"
                     /clone_lib="RPCI-6"
     repeat_region   1. .665
                     /note="L1P repeat: matches 2. .667 of consensus"
     repeat_region   825. .1034
                     /note="MER58A repeat: matches 2. .216 of consensus"
     repeat_region   1103. .1634
                     /note="L1ME repeat: matches 572. .1075 of consensus"
     repeat_region   1635. .2006
                     /note="MLT1A1 repeat: matches 1. .365 of consensus"
     repeat_region   2007. .2521
                     /note="L1ME repeat: matches 1075. .1707 of consensus"
     repeat_region   2526. .2715
                     /note="L1MEc repeat: matches 1961. .2159 of consensus"
     repeat_region   2761. .4193
                     /note="Tigger1 repeat: matches 902. .2357 of consensus"
     repeat_region   4195. .5277
                     /note="L1PA16 repeat: matches 5057. .6133 of consensus"
     repeat_region   5291. .6213
                     /note="Tigger1 repeat: matches 1. .931 of consensus"
     repeat_region   6576. .8350
                     /note="L1MA8 repeat: matches 4428. .6279 of consensus"
     repeat_region   8501. .11123
                     /note="L1PA10 repeat: matches 3561. .6159 of consensus"
     repeat_region   11119. .11399
                     /note="L1M4 repeat: matches 2315. .2614 of consensus"
     repeat_region   11408. .17541
                     /note="L1PA6 repeat: matches 11. .6143 of consensus"
     repeat_region   17545. .17592
                     /note="24 copies 2 mer aa 77% conserved"
     repeat_region   17716. .17752
                     /note="L1PA16 repeat: matches 6117. .6153 of consensus"
     repeat_region   18281. .18366
                     /note="L1MB8 repeat: matches 6083. .6169 of consensus"
     repeat_region   20394. .20549
                     /note="MER97c repeat: matches 19. .177 of consensus"
     repeat_region   20895. .21189
                     /note="L2 repeat: matches 2428. .2709 of consensus"
     repeat_region   21669. .21960
                     /note="AluJo repeat: matches 1. .310 of consensus"
     repeat_region   21977. .22111
                     /note="MIR repeat: matches 107. .250 of consensus"
     repeat_region   23514. .23619
                     /note="53 copies 2 mer at 83% conserved"
     repeat_region   23520. .23619
                     /note="25 copies 4 mer atat 85% conserved"
     repeat_region   23621. .23821
                     /note="AluJo repeat: matches 92. .283 of consensus"
     repeat_region   24261. .24306
                     /note="L2 repeat: matches 2659. .2705 of consensus"
     repeat_region   24403. .24543
                     /note="MIR repeat: matches 116. .262 of consensus"
     repeat_region   24579. .24823
                     /note="L2 repeat: matches 2255. .2481 of consensus"
     repeat_region   25586. .25873
                     /note="MER34 repeat: matches 9. .297 of consensus"
     repeat_region   25874. .26288
                     /note="MSTB repeat: matches 1. .426 of consensus"
     repeat_region   26289. .26372
                     /note="MER34 repeat: matches 297. .380 of consensus"
     repeat_region   26373. .26768
                     /note="MSTA repeat: matches 1. .426 of consensus"
     repeat_region   26769. .26932
                     /note="MER34 repeat: matches 380. .543 of consensus"
     repeat_region   27373. .27775
                     /note="L2 repeat: matches 1058. .1475 of consensus"
     repeat_region   28030. .28249
                     /note="MIR repeat: matches 1. .243 of consensus"
     gene            complement(28358. .117309)
                     /gene="dA141H5.1"
     mRNA            complement(join(28358. .30863,33838. .33927,35984. .36151,
                     43096. .43305,48665. .48833,55156. .55223,74337. .74496,
                     75887. .76032,114163. .114256,117197. .>117309))
                     /gene="dA141H5.1"
                     /product="dA141H5.1 (novel protein with von Willebrand
                     factor type C domains)"
                     /note="match: cDNAs: Em:AF039656
                     match: ESTs: Em:AA302430 Em:AI246227 Em:H77289 Em:AA156742
                     Em:AA724151 Em:AA843324 Em:AA057564 Em:AI337909 Em:N39848
                     Em:AA555017 Em:AA135450 Em:N26819 Em:AI123481 Em:AI139649
                     Em:AA018704 Em:R52421 Em:AA724185 Em:AI138345 Em:AA057563
                     Em:AI004265 Em:AA150072 Em:AA244080 Em:AA854359
                     Em:AI420184 Em:R59551 Em:W52860 Em:AA724504 Em:AA635439
                     Em:AA018487 Em:Z42301 Em:AA327755 Em:C00874 Em:N43181
                     Em:AA177053 Em:AA015694 Em:W96276 Em:AA029156 Em:AA463650
                     Em:AA013040 Em:W28849 Em:AA918837"
                     /evidence=not_experimental
     polyA_signal    complement(28361. .28366)
                     /gene="dA141H5.1"
     polyA_site      complement(28384)
                     /gene="dA141H5.1"
     repeat_region   28960. .29007
                     /note="24 copies 2 mer ca 100% conserved"
     misc_feature    complement(28995. .29451)
                     /gene="dA141H5.1"
                     /note="match: GSS: Em:AQ442614"
     misc_feature    29462. .29764
                     /note="match: GSS: Em:AQ618076"
     repeat_region   30162. .30217
                     /note="28 copies 2 mer tg 89% conserved"
     CDS             complement(join(30733. .30863,33838. .33927,35984. .36151,
                     43096. .43305,48665. .48833,55156. .55223,74337. .74496,
                     75887. .76032,114163. .114256,117197. .>117309))
                     /gene="dA141H5.1"
                     /note="continued from bA814C6.1.1 in Em:AL591489
                     match: proteins: Tr:O57472 Sw:Q24025 Tr:O57465 Sw:Q91713"
                     /codon_start=3
                     /evidence=not_experimental
                     /product="dA141H5.1 (novel protein with von Willebrand
                     factor type C domains)"
                     /protein_id="CAD92488.1"
                     /db_xref="GI:31455411"
                     /translation="SETYCMFQDKKYRVGERWHPYLEPYGLVYCVNCICSENGNVLCS
                     RVRCPNVHCLSPVHIPHLCCPRCPEDSLPPVNNKVTSKSCEYNGTTYQHGELFVAEGL
                     FQNRQPNQCTQCSCSVLSEAFTLILDASNLNSVSPWQEGNVYCGLKTCPKLTCAFPVS
                     VPDSCCRVCRGDGELSWEHSDGDIFRQPANREARHSYHRSHYDPPPSRQAGGLSRFPG
                     ARSHRGALMDSQQASGTIVQIVINNKHKHGQVCVSNGKTYSHGESWHPNLRAFGIVEC
```

```
                     VLCTCNVTKQECKKIHCPNRYPCKYPQKIDGKCCKVCPGKKAKEELPGQSFDNKGYFC
                     GEETMPVYESVFMEDGETTRKIALETERPPQVEVHVWTIRKGILQHFHIEKISKRMFE
                     ELPHFKLVTRTTLSQWKIFTEGEAQISQMCSSRVCRTELEDLVKVLYLERSEKGHC"
     misc_feature    complement(31798. .32170)
                     /gene="dA141H5.1"
                     /note="match: GSS: Em:AQ694976"
     repeat_region   32245. .32729
                     /note="L2 repeat: matches 1728. .2263 of consensus"
     repeat_region   34978. .35165
                     /note="MER5A repeat: matches 1. .189 of consensus"
     misc_feature    35093. .35438
                     /note="match: GSS: Em:AQ205639"
     repeat_region   35552. .35734
                     /note="MER5B repeat: matches 6. .177 of consensus"
     misc_feature    37351. .37723
                     /note="match: GSS: Em:B99326"
     repeat_region   38099. .38585
                     /note="LTR32 repeat: matches 15. .471 of consensus"
     repeat_region   38934. .39248
                     /note="L2 repeat: matches 2418. .2748 of consensus"
     repeat_region   39760. .40064
                     /note="AluJb repeat: matches 1. .301 of consensus"
     repeat_region   40389. .40698

  Query Match             8.7%;  Score 356.6;  DB 9;  Length 121600;
  Best Local Similarity   74.7%;  Pred. No. 2.6e-60;
  Matches  543;  Conservative    0;  Mismatches  169;  Indels   15;  Gaps    7;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAA--- 57
              || || |||||| ||||   ||||  |||||||||| |||||||||||||| |||||
Db      44776 AGATCGAGACCATCCTGGCTAACACAGTGAAACCCCGTCTCTACTAAAAATCCAAAAAAA 44717

Qy         58 --TTAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGG 115
                |||| | ||| |||||     | |||||| |||||||||||  |||||| ||||||||
Db      44716 AGTTAGCCGGGCGTGGTGGCGGGCACCTGTAGTCCCAGCTACTCTGGAGGCTGAGGCAGG 44657

Qy        116 AGAATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCAC 174
              |||||  | ||||| ||  || || |||| |||| | ||| | ||||   |||| |||||
Db      44656 AGAATGGCGTGAACTTG--GGAGGCGGAGCTTGCAGTGAGCCGAGATAGCGCCACTGCAC 44599

Qy        175 TCCAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATA 234
              |||||  |   ||| || ||||||| |   ||||||| |||||||| |      | ||||||
Db      44598 TCCAGTCTGGGCGAC-AGAGCGAGACTCCATCTCAAAAAAAAAATA----TATATATATA 44544

Qy        235 TATATATATATATATATACACACACACACACATATATATATACACATATATATACACATA 294
              || |||||||| ||||||| | | | | | ||||||| ||| | | ||||||| | |||
Db      44543 TAGATATATATAGATATACAGATATATAGATATATATAGATATATAGATATATATAGATA 44484

Qy        295 TATACATATATACACATATATACACATATATATACACATACATATGTACACATATATATA 354
              |||| ||||||| | ||||||| | ||||||| | | ||| |||| || | ||||||| |
Db      44483 TATAGATATATATAGATATATAGATATATATAGATATATAGATATATATAGATATATAGA 44424

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              | ||||  ||| | | ||||| | | ||||| | | ||||| | | ||||| | | | |
Db      44423 TATATATAGATATATAGATATATATAGATATATAGATATATAGATATATATAGATATATA 44364

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              ||||| | | ||||| | ||||||| | ||||| | ||||||| | ||||||| |||||
Db      44363 GATATATATAGATATATAGATATATATAGATATATAGATATATAGATATATATAGATATA 44304

Qy        475 TACACATATATATAATATACACACATATATATACACATATATACACACATATATACACAT 534
              || |  |||| || ||||| | | ||||| ||| | ||| ||| | | |||||| | ||
Db      44303 TAGATGTATAGAT-ATATATAGATATATAGATATATATAGATATATAGATATATAGATAT 44245

Qy        535 ATATACACATATATATACACATATATACACATATATACATATATACACATATATATACAT 594
              ||||| | ||||| ||| | ||| ||| | ||||| | ||||| | | ||| ||||| |
Db      44244 ATATAGATATATAGATATATATAGATATATATATAGATATATAGATATATAGATATATAG 44185

Qy        595 ATATACACATATATACATATATACA-CATATATACATATATACACACATATATACACATA 653
              ||||| | ||||||| ||||||| |  ||||||| ||||||| | | ||||| | | |||
Db      44184 ATATATAGATATATAGATATATAGATGATATATAGATATATATAGATATATAGATATATA 44125

Qy        654 CATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATA 713
```

```
              ||||| | | | ||||||||| ||| ||||| | ||||||| |  |||| |||| ||||
Db      44124 GATATATAGATATATAGATATATATAGATATATAGATATATAGATATATAGATATATA 44065

Qy        714 TATATAT 720
              |||||||
Db      44064 GATATAT 44058
```

```
RESULT 32
CR788239/c
LOCUS       CR788239              127172 bp    DNA     linear   HTG 18-SEP-2004
DEFINITION  Danio rerio clone CH211-105M19, *** SEQUENCING IN PROGRESS ***, 26
            unordered pieces.
ACCESSION   CR788239
VERSION     CR788239.1  GI:52345234
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 127172)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zC105M19
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 116372 bases at least Q40
            Consensus quality: 119969 bases at least Q30
            Consensus quality: 122718 bases at least Q20
            Insert size: 124672; sum-of-contigs
            Insert size: 168914; 5.4% error; agarose-fp
            Quality coverage: 2.88x in Q20 bases; sum-of-contigs Quality
            coverage: 2.51x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 26 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     3262: contig of 3262 bp in length
            *     3263     3362: gap of 100 bp
            *     3363     5810: contig of 2448 bp in length
            *     5811     5910: gap of 100 bp
            *     5911     8259: contig of 2349 bp in length
            *     8260     8359: gap of 100 bp
            *     8360    11266: contig of 2907 bp in length
            *    11267    11366: gap of 100 bp
            *    11367    14728: contig of 3362 bp in length
            *    14729    14828: gap of 100 bp
            *    14829    18192: contig of 3364 bp in length
            *    18193    18292: gap of 100 bp
            *    18293    25841: contig of 7549 bp in length
            *    25842    25941: gap of 100 bp
            *    25942    28449: contig of 2508 bp in length
            *    28450    28549: gap of 100 bp
            *    28550    32988: contig of 4439 bp in length
            *    32989    33088: gap of 100 bp
            *    33089    37208: contig of 4120 bp in length
            *    37209    37308: gap of 100 bp
```

```
            *    37309    41754: contig of 4446 bp in length
            *    41755    41854: gap of 100 bp
            *    41855    48505: contig of 6651 bp in length
            *    48506    48605: gap of 100 bp
            *    48606    69764: contig of 21159 bp in length
            *    69765    69864: gap of 100 bp
            *    69865    77035: contig of 7171 bp in length
            *    77036    77135: gap of 100 bp
            *    77136    80534: contig of 3399 bp in length
            *    80535    80634: gap of 100 bp
            *    80635    85265: contig of 4631 bp in length
            *    85266    85365: gap of 100 bp
            *    85366    87432: contig of 2067 bp in length
            *    87433    87532: gap of 100 bp
            *    87533    92877: contig of 5345 bp in length
            *    92878    92977: gap of 100 bp
            *    92978   100746: contig of 7769 bp in length
            *   100747   100846: gap of 100 bp
            *   100847   103064: contig of 2218 bp in length
            *   103065   103164: gap of 100 bp
            *   103165   106145: contig of 2981 bp in length
            *   106146   106245: gap of 100 bp
            *   106246   108396: contig of 2151 bp in length
            *   108397   108496: gap of 100 bp
            *   108497   114275: contig of 5779 bp in length
            *   114276   114375: gap of 100 bp
            *   114376   121458: contig of 7083 bp in length
            *   121459   121558: gap of 100 bp
            *   121559   124719: contig of 3161 bp in length
            *   124720   124819: gap of 100 bp
            *   124820   127172: contig of 2353 bp in length.
FEATURES             Location/Qualifiers
     source          1. .127172
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-105M19"
                     /clone_lib="CHORI-211"
     misc_feature    1. .3262
                     /note="assembly_fragment:00259
                     fragment_chain:1"
     misc_feature    3363. .5810
                     /note="assembly_fragment:00160
                     fragment_chain:1
                     clone_end:T7
                     vector_side:left"
     misc_feature    5911. .8259
                     /note="assembly_fragment:00081
                     fragment_chain:2"
     misc_feature    8360. .11266
                     /note="assembly_fragment:00224
                     fragment_chain:2"
     misc_feature    11367. .14728
                     /note="assembly_fragment:00278
                     fragment_chain:2"
     misc_feature    14829. .18192
                     /note="assembly_fragment:00208
                     fragment_chain:2"
     misc_feature    18293. .25841
                     /note="assembly_fragment:00560
                     fragment_chain:2"
     misc_feature    25942. .28449
                     /note="assembly_fragment:00192
                     fragment_chain:2"
     misc_feature    28550. .32988
                     /note="assembly_fragment:00144
                     fragment_chain:2"
     misc_feature    33089. .37208
                     /note="assembly_fragment:00176
                     fragment_chain:2"
     misc_feature    37309. .41754
                     /note="assembly_fragment:00343
                     fragment_chain:2"
     misc_feature    41855. .48505
                     /note="assembly_fragment:00443
                     fragment_chain:2"
     misc_feature    48606. .69764
                     /note="assembly_fragment:00603
                     fragment_chain:2"
     misc_feature    69865. .77035
                     /note="assembly_fragment:00417
                     fragment_chain:3"
     misc_feature    77136. .80534
                     /note="assembly_fragment:00367
                     fragment_chain:3"
     misc_feature    80635. .85265
                     /note="assembly_fragment:00391
                     fragment_chain:3"
     misc_feature    85366. .87432
                     /note="assembly_fragment:00132
                     fragment_chain:3"
     misc_feature    87533. .92877
                     /note="assembly_fragment:00321
                     fragment_chain:3"
     misc_feature    92978. .100746
                     /note="assembly_fragment:00479
                     fragment_chain:3"
     misc_feature    100847. .103064
                     /note="assembly_fragment:00049
                     fragment_chain:3"
     misc_feature    103165. .106145
                     /note="assembly_fragment:00241
                     fragment_chain:3"
     misc_feature    106246. .108396
                     /note="assembly_fragment:00090
                     fragment_chain:3"
     misc_feature    108497. .114275
                     /note="assembly_fragment:00299
                     fragment_chain:4"
     misc_feature    114376. .121458
                     /note="assembly_fragment:00518
                     fragment_chain:4"
     misc_feature    121559. .124719
                     /note="assembly_fragment:00110"
     misc_feature    124820. .127172
                     /note="assembly_fragment:00100
                     clone_end:SP6
                     vector_side:right"
ORIGIN

  Query Match             8.7%;  Score 356.6;  DB 2;  Length 127172;
  Best Local Similarity   64.9%;  Pred. No. 2.5e-60;
  Matches  454;  Conservative    0;  Mismatches  244;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||| | | | | | ||||||||||| | |||||||
Db      42349 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 42290

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db      42289 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 42230

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              ||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db      42229 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 42170

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | |||||||| | ||||||| | ||||| | ||||||| | |||||
Db      42169 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 42110

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db      42109 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 42050

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
```

```
Db      42049 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 41990

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | |||.||| ||||||| | | |||||
Db      41989 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 41930

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db      41929 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 41870

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||
Db      41869 TATATATATATATATNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 41810

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
                                                                        | |
Db      41809 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTTATA 41750

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885
              ||       | ||| ||| || | ||   | |  ||| | ||   ||  ||   |  ||
Db      41749 TATATATATATATATATATATATATATATATATATATATATATATTATATATATATATAT 41690

Qy        886 CACTATATATATATAGTGAAAGCAGTGTAAACCTAAA 924
              |||||||||||||   | |   | || |  || |
Db      41689 ATATATATATATATATATATATATATATATATATATA 41651
```

```
RESULT 33
CR753865
LOCUS       CR753865              282240 bp    DNA     linear   HTG 04-SEP-2004
DEFINITION  Danio rerio clone DKEY-172B2, WORKING DRAFT SEQUENCE, 17 unordered
            pieces.
ACCESSION   CR753865
VERSION     CR753865.4  GI:51890091
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 282240)
  AUTHORS   Burton,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 4, 2004 this sequence version replaced gi:51870573.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK172B2
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 274515 bases at least Q40
            Consensus quality: 275937 bases at least Q30
            Consensus quality: 276861 bases at least Q20
            Insert size: 280640; sum-of-contigs
            Insert size: 261465; 4.0% error; agarose-fp
            Quality coverage: 9.23x in Q20 bases; sum-of-contigs Quality
            coverage: 10.00x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 17 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     6899: contig of 6899 bp in length
            *    6900     6999: gap of 100 bp
            *    7000    20575: contig of 13576 bp in length
            *   20576    20675: gap of 100 bp
            *   20676    38280: contig of 17605 bp in length
            *   38281    38380: gap of 100 bp
            *   38381    76364: contig of 37984 bp in length
            *   76365    76464: gap of 100 bp
            *   76465    95438: contig of 18974 bp in length
            *   95439    95538: gap of 100 bp
            *   95539   114600: contig of 19062 bp in length
            *  114601   114700: gap of 100 bp
            *  114701   134153: contig of 19453 bp in length
            *  134154   134253: gap of 100 bp
            *  134254   167898: contig of 33645 bp in length
            *  167899   167998: gap of 100 bp
            *  167999   171913: contig of 3915 bp in length
            *  171914   172013: gap of 100 bp
            *  172014   176193: contig of 4180 bp in length
            *  176194   176293: gap of 100 bp
            *  176294   206625: contig of 30332 bp in length
            *  206626   206725: gap of 100 bp
            *  206726   209648: contig of 2923 bp in length
            *  209649   209748: gap of 100 bp
            *  209749   213682: contig of 3934 bp in length
            *  213683   213782: gap of 100 bp
            *  213783   230655: contig of 16873 bp in length
            *  230656   230755: gap of 100 bp
            *  230756   239434: contig of 8679 bp in length
            *  239435   239534: gap of 100 bp
            *  239535   241595: contig of 2061 bp in length
            *  241596   241695: gap of 100 bp
            *  241696   282240: contig of 40545 bp in length.
FEATURES             Location/Qualifiers
     source          1. .282240
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-172B2"
                     /clone_lib="DanioKey"
     misc_feature    1. .6899
                     /note="assembly_fragment:00130
                     fragment_chain:1"
     misc_feature    7000. .20575
                     /note="assembly_fragment:00501
                     fragment_chain:1"
     misc_feature    20676. .38280
                     /note="assembly_fragment:00282
                     fragment_chain:1"
     misc_feature    38381. .76364
                     /note="assembly_fragment:02629
                     fragment_chain:1"
     misc_feature    76465. .95438
                     /note="assembly_fragment:00723
                     fragment_chain:1"
     misc_feature    95539. .114600
                     /note="assembly_fragment:01220
                     fragment_chain:1"
     misc_feature    114701. .134153
                     /note="assembly_fragment:01489
                     fragment_chain:1"
     misc_feature    134254. .167898
                     /note="assembly_fragment:01785
                     fragment_chain:1"
     misc_feature    167999. .171913
                     /note="assembly_fragment:00062
                     fragment_chain:1"
     misc_feature    172014. .176193
                     /note="assembly_fragment:00105
                     fragment_chain:1"
     misc_feature    176294. .206625
                     /note="assembly_fragment:02143
                     fragment_chain:1"
```

```
     misc_feature    206726. .209648
                     /note="assembly_fragment:00050
                     fragment_chain:2"
     misc_feature    209749. .213682
                     /note="assembly_fragment:00082
                     fragment_chain:2"
     misc_feature    213783. .230655
                     /note="assembly_fragment:00958
                     fragment_chain:2"
     misc_feature    230756. .239434
                     /note="assembly_fragment:00171
                     fragment_chain:2"
     misc_feature    239535. .241595
                     /note="assembly_fragment:00042"
     misc_feature    241696. .282240
                     /note="assembly_fragment:03238.0"
ORIGIN

  Query Match             8.6%;  Score 355.4;  DB 2;  Length 282240;
  Best Local Similarity   43.6%;  Pred. No. 3.9e-60;
  Matches  564;  Conservative    0;  Mismatches  728;  Indels    2;  Gaps    2;

Qy        182 CAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATAT 241
              || |||| || |   ||||         | || || || ||      | ||||||||||||||
Db       5750 CATGCAAAAAAAAATAAATAAATAAAAATAATAATAATAATAATATATATATATATATAT 5809

Qy        242 ATATATATATACACACACACACATATATATATACACATATATATACACATATATACAT 301
              |||||||||| | | | | | | |||||||||| | ||||||||| | ||||||| ||
Db       5810 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 5869

Qy        302 ATATACACATATATACACATATATATACACATACATATGTACACATATATATACACATAT 361
              ||||| | ||||||| | ||||||||| | ||| |||| || | ||||||||| | ||||
Db       5870 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 5929

Qy        362 GTATACACATATATACACATATATACACACATATATACACATATATACACACATATATAC 421
              |||| | ||||||| | |||||||| | | ||||||| | ||||||| | | |||||||
Db       5930 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 5989

Qy        422 ACATATATACACATATATACACATATACACATATATACACATATATACATATATACACAT 481
              | ||||||| | ||||||| | ||||| | ||||||| | ||||||| ||||||| | ||
Db       5990 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 6049

Qy        482 ATATATAATATACACACATATATATACACATATATACACACATATATACACATATATACA 541
              |||||| ||||| | | |||||||||| | ||||||| | | ||||||| | |||||||
Db       6050 ATATAT-ATATATATATATATATATATATATATATATATATATATATATATATATATA 6108

Qy        542 CATATATATACACATATATACACATATATACATATATACACATATATATACATATATACA 601
              |||||||||| | |||||||| | |||||||| ||||||| | |||||||| ||||||| |
Db       6109 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 6168

Qy        602 CATATATACATATATACACATATATACATATATACACACATATATACACATACATATACA 661
              ||||||| ||||||| | ||||||| ||||||| | | ||||||| | ||| ||||| |
Db       6169 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 6228

Qy        662 CACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATATG 721
              | | ||| ||||| |||||||||| | |||||||||  |||||||| ||||||||||
Db       6229 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 6288

Qy        722 CTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTT 781
                  |  |||          |  |   ||     |    ||| |   | | | |  |
Db       6289 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 6348

Qy        782 GACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACA 841
              | |   | |||   | |  | || |     |     | |||     | ||| |
Db       6349 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 6408

Qy        842 TAAATGTGTAATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTAACACTATATATATATA 900
              || || | ||  || |  ||| | || |  ||  ||   |  ||    ||||||||||
Db       6409 TATATATATATAAATATATATATATATATATATATATATATANATATATATATATA 6468

Qy        901 TAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTATAGAAGAATGG 960
              |  |      | || |  |  |         || || || |   | || |
Db       6469 AATATATNTNTATNTANATAANTAAAAATATNNNNNAATATANATATANNAAAANANANN 6528

Qy        961 CTAAGATGCCAAAGCTCAGTGTATGTGGCAGAGGCATGGTGTAGGGTGTGTCCAGGTTCA 1020
              |  |  |                                          | | |
Db       6529 NNNATTNNNAANNNNNTNTNNNNNTANNNNNNNTNNATNNNNNNNTNNTTTNNANANNNN 6588

Qy       1021 TATATTGCATTAAGTGTGAGAACACCCTGGAGTATGAACCAAGAAAATGCAAAAGCCAGA 1080
                                      ||          |  |  ||  |           |
Db       6589 NNNNNNNNNNNNNNNNAANNAAANNNNANNNNNAANNAAAAANNANNNNNNNNNNNANN 6648

Qy       1081 AGTGATGGAGGAAATGAGACACAATAATGAAGATATTGAGAGGAGGGTGTGGGCCTAGAG 1140
                  |
Db       6649 NTNNNNNNANNNANNNNTNNNNNNNNNNNNTNNNNNNNNNNNNNNNNNNTNNNNNNN 6708

Qy       1141 TGAAGCTTTTCGTGCCAGTACTTCTTTTGAAGGCCCAGTTCTCTTCTCTCTCGGGGGCTC 1200
                ||                                 | |       |
Db       6709 NNTNAATTNNNNNNNNNNNNNNNNNNTNNNNNNAANTNNNNNNNNNTANTNNNNNNNN 6768

Qy       1201 CTTCATCTCTCATAGAGTCCACAGCTTTTAAGGGCCAACACTTGAGGTCAGCCTGGCTCT 1260
                      |                 ||
Db       6769 NNNNNNNNNNANNNNNNNNNANNNNNNAANNNNNNNNNNNNNNNNAANTNNNNNNNNN 6828

Qy       1261 CTCATTTGAGCTGGATAGAACATTTTAGAGCACCATCTATTCTTCAAGAGGAAGTTTAAA 1320
                    |     ||                          |
Db       6829 NNNNNNTNNNANTAATNNNNNTNNNNNNNANNNNNNNNNNNNTNNNNNNNNTNNNNNNT 6888

Qy       1321 AATAAAAGAACCTTGAAGAGGAAAAAATGTAGACATTCAATCTAACCTTTTCATTTTACT 1380
                  ||
Db       6889 NNNNNNNNAAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 6948

Qy       1381 AGCCAAAGCTAAATAGAATGCAGGTTACCTGTTTTTCAGCCAGGCACCATCATTTCCTAA 1440
                                                                |   ||
Db       6949 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTATAAATAT 7008

Qy       1441 TTGTTATAAAATTTATTATTATTGTTGTTATTAT 1474
                | ||   | || |||  | ||  |   |  |||
Db       7009 ATATTTATATATATATATTAATATTATATTATAT 7042


RESULT 34
AC024180/c
LOCUS       AC024180               149969 bp    DNA     linear   HTG 01-SEP-2000
DEFINITION  Homo sapiens chromosome 3 clone RP11-172A10, WORKING DRAFT
            SEQUENCE, 17 unordered pieces.
ACCESSION   AC024180
VERSION     AC024180.3  GI:8569838
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 149969)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 149969)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-FEB-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Jun 16, 2000 this sequence version replaced gi:7630874.

            ---------------- Genome Center ----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0172A10
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 100%
            Sequencing vector: plasmid; 0%
            Chemistry: Dye-primer ET; 97% of reads
            Chemistry: Dye-terminator Big Dye; 3% of reads
```

```
            Assembly program: Phrap; version 0.990319
            Consensus quality: 141584 bases at least Q40
            Consensus quality: 144367 bases at least Q30
            Consensus quality: 145803 bases at least Q20
            Insert size: 150000; agarose-fp
            Insert size: 148369; sum-of-contigs
            Quality coverage: 3.76 in Q20 bases; agarose-fp
            Quality coverage: 4.04 in Q20 bases; sum-of-contigs
            ----------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 17 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1731: contig of 1731 bp in length
            *     1732     1831: gap of unknown length
            *     1832     4275: contig of 2444 bp in length
            *     4276     4375: gap of unknown length
            *     4376     6562: contig of 2187 bp in length
            *     6563     6662: gap of unknown length
            *     6663     9713: contig of 3051 bp in length
            *     9714     9813: gap of unknown length
            *     9814    13035: contig of 3222 bp in length
            *    13036    13135: gap of unknown length
            *    13136    17265: contig of 4130 bp in length
            *    17266    17365: gap of unknown length
            *    17366    23146: contig of 5781 bp in length
            *    23147    23246: gap of unknown length
            *    23247    28187: contig of 4941 bp in length
            *    28188    28287: gap of unknown length
            *    28288    35880: contig of 7593 bp in length
            *    35881    35980: gap of unknown length
            *    35981    46076: contig of 10096 bp in length
            *    46077    46176: gap of unknown length
            *    46177    56492: contig of 10316 bp in length
            *    56493    56592: gap of unknown length
            *    56593    67536: contig of 10944 bp in length
            *    67537    67636: gap of unknown length
            *    67637    79604: contig of 11968 bp in length
            *    79605    79704: gap of unknown length
            *    79705    94913: contig of 15209 bp in length
            *    94914    95013: gap of unknown length
            *    95014   112048: contig of 17035 bp in length
            *   112049   112148: gap of unknown length
            *   112149   130002: contig of 17854 bp in length
            *   130003   130102: gap of unknown length
            *   130103   149969: contig of 19867 bp in length.
FEATURES             Location/Qualifiers
     source          1. .149969
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-172A10"
     misc_feature    1. .1731
                     /note="assembly_name:Contig10"
     misc_feature    1832. .4275
                     /note="assembly_name:Contig11"
     misc_feature    4376. .6562
                     /note="assembly_name:Contig12
                     clone_end:T7
                     vector_side:left"
     misc_feature    6663. .9713
                     /note="assembly_name:Contig13"
     misc_feature    9814. .13035
                     /note="assembly_name:Contig14"
     misc_feature    13136. .17265
                     /note="assembly_name:Contig15"
     misc_feature    17366. .23146
                     /note="assembly_name:Contig16"
     misc_feature    23247. .28187
                     /note="assembly_name:Contig17
                     clone_end:SP6
                     vector_side:left"
     misc_feature    28288. .35880
                     /note="assembly_name:Contig18"
     misc_feature    35981. .46076
                     /note="assembly_name:Contig19"
     misc_feature    46177. .56492
                     /note="assembly_name:Contig20"
     misc_feature    56593. .67536
                     /note="assembly_name:Contig21"
     misc_feature    67637. .79604
                     /note="assembly_name:Contig22"
     misc_feature    79705. .94913
                     /note="assembly_name:Contig23"
     misc_feature    95014. .112048
                     /note="assembly_name:Contig24"
     misc_feature    112149. .130002
                     /note="assembly_name:Contig25"
     misc_feature    130103. .149969
                     /note="assembly_name:Contig26"
ORIGIN

  Query Match             8.6%;  Score 355.2;  DB 2;  Length 149969;
  Best Local Similarity   68.3%;  Pred. No. 4.7e-60;
  Matches  596;  Conservative    0;  Mismatches  263;  Indels   13;  Gaps    7;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              || ||||||||| ||||    |||| ||||||||||||||||||||||||||| ||||||
Db      74523 AGATCAAGACCATCCTGGCTAACACGGTGAAACCCCATCTCTACTAAAAATACAAAAATT 74464

Qy         60 AGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGAA 119
              || ||||||||||      | |||||| |||||||||||| |||||| ||||||||||||
Db      74463 AGCCAGGCATGGTGGCGGGCACCTGTAGTCCCAGCTACTTGGGAGGCTGAGGCAGGAGAA 74404

Qy        120 TCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAGC-AAGATCATGCCATTGCACTCCA 178
              |  | ||||||||  || ||||||| |||| | |||| |||||   |||| |||||||||
Db      74403 TGGCCTGAACCTG--GGAGGTGGAGCTTGCAGTGAGCTAAGATGGCGCCACTGCACTCCA 74346

Qy        179 GCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATA 238
              |||  |||||| ||||| | | | |   |      |||||| | |      |  |||| ||||
Db      74345 GCCTGGGCAAC-AGAGCAAGAAGACTCCATCTCAAAAAATATATTATGTACATATGTATA 74287

Qy        239 TATATATATATATACACACACACACATATATATATACACATATATATACACATATATA 298
              | | | |||| |||   | | |||||||| |||| ||| | |  ||| |||| |||| ||
Db      74286 TGTGTCTATACATATGTATATACACACATGTATACATATATACTTATGTACATATATGTA 74227

Qy        299 CATATATACACATATATACACATATATATACACATACATATGTACACATATATATACACA 358
              |||  ||| |||||| || | | ||||| | |  || |||  || |||||| |||| |||
Db      74226 TATACATATACATATGTATATACATATACATATGTATATACATATACATATGTATATACA 74167

Qy        359 TATGTATACACATATATACACATATATACACACATATATACACATATATACACACATATA 418
              ||||||||    ||||| || ||| |||| | || ||||||||||||| | |    |||||
Db      74166 CATGTATAC----ATATATACTTATGTACATATATGTATACACATATACATATGTATATA 74111

Qy        419 TACACATATATACACATATATACACATATACACATATATACACATATATACATATATACA 478
              ||||||  ||||||||||||    |  || |||||| ||||||||||| | || |||| |
Db      74110 CACACATGTATACACATATACTTATGTACACACATGTATACACATATACATATGTATATA 74051

Qy        479 CATATATATA---ATATACACACATATATATACACATATATACACACATATATACACATA 535
              || | || ||   | ||| ||| || |||||||||| || || |||||||||| | |  ||
Db      74050 CACACATGTATACACATATACATATGTATATACACACATGTATACACATATACATATGTA 73991

Qy        536 TATACACATATATATACACATATATACACATATATACATATATACACATATATATACATA 595
              |||||||| || |||||||||||| | |  |||||||| | ||  | | | |||||||||
Db      73990 TATACACACATGTATACACATATACATATGTATATACACACATGTATACACATATACATA 73931

Qy        596 TATACACATATATACATATATACACATATATACATATATACACACATATATACACATACA 655
              | || | | | |  |||| ||| ||| |||  || |||||   ||||| ||| |||||
Db      73930 TGTATATACACACATGTATACACATATACATATGTACATACATGTATATACACATATACA 73871

Qy        656 TATACACACACATAGATATACATATATATACAC-ATATATATACGTATATATATGTATAT 714
```

```
                |||  ||| | ||  |||||||| |||   ||  || ||||| ||||||| || || ||
Db      73870 TATGTACATATATGTATATACATACATACTTACTGTACATATATGTATATACATATACAT 73811

Qy        715 ATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAG 774
              || |||   |     |  |||         |   ||   |     ||   ||| |   | |
Db      73810 ATGTATATACACCCATGTATACATATATACTTATGTACATATATGTATATACATATATAC 73751

Qy        775 AGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGG 834
                |  |  | |   |  ||      |    || |  |  |  |    | | |||
Db      73750 TTATGTACATATATGAATACATATATGTAGATACATGTGTATATGTGTAGATACATATAT 73691

Qy        835 AAATACATAAATGTGTAATAAAACCTATCTGT 866
                |||||||  ||| ||   |    ||| |||
Db      73690 GAATACATATGTGTATATAGATGTGTATGTGT 73659


RESULT 35
CR759776/c
LOCUS       CR759776             151047 bp    DNA     linear   HTG 08-SEP-2004
DEFINITION  Danio rerio clone DKEY-205B10, *** SEQUENCING IN PROGRESS ***, 5
            unordered pieces.
ACCESSION   CR759776
VERSION     CR759776.3  GI:51965222
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 151047)
  AUTHORS   McLay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 9, 2004 this sequence version replaced gi:51890366.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK205B10
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 149837 bases at least Q40
            Consensus quality: 150168 bases at least Q30
            Consensus quality: 150366 bases at least Q20
            Insert size: 150647; sum-of-contigs
            Insert size: 157117; 1.3% error; agarose-fp
            Quality coverage: 8.56x in Q20 bases; sum-of-contigs Quality
            coverage: 8.40x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    21342: contig of 21342 bp in length
            *    21343    21442: gap of 100 bp
            *    21443    23542: contig of 2100 bp in length
            *    23543    23642: gap of 100 bp
            *    23643    56824: contig of 33182 bp in length
            *    56825    56924: gap of 100 bp
            *    56925   140273: contig of 83349 bp in length
            *   140274   140373: gap of 100 bp
            *   140374   151047: contig of 10674 bp in length.
FEATURES             Location/Qualifiers
     source          1. .151047
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-205B10"
                     /clone_lib="DanioKey"
     misc_feature    1. .21342
                     /note="assembly_fragment:00167
                     fragment_chain:1"
     misc_feature    21443. .23542
                     /note="assembly_fragment:00022
                     fragment_chain:1"
     misc_feature    23643. .56824
                     /note="assembly_fragment:00416
                     fragment_chain:1"
     misc_feature    56925. .140273
                     /note="assembly_fragment:00819
                     fragment_chain:1"
     misc_feature    140374. .151047
                     /note="assembly_fragment:00054"
ORIGIN

  Query Match             8.6%;  Score 355.2;  DB 2;  Length 151047;
  Best Local Similarity   64.9%;  Pred. No. 4.7e-60;
  Matches  470;  Conservative    0;  Mismatches  248;  Indels    6;  Gaps    2;

Qy        227 TTTATATATATATATATATATATATACACACACACACATATATATATACACATATATA 286
              | |||||||||||||||||||||||||| | | | | | |||||||||| | |||||||
Db     140869 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 140810

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db     140809 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 140750

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db     140749 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 140690

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db     140689 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 140630

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | |||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db     140629 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 140570

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db     140569 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 140510

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db     140509 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 140450

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | ||||| ||||| ||||||||| | |||||||||  ||||||
Db     140449 TATATATATATATATATACATATATATATATATATATATATATATATATATATATATATA 140390

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||
Db     140389 TATATATATATATATANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 140330

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
                                                                      |||
Db     140329 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNAAAA 140270

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA-----AGGTAGTTAAGAA 880
              | |    | ||| ||| || | || | | |  ||| | ||     |  ||  ||   |
Db     140269 AAAAAAATTATATATATATTTATATTTATATATATATATATATATATATATATATATA 140210

Qy        881 GGTAACACTATATATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATT 940
                ||    |||||||||||||   | || | |  |||  | |||   |   | | ||
Db     140209 TATATATATATATATATATATATATAGTACTTAAAAAATGAAGACATGAAGACAGCTT 140150
```

```
Qy        941 TAAA 944
              ||||
Db     140149 TAAA 140146


RESULT 36
AC006441/c
LOCUS       AC006441               162996 bp    DNA     linear   PRI 14-MAY-1999
DEFINITION  Homo sapiens chromosome 17, clone hRPC.56_K_13, complete sequence.
ACCESSION   AC006441
VERSION     AC006441.13  GI:4760435
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 162996)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone hRPC.56_K_13
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 162996)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JAN-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 162996)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-MAY-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 8, 1999 this sequence version replaced gi:4755216.
            All repeats were identified using RepeatMasker: Smit, A.F.A. &
            Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html.
FEATURES             Location/Qualifiers
     source          1. .162996
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="hRPC.56_K_13"
                     /clone_lib="Peter de Jong/ human PAC library"
     repeat_region   90. .376
                     /rpt_family="AluSx"
     repeat_region   1084. .1390
                     /rpt_family="AluSq"
     unsure          1092. .1093
                     /note="Unresolved compression. Probably GC."
     repeat_region   1424. .1557
                     /rpt_family="AluSg"
     repeat_region   complement(2260. .2414)
                     /rpt_family="AluSc"
     repeat_region   2562. .2857
                     /rpt_family="AluY"
     repeat_region   3102. .3129
                     /rpt_family="(CAAAAA)n"
     repeat_region   complement(3499. .3790)
                     /rpt_family="AluSg"
     repeat_region   complement(3931. .4252)
                     /rpt_family="LTR28"
     repeat_region   complement(4253. .4693)
                     /rpt_family="Tigger1"
     repeat_region   4694. .4996
                     /rpt_family="AluSq"
     repeat_region   complement(4997. .5181)
                     /rpt_family="Tigger1"
     repeat_region   complement(5186. .5466)
                     /rpt_family="AluSg"
     repeat_region   complement(5472. .6215)
                     /rpt_family="Tigger1"
     repeat_region   7119. .7436
                     /rpt_family="AluSq"
     repeat_region   7451. .7521
                     /rpt_family="MADE1"
     repeat_region   7945. .8251
                     /rpt_family="AluSq"
     repeat_region   complement(8354. .8748)
                     /rpt_family="MER21A"
     repeat_region   8743. .9125
                     /rpt_family="LTR1B"
     repeat_region   complement(9126. .9419)
                     /rpt_family="AluSg"
     repeat_region   9420. .9581
                     /rpt_family="LTR1B"
     repeat_region   complement(9784. .9915)
                     /rpt_family="MER21B"
     repeat_region   complement(9916. .10213)
                     /rpt_family="AluSg"
     repeat_region   complement(10214. .10249)
                     /rpt_family="MER21B"
     repeat_region   complement(10305. .10409)
                     /rpt_family="MER21A"
     repeat_region   complement(10410. .11116)
                     /rpt_family="LTR8"
     repeat_region   complement(11117. .11155)
                     /rpt_family="MER21A"
     repeat_region   11215. .11433
                     /rpt_family="MER33"
     repeat_region   11434. .11718
                     /rpt_family="AluSx"
     repeat_region   11719. .11827
                     /rpt_family="MER33"
     repeat_region   complement(11851. .12153)
                     /rpt_family="AluSx"
     repeat_region   complement(12155. .12281)
                     /rpt_family="MIR"
     repeat_region   12282. .12619
                     /rpt_family="MER1B"
     repeat_region   complement(12620. .12649)
                     /rpt_family="MIR"
     repeat_region   12656. .12809
                     /rpt_family="L1MA2"
     repeat_region   complement(12941. .13659)
                     /rpt_family="LTR1B"
     repeat_region   13662. .13779
                     /rpt_family="L1MA3"
     repeat_region   13780. .14075
```

```
                     /rpt_family="AluJb"
     repeat_region   14076. .14240
                     /rpt_family="L1MA3"
     repeat_region   complement(14333. .14627)
                     /rpt_family="AluY"
     repeat_region   14628. .14651
                     /rpt_family="AT_rich"
     repeat_region   complement(14691. .14761)
                     /rpt_family="MIR"
     repeat_region   15905. .16202
                     /rpt_family="AluJo"
     repeat_region   complement(16333. .16724)
                     /rpt_family="MSTC"
     repeat_region   complement(16777. .16993)
                     /rpt_family="MIR"
     repeat_region   complement(17020. .17125)
                     /rpt_family="L1MC5"
     repeat_region   complement(17141. .17428)
                     /rpt_family="AluSg1"
     repeat_region   complement(17432. .17570)
                     /rpt_family="AluSq/x"
     repeat_region   complement(17582. .17653)
                     /rpt_family="L1MC5"
     repeat_region   17954. .18218
                     /rpt_family="AluSp"
     repeat_region   18219. .18289
                     /rpt_family="(GGGAA)n"
     repeat_region   18359. .18405
                     /rpt_family="(TG)n"
     repeat_region   18639. .18676
                     /rpt_family="(TG)n"
     repeat_region   18878. .19179
                     /rpt_family="AluJb"
     repeat_region   19202. .19503
                     /rpt_family="AluSx"
     repeat_region   complement(20063. .20356)
                     /rpt_family="AluSx"
     repeat_region   20835. .20877
                     /rpt_family="L2"
     repeat_region   20878. .20952
                     /rpt_family="GA-rich"
     repeat_region   complement(21223. .21533)
                     /rpt_family="AluSx"
     repeat_region   complement(21991. .22245)
                     /rpt_family="L2"
     repeat_region   complement(22402. .22474)
                     /rpt_family="MIR"
     repeat_region   complement(22472. .22790)
                     /rpt_family="L2"
     repeat_region   complement(22924. .23524)
                     /rpt_family="L2"
     repeat_region   23525. .23829
                     /rpt_family="AluJb"
     repeat_region   complement(23830. .23861)
                     /rpt_family="L2"
     repeat_region   23936. .24244
                     /rpt_family="AluSx"
     repeat_region   24247. .24443
                     /rpt_family="AluSg/x"
     repeat_region   25664. .25971
                     /rpt_family="AluY"

  Query Match             8.6%;  Score 355.2;  DB 9;  Length 162996;
  Best Local Similarity   66.4%;  Pred. No. 4.6e-60;
  Matches  616;  Conservative    0;  Mismatches  303;  Indels    9;  Gaps    7;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC--AAAAT 58
              || || |||||| ||||   |||| ||||||||||  ||||||||||| ||||  |||||
Db      79561 AGATCGAGACCAACCTGGCTAACACGGTGAAACCCTGTCTCTACTAAATATACAAAAAT 79502

Qy         59 TAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGA 118
                | | ||  || ||     |||||||| |||||||||||| |||| | ||||||||||||
Db      79501 AAGCCGGGAGTGATGGCGGGCGCCTGTAGTCCCAGCTACTTGGGAGACTGAGGCAGGAGA 79442

Qy        119 ATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCC 177
              ||  | | ||||  | || ||||||| |||  | ||| ||||||| ||||| ||||||||
Db      79441 ATGGCGTTAACC--CGGGAGGTGGAGCTTGTAGTGAGCCAAGATCGTGCCACTGCACTCC 79384

Qy        178 AGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATAT 237
              ||||  ||  || ||||||| |      |||| ||||||||||||     | |||||||||
Db      79383 AGCCTGGGTGAC-AGAGCGAGACTCTGGCTCAAAAAAAAAAAAAAATATATATATATATAT 79325

Qy        238 ATATATATATATATACACACACACACATATATATATACACATATATATACACATATAT 297
              | ||||||| ||||| | ||| | | | | |||||||||| | | ||| ||| | | ||||
Db      79324 AAATATATAAATATATATACAAATATATAAATATATATAAATACATAAATATAAAAATAT 79265

Qy        298 ACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACAC 357
              | | ||||| | | ||||| | | | ||||| | |||||||  || | || |||||| |
Db      79264 AAAAATATATATAAATATATATA-AAATATATAAATACATAAATATATATTTATATATAA 79206

Qy        358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417
              |||  || | | |||| || | ||| ||||| | ||| ||| | ||| ||| | | ||||
Db      79205 ATACATAAATATATATTTATATATAAATACATAAATAAATATATATACATATATAAATAT 79146

Qy        418 ATACACATATATACACATATATACACATATACACATATATACACATATATACATATATAC 477
              ||| | | ||| | | ||||| | | ||| | | ||||||| | ||||| | |||||||
Db      79145 ATAAATACATAAAAATATATAAATATATAAAAATATATATAAATATATAAAAATATATAT 79086

Qy        478 ACATATATATAATATACACACATATATATACACATATATACACACATATATACACATATA 537
              | ||||||| || | | | | ||| ||||| ||| ||||| | | ||||| | | || ||
Db      79085 AAATATATACAA-AAATATATATAAATATATACAAATATATATAAATATAAAAATATGTA 79027

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATACATATA 597
              || | ||||||| | | ||||||| | ||||| | ||||||| | ||||||| | |||||
Db      79026 TAAATATATATAAAAATATATATAAATATATAAAAATATATATAAATATATAAATATATA 78967

Qy        598 TACACATATATACATATATACACATATATACATATATACACACATATATACACATACATA 657
              || | ||||| | ||||||| | ||||| | ||||||| | | ||| ||| | ||| | |
Db      78966 TAAATATATAAATATATATAAATATATAAATATATATAAATATATAGATATATATATAAA 78907

Qy        658 TACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATA 717
              || | | |  || ||||| |||||||| | | |||||||     |||| |||| |||  |||
Db      78906 TATATATATCTATATATAAATATATAGATATATATATAACTATATAAATATATATGAATA 78847

Qy        718 TATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGA 777
              |||     |   |||      | | |           |   ||| | ||| | | |
Db      78846 TATATAACTATATAAATATATATGAATATATATAACTATATAAATATATATGAATATATA 78787

Qy        778 GTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAA 837
                | |  | | |||    | |   | | |                 |   |       ||
Db      78786 TAACTATATAAATATATATGAATATATATAACTATATAAATATATATGAATATATATAAC 78727

Qy        838 TACATAAATGTGTAATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTAACACTATATATA 896
              || |||||| | ||  || |  || || || |  ||  | |  |  ||  |||||||| |
Db      78726 TATATAAATATATATGAATATATAACTATATAAATATATGAATATATATAACTATATAAA 78667

Qy        897 TATATAGTGAAAGCAGTGTAAACCTAAA 924
              ||||||   | |      || |  ||||
Db      78666 TATATATGAATATATAAATATATATAAA 78639


RESULT 37
AL590381/c
LOCUS       AL590381              182430 bp    DNA     linear   PRI 16-MAY-2001
DEFINITION  Human DNA sequence from clone RP11-373O10 on chromosome 9, complete
            sequence.
ACCESSION   AL590381
VERSION     AL590381.4  GI:13872470
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 182430)
  AUTHORS   Griffiths,C.
  TITLE     Direct Submission
```

```
JOURNAL   Submitted (16-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire,
          CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
          requests: clonerequest@sanger.ac.uk
COMMENT   On Apr 27, 2001 this sequence version replaced gi:13752046.
          During sequence assembly data is compared from overlapping clones.
          Where differences are found these are annotated as variations
          together with a note of the overlapping clone name. Note that the
          variation annotation may not be found in the sequence submission
          corresponding to the overlapping clone, as we submit sequences with
          only a small overlap as described above.
          This sequence was finished as follows unless otherwise noted: all
          regions were either double-stranded or sequenced with an alternate
          chemistry or covered by high quality data (i.e., phred quality >=
          30); an attempt was made to resolve all sequencing problems, such
          as compressions and repeats; all regions were covered by at least
          one plasmid subclone or more than one M13 subclone; and the
          assembly was confirmed by restriction digest. The following
          abbreviations are used to associate primary accession numbers given
          in the feature table with their source databases: Em:, EMBL; Sw:,
          SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
          database can be found at
          http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
          was generated from part of bacterial clone contigs of human
          chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
          Group.  Further information can be found at
          http://www.sanger.ac.uk/HGP/Chr9
          RP11-373O10 is from the library RPCI-11.2 constructed by the group
          of Pieter de Jong. For further details see
          http://www.chori.org/bacpac/home.htm
          VECTOR: pBACe3.6
          This sequence is the entire insert of clone RP11-373O10 The true
          left end of clone RP11-82L2 is at 128987 in this sequence. The true
          right end of clone RP11-346G21 is at 9227 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .182430
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /clone="RP11-373O10"
                     /clone_lib="RPCI-11.2"
     repeat_region   8. .317
                     /note="L1M4 repeat: matches 2333. .2656 of consensus"
     repeat_region   348. .683
                     /note="L1M4 repeat: matches 2747. .3082 of consensus"
     repeat_region   763. .1062
                     /note="AluSc repeat: matches 1. .296 of consensus"
     repeat_region   1568. .1826
                     /note="L1M4c repeat: matches 1694. .1953 of consensus"
     repeat_region   2178. .2434
                     /note="L1MB8 repeat: matches 5132. .5403 of consensus"
     repeat_region   2454. .2584
                     /note="MLT1E repeat: matches 35. .157 of consensus"
     repeat_region   2638. .2888
                     /note="MLT1E repeat: matches 312. .561 of consensus"
     repeat_region   2895. .3648
                     /note="L1MB8 repeat: matches 5403. .6171 of consensus"
     repeat_region   4203. .4252
                     /note="L1MC2 repeat: matches 6263. .6312 of consensus"
     repeat_region   4262. .4578
                     /note="L1M4 repeat: matches 2683. .3032 of consensus"
     repeat_region   4866. .5228
                     /note="L1PB3 repeat: matches 5785. .6149 of consensus"
     repeat_region   5844. .5891
                     /note="24 copies 2 mer ta 75% conserved"
     repeat_region   5893. .6306
                     /note="23 copies 18 mer 82% conserved"
     repeat_region   5909. .6298
                     /note="13 copies 30 mer 83% conserved"
     repeat_region   6301. .6597
                     /note="AluY repeat: matches 1. .296 of consensus"
     repeat_region   6743. .6839
                     /note="L1M4 repeat: matches 4041. .4137 of consensus"
     repeat_region   7150. .7519
                     /note="THE1C repeat: matches 1. .371 of consensus"
     repeat_region   8154. .8217
                     /note="THE1B repeat: matches 1. .62 of consensus"
     repeat_region   8237. .8467
                     /note="MIR repeat: matches 2. .253 of consensus"
     repeat_region   8679. .8814
                     /note="AluJo/FRAM repeat: matches 159. .295 of consensus"
     repeat_region   9551. .9910
                     /note="THE1B repeat: matches 1. .363 of consensus"
     repeat_region   11221. .11700
                     /note="L1MC4 repeat: matches 7390. .7847 of consensus"
     repeat_region   13929. .14162
                     /note="L1PA4 repeat: matches 1430. .1662 of consensus"
     repeat_region   14156. .18641
                     /note="L1PA4 repeat: matches 1666. .6146 of consensus"
     repeat_region   18690. .19243
                     /note="L1ME3A repeat: matches 5523. .6091 of consensus"
     repeat_region   19861. .20102
                     /note="L1MEc repeat: matches 1817. .2070 of consensus"
     repeat_region   20203. .20330
                     /note="L1MEc repeat: matches 2167. .2303 of consensus"
     repeat_region   20430. .21121
                     /note="L1 repeat: matches 2894. .3643 of consensus"
     repeat_region   21332. .22100
                     /note="L1M4 repeat: matches 3912. .4721 of consensus"
     repeat_region   22457. .22551
                     /note="L1MC/D repeat: matches 5374. .5466 of consensus"
     repeat_region   23344. .23992
                     /note="L1PA4 repeat: matches 5496. .6144 of consensus"
     repeat_region   26070. .26233
                     /note="MLT1G repeat: matches 12. .179 of consensus"
     repeat_region   26946. .27101
                     /note="LTR1 repeat: matches 631. .785 of consensus"
     repeat_region   27115. .27210
                     /note="LTR1 repeat: matches 1. .95 of consensus"
     repeat_region   27634. .27945
                     /note="AluSx repeat: matches 1. .312 of consensus"
     repeat_region   29014. .29302
                     /note="AluJb repeat: matches 1. .290 of consensus"
     repeat_region   30332. .30492
                     /note="L1MEc repeat: matches 1749. .1897 of consensus"
     repeat_region   31684. .32009
                     /note="MLT1A2 repeat: matches 34. .374 of consensus"
     repeat_region   32010. .32313
                     /note="AluY repeat: matches 1. .305 of consensus"
     repeat_region   33320. .33629
                     /note="AluSx repeat: matches 1. .308 of consensus"
     repeat_region   33683. .33975
                     /note="AluSc repeat: matches 1. .295 of consensus"
     repeat_region   33986. .34132
                     /note="MIR repeat: matches 100. .243 of consensus"
     repeat_region   34156. .34309
                     /note="L2 repeat: matches 2575. .2734 of consensus"
     repeat_region   34739. .34862
                     /note="MIR repeat: matches 2. .139 of consensus"
     repeat_region   35497. .35578
                     /note="41 copies 2 mer tt 67% conserved"
     repeat_region   36076. .36464
                     /note="L1M4c repeat: matches 1599. .1989 of consensus"
     repeat_region   36471. .36974
                     /note="MLT1D repeat: matches 3. .499 of consensus"
     repeat_region   36977. .37504
                     /note="L1M4c repeat: matches 1001. .1604 of consensus"
     repeat_region   37507. .38001
                     /note="L1ME2 repeat: matches 5652. .6156 of consensus"
     repeat_region   38998. .39313
                     /note="AluSx repeat: matches 1. .289 of consensus"
     repeat_region   39775. .40319
                     /note="MLT2D repeat: matches 1. .553 of consensus"
     repeat_region   41011. .41583
                     /note="MLT2D repeat: matches 1. .545 of consensus"
     repeat_region   42023. .42458
```

```
                     /note="MLT2G repeat: matches 1. .386 of consensus"
     repeat_region   42910. .42964
                     /note="11 copies 5 mer atata 72% conserved"
     repeat_region   43137. .43251
                     /note="L1MB6 repeat: matches 5937. .6053 of consensus"
     repeat_region   43240. .43428
                     /note="L1MB6 repeat: matches 5584. .5776 of consensus"
     repeat_region   43431. .44672
                     /note="L1MA4 repeat: matches 5053. .6289 of consensus"
     repeat_region   44661. .44816
                     /note="L1MA4 repeat: matches 4931. .5090 of consensus"
     repeat_region   44810. .45261
                     /note="L1 repeat: matches 3313. .3738 of consensus"
     repeat_region   45270. .45407
                     /note="L1M4 repeat: matches 5474. .5595 of consensus"
     repeat_region   45415. .45634
                     /note="AluSg repeat: matches 1. .301 of consensus"
     repeat_region   45635. .45677
                     /note="AluSg repeat: matches 257. .299 of consensus"
     repeat_region   45680. .45891
                     /note="L1M4 repeat: matches 5254. .5470 of consensus"
     repeat_region   46033. .47277
                     /note="L1PA14 repeat: matches 4885. .6149 of consensus"
     repeat_region   47274. .47631
                     /note="L1MB8 repeat: matches 5816. .6169 of consensus"
     repeat_region   47654. .47961
                     /note="AluSx repeat: matches 1. .308 of consensus"
     repeat_region   47964. .49294
                     /note="L1MA4 repeat: matches 4951. .6299 of consensus"
     repeat_region   49298. .49946
                     /note="L1MD3 repeat: matches 6922. .7588 of consensus"
     repeat_region   51174. .51342
                     /note="MER90 repeat: matches 4. .169 of consensus"
     repeat_region   52038. .52075
                     /note="19 copies 2 mer tt 81% conserved"
     repeat_region   52894. .53118
                     /note="MIR repeat: matches 14. .245 of consensus"
     repeat_region   53389. .53773
                     /note="L1ME1 repeat: matches 5778. .6159 of consensus"
     repeat_region   54054. .54383
                     /note="L2 repeat: matches 2370. .2709 of consensus"

  Query Match             8.6%;  Score 354.8;  DB 9;  Length 182430;
  Best Local Similarity   73.8%;  Pred. No. 5.4e-60;
  Matches  549;  Conservative    0;  Mismatches  177;  Indels   18;  Gaps    7;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAA---A 57
              ||  |  |||||| |||    ||||  |||||||||| ||||||| ||||||||  ||   |
Db       6524 AGATTGAGACCAACCTAGCGAACACTGTGAAACCCCGTCTCTACCAAAAATACCAAAATA 6465

Qy         58 TTAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAG 117
              ||||   |||  ||||     ||||||||  ||| ||||||| |||||| ||||||||||
Db       6464 TTAGCTGGGCGCGGTGGCCGGCGCCTGTAGGCCCGGCTACTTGGGAGGCTGAGGCAGGAG 6405

Qy        118 AATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTC 176
              |||    ||||||||  || ||  ||| |||| | ||| | |||    |||| |||||||
Db       6404 AATGGTGTGAACCTG--GGAGGCAGAGCTTGCAGTGAGCCGAGACTGCGCCACTGCACTC 6347

Qy        177 CAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATA 236
              |||||  |||||||       | | |      | ||||||||||     | |||| |||
Db       6346 CAGCCTGGGCAACAGAGCAAGACTCTGTCTCAAAAAAAAAAAAAAAATATATATATACATA 6287

Qy        237 TATATATATATATATACACACACACACACATATATATATACACATATATATACACATATA 296
              |||||||| | ||||| | ||||| | | ||||| | ||| | ||| | ||| | ||| |
Db       6286 TATATATACACATATATATACACATATATATATACACATATATATACACATATATATACA 6227

Qy        297 TACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACA 356
              |||||||||| ||| ||||| | | ||||||||||  ||||||  || ||||||||||||
Db       6226 CACATATATATACACATATATATACATATATACAC--ACATATATATACATATATATACA 6169

Qy        357 CATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATA 416
              ||||| |||||||| |||||||||||||| ||| ||||||||| ||||||||| | |||
Db       6168,CATATATATACACACATATACACATATATATACATATATATACATATATATACATATATA 6109

Qy        417 TATACACATATATACACATATATACACATATACAC------ATATATACACATATATACA 470
              | ||| ||||| ||| |||||||||||||| ||      ||||| ||| |||||||||
Db       6108 CACACATATATACACATATATATACACATATATACACATATATATACACATATATATACA 6049

Qy        471 TATATACACATATATATA-ATATACACACATATATATACACATATATACACACATATATA 529
              |||||||||||||||||| ||||| | ||| |||||||||||||| ||| | ||||| |
Db       6048 TATATACACATATATATACATATATATACACATATATACACATATACACATATATATACA 5989

Qy        530 CACATATATACACATATATATACACATATATACACATATATACATATATACACATATATA 589
              | |||| |||| |||||| |||| |||||||||||||||||  | |||||   | | |||
Db       5988 TATATATGTACATATATATGTACATATATATACACATATATGTACATATATGTACACATA 5929

Qy        590 TACATATATACACATATATACATATATACACATATATACATATATACACACATATATACA 649
              || | ||||| ||||||||||||||||||| |||||   |||||||||| ||||| | |
Db       5928 TATACATATATACATATATACATATATACATATATA---ATATATACACATATATACATA 5872

Qy        650 CATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATATG 709
              |||  ||||||| | | | ||| | | ||    |   ||||  || |  |    |||
Db       5871 TATATATATACAGAGAGAGAGAGAGAGATGGTGAATTGCATAAGTAAATTTTCTGATAAT 5812

Qy        710 TATATATATATGCTCCAGAGTTCA 733
                  ||   | | | |||| | |
Db       5811 GGATAATTATTTCACTAGAGATAA 5788


RESULT 38
CR753861/c
LOCUS       CR753861              252450 bp    DNA     linear   HTG 11-OCT-2004
DEFINITION  Danio rerio clone DKEY-192K11, WORKING DRAFT SEQUENCE, 9 unordered
            pieces.
ACCESSION   CR753861
VERSION     CR753861.4  GI:54035031
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 252450)
  AUTHORS   McLaren,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 11, 2004 this sequence version replaced gi:51861477.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK192K11
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 247462 bases at least Q40
            Consensus quality: 248434 bases at least Q30
            Consensus quality: 249070 bases at least Q20
            Insert size: 251650; sum-of-contigs
            Insert size: 239884; 3.5% error; agarose-fp
            Quality coverage: 10.19x in Q20 bases; sum-of-contigs Quality
            coverage: 10.69x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     5351: contig of 5351 bp in length
```

```
            *     5352     5451: gap of 100 bp
            *     5452    19343: contig of 13892 bp in length
            *    19344    19443: gap of 100 bp
            *    19444    58067: contig of 38624 bp in length
            *    58068    58167: gap of 100 bp
            *    58168    75897: contig of 17730 bp in length
            *    75898    75997: gap of 100 bp
            *    75998   166203: contig of 90206 bp in length
            *   166204   166303: gap of 100 bp
            *   166304   216140: contig of 49837 bp in length
            *   216141   216240: gap of 100 bp
            *   216241   224264: contig of 8024 bp in length
            *   224265   224364: gap of 100 bp
            *   224365   228110: contig of 3746 bp in length
            *   228111   228210: gap of 100 bp
            *   228211   252450: contig of 24240 bp in length.
FEATURES             Location/Qualifiers
     source          1. .252450
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-192K11"
                     /clone_lib="DanioKey"
     misc_feature    1. .5351
                     /note="assembly_fragment:00110
                     fragment_chain:1"
     misc_feature    5452. .19343
                     /note="assembly_fragment:00179
                     fragment_chain:1"
     misc_feature    19444. .58067
                     /note="assembly_fragment:00986
                     fragment_chain:1"
     misc_feature    58168. .75897
                     /note="assembly_fragment:00384
                     fragment_chain:1"
     misc_feature    75998. .166203
                     /note="assembly_fragment:02209
                     fragment_chain:2"
     misc_feature    166304. .216140
                     /note="assembly_fragment:01459
                     fragment_chain:2"
     misc_feature    216241. .224264
                     /note="assembly_fragment:00050
                     fragment_chain:2"
     misc_feature    224365. .228110
                     /note="assembly_fragment:00015"
     misc_feature    228211. .252450
                     /note="assembly_fragment:00628.0"
ORIGIN

  Query Match             8.6%;  Score 354;  DB 2;  Length 252450;
  Best Local Similarity   64.7%;  Pred. No. 7.4e-60;
  Matches  452;  Conservative    0;  Mismatches  246;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db     166799 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 166740

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db     166739 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 166680

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db     166679 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 166620

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db     166619 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 166560

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db     166559 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 166500

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db     166499 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 166440

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||| | ||||||| | ||||||| ||||||| | |  ||||
Db     166439 TATATATATATATATATATATACATATATATATATATATATATATATATATTTATA 166380

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || ||||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db     166379 TATACATATATATAGATATATATATATATATATATATATATATATATATATATATATATA 166320

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||  ||||||||||||
Db     166319 TAAATATATATATATGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 166260

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
                                                                      |||
Db     166259 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNAAAA 166200

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885
               |      | | | ||||||   ||   | |  ||| | ||   |  |||   |  ||
Db     166199 AATATTTATATAAAAATAAATANATATATATATATTTATATTTTTTTTATTTATATAT 166140

Qy        886 CACTATATATATATATAGTGAAAGCAGTGTAAACCTAAA 924
                |||||||||||||||   | |   | || |  || |
Db     166139 ATATATATATATATATATATATATATATATATATATA 166101


RESULT 39
CR790381
LOCUS       CR790381              163651 bp    DNA     linear   HTG 24-SEP-2004
DEFINITION  Danio rerio clone DKEY-202K24, *** SEQUENCING IN PROGRESS ***, 10
            unordered pieces.
ACCESSION   CR790381
VERSION     CR790381.3  GI:52673626
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 163651)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 24, 2004 this sequence version replaced gi:52626686.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK202K24
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 159309 bases at least Q40
            Consensus quality: 160146 bases at least Q30
            Consensus quality: 160896 bases at least Q20
            Insert size: 162751; sum-of-contigs
            Insert size: 165246; 2.0% error; agarose-fp
            Quality coverage: 8.55x in Q20 bases; sum-of-contigs Quality
            coverage: 8.63x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
```

```
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     4775: contig of 4775 bp in length
            *    4776     4875: gap of 100 bp
            *    4876    26055: contig of 21180 bp in length
            *   26056    26155: gap of 100 bp
            *   26156    28584: contig of 2429 bp in length
            *   28585    28684: gap of 100 bp
            *   28685    34192: contig of 5508 bp in length
            *   34193    34292: gap of 100 bp
            *   34293    48345: contig of 14053 bp in length
            *   48346    48445: gap of 100 bp
            *   48446    55079: contig of 6634 bp in length
            *   55080    55179: gap of 100 bp
            *   55180   119903: contig of 64724 bp in length
            *  119904   120003: gap of 100 bp
            *  120004   140199: contig of 20196 bp in length
            *  140200   140299: gap of 100 bp
            *  140300   150280: contig of 9981 bp in length
            *  150281   150380: gap of 100 bp
            *  150381   163651: contig of 13271 bp in length.
FEATURES             Location/Qualifiers
     source          1. .163651
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-202K24"
                     /clone_lib="DanioKey"
     misc_feature    1. .4775
                     /note="assembly_fragment:00038
                     fragment_chain:1"
     misc_feature    4876. .26055
                     /note="assembly_fragment:00583
                     fragment_chain:1"
     misc_feature    26156. .28584
                     /note="assembly_fragment:00027
                     fragment_chain:1"
     misc_feature    28685. .34192
                     /note="assembly_fragment:00076
                     fragment_chain:1"
     misc_feature    34293. .48345
                     /note="assembly_fragment:00435
                     fragment_chain:1"
     misc_feature    48446. .55079
                     /note="assembly_fragment:00118
                     fragment_chain:1"
     misc_feature    55180. .119903
                     /note="assembly_fragment:01052
                     fragment_chain:1"
     misc_feature    120004. .140199
                     /note="assembly_fragment:00806
                     fragment_chain:1"
     misc_feature    140300. .150280
                     /note="assembly_fragment:00174
                     fragment_chain:1"
     misc_feature    150381. .163651
                     /note="assembly_fragment:00293.0"
ORIGIN

  Query Match             8.6%;  Score 353.6;  DB 2;  Length 163651;
  Best Local Similarity   68.5%;  Pred. No. 9.5e-60;
  Matches  502;  Conservative    0;  Mismatches  230;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db     139185 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 139244

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db     139245 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 139304

Qy    ,   347,TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
```

```
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db     139305 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 139364

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db     139365 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 139424

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db     139425 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 139484

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db     139485 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 139544

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db     139545 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 139604

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| |||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db     139605 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 139664

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||||        | |||       | | |  ||    |   |||
Db     139665 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 139724

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
              |   | | | |  | | |  | |||    | |   | || |     |      | |
Db     139725 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 139784

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885
              ||       | ||| ||| || | ||   | |  ||| | ||    |  ||  ||  ||
Db     139785 TATATATATATATATATATATATATATATATATATATATATATATATAATATATCAATATAT 139844

Qy        886 CACTATATATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAAT 945
              | ||||||||||||||   | |    | ||||  ||||        |   || ||||
Db     139845 AAATATATATATATATATATATATATATATAAAAATAAATATATATATATAAATNTAAAA 139904

Qy        946 GTTATAGAAGAAT 958
                || | |  |||
Db     139905 AATAAATATTAAT 139917
```

```
RESULT 40
AC141288/c
LOCUS       AC141288              164994 bp    DNA     linear   HTG 11-MAR-2003
DEFINITION  Homo sapiens chromosome 16 clone RP11-357A20, WORKING DRAFT
            SEQUENCE, 2 unordered pieces.
ACCESSION   AC141288
VERSION     AC141288.1  GI:28913068
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 164994)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Human Chromosome 16
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 164994)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-MAR-2003) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 543473
            Center clone name: RPCI-11_357A20
```

```
            -------
            Summary Statistics
            Consensus quality: 164710 bases at least Q40
            Consensus quality: 164791 bases at least Q30
            Consensus quality: 164821 bases at least Q20
            Estimated insert size: 175000; agarose-fp estimation
            Estimated insert size: 164894; sum-of-contigs estimation
            Quality coverage:  8.51 in Q20 bases; agarose-fp estimation
            Quality coverage:  9.03 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    66545: contig of 66545 bp in length
            *   66546    66645: gap of unknown length
            *   66646   164994: contig of 98349 bp in length.
FEATURES             Location/Qualifiers
     source          1. .164994
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone="RP11-357A20"
                     /clone_lib="RPCI human BAC library 11"
ORIGIN

  Query Match             8.6%;  Score 353.6;  DB 2;  Length 164994;
  Best Local Similarity   72.6%;  Pred. No. 9.5e-60;
  Matches  527;  Conservative    0;  Mismatches  189;  Indels   10;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATACAAAA--- 57
              || || |||||| ||||   ||||  |||||||||| ||||||||||||||| |||||
Db     149586 AGATCGAGACCATCCTGGCTAACACAGTGAAACCCCGTCTCTACTAAAAATCCAAAAAAA 149527

Qy         58 --TTAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGG 115
                |||| | ||| |||||     | |||||| |||||||||||  |||||| ||||||||
Db     149526 AGTTAGCCGGGCGTGGTGGCGGGCACCTGTAGTCCCAGCTACTCTGGAGGCTGAGGCAGG 149467

Qy        116 AGAATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCAC 174
              |||||  | ||||| ||  || ||  ||| |||| | ||| | ||||  ||||| |||||
Db     149466 AGAATGGCGTGAACTTG--GGAGGCAGAGCTTGCAGTGAGCCGAGATGGTGCCACTGCAC 149409

Qy        175 TCCAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATA 234
              ||||| |  ||| || ||||||| |   ||||||| |||||| | |      | |||| |
Db     149408 TCCAGTCTGGGCGAC-AGAGCGAGACTCCATCTCAAAAAAAATATATATATATATAGA 149350

Qy        235 TATATATATATATATACACACACACACATATATATATACACATATATATACACATA 294
              |||||||| ||||||| | | | | | | ||| |||||| | ||||||| | | |||
Db     149349 TATATATAGATATATAGATATATATAGATATATAGATATATATAGATATATAGATATATA 149290

Qy        295 TATACATATATACACATATATACACATATATATACACATACATATGTACACATATATATA 354
              || | ||||| | | ||||||| | ||||||| | | ||| | || || | ||||||| |
Db     149289 TAGATATATAGATATATATATATAGATATATAGATATATATAGATATATAGATATATAGA 149230

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
              | ||||  ||| | | ||||| | ||||||| | | ||| ||| | |||||||| | | |
Db     149229 TATATATAGATATATAGATATATAGATATATAGATATATAGATATAGATATATAGATATA 149170

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              |||| | | ||| ||| | | ||||| | | ||| | | ||||| | ||||||| |  ||
Db     149169 TATAGATATATAGATATATAGATATATATAGATATATAGATATATAGATATATATAGTTA 149110

Qy        475 TACACATATATATAATATACACACATATATATACACATATATACACACATATATACACAT 534
              || | |||||||  ||||| | | ||||| | ||| ||| ||| | | | ||||| | ||
Db     149109 TATAGATATATA-GATATATATAGATATAGAGACATATAGATATATATAGATATAGAGAT 149051

Qy        535 ATATACACATATATATACACATATATACACATATATACATATATACACATATATATACAT 594
              |||||  | |||||  | | |  ||| ||| | |  ||||| | |  ||| |  ||| ||||| |
Db     149050 ATATAGATATATAGAGATATATAGATATATAGATATATAGAGATATATATAGATATATAG 148991

Qy        595 ATATACACATATATACATATATACACATATATACATATATACACACATATATACACATAC 654
              |||||  | |||||||  |||||||  | |||||||  |||||||  |  | |||||   | |||
Db     148990 ATATATAGATATATAGATATATATAGATATATAGATATATAGATATATATAGTTATATAG 148931

Qy        655 ATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATAT 714
              |||||  |  |  |||  | |||  | |  | |||  | |||||||||   ||||  | ||  ||||
Db     148930 ATATATAGATATATATAGATATAGAGACATATAGATATATATAGATATAGAGATATATAG 148871

Qy        715 ATATAT 720
              ||||||
Db     148870 ATATAT 148865


RESULT 41
CR376741/c
LOCUS       CR376741              165979 bp    DNA     linear   HTG 24-SEP-2004
DEFINITION  Danio rerio clone DKEY-114B16, *** SEQUENCING IN PROGRESS ***, 12
            unordered pieces.
ACCESSION   CR376741
VERSION     CR376741.5  GI:52673643
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 165979)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 24, 2004 this sequence version replaced gi:52546082.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK114B16
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 159153 bases at least Q40
            Consensus quality: 160217 bases at least Q30
            Consensus quality: 161345 bases at least Q20
            Insert size: 164879; sum-of-contigs
            Insert size: 169971; 1.8% error; agarose-fp
            Quality coverage: 8.37x in Q20 bases; sum-of-contigs Quality
            coverage: 8.39x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 12 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     9842: contig of 9842 bp in length
            *    9843     9942: gap of 100 bp
            *    9943    22622: contig of 12680 bp in length
            *   22623    22722: gap of 100 bp
            *   22723    45138: contig of 22416 bp in length
            *   45139    45238: gap of 100 bp
            *   45239    70145: contig of 24907 bp in length
            *   70146    70245: gap of 100 bp
            *   70246    73805: contig of 3560 bp in length
            *   73806    73905: gap of 100 bp
            *   73906    77192: contig of 3287 bp in length
            *   77193    77292: gap of 100 bp
            *   77293   101185: contig of 23893 bp in length
```

```
           *   101186   101285: gap of 100 bp
           *   101286   127529: contig of 26244 bp in length
           *   127530   127629: gap of 100 bp
           *   127630   143030: contig of 15401 bp in length
           *   143031   143130: gap of 100 bp
           *   143131   145591: contig of 2461 bp in length
           *   145592   145691: gap of 100 bp
           *   145692   159780: contig of 14089 bp in length
           *   159781   159880: gap of 100 bp
           *   159881   165979: contig of 6099 bp in length.
FEATURES             Location/Qualifiers
     source          1. .165979
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-114B16"
                     /clone_lib="DanioKey"
     misc_feature    1. .9842
                     /note="assembly_fragment:00238
                     fragment_chain:1
                     clone_end:T7
                     vector_side:left"
     misc_feature    9943. .22622
                     /note="assembly_fragment:00377
                     fragment_chain:1"
     misc_feature    22723. .45138
                     /note="assembly_fragment:00739
                     fragment_chain:2"
     misc_feature    45239. .70145
                     /note="assembly_fragment:01657
                     fragment_chain:2"
     misc_feature    70246. .73805
                     /note="assembly_fragment:00125
                     fragment_chain:2"
     misc_feature    73906. .77192
                     /note="assembly_fragment:00093
                     fragment_chain:2"
     misc_feature    77293. .101185
                     /note="assembly_fragment:01273
                     fragment_chain:2"
     misc_feature    101286. .127529
                     /note="assembly_fragment:02046
                     fragment_chain:3"
     misc_feature    127630. .143030
                     /note="assembly_fragment:00558
                     fragment_chain:3"
     misc_feature    143131. .145591
                     /note="assembly_fragment:00062
                     fragment_chain:3"
     misc_feature    145692. .159780
                     /note="assembly_fragment:00999
                     fragment_chain:3"
     misc_feature    159881. .165979
                     /note="assembly_fragment:00165.0"
ORIGIN

  Query Match             8.6%;  Score 353;  DB 2;  Length 165979;
  Best Local Similarity   68.8%;  Pred. No. 1.2e-59;
  Matches  499;  Conservative    0;  Mismatches  225;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db      74825 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 74766

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db      74765 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 74706

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db      74705 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 74646

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | |||||| | ||||||| | |||||
Db      74645 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 74586

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db      74585 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 74526

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db      74525 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 74466

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db      74465 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 74406

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db      74405 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 74346

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||||          | |||          | |   ||    |   |||
Db      74345 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 74286

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
              |   | | | |  | | |   | |||   | |  | || |     |    | |
Db      74285 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 74226

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885
              ||        | ||| |||||| | ||   |     ||| | ||      ||||  |  |||
Db      74225 TATATATATATATATATAAATATATATATATTTATATATATATATATATTAATTATATAA 74166

Qy        886 CACTATATATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAAT 945
              |||||||||||||| | | |  | | || |  ||||  |       |  |   ||||
Db      74165 ATATATATATATATATAATTATAAAATTATATAAATAAATAAAATTAATAAAAAAAAAAT 74106

Qy        946 GTTAT 950
              | ||
Db      74105 ATAAT 74101


RESULT 42
CR627491/c
LOCUS       CR627491               248167 bp    DNA     linear   HTG 07-AUG-2004
DEFINITION  Danio rerio clone DKEY-163B6, *** SEQUENCING IN PROGRESS ***, 11
            unordered pieces.
ACCESSION   CR627491
VERSION     CR627491.2  GI:51036473
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 248167)
  AUTHORS   McLay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-AUG-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 7, 2004 this sequence version replaced gi:50978439.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK163B6
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 244397 bases at least Q40
            Consensus quality: 245102 bases at least Q30
            Consensus quality: 245733 bases at least Q20
```

```
            Insert size: 247167; sum-of-contigs
            Insert size: 234721; 4.3% error; agarose-fp
            Quality coverage: 5.83x in Q20 bases; sum-of-contigs Quality
            coverage: 6.34x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     5680: contig of 5680 bp in length
            *     5681     5780: gap of 100 bp
            *     5781    34320: contig of 28540 bp in length
            *    34321    34420: gap of 100 bp
            *    34421    48965: contig of 14545 bp in length
            *    48966    49065: gap of 100 bp
            *    49066    71965: contig of 22900 bp in length
            *    71966    72065: gap of 100 bp
            *    72066    76303: contig of 4238 bp in length
            *    76304    76403: gap of 100 bp
            *    76404   114940: contig of 38537 bp in length
            *   114941   115040: gap of 100 bp
            *   115041   163084: contig of 48044 bp in length
            *   163085   163184: gap of 100 bp
            *   163185   176987: contig of 13803 bp in length
            *   176988   177087: gap of 100 bp
            *   177088   189934: contig of 12847 bp in length
            *   189935   190034: gap of 100 bp
            *   190035   220399: contig of 30365 bp in length
            *   220400   220499: gap of 100 bp
            *   220500   248167: contig of 27668 bp in length.
FEATURES             Location/Qualifiers
     source          1. .248167
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-163B6"
                     /clone_lib="DanioKey"
     misc_feature    1. .5680
                     /note="assembly_fragment:00062
                     fragment_chain:1"
     misc_feature    5781. .34320
                     /note="assembly_fragment:01271
                     fragment_chain:1"
     misc_feature    34421. .48965
                     /note="assembly_fragment:00097
                     fragment_chain:1"
     misc_feature    49066. .71965
                     /note="assembly_fragment:00433
                     fragment_chain:1"
     misc_feature    72066. .76303
                     /note="assembly_fragment:00035
                     fragment_chain:2"
     misc_feature    76404. .114940
                     /note="assembly_fragment:01575
                     fragment_chain:2"
     misc_feature    115041. .163084
                     /note="assembly_fragment:01979
                     fragment_chain:2"
     misc_feature    163185. .176987
                     /note="assembly_fragment:00318
                     fragment_chain:3"
     misc_feature    177088. .189934
                     /note="assembly_fragment:00207
                     fragment_chain:3"
     misc_feature    190035. .220399
                     /note="assembly_fragment:00968
                     fragment_chain:3"
     misc_feature    220500. .248167
                     /note="assembly_fragment:00686.0"
ORIGIN

  Query Match             8.6%;  Score 353;  DB 2;  Length 248167;
  Best Local Similarity   65.2%;  Pred. No. 1.2e-59;
  Matches  462;  Conservative    0;  Mismatches  245;  Indels    2;  Gaps    2;

Qy        218 AAAAGGCATTTTATATATATATATATATATATACACACACACACATATATATATAC 277
              ||||  ||  |||||||||||||||||||||||| | | | | | ||||||||||
Db     115544 AAAAAACAGCATATATATATATATATATATATATATATATATATATATATATAT 115485

Qy        278 ACATATATATACACATATATACATATATACACATATATACACATATATATACACATACAT 337
              | ||||||||| | ||||||| ||||||| | ||||||| | ||||||||| | ||| ||
Db     115484 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 115425

Qy        338 ATGTACACATATATATACACATATGTATACACATATATACACATATATACACACATATAT 397
              || || | ||||||||| | |||| |||| | ||||||| | ||||||| | | ||||||
Db     115424 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 115365

Qy        398 ACACATATATACACACATATATACACATATATACACATATATACACATATACACATATAT 457
              | | ||||||| | | ||||||| | ||||||| | ||||||| | ||||| | ||||||
Db     115364 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 115305

Qy        458 ACACATATATACATATATACACATATATATA-ATATACACACATATATATACACATATAT 516
              | | ||||||| ||||||| | ||||||||| ||||| | | ||||||||| | ||||||
Db     115304 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 115245

Qy        517 ACACACATATATACACATATATACACATATATATACACATATATACACATATATACATAT 576
              | | | ||||||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||
Db     115244 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 115185

Qy        577 ATACACATATATATACATATATACACATATATACATATATACACATATATACATATATAC 636
              ||| | ||||||||| ||||||| | ||||||| ||||||| | ||||||| |||||||
Db     115184 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 115125

Qy        637 ACACATATATACACATACATATACACACATAGATATACATATATATACACATATATAT 696
              | | ||||||| | ||| ||||| | | | ||| ||||| ||||||||| | ||||||||
Db     115124 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 115065

Qy        697 ACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCA 756
              |  ||||||||| ||||||||||
Db     115064 ATATATATATATATATATATATATNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 115005

Qy        757 CTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTT 816

Db     115004 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 114945

Qy        817 GATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA-AGGTAGTT 875
                  |||||        | ||| |||||| | ||    | |  ||| | || |  ||  |
Db     114944 NNNNAAAAATATATATATATATATATAAATATATATATATATATATATATATATATAT 114885

Qy        876 AAGAAGGTAACACTATATATATATATAGTGAAAGCAGTGTAAACCTAAA 924
              |    |  ||     |||||||||||||    | |     | ||| |   || |
Db     114884 ATATATATATATATATATATATATATATATATATATATATATATA 114836


RESULT 43
AL731541
LOCUS       AL731541              178254 bp    DNA     linear   PRI 20-JUN-2002
DEFINITION  Human DNA sequence from clone RP11-472K8 on chromosome 10, complete
            sequence.
ACCESSION   AL731541 AC016823
VERSION     AL731541.6  GI:21537521
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 178254)
  AUTHORS   Heath,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAY-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
```

```
COMMENT     On Jun 21, 2002 this sequence version replaced gi:21065395.
            Draft Sequence Produced by Genome Therapeutics Corp, 100 Beaver
            Street, Waltham, MA 02453, USA
            http://www.genomecorp.com
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 10, constructed by the Sanger Centre Chromosome 10
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr10
            RP11-472K8 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6.
FEATURES              Location/Qualifiers
     source           1. .178254
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:9606"
                      /chromosome="10"
                      /clone="RP11-472K8"
                      /clone_lib="RPCI-11.2"
ORIGIN

  Query Match             8.6%;  Score 352.4;  DB 9;  Length 178254;
  Best Local Similarity   70.2%;  Pred. No. 1.6e-59;
  Matches  559;  Conservative    0;  Mismatches  226;  Indels   11;  Gaps    6;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC-AAAATT 59
              ||||  ||||||||||||  ||||||||||||||||||| ||| ||||||||||| ||||||
Db     104371 AGTTTGAGACCAGCCTGGCCAACATGGTGAAACCCCAGCTCCACTAAAAATACAAAAATT 104430

Qy         60 AGACAGGCATGGTGATACACGCCTGTA-----ATCCCAGCTACTTCGGAGGCCGAGGCAG 114
              || ||||||| ||| | ||||||||||     |||| |||||||  |||||| |||||||
Db     104431 AGCCAGGCATTGTGGTGCACGCCTGTAATCTCATCCTAGCTACTCGGGAGGCTGAGGCAG 104490

Qy        115 GAGAATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCA 173
              ||||||  ||||||  | | || || |||||||| | ||| | |||||  |||| ||||
Db     104491 AAGAATCGCTTGAA--TCCGGGAGGCGGAGGTTGCAGTGAGCCCAGATCGCGCCACTGCA 104548

Qy        174 CTCCAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATAT 233
              ||||||||  || ||| |||| ||||     | ||||||| | | |         | |
Db     104549 CTCCAGCCTGGGTAAC-AGAGTGAAACCGTGTTTCAGAAAATATACACACACACACACAC 104607

Qy        234 ATATATATATATATATATACACACACACACATATATATATACACATATATATACACAT 293
              | | | | ||||| | |||||||||| | ||||| | | ||||| ||||| |||||
Db     104608 ACACACACACATATACACACACACACACATATATATACACACACACACATATACACACAC 104667

Qy        294 ATATACATATATACACATATATACACATATATATACACATACATATGTACACATATATAT 353
              | ||| ||| | ||||| ||||| ||| | | | | | ||||| |   ||| |||||||
Db     104668 ACATATATACACACACACATATATACACACACACATATATACACACACACATATATATAC 104727

Qy        354 ACACATATGTATACACATATATACACATATATACACACATATATACACATATATACACAC 413
              ||||| || |||| ||| | | ||| |||||||||||||| | ||||| ||||||||||||
Db     104728 ACACACATATATATACACACACACATATATATACACACACACATACATATATATACACAC 104787

Qy        414 ATATATACACATATATACACATATATACACATATACACATATATACACATATATACATAT 473
              | | ||| | | | | ||||| ||||| ||| | |||||||||||| ||| | | ||||||
Db     104788 ACACATATATACACACACACACATATATACACACACACATATATATACACACACACATAT 104847

Qy        474 ATACACATATATATAATATACACACATATATATACACATATATACACACATATATACACA 533
              ||| ||| | | || ||| ||||||| ||| ||||| |||| |    | ||||||  | |
Db     104848 ATATACACACACAT-ATACACACACACATACATACATATATGTGTGTATATATATGTATA 104906

Qy        534 TATATACACATATATATACACATATATACACATATATACATATATACACATATATATACA 593
              || ||| | || | | || | |||| || | | ||||| || | |   | |||||| || |
Db     104907 TACATATATATGTGTGTATATATATGTATATACATATATATGTGTGTATATATATGTATA 104966

Qy        594 TATATACACATATATACATATATACACATATATACATATATACACACATATATACACATA 653
              || ||| | || | |  ||||||    |||||||  |||||||| | |||| ||   |||
Db     104967 TACATATATATGTGTGTATATATGTGTATATATATGTATATACATATATATGTATGTATA 105026

Qy        654 CATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATA 713
              ||||  | | ||||| ||||  || ||||||    |||||||| ||||||| |   | |
Db     105027 TATATGTATATACATATATATGTATGTATATATGTGTATATATATGTATATACACACACA 105086

Qy        714 TATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAA 773
              |||||||          | | | |||          | |  |  |||  |   || |   |
Db     105087 TATATATATATATATATCCCTGAGATAATAATACTTAAAAAGACTTAGTCCAGTGCCCAG 105146

Qy        774 GAGAGTTTGACAGCAG 789
                | |||  ||   |||
Db     105147 CACAGTGGGAACTCAG 105162


RESULT 44
CR752655
LOCUS       CR752655              204486 bp    DNA     linear   HTG 22-SEP-2004
DEFINITION  Danio rerio clone DKEY-107O6, WORKING DRAFT SEQUENCE, 11 unordered
            pieces.
ACCESSION   CR752655
VERSION     CR752655.5  GI:52546093
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 204486)
  AUTHORS   Burton,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 23, 2004 this sequence version replaced gi:52353233.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK107O6
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 201681 bases at least Q40
            Consensus quality: 202171 bases at least Q30
            Consensus quality: 202651 bases at least Q20
            Insert size: 203486; sum-of-contigs
            Insert size: 204129; 3.5% error; agarose-fp
            Quality coverage: 11.12x in Q20 bases; sum-of-contigs Quality
            coverage: 11.09x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
```

```
          * be preserved.
          *       1    10267: contig of 10267 bp in length
          *   10268    10367: gap of 100 bp
          *   10368    14914: contig of 4547 bp in length
          *   14915    15014: gap of 100 bp
          *   15015    25928: contig of 10914 bp in length
          *   25929    26028: gap of 100 bp
          *   26029    68196: contig of 42168 bp in length
          *   68197    68296: gap of 100 bp
          *   68297   108045: contig of 39749 bp in length
          *  108046   108145: gap of 100 bp
          *  108146   114266: contig of 6121 bp in length
          *  114267   114366: gap of 100 bp
          *  114367   118079: contig of 3713 bp in length
          *  118080   118179: gap of 100 bp
          *  118180   139935: contig of 21756 bp in length
          *  139936   140035: gap of 100 bp
          *  140036   151139: contig of 11104 bp in length
          *  151140   151239: gap of 100 bp
          *  151240   189955: contig of 38716 bp in length
          *  189956   190055: gap of 100 bp
          *  190056   204486: contig of 14431 bp in length.
FEATURES             Location/Qualifiers
     source          1. .204486
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-107O6"
                     /clone_lib="DanioKey"
     misc_feature    1. .10267
                     /note="assembly_fragment:00170
                     fragment_chain:1"
     misc_feature    10368. .14914
                     /note="assembly_fragment:00029
                     fragment_chain:1"
     misc_feature    15015. .25928
                     /note="assembly_fragment:00569
                     fragment_chain:1"
     misc_feature    26029. .68196
                     /note="assembly_fragment:03030
                     fragment_chain:1"
     misc_feature    68297. .108045
                     /note="assembly_fragment:02194
                     fragment_chain:1"
     misc_feature    108146. .114266
                     /note="assembly_fragment:00286
                     fragment_chain:2"
     misc_feature    114367. .118079
                     /note="assembly_fragment:00088
                     fragment_chain:2"
     misc_feature    118180. .139935
                     /note="assembly_fragment:00956
                     fragment_chain:2"
     misc_feature    140036. .151139
                     /note="assembly_fragment:00410
                     fragment_chain:2"
     misc_feature    151240. .189955
                     /note="assembly_fragment:01419
                     fragment_chain:2"
     misc_feature    190056. .204486
                     /note="assembly_fragment:00735.0"
ORIGIN

  Query Match             8.6%;  Score 352.4;  DB 2;  Length 204486;
  Best Local Similarity   69.8%;  Pred. No. 1.6e-59;
  Matches  504;  Conservative    0;  Mismatches  216;  Indels    2;  Gaps    2;

Qy        209 AGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACATA 268
              ||||||||||||||     | |||||||||||||||||||||||| | | | | | | |||
Db     107299 AGAAAAAAAAAAAAAAATATATATATATATATATATATATATATATATATATATATATA 107358

Qy        269 TATATATACACATATATATACACATATATACATATATACACATATATACACATATATATA 328
              |||||||| | ||||||||| | ||||||| ||||||| | ||||||| | |||||||||

Db     107359 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 107418

Qy        329 CACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACA 388
              | ||| |||| || | ||||||||| | |||| |||| | ||||||| | ||||||| |
Db     107419 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 107478

Qy        389 CACATATATACACATATATACACACATATATACACATATATACACATATATACACATATA 448
              | ||||||| | ||||||| | | ||||||| | ||||||| | ||||||| | |||||
Db     107479 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 107538

Qy        449 CACATATATACACATATATACATATATACACATATATATA-ATATACACACATATATATA 507
              | ||||||| | ||||||| ||||||| | ||||||||| ||||| | | |||||||||
Db     107539 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 107598

Qy        508 CACATATATACACACATATATACACATATATACACATATATATACACATATATACACATA 567
              | ||||||| | | ||||||| | ||||||| | ||||||||| | ||||||| | |||
Db     107599 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 107658

Qy        568 TATACATATATACACATATATATACATATATACACATATATACATATATACACATATATA 627
              |||| ||||||| | |||||||| ||||||| | ||||||| ||||||| | |||||||
Db     107659 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 107718

Qy        628 CATATATACACACATATATACACATACATATACACACACATAGATATACATATATATACA 687
              ||||||| | | ||||||| | ||| ||||| | | | ||| ||||| |||||||||
Db     107719 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 107778

Qy        688 CATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGT 747
              |||||||||  |||||||||  ||||||||||          | |||          |
Db     107779 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 107838

Qy        748 TGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACA 807
              |   ||   |   ||| |   | | | |  | |   | |||   | |   | |
Db     107839 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 107898

Qy        808 TTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA 867
              | |   |     | |||     | ||| ||| || | ||  | |  ||| | ||
Db     107899 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 107958

Qy        868 -AGGTAGTTAAGAAGGTAACACTATATATATATAGTGAAAGCAGTGTAAACCTAAAGG 926
               |  ||  ||  ||  ||     |||||| |||||| | ||| || | || |  || |
Db     107959 TATATATATATAAATATATATATATACATATATATATAAATCAATATATATATATATA 108018

Qy        927 AT 928
              ||
Db     108019 AT 108020


RESULT 45
CR392006/c
LOCUS       CR392006              178427 bp    DNA     linear   HTG 23-AUG-2004
DEFINITION  Danio rerio clone DKEYP-57F11, *** SEQUENCING IN PROGRESS ***, 9
            unordered pieces.
ACCESSION   CR392006
VERSION     CR392006.3  GI:51510714
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 178427)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-AUG-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 23, 2004 this sequence version replaced gi:46879107.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
```

```
          Center project name: zKp57F11
          ------------ Summary Statistics
          Assembly program: XGAP4; version 4.5
          Chemistry: Dye-terminator; 100% of reads
          Consensus quality: 173955 bases at least Q40
          Consensus quality: 174654 bases at least Q30
          Consensus quality: 175292 bases at least Q20
          Insert size: 177627; sum-of-contigs
          Insert size: 177796; 2.7% error; agarose-fp
          Quality coverage: 7.97x in Q20 bases; sum-of-contigs Quality
          coverage: 8.05x in Q20 bases; agarose-fp
          --------------.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 9 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1    11505: contig of 11505 bp in length
          *    11506    11605: gap of 100 bp
          *    11606    24931: contig of 13326 bp in length
          *    24932    25031: gap of 100 bp
          *    25032    31819: contig of 6788 bp in length
          *    31820    31919: gap of 100 bp
          *    31920    50297: contig of 18378 bp in length
          *    50298    50397: gap of 100 bp
          *    50398    83841: contig of 33444 bp in length
          *    83842    83941: gap of 100 bp
          *    83942   122559: contig of 38618 bp in length
          *   122560   122659: gap of 100 bp
          *   122660   166383: contig of 43724 bp in length
          *   166384   166483: gap of 100 bp
          *   166484   175663: contig of 9180 bp in length
          *   175664   175763: gap of 100 bp
          *   175764   178427: contig of 2664 bp in length.
FEATURES             Location/Qualifiers
     source          1. .178427
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEYP-57F11"
                     /clone_lib="DanioKeypilot"
     misc_feature    1. .11505
                     /note="assembly_fragment:00415
                     fragment_chain:1"
     misc_feature    11606. .24931
                     /note="assembly_fragment:00266
                     fragment_chain:1"
     misc_feature    25032. .31819
                     /note="assembly_fragment:00041
                     fragment_chain:1"
     misc_feature    31920. .50297
                     /note="assembly_fragment:00579
                     fragment_chain:1"
     misc_feature    50398. .83841
                     /note="assembly_fragment:00848
                     fragment_chain:1"
     misc_feature    83942. .122559
                     /note="assembly_fragment:01340
                     fragment_chain:1"
     misc_feature    122660. .166383
                     /note="assembly_fragment:01877
                     fragment_chain:1"
     misc_feature    166484. .175663
                     /note="assembly_fragment:00124
                     fragment_chain:1"
     misc_feature    175764. .178427
                     /note="assembly_fragment:00006"
ORIGIN

  Query Match,            8.5%;  Score 352.2;  DB 2;  Length 178427;
  Best Local Similarity   60.3%;  Pred. No. 1.8e-59;
  Matches  502;  Conservative    0;  Mismatches  328;  Indels    3;  Gaps    1;

Qy        194 GCGAAATGTCATCTCAGAAAAAAAAAAAGGCATTTTATATATATATATATATATATATAC 253
              || ||||  |   | |||  ||||||||   |  ||||||||||||||||||||||||
Db      12310 GCAAAATAACTGGTAAGATGAAAAAAAAAAAAAAATATATATATATATATATATATATAT 12251

Qy        254 ACACACACACACATATATATATACACATATATATACACATATATACACATATATACACATAT 313
              | | | | | | ||||||||||| | ||||||||| | ||||||| ||||||| | ||||
Db      12250 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 12191

Qy        314 ATACACATATATATACACATACATATGTACACATATATATACACATATGTATACACATAT 373
              ||| | ||||||||| | ||| |||| || | ||||||||| | |||| |||| | ||||
Db      12190 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 12131

Qy        374 ATACACATATATACACACATATATACACATATATACACACATATATACACATATATACAC 433
              ||| | ||||||| | | ||||||| | ||||||| | | ||||||| | ||||||| |
Db      12130 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 12071

Qy        434 ATATATACACATATACACATATATACACATATATACATATATACACATATATATAATATA 493
              ||||||| | ||||| | ||||||| | ||||||| ||||||| | ||||||   |||||
Db      12070 ATATATATATATATATATATATATATATATATATATATATATATATATATATAT---ATATA 12014

Qy        494 CACACATATATATACACATATATACACACATATATACACATATATACACATATATATACA 553
              | | ||||||||| | ||||||| | | ||||||| | ||||||| | ||||||||| |
Db      12013 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 11954

Qy        554 CATATATACACATATATACATATATACACATATATATACATATATACACATATATACATA 613
              ||||||| | ||||||| ||||||| | ||||||||| ||||||| | ||||||| |||
Db      11953 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 11894

Qy        614 TATACACATATATACATATATACACACATATATACACATACATATACACACACATAGATA 673
              |||| | ||||||| ||||||| | | ||||||| | ||| ||||| | | | ||| |||
Db      11893 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 11834

Qy        674 TACATATATATACACATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCA 733
              || ||||||||| | |||||||||  ||||||||| ||||||||||         | |
Db      11833 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 11774

Qy        734 TAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTA 793
              ||          | |  ||     |   ||| |   | | | |  |  | |    | ||
Db      11773 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 11714

Qy        794 TTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAAT 853
              |    | |  | || |     |       | |||       | ||| ||| || | ||
Db      11713 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 11654

Qy        854 AAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATATATATATAGTGAAAGCAGT 913
              | |  ||||     |  ||  ||   |  ||    ||||||||||||
Db      11653 TATATATATCATATATATATATATATATATATATATATATATATATATNNNNNNNNNNNN 11594

Qy        914 GTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTATAGAAGAATGGCTAAGATGCCAAA 973

Db      11593 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 11534

Qy        974 GCTCAGTGTATGTGGCAGAGGCATGGTGTAGGGTGTGTCCAGGTTCATATATT 1026
                                      |       | |   |  |  |||||||
Db      11533 NNNNNNNNNNNNNNNNNNNNNNNNNNNTTTATAATATATAATTATATATATT 11481


RESULT 46
CR548632/c
LOCUS       CR548632              184630 bp    DNA     linear   HTG 08-JUL-2004
DEFINITION  Danio rerio clone DKEY-175H17, WORKING DRAFT SEQUENCE, 8 unordered
            pieces.
ACCESSION   CR548632
VERSION     CR548632.3  GI:50080050
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
```

```
REFERENCE   1  (bases 1 to 184630)
  AUTHORS   Burton,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUL-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 8, 2004 this sequence version replaced gi:50057714.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK175H17
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 181205 bases at least Q40
            Consensus quality: 181924 bases at least Q30
            Consensus quality: 182467 bases at least Q20
            Insert size: 183930; sum-of-contigs
            Insert size: 181082; 1.5% error; agarose-fp
            Quality coverage: 10.04x in Q20 bases; sum-of-contigs Quality
            coverage: 10.36x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    14714: contig of 14714 bp in length
            *    14715    14814: gap of 100 bp
            *    14815    95419: contig of 80605 bp in length
            *    95420    95519: gap of 100 bp
            *    95520   132623: contig of 37104 bp in length
            *   132624   132723: gap of 100 bp
            *   132724   165602: contig of 32879 bp in length
            *   165603   165702: gap of 100 bp
            *   165703   169699: contig of 3997 bp in length
            *   169700   169799: gap of 100 bp
            *   169800   174998: contig of 5199 bp in length
            *   174999   175098: gap of 100 bp
            *   175099   177397: contig of 2299 bp in length
            *   177398   177497: gap of 100 bp
            *   177498   184630: contig of 7133 bp in length.
FEATURES             Location/Qualifiers
     source          1. .184630
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-175H17"
                     /clone_lib="DanioKey"
     misc_feature    1. .14714
                     /note="assembly_fragment:00296
                     fragment_chain:1"
     misc_feature    14815. .95419
                     /note="assembly_fragment:01430
                     fragment_chain:1"
     misc_feature    95520. .132623
                     /note="assembly_fragment:00810
                     fragment_chain:1"
     misc_feature    132724. .165602
                     /note="assembly_fragment:00444
                     fragment_chain:1"
     misc_feature    165703. .169699
                     /note="assembly_fragment:00028
                     fragment_chain:1"
     misc_feature    169800. .174998
                     /note="assembly_fragment:00068
                     fragment_chain:2"
     misc_feature    175099. .177397
                     /note="assembly_fragment:00017
                     fragment_chain:2"
     misc_feature    177498. .184630
                     /note="assembly_fragment:00176.0"
ORIGIN

  Query Match             8.5%;  Score 352.2;  DB 2;  Length 184630;
  Best Local Similarity   69.2%;  Pred. No. 1.7e-59;
  Matches  491;  Conservative    0;  Mismatches  218;  Indels    1;  Gaps    1;

Qy        216 AAAAAAGGCATTTTATATATATATATATATATATACACACACACACATATATATAT 275
              ||| ||    ||||||||||||||||||||||||||||| | | | | | ||||||||||
Db        744 AAATAAAAGATTTTATATATATATATATATATATATATATATATATATATATATAT 685

Qy        276 ACACATATATATACACATATATACATATATACACATATATACACATATATATACACATAC 335
              | | ||||||||| | ||||||| ||||||| | ||||||| | ||||||||| | |||
Db        684 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 625

Qy        336 ATATGTACACATATATATACACATATGTATACACATATATACACATATATACACACATAT 395
              |||| || | ||||||||| | |||| |||| | ||||||| | ||||||| | | ||||
Db        624 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 565

Qy        396 ATACACATATATACACACATATATACACATATATACACATATATACACATATACACATAT 455
              ||| | ||||||| | ||||||| | ||||||| | ||||||| | ||||| | ||||
Db        564 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 505

Qy        456 ATACACATATATACATATATACACATATATATA-ATATACACACATATATATACACATAT 514
              ||| | ||||||| ||||||| | ||||||||| ||||| | | ||||||||| | ||||
Db        504 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 445

Qy        515 ATACACACATATATACACATATATACACATATATATACACATATATACACATATATACAT 574
              ||| | | |||||||| | ||||||| | ||||||||| | ||||||| | ||||||| ||
Db        444 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 385

Qy        575 ATATACACATATATATACATATATACACATATATACATATATACACATATATACATATAT 634
              ||||| | ||||||||| ||||||| | ||||||| ||||||| | ||||||| ||||||
Db        384 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 325

Qy        635 ACACACATATATACACATACATATACACACACATAGATATACATATATATACACATATAT 694
              | | | ||||||| | ||| ||||| | | | ||| ||||| ||||||||| | ||||||
Db        324 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 265

Qy        695 ATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTAC 754
              |||  |||||||| ||||||||||          | |||         |  |   ||
Db        264 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 205

Qy        755 CACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGG 814
                  |   ||| |   | | | |  | |     | |||    | |   | || |
Db        204 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 145

Qy        815 TTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGT 874
                |  | |||       | ||| ||| || | ||   | |  ||| | | |  ||
Db        144 ATATATATATATATATATATATATATATATATATATATATATATATATNTATTATATATA 85

Qy        875 TAAGAAGGTAACACTATATATATATAGTGAAAGCAGTGTAAACCTAAA 924
              |||  |  ||    |||| ||||||  |   | |    | || |  || |
Db         84 TAAATATATATATATATANATATATNNATATATATATATATATATNTATA 35


RESULT 47
CR762385
LOCUS       CR762385              195644 bp    DNA     linear   HTG 19-SEP-2004
DEFINITION  Danio rerio clone DKEY-245L10, *** SEQUENCING IN PROGRESS ***, 7
            unordered pieces.
ACCESSION   CR762385
VERSION     CR762385.3  GI:52346294
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
```

```
REFERENCE   1  (bases 1 to 195644)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 19, 2004 this sequence version replaced gi:51988239.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK245L10
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 193251 bases at least Q40
            Consensus quality: 193792 bases at least Q30
            Consensus quality: 194212 bases at least Q20
            Insert size: 195044; sum-of-contigs
            Insert size: 196613; 1.8% error; agarose-fp
            Quality coverage: 8.31x in Q20 bases; sum-of-contigs Quality
            coverage: 8.42x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    38651: contig of 38651 bp in length
            *    38652    38751: gap of 100 bp
            *    38752    58158: contig of 19407 bp in length
            *    58159    58258: gap of 100 bp
            *    58259   117292: contig of 59034 bp in length
            *   117293   117392: gap of 100 bp
            *   117393   123058: contig of 5666 bp in length
            *   123059   123158: gap of 100 bp
            *   123159   150654: contig of 27496 bp in length
            *   150655   150754: gap of 100 bp
            *   150755   160809: contig of 10055 bp in length
            *   160810   160909: gap of 100 bp
            *   160910   195644: contig of 34735 bp in length.
FEATURES             Location/Qualifiers
     source          1. .195644
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-245L10"
                     /clone_lib="DanioKey"
     misc_feature    1. .38651
                     /note="assembly_fragment:01153
                     fragment_chain:1"
     misc_feature    38752. .58158
                     /note="assembly_fragment:00183
                     fragment_chain:1"
     misc_feature    58259. .117292
                     /note="assembly_fragment:01608
                     fragment_chain:1"
     misc_feature    117393. .123058
                     /note="assembly_fragment:00020
                     fragment_chain:2"
     misc_feature    123159. .150654
                     /note="assembly_fragment:00406
                     fragment_chain:2"
     misc_feature    150755. .160809
                     /note="assembly_fragment:00070.0"
     misc_feature    160910. .195644
                     /note="assembly_fragment:00721"
ORIGIN
```

```
  Query Match             8.5%;  Score 352.2;  DB 2;  Length 195644;
  Best Local Similarity   59.1%;  Pred. No. 1.7e-59;
  Matches  518;  Conservative    0;  Mismatches  358;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | |||||||||||||||||||||||| | | | | | |||||||||| | |||||||
Db     122359 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 122418

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db     122419 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 122478

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db     122479 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 122538

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db     122539 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 122598

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db     122599 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 122658

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db     122659 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 122718

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db     122719 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 122778

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db     122779 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 122838

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| |||||||||||          | |||          |  |   ||    |   |||
Db     122839 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 122898

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
              |  | | | | |  | |  | |||   | |   | || |     |   | |  |
Db     122899 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 122958

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885
              ||         | ||| ||| || | ||   | |  ||| |   |  || ||  ||  |
Db     122959 TATATATAAATATAAATATATATATATATATATATATAATATATATATATAATAAATA 123018

Qy        886 CACTATATATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAAT 945
              | |||||||||||||| | | |    | ||| |  ||
Db     123019 AAATATATATATATAATTATATATATATATAAATATTTNNNNNNNNNNNNNNNNNNNN 123078

Qy        946 GTTATAGAAGAATGGCTAAGATGCCAAAGCTCAGTGTATGTGGCAGAGGCATGGTGTAGG 1005
                .
Db     123079 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 123138

Qy       1006 GTGTGTCCAGGTTCATATATTGCATTAAGTGTGAGAACACCCTGGAGTATGAACCAAGAA 1065
                                  | ||| | | | |     | | |  | | |  | | ||
Db     123139 NNNNNNNNNNNNNNNNNNNCGTTTTATGGATAATATTTTGCGGAACAAGGCCCTTAAAA 123198

Qy       1066 AATGCAAAAGCCAGAAGTGATGGAGGAAATGAGACAC 1102
              | ||   | |  | |  | || | |   | |  || |
Db     123199 AGTGGGGAGGGAACACCTTATTGGGTCTAGGTAACCC 123235
```

```
RESULT 48
CR749751/c
LOCUS       CR749751              224393 bp    DNA     linear   HTG 27-AUG-2004
DEFINITION  Danio rerio clone DKEY-24H18, WORKING DRAFT SEQUENCE, 10 unordered
            pieces.
ACCESSION   CR749751
VERSION     CR749751.4  GI:51591749
```

```
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 224393)
  AUTHORS   Burton,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-AUG-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 27, 2004 this sequence version replaced gi:51510864.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK24H18
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 221304 bases at least Q40
            Consensus quality: 221921 bases at least Q30
            Consensus quality: 222377 bases at least Q20
            Insert size: 223493; sum-of-contigs
            Insert size: 225057; 1.0% error; agarose-fp
            Quality coverage: 10.31x in Q20 bases; sum-of-contigs Quality
            coverage: 10.28x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2873: contig of 2873 bp in length
            *     2874     2973: gap of 100 bp
            *     2974    66389: contig of 63416 bp in length
            *    66390    66489: gap of 100 bp
            *    66490   106977: contig of 40488 bp in length
            *   106978   107077: gap of 100 bp
            *   107078   145207: contig of 38130 bp in length
            *   145208   145307: gap of 100 bp
            *   145308   153395: contig of 8088 bp in length
            *   153396   153495: gap of 100 bp
            *   153496   167269: contig of 13774 bp in length
            *   167270   167369: gap of 100 bp
            *   167370   175339: contig of 7970 bp in length
            *   175340   175439: gap of 100 bp
            *   175440   179204: contig of 3765 bp in length
            *   179205   179304: gap of 100 bp
            *   179305   200083: contig of 20779 bp in length
            *   200084   200183: gap of 100 bp
            *   200184   224393: contig of 24210 bp in length.
FEATURES             Location/Qualifiers
     source          1. .224393
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-24H18"
                     /clone_lib="DanioKey"
     misc_feature    1. .2873
                     /note="assembly_fragment:00089
                     fragment_chain:1"
     misc_feature    2974. .66389
                     /note="assembly_fragment:02891
                     fragment_chain:1"
     misc_feature    66490. .106977
                     /note="assembly_fragment:01374
                     fragment_chain:1"
     misc_feature    107078. .145207
                     /note="assembly_fragment:02083
                     fragment_chain:1"
     misc_feature    145308. .153395
                     /note="assembly_fragment:00265
                     fragment_chain:1"
     misc_feature    153496. .167269
                     /note="assembly_fragment:00404
                     fragment_chain:2"
     misc_feature    167370. .175339
                     /note="assembly_fragment:00155
                     fragment_chain:2"
     misc_feature    175440. .179204
                     /note="assembly_fragment:00108
                     fragment_chain:2"
     misc_feature    179305. .200083
                     /note="assembly_fragment:00671"
     misc_feature    200184. .224393
                     /note="assembly_fragment:00942.0"
ORIGIN

  Query Match             8.5%;  Score 352.2;  DB 2;  Length 224393;
  Best Local Similarity   79.4%;  Pred. No. 1.7e-59;
  Matches  417;  Conservative    0;  Mismatches  108;  Indels    0;  Gaps    0;

Qy        196 GAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACAC 255
              | ||| | || ||    |   ||||||   | | ||||||||||||||||||||||| |
Db      19100 GTAATTTTATTTCTTTAGTTAAAAAAAAAAATATATATATATATATATATATATATAT 19041

Qy        256 ACACACACACATATATATATACACATATATATACACATATATACATATATACACATATAT 315
              | | | | | |||||||||||| | ||||||||| | ||||||| ||||||| | ||||||
Db.     19040 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 18981

Qy        316 ACACATATATATACACATACATATGTACACATATATATACACATATGTATACACATATAT 375
              | | ||||||||| | ||| |||| || | ||||||||| | |||| |||| | ||||||
Db      18980 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 18921

Qy        376 ACACATATATACACACATATATACACATATATACACACATATATACACATATATACACAT 435
              | | ||||||| | | ||||||| | ||||||| | | ||||||| | ||||||| | ||
Db      18920 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 18861

Qy        436 ATATACACATATACACATATATACACATATATACATATATACACATATATATAATATACA 495
              ||||| | ||||| | ||||||| | ||||||| ||||||||| |||||||| ||||| |
Db      18860 ATATATATATATATATATATATATATATATATATATATATACATATATATATCATATATA 18801

Qy        496 CACATATATATACACATATATACACACATATATACACATATATACACATATATATACACA 555
              | |||||||||| | ||||||| | |||||||||||| ||||||| | ||||||| | | |
Db      18800 TATATATATATATATATATATATATATACATATATACATATATATATATATATACATATA 18741

Qy        556 TATATACACATATATACATATATACACATATATATACATATATACACATATATACATATA 615
              |||||| | ||||||| ||||||| | ||||||||| ||||||| |  |||||| |||||
Db      18740 TATATATATATATATATATATATATATATATATATATATATATATATATATGTATATATATA 18681

Qy        616 TACACATATATACATATATACACACATATATACACATACATATACACACACATAGATATA 675
              || | ||||||| ||||||| | | ||||||| | ||| ||||| | | | ||| |||||
Db      18680 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 18621

Qy        676 CATATATATACACATATATATACGTATATATATGTATATATATAT 720
              ||||||||| | |||||||||  ||||||||| ||||||||||
Db      18620 TATATATATATATATATATATATATATATATATATATATATATAT 18576


RESULT 49
CR762494
LOCUS       CR762494              238804 bp    DNA     linear   HTG 10-OCT-2004
DEFINITION  Danio rerio clone DKEYP-78B11, WORKING DRAFT SEQUENCE, 7 unordered
            pieces.
ACCESSION   CR762494
VERSION     CR762494.3  GI:52345306
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
```

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 238804)
  AUTHORS   McLaren,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 18, 2004 this sequence version replaced gi:52214011.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zKp78B11
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 236234 bases at least Q40
            Consensus quality: 236747 bases at least Q30
            Consensus quality: 237190 bases at least Q20
            Insert size: 238204; sum-of-contigs
            Insert size: 231499; 2.2% error; agarose-fp
            Quality coverage: 11.55x in Q20 bases; sum-of-contigs Quality
            coverage: 11.97x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    21508: contig of 21508 bp in length
            *   21509    21608: gap of 100 bp
            *   21609    47490: contig of 25882 bp in length
            *   47491    47590: gap of 100 bp
            *   47591    54540: contig of 6950 bp in length
            *   54541    54640: gap of 100 bp
            *   54641   140186: contig of 85546 bp in length
            *  140187   140286: gap of 100 bp
            *  140287   163133: contig of 22847 bp in length
            *  163134   163233: gap of 100 bp
            *  163234   166432: contig of 3199 bp in length
            *  166433   166532: gap of 100 bp
            *  166533   238804: contig of 72272 bp in length.
FEATURES             Location/Qualifiers
     source          1. .238804
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEYP-78B11"
                     /clone_lib="DanioKeypilot"
     misc_feature    1. .21508
                     /note="assembly_fragment:00516
                     fragment_chain:1
                     clone_end:SP6
                     vector_side:left"
     misc_feature    21609. .47490
                     /note="assembly_fragment:00888
                     fragment_chain:1"
     misc_feature    47591. .54540
                     /note="assembly_fragment:00086
                     fragment_chain:2"
     misc_feature    54641. .140186
                     /note="assembly_fragment:01354
                     fragment_chain:2"
     misc_feature    140287. .163133
                     /note="assembly_fragment:00178
                     fragment_chain:2"
     misc_feature    163234. .166432
                     /note="assembly_fragment:00035
                     fragment_chain:3"
     misc_feature    166533. .238804
                     /note="assembly_fragment:02624
                     fragment_chain:3"
ORIGIN

  Query Match             8.5%;  Score 352;  DB 2;  Length 238804;
  Best Local Similarity   68.8%;  Pred. No. 1.8e-59;
  Matches  512;  Conservative    0;  Mismatches  230;  Indels    2;  Gaps    2;

Qy        182 CAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATAT 241
              ||   || ||  |   |||   |||  || ||||||||||    |   |||||||||||||
Db     165682 CAAAAAAAAAAAAAAAAAAAACATGACATAAAAAAAAAAAAAAAAAAATATATATATATAT 165741

Qy        242 ATATATATATACACACACACACATATATATATACACATATATATACACATATATACAT 301
              |||||||||| | | | | | |||||||||| | |||||||| | |||||| ||
Db     165742 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 165801

Qy        302 ATATACACATATATACACATATATATACACATACATATGTACACATATATATACACATAT 361
              ||||| | ||||||| | |||||||| | ||| |||| || | |||||||| | ||||
Db     165802 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 165861

Qy        362 GTATACACATATATACACATATATACACACATATATACACATATATACACACATATATAC 421
              |||| | ||||||| | ||||||| | | ||||||| | ||||||| | | |||||||
Db     165862 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 165921

Qy        422 ACATATATACACATATATACACATATACACATATATACACATATATACATATATACACAT 481
              | ||||||| | ||||||| | ||||| | ||||||| | ||||||| ||||||| | ||
Db     165922 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 165981

Qy        482 ATATATAATATACACACATATATATACACATATATACACACATATATACACATATATACA 541
              |||||| ||||| | | ||||||||| | ||||||| | | ||||||| | ||||||| |
Db     165982 ATATAT-ATATATATATATATATATATATATATATATATATATATATATATATATATATA 166040

Qy        542 CATATATATACACATATATACACATATATACATATATACACATATATATACATATATACA 601
              ||||||||| | ||||||| | ||||||| ||||||| | ||||||||| ||||||| |
Db     166041 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 166100

Qy        602 CATATATACATATATACACATATATACATATATACACACATATATACACATACATATACA 661
              ||||||| ||||||| | ||||||| ||||||| | | ||||||| | ||| ||||| |
Db     166101 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 166160

Qy        662 CACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATATG 721
              | | ||| ||||| |||||||||| | |||||||||  |||||||| ||||||||||
Db     166161 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 166220

Qy        722 CTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTT 781
                  |  |||          |  |   ||     |   ||| |   | | | |  |
Db     166221 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 166280

Qy        782 GACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACA 841
              | |    | |||    | |   | || |      |       | |||       | ||| |
Db     166281 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 166340

Qy        842 TAAATGTGTAATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTAACACTATATATATATA 900
              || || | ||   | |  ||| | || |  ||  ||   |  ||    ||||||||||||
Db     166341 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 166400

Qy        901 TAGTGAAAGCAGTGTAAACCTAAA 924
              ||   |||    | || |  || |
Db     166401 TATATAAATATATATATATATA 166424


RESULT 50
CR790372/c
LOCUS       CR790372              299106 bp    DNA     linear   HTG 05-OCT-2004
DEFINITION  Danio rerio clone DKEY-70K13, *** SEQUENCING IN PROGRESS ***, 25
            unordered pieces.
ACCESSION   CR790372
VERSION     CR790372.4  GI:53828302
KEYWORDS    HTG; HTGS_PHASE1.
```

```
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 299106)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 5, 2004 this sequence version replaced gi:53755940.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK70K13
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 289331 bases at least Q40
            Consensus quality: 291856 bases at least Q30
            Consensus quality: 293478 bases at least Q20
            Insert size: 296706; sum-of-contigs
            Insert size: 225769; 2.9% error; agarose-fp
            Quality coverage: 6.22x in Q20 bases; sum-of-contigs Quality
            coverage: 9.16x in Q20 bases; agarose-fp
            --------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 25 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     3928: contig of 3928 bp in length
            *     3929     4028: gap of 100 bp
            *     4029     8417: contig of 4389 bp in length
            *     8418     8517: gap of 100 bp
            *     8518    12283: contig of 3766 bp in length
            *    12284    12383: gap of 100 bp
            *    12384    16946: contig of 4563 bp in length
            *    16947    17046: gap of 100 bp
            *    17047    19834: contig of 2788 bp in length
            *    19835    19934: gap of 100 bp
            *    19935    26863: contig of 6929 bp in length
            *    26864    26963: gap of 100 bp
            *    26964    29135: contig of 2172 bp in length
            *    29136    29235: gap of 100 bp
            *    29236    43722: contig of 14487 bp in length
            *    43723    43822: gap of 100 bp
            *    43823   109813: contig of 65991 bp in length
            *   109814   109913: gap of 100 bp
            *   109914   140749: contig of 30836 bp in length
            *   140750   140849: gap of 100 bp
            *   140850   156362: contig of 15513 bp in length
            *   156363   156462: gap of 100 bp
            *   156463   165292: contig of 8830 bp in length
            *   165293   165392: gap of 100 bp
            *   165393   243285: contig of 77893 bp in length
            *   243286   243385: gap of 100 bp
            *   243386   245562: contig of 2177 bp in length
            *   245563   245662: gap of 100 bp
            *   245663   250094: contig of 4432 bp in length
            *   250095   250194: gap of 100 bp
            *   250195   256405: contig of 6211 bp in length
            *   256406   256505: gap of 100 bp
            *   256506   258788: contig of 2283 bp in length
            *   258789   258888: gap of 100 bp
            *   258889   262780: contig of 3892 bp in length
            *   262781   262880: gap of 100 bp
            *   262881   265140: contig of 2260 bp in length
            *   265141   265240: gap of 100 bp
            *   265241   269635: contig of 4395 bp in length
            *   269636   269735: gap of 100 bp
            *   269736   272991: contig of 3256 bp in length
            *   272992   273091: gap of 100 bp
            *   273092   291385: contig of 18294 bp in length
            *   291386   291485: gap of 100 bp
            *   291486   293629: contig of 2144 bp in length
            *   293630   293729: gap of 100 bp
            *   293730   296461: contig of 2732 bp in length
            *   296462   296561: gap of 100 bp
            *   296562   299106: contig of 2545 bp in length.
FEATURES             Location/Qualifiers
     source          1. .299106
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-70K13"
                     /clone_lib="DanioKey"
     misc_feature    1. .3928
                     /note="assembly_fragment:00025
                     fragment_chain:1"
     misc_feature    4029. .8417
                     /note="assembly_fragment:00178
                     fragment_chain:1"
     misc_feature    8518. .12283
                     /note="assembly_fragment:00051
                     fragment_chain:1"
     misc_feature    12384. .16946
                     /note="assembly_fragment:00068
                     fragment_chain:1"
     misc_feature    17047. .19834
                     /note="assembly_fragment:03721
                     fragment_chain:1"
     misc_feature    19935. .26863
                     /note="assembly_fragment:00206
                     fragment_chain:1"
     misc_feature    26964. .29135
                     /note="assembly_fragment:03693
                     fragment_chain:1"
     misc_feature    29236. .43722
                     /note="assembly_fragment:00405
                     fragment_chain:2"
     misc_feature    43823. .109813
                     /note="assembly_fragment:01445
                     fragment_chain:2"
     misc_feature    109914. .140749
                     /note="assembly_fragment:01030
                     fragment_chain:2"
     misc_feature    140850. .156362
                     /note="assembly_fragment:00596
                     fragment_chain:3"
     misc_feature    156463. .165292
                     /note="assembly_fragment:00282
                     fragment_chain:3"
     misc_feature    165393. .243285
                     /note="assembly_fragment:02358
                     fragment_chain:3"
     misc_feature    243386. .245562
                     /note="assembly_fragment:03730
                     fragment_chain:4"
     misc_feature    245663. .250094
                     /note="assembly_fragment:00104
                     fragment_chain:4"
     misc_feature    250195. .256405
                     /note="assembly_fragment:00151
                     fragment_chain:4"
     misc_feature    256506. .258788
                     /note="assembly_fragment:00124
                     fragment_chain:5"
     misc_feature    258889. .262780
```

```
                     /note="assembly_fragment:00234
                     fragment_chain:5"
     misc_feature    262881. .265140
                     /note="assembly_fragment:03713
                     fragment_chain:6"
     misc_feature    265241. .269635
                     /note="assembly_fragment:00038
                     fragment_chain:6"
     misc_feature    269736. .272991
                     /note="assembly_fragment:00085"
     misc_feature    273092. .291385
                     /note="assembly_fragment:00801.0"
     misc_feature    291486. .293629
                     /note="assembly_fragment:03699"
     misc_feature    293730. .296461
                     /note="assembly_fragment:03706"
     misc_feature    296562. .299106
                     /note="assembly_fragment:03750"
ORIGIN

  Query Match             8.5%;  Score 351.8;  DB 2;  Length 299106;
  Best Local Similarity   66.8%;  Pred. No. 2e-59;
  Matches  531;  Conservative    0;  Mismatches  262;  Indels    2;  Gaps    2;

Qy        171 GCACTCCAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTA 230
              | | ||||     | || |||    | |    ||    |    | ||||||||||      | ||
Db     166223 GTATTCCAATTTAAGCTACAGATGGGGCCTGCGCTAAAAAAAAAAAAAAAAAATATATATA 166164

Qy        231 TATATATATATATATATATATACACACACACACACATATATATATACACATATATATACA 290
              |||||||||||||||||||||| | | | | | | |||||||||||| | ||||||||| |
Db     166163 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 166104

Qy        291 CATATATACATATATACACATATATACACATATATATACACATACATATGTACACATATA 350
              ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |||||
Db     166103 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 166044

Qy        351 TATACACATATGTATACACATATATACACATATATACACACATATATACACATATATACA 410
              |||| | |||| |||| | ||||||| | ||||||| | | ||||||| | ||||||| |
Db     166043 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 165984

Qy        411 CACATATATACACATATATACACATATATACACATATACACATATATACACATATATACA 470
              | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | ||||||| |
Db     165983 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 165924

Qy        471 TATATACACATATATATA-ATATACACACATATATATACACATATATACACACATATATA 529
              |||||| | |||||||| ||||| | | |||||||| | ||||||| | | |||||||
Db     165923 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 165864

Qy        530 CACATATATACACATATATATACACATATATACACATATATACATATATACACATATATA 589
              | ||||||| | |||||||| | ||||||| | ||||||| ||||||| | |||||||
Db     165863 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 165804

Qy        590 TACATATATACACATATATACATATATACACATATATACATATATACACACATATATACA 649
              || ||||||| | ||||||| ||||||| | ||||||| ||||||| | | ||||||| |
Db     165803 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 165744

Qy        650 CATACATATACACACACATAGATATACATATATATACACATATATATACGTATATATATG 709
              ||| ||||| | | | ||| ||||| |||||||| | |||||||||  |||||||||
Db     165743 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 165684

Qy        710 TATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGG 769
              ||||||||||          | |||          | | ||    |     ||| |
Db     165683 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 165624

Qy        770 AAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGT 829
              | | | |  | | |    | |||    | |    | || |      |      | |||
Db     165623 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 165564

Qy        830 AGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTAACAC 888
                  | ||| ||| || | ||   | |  ||| | || |  ||  ||   |  ||
Db     165563 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 165504

Qy        889 TATATATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTT 948
              |||||||||||||    | |    | || |  || |         |   || ||||| |
Db     165503 TATATATATATATATATATATATATATATATATATATATATATATATATATATAAATATA 165444

Qy        949 ATAGAAGAATGGCTA 963
              || |   ||   ||
Db     165443 TTATATTTATAAATA 165429


RESULT 51
CR762407/c
LOCUS       CR762407              149106 bp    DNA     linear   HTG 13-SEP-2004
DEFINITION  Danio rerio clone DKEY-262N17, *** SEQUENCING IN PROGRESS ***, 5
            unordered pieces.
ACCESSION   CR762407
VERSION     CR762407.2  GI:52000335
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 149106)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 13, 2004 this sequence version replaced gi:51988183.
            ------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            ------------- Project Information
            Center project name: zK262N17
            ------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 147385 bases at least Q40
            Consensus quality: 147795 bases at least Q30
            Consensus quality: 148053 bases at least Q20
            Insert size: 148706; sum-of-contigs
            Insert size: 147741; 3.4% error; agarose-fp
            Quality coverage: 9.30x in Q20 bases; sum-of-contigs Quality
            coverage: 9.36x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    36781: contig of 36781 bp in length
            *   36782    36881: gap of 100 bp
            *   36882    62357: contig of 25476 bp in length
            *   62358    62457: gap of 100 bp
            *   62458    68981: contig of 6524 bp in length
            *   68982    69081: gap of 100 bp
            *   69082    89839: contig of 20758 bp in length
            *   89840    89939: gap of 100 bp
            *   89940   149106: contig of 59167 bp in length.
FEATURES             Location/Qualifiers
     source          1. .149106
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-262N17"
                     /clone_lib="DanioKey"
     misc_feature    1. .36781
                     /note="assembly_fragment:00719
                     fragment_chain:1"
```

```
     misc_feature    36882. .62357
                     /note="assembly_fragment:00059
                     fragment_chain:1"
     misc_feature    62458. .68981
                     /note="assembly_fragment:00007"
     misc_feature    69082. .89839
                     /note="assembly_fragment:00379.0"
     misc_feature    89940. .149106
                     /note="assembly_fragment:01240"
ORIGIN

  Query Match             8.5%;  Score 351.6;  DB 2;  Length 149106;
  Best Local Similarity   68.8%;  Pred. No. 2.4e-59;
  Matches  497;  Conservative    0;  Mismatches  224;  Indels    1;  Gaps    1;

Qy        197 AAATGTCATCTCAGAAAAAAAAAAAGGCATTTTATATATATATATATATATATATACACA 256
              || |||  | ||| ||||||||||    |    | ||||||||||||||||||||| | |
Db      90716 AACTGTATTTTCAAAAAAAAAAAAAAAAAAAAAAAAATATATATATATATATATATATA 90657

Qy        257 CACACACACATATATATATACACATATATATACACATATATACATATATACACATATATA 316
              | | | | |||||||||||| | ||||||||| | ||||||| ||||||| | |||||||
Db      90656 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 90597

Qy        317 CACATATATATACACATACATATGTACACATATATATACACATATGTATACACATATATA 376
              | ||||||||| | ||| |||| || | ||||||||| | |||| |||| | |||||||
Db      90596 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 90537

Qy        377 CACATATATACACACATATATACACATATATACACACATATATACACATATATACACATA 436
              | ||||||| | | ||||||| | ||||||| | | ||||||| | ||||||| | |||
Db      90536 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 90477

Qy        437 TATACACATATACACATATATACACATATATACATATATACACATATATATAATATACAC 496
              |||| | ||||| | ||||||| | ||||||| ||||||| | |||||||| ||||| |
Db      90476 TATATATATATATATATATATATATATATATATATATATATATATATATATAT-ATATATAT 90418

Qy        497 ACATATATATACACATATATACACACATATATACACATATATACACATATATATACACAT 556
              | ||||||||| | ||||||| | | ||||||| | ||||||| | ||||||||| | ||
Db      90417 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 90358

Qy        557 ATATACACATATATACATATATACACATATATATACATATATACACATATATACATATAT 616
              ||||| | ||||||| ||||||| | ||||||||| ||||||| | ||||||| ||||||
Db      90357 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 90298

Qy        617 ACACATATATACATATATACACACATATATACACATACATATACACACACATAGATATAC 676
              | | ||||||| ||||||| | | ||||||| | ||| ||||| | | | ||| |||||
Db      90297 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 90238

Qy        677 ATATATATACACATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAA 736
              ||||||||| | |||||||||  |||||||| ||||||||||          |  |||
Db      90237 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 90178

Qy        737 GAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTG 796
                | |  ||   |  ||| |  | | | |  | | |    | |||
Db      90177 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 90118

Qy        797 TGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAA 856
              | |   | || |     |      | |||       | ||| ||| || | ||   |
Db      90117 ATATATATATATATATATATATATAAATATATATATATATATATATATATATATATAT 90058

Qy        857 ACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATATATATATAGTGAAAGCAGTGTA 916
              | ||| |   |  || ||   |  ||     |||||| ||| |||   | |  | | ||
Db      90057 ATATATATAATATATATATATATATATATACATATATACATAAATACACATAAAATTATA 89998

Qy        917 AA 918
              ||
Db      89997 TA 89996


RESULT 52
CR762414
LOCUS       CR762414              171302 bp    DNA     linear   HTG 13-SEP-2004
DEFINITION  Danio rerio clone DKEY-197K18, *** SEQUENCING IN PROGRESS ***, 6
            unordered pieces.
ACCESSION   CR762414
VERSION     CR762414.2  GI:52000341
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 171302)
  AUTHORS   McLay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 13, 2004 this sequence version replaced gi:51988190.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK197K18
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 168888 bases at least Q40
            Consensus quality: 169406 bases at least Q30
            Consensus quality: 169716 bases at least Q20
            Insert size: 170802; sum-of-contigs
            Insert size: 170040; 4.9% error; agarose-fp
            Quality coverage: 7.32x in Q20 bases; sum-of-contigs Quality
            coverage: 7.59x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    15887: contig of 15887 bp in length
            *    15888    15987: gap of 100 bp
            *    15988    42962: contig of 26975 bp in length
            *    42963    43062: gap of 100 bp
            *    43063    58072: contig of 15010 bp in length
            *    58073    58172: gap of 100 bp
            *    58173    60197: contig of 2025 bp in length
            *    60198    60297: gap of 100 bp
            *    60298   112076: contig of 51779 bp in length
            *   112077   112176: gap of 100 bp
            *   112177   171302: contig of 59126 bp in length.
FEATURES             Location/Qualifiers
     source          1. .171302
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-197K18"
                     /clone_lib="DanioKey"
     misc_feature    1. .15887
                     /note="assembly_fragment:00045
                     fragment_chain:1"
     misc_feature    15988. .42962
                     /note="assembly_fragment:00332
                     fragment_chain:1"
     misc_feature    43063. .58072
                     /note="assembly_fragment:00177
                     fragment_chain:2"
     misc_feature    58173. .60197
                     /note="assembly_fragment:00025
                     fragment_chain:2"
     misc_feature    60298. .112076
                     /note="assembly_fragment:00601.0"
```

```
     misc_feature    112177. .171302
                     /note="assembly_fragment:01241"
ORIGIN

  Query Match             8.5%;  Score 351.4;  DB 2;  Length 171302;
  Best Local Similarity   69.0%;  Pred. No. 2.5e-59;
  Matches  486;  Conservative    0;  Mismatches  217;  Indels    1;  Gaps    1;

Qy        226 TTTTATATATATATATATATATATATACACACACACACACATATATATATACACATATAT 285
              || ||||||||||||||||||||||| | | | | | | |||||||||||| | ||||||
Db      42255 TTATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 42314

Qy        286 ATACACATATATACATATATACACATATATACACATATATATACACATACATATGTACAC 345
              ||| | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || |
Db      42315 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 42374

Qy        346 ATATATATACACATATGTATACACATATATACACATATATACACACATATATACACATAT 405
              ||||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | ||||
Db      42375 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 42434

Qy        406 ATACACACATATATACACATATATACACATATATACACATATACACATATATACACATAT 465
              ||| | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | ||||
Db      42435 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 42494

Qy        466 ATACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACAT 524
              ||| ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | ||
Db      42495 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 42554

Qy        525 ATATACACATATATACACATATATATACACATATATACACATATATACATATATACACAT 584
              ||||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | ||
Db      42555 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 42614

Qy        585 ATATATACATATATACACATATATACATATATACACATATATACATATATACACACATAT 644
              ||||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | ||||
Db      42615 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 42674

Qy        645 ATACACATACATATATACACACACATAGATATACATATATATACACATATATATACGTATAT 704
              ||| | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  |||||
Db      42675 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 42734

Qy        705 ATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGAT 764
              |||| ||||||||||||          | |||          | |    ||    |    ||
Db      42735 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 42794

Qy        765 AGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAA 824
              | |   | | | |  | | |    | |||    | |    | || | |         |
Db      42795 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 42854

Qy        825 ATAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTA 884
              |||  |    | ||| |||||| | ||  || |  ||| | ||    |  ||  ||   |
Db      42855 ATATAAAAAAATATATATAAATATATATAAATATATATATNTATATAAAATANAAANNAA 42914

Qy        885 ACACTATATATATATATAGTGAAAGCAGTGTAAACCTAAAGGAT 928
                | |||| ||  | ||| | ||   | |       || |  ||
Db      42915 TAAATATAAATTAANATAATAAANNAAATAATNNATTATAATAT 42958


RESULT 53
CR749761
LOCUS       CR749761              216259 bp    DNA     linear   HTG 10-OCT-2004
DEFINITION  Danio rerio clone CH211-215F19, WORKING DRAFT SEQUENCE, 10
            unordered pieces.
ACCESSION   CR749761
VERSION     CR749761.4  GI:54021792
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 216259)
  AUTHORS   McLaren,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 10, 2004 this sequence version replaced gi:51988273.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zC215F19
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 212820 bases at least Q40
            Consensus quality: 213560 bases at least Q30
            Consensus quality: 214140 bases at least Q20
            Insert size: 215359; sum-of-contigs
            Insert size: 205140; 3.4% error; agarose-fp
            Quality coverage: 7.94x in Q20 bases; sum-of-contigs Quality
            coverage: 8.53x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    48219: contig of 48219 bp in length
            *   48220    48319: gap of 100 bp
            *   48320    65856: contig of 17537 bp in length
            *   65857    65956: gap of 100 bp
            *   65957    86815: contig of 20859 bp in length
            *   86816    86915: gap of 100 bp
            *   86916   108880: contig of 21965 bp in length
            *  108881   108980: gap of 100 bp
            *  108981   118104: contig of 9124 bp in length
            *  118105   118204: gap of 100 bp
            *  118205   185027: contig of 66823 bp in length
            *  185028   185127: gap of 100 bp
            *  185128   193579: contig of 8452 bp in length
            *  193580   193679: gap of 100 bp
            *  193680   207482: contig of 13803 bp in length
            *  207483   207582: gap of 100 bp
            *  207583   211897: contig of 4315 bp in length
            *  211898   211997: gap of 100 bp
            *  211998   216259: contig of 4262 bp in length.
FEATURES             Location/Qualifiers
     source          1. .216259
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-215F19"
                     /clone_lib="CHORI-211"
     misc_feature    1. .48219
                     /note="assembly_fragment:01191
                     fragment_chain:1
                     clone_end:T7
                     vector_side:left"
     misc_feature    48320. .65856
                     /note="assembly_fragment:00391
                     fragment_chain:1
                     clone_end:SP6
                     vector_side:right"
     misc_feature    65957. .86815
                     /note="assembly_fragment:00649
                     fragment_chain:2"
     misc_feature    86916. .108880
                     /note="assembly_fragment:00908
                     fragment_chain:2"
     misc_feature    108981. .118104
```

```
                     /note="assembly_fragment:00069
                     fragment_chain:2"
     misc_feature    118205. .185027
                     /note="assembly_fragment:01823
                     fragment_chain:2"
     misc_feature    185128. .193579
                     /note="assembly_fragment:00271
                     fragment_chain:2"
     misc_feature    193680. .207482
                     /note="assembly_fragment:00167
                     fragment_chain:3"
     misc_feature    207583. .211897
                     /note="assembly_fragment:00018
                     fragment_chain:3"
     misc_feature    211998. .216259
                     /note="assembly_fragment:00040
                     fragment_chain:3"
ORIGIN

  Query Match             8.5%;  Score 351.4;  DB 2;  Length 216259;
  Best Local Similarity   69.0%;  Pred. No. 2.5e-59;
  Matches  492;  Conservative    0;  Mismatches  220;  Indels    1;  Gaps    1;

Qy        191 AGAGCGAAATGTCATCTCAGAAAAAAAAAAAGGCATTTTATATATATATATATATATATA 250
              ||||   |||         | ||||||||||||     | ||||||||||||||||||||
Db     215544 AGAGATTAATAAGTCTGTAAAAAAAAAAAAAAATATATATATATATATATATATATATATA 215603

Qy        251 TACACACACACACACATATATATATACACATATATATACACATATATACATATATACACA 310
              || | | | | | | |||||||||||| | |||||||| | ||||||| ||||||| | |
Db     215604 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 215663

Qy        311 TATATACACATATATATACACATACATATGTACACATATATATACACATATGTATACACA 370
              |||||| | ||||||||| | ||| |||| || | ||||||||| | |||| |||| | |
Db     215664 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 215723

Qy        371 TATATACACATATATACACACATATATACACATATATACACACATATATACACATATATA 430
              |||||| | ||||||| | | ||||||| | ||||||| | | ||||||| | |||||||
Db     215724 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 215783

Qy        431 CACATATATACACATATACACATATATACACATATATACATATATACACATATATATA-A 489
              | |||||||| | ||||| | ||||||| | ||||||| ||||||| | ||||||||| |
Db     215784 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 215843

Qy        490 TATACACACATATATATACACATATATACACACATATATACACATATATACACATATATA 549
              |||| | | ||||||||| | ||||||| | | ||||||| | ||||||| | |||||||
Db     215844 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 215903

Qy        550 TACACATATATACACATATATACATATATACACATATATATACATATATACACATATATA 609
              ||.| ||||||| | ||||||| ||||||| | ||||||||| ||||||| | |||||||
Db     215904 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 215963

Qy        610 CATATATACACATATATACATATATACACACATATATACACATACATATACACACACATA 669
              ||||||| | ||||||| ||||||| | | ||||||| | ||| ||||| | | | |||
Db     215964 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 216023

Qy        670 GATATACATATATATACACATATATATACGTATATATATGTATATATATATGCTCCAGAG 729
              ||||| |||||||||| | |||||||||  |||||||||| ||||||||||
Db     216024 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 216083

Qy        730 TTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAG 789
              | |||          | |   ||     |     ||| |   | | | |   | | |
Db     216084 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 216143

Qy        790 TGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTG 849
              | |||    | |   | || |      |      | |||       | ||| ||| | |
Db     216144 TATATATATATATATATATATATATATATATATATATATATATATATATATATATAAATA 216203

Qy        850 TAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATATATATATA 902
              || | | |  ||| | ||    | |  |     ||  | ||||| || |||||
Db     216204 TATTTATATATATTTATATANAAATATAANNANTATAAATATATTTAAATATA 216256


RESULT 54
```

```
CR548622
LOCUS       CR548622              200752 bp    DNA     linear   HTG 06-JUL-2004
DEFINITION  Danio rerio clone DKEY-189A9, WORKING DRAFT SEQUENCE, 8 unordered
            pieces.
ACCESSION   CR548622
VERSION     CR548622.2  GI:50057705
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 200752)
  AUTHORS   Burton,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUL-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 8, 2004 this sequence version replaced gi:49658985.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK189A9
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 198295 bases at least Q40
            Consensus quality: 198730 bases at least Q30
            Consensus quality: 199141 bases at least Q20
            Insert size: 200052; sum-of-contigs
            Insert size: 193380; 3.4% error; agarose-fp
            Quality coverage: 9.52x in Q20 bases; sum-of-contigs Quality
            coverage: 9.92x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     5799: contig of 5799 bp in length
            *     5800     5899: gap of 100 bp
            *     5900    61323: contig of 55424 bp in length
            *    61324    61423: gap of 100 bp
            *    61424    94914: contig of 33491 bp in length
            *    94915    95014: gap of 100 bp
            *    95015   169245: contig of 74231 bp in length
            *   169246   169345: gap of 100 bp
            *   169346   171607: contig of 2262 bp in length
            *   171608   171707: gap of 100 bp
            *   171708   174537: contig of 2830 bp in length
            *   174538   174637: gap of 100 bp
            *   174638   183583: contig of 8946 bp in length
            *   183584   183683: gap of 100 bp
            *   183684   200752: contig of 17069 bp in length.
FEATURES             Location/Qualifiers
     source          1. .200752
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-189A9"
                     /clone_lib="DanioKey"
     misc_feature    1. .5799
                     /note="assembly_fragment:00060
                     fragment_chain:1"
     misc_feature    5900. .61323
                     /note="assembly_fragment:00964
                     fragment_chain:1"
```

```
     misc_feature    61424. .94914
                     /note="assembly_fragment:00439
                     fragment_chain:1"
     misc_feature    95015. .169245
                     /note="assembly_fragment:01748
                     fragment_chain:2"
     misc_feature    169346. .171607
                     /note="assembly_fragment:00009
                     fragment_chain:2"
     misc_feature    171708. .174537
                     /note="assembly_fragment:00025"
     misc_feature    174638. .183583
                     /note="assembly_fragment:00114"
     misc_feature    183684. .200752
                     /note="assembly_fragment:00228.0"
ORIGIN

  Query Match             8.5%;  Score 351.2;  DB 2;  Length 200752;
  Best Local Similarity   63.0%;  Pred. No. 2.7e-59;
  Matches  466;  Conservative    0;  Mismatches  273;  Indels    1;  Gaps    1;

Qy        186 CAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATAT 245
              ||| || |   | |  | |    | ||||||||||||   | | ||||||||||||||||||
Db      60737 CAAAAATAATAATAAATAAATAAAAAAAAAAAAAAAAAAAAAAATATATATATATATATAT 60796

Qy        246 ATATATACACACACACACATATATATATACACATATATATACACATATATACATATAT 305
              ||||||| | | | | | | |||||||||||| | ||||||||| | ||||||| ||||||
Db      60797 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 60856

Qy        306 ACACATATATACACATATATATACACATACATATGTACACATATATATACACATATGTAT 365
              | | ||||||| | ||||||||| | ||| |||| || | ||||||||| | |||| |||
Db      60857 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 60916

Qy        366 ACACATATATACACATATATACACACATATATACACATATATACACACATATATACACAT 425
              | | ||||||| | ||||||| | | ||||||| | ||||||| | | ||||||| | ||
Db      60917 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 60976

Qy        426 ATATACACATATATACACATATACACATATATACACATATATACATATATACACATATAT 485
              ||||| | ||||||| | ||||| | ||||||| | ||||||| ||||||| | ||||||
Db      60977 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 61036

Qy        486 ATA-ATATACACACATATATATACACATATATACACACATATATACACATATATACACAT 544
              ||| ||||| | | ||||||||| | ||||||| | | ||||||| | ||||||| | ||
Db      61037 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 61096

Qy        545 ATATATACACATATATACACATATATACATATATACACATATATATACATATATACACAT 604
              ||||||| | |||||||| | ||||||| ||||||| | ||||||||| ||||||| | ||
Db      61097 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 61156

Qy        605 ATATACATATATACACATATATACATATATACACACATATATACACATACATATACACAC 664
              ||||| ||||||| | ||||||| ||||||| | | ||||||| | ||| ||||| | |
Db      61157 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 61216

Qy        665 ACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATATGCTC 724
              | ||| ||||| ||||||||| | |||||||||  ||||||||| ||||||||||
Db      61217 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 61276

Qy        725 CAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGAC 784
              |  |||          |  |    ||     |    ||| |   |
Db      61277 ATATATATATATATATATATATATATATATATATATATATATATATANNNNNNNNNNNNN 61336

Qy        785 AGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAA 844

Db      61337 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 61396

Qy        845 ATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATATATATATAGT 904
                                          |  ||   |  ||   ||||||||||||||
Db      61397 NNNNNNNNNNNNNNNNNNNNNNNNNNNATATATATATATATATATATATATATATATA 61456

Qy        905 GAAAGCAGTGTAAACCTAAA 924
              | |    | || |  || |
Db      61457 TATATATATATATATATATA 61476
```

```
RESULT 55
CR786577/c
LOCUS       CR786577              156578 bp    DNA     linear   HTG 17-SEP-2004
DEFINITION  Danio rerio clone CH211-285C10, *** SEQUENCING IN PROGRESS ***, 11
            unordered pieces.
ACCESSION   CR786577
VERSION     CR786577.1  GI:52313179
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 156578)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zC285C10
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 152392 bases at least Q40
            Consensus quality: 153620 bases at least Q30
            Consensus quality: 154402 bases at least Q20
            Insert size: 155578; sum-of-contigs
            Insert size: 156602; 1.5% error; agarose-fp
            Quality coverage: 6.30x in Q20 bases; sum-of-contigs Quality
            coverage: 6.26x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    28552: contig of 28552 bp in length
            *    28553    28652: gap of 100 bp
            *    28653    32448: contig of 3796 bp in length
            *    32449    32548: gap of 100 bp
            *    32549    46601: contig of 14053 bp in length
            *    46602    46701: gap of 100 bp
            *    46702    54418: contig of 7717 bp in length
            *    54419    54518: gap of 100 bp
            *    54519    79156: contig of 24638 bp in length
            *    79157    79256: gap of 100 bp
            *    79257    90677: contig of 11421 bp in length
            *    90678    90777: gap of 100 bp
            *    90778    95922: contig of 5145 bp in length
            *    95923    96022: gap of 100 bp
            *    96023   100602: contig of 4580 bp in length
            *   100603   100702: gap of 100 bp
            *   100703   125015: contig of 24313 bp in length
            *   125016   125115: gap of 100 bp
            *   125116   150124: contig of 25009 bp in length
            *   150125   150224: gap of 100 bp
            *   150225   156578: contig of 6354 bp in length.
FEATURES             Location/Qualifiers
     source          1. .156578
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-285C10"
```

```
                     /clone_lib="CHORI-211"
     misc_feature    1. .28552
                     /note="assembly_fragment:01088
                     fragment_chain:1
                     clone_end:SP6
                     vector_side:left"
     misc_feature    28653. .32448
                     /note="assembly_fragment:00002
                     fragment_chain:1"
     misc_feature    32549. .46601
                     /note="assembly_fragment:00430
                     fragment_chain:1"
     misc_feature    46702. .54418
                     /note="assembly_fragment:00210
                     fragment_chain:1"
     misc_feature    54519. .79156
                     /note="assembly_fragment:00576
                     fragment_chain:1"
     misc_feature    79257. .90677
                     /note="assembly_fragment:00303
                     fragment_chain:1"
     misc_feature    90778. .95922
                     /note="assembly_fragment:00082
                     fragment_chain:1"
     misc_feature    96023. .100602
                     /note="assembly_fragment:00039
                     fragment_chain:1"
     misc_feature    100703. .125015
                     /note="assembly_fragment:00813
                     fragment_chain:1"
     misc_feature    125116. .150124
                     /note="assembly_fragment:01376"
     misc_feature    150225. .156578
                     /note="assembly_fragment:00130
                     clone_end:T7
                     vector_side:right"
ORIGIN

  Query Match             8.5%;  Score 351;  DB 2;  Length 156578;
  Best Local Similarity   69.8%;  Pred. No. 3.1e-59;
  Matches  488;  Conservative    0;  Mismatches  210;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db     100582 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 100523

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db     100522 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 100463

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db     100462 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 100403

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db     100402 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 100343

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db     100342 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 100283

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db     100282 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 100223

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db     100222 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 100163

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db     100162 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 100103

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||||          | |||          |  |   ||    |   |||
Db     100102 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 100043

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
              |   | | | |  | | |    | |||    | |   | || |      |      | |
Db     100042 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 99983

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885
              ||       | ||| ||| || | ||   | |  ||| | ||      |||| ||||||
Db      99982 TATATATATATATATATATATATATATATATATATATATATATATATATTAATAAGGTAT 99923

Qy        886 CACTATATATATATATAGTGAAAGCAGTGTAAACCTAAA 924
                |||||||||||||   | |    | || |  || |
Db      99922 ATATATATATATATATATATATATATATATATATATA 99884


RESULT 56
CR762422/c
LOCUS       CR762422               159603 bp    DNA     linear   HTG 13-SEP-2004
DEFINITION  Danio rerio clone CH211-184I23, *** SEQUENCING IN PROGRESS ***, 13
            unordered pieces.
ACCESSION   CR762422
VERSION     CR762422.2  GI:52000347
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 159603)
  AUTHORS   McLay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 13, 2004 this sequence version replaced gi:51988198.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zC184I23
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 153793 bases at least Q40
            Consensus quality: 154824 bases at least Q30
            Consensus quality: 155824 bases at least Q20
            Insert size: 158403; sum-of-contigs
            Insert size: 158676; 6.9% error; agarose-fp
            Quality coverage: 8.08x in Q20 bases; sum-of-contigs Quality
            coverage: 8.13x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 13 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    33117: contig of 33117 bp in length
            *    33118    33217: gap of 100 bp
            *    33218    38465: contig of 5248 bp in length
            *    38466    38565: gap of 100 bp
            *    38566    47730: contig of 9165 bp in length
            *    47731    47830: gap of 100 bp
            *    47831    83298: contig of 35468 bp in length
            *    83299    83398: gap of 100 bp
            *    83399    86467: contig of 3069 bp in length
```

```
           *    86468    86567: gap of 100 bp
           *    86568    88634: contig of 2067 bp in length
           *    88635    88734: gap of 100 bp
           *    88735   103656: contig of 14922 bp in length
           *   103657   103756: gap of 100 bp
           *   103757   120853: contig of 17097 bp in length
           *   120854   120953: gap of 100 bp
           *   120954   132618: contig of 11665 bp in length
           *   132619   132718: gap of 100 bp
           *   132719   146433: contig of 13715 bp in length
           *   146434   146533: gap of 100 bp
           *   146534   152937: contig of 6404 bp in length
           *   152938   153037: gap of 100 bp
           *   153038   156530: contig of 3493 bp in length
           *   156531   156630: gap of 100 bp
           *   156631   159603: contig of 2973 bp in length.
FEATURES             Location/Qualifiers
     source          1. .159603
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-184I23"
                     /clone_lib="CHORI-211"
     misc_feature    1. .33117
                     /note="assembly_fragment:01129
                     fragment_chain:1
                     clone_end:SP6
                     vector_side:left"
     misc_feature    33218. .38465
                     /note="assembly_fragment:00128
                     fragment_chain:1"
     misc_feature    38566. .47730
                     /note="assembly_fragment:00252
                     fragment_chain:1"
     misc_feature    47831. .83298
                     /note="assembly_fragment:01535
                     fragment_chain:1"
     misc_feature    83399. .86467
                     /note="assembly_fragment:00044
                     fragment_chain:1"
     misc_feature    86568. .88634
                     /note="assembly_fragment:00029
                     fragment_chain:2"
     misc_feature    88735. .103656
                     /note="assembly_fragment:00682
                     fragment_chain:2"
     misc_feature    103757. .120853
                     /note="assembly_fragment:00894"
     misc_feature    120954. .132618
                     /note="assembly_fragment:00349
                     fragment_chain:3"
     misc_feature    132719. .146433
                     /note="assembly_fragment:00481
                     fragment_chain:3"
     misc_feature    146534. .152937
                     /note="assembly_fragment:00172
                     fragment_chain:3"
     misc_feature    153038. .156530
                     /note="assembly_fragment:00068
                     fragment_chain:3"
     misc_feature    156631. .159603
                     /note="assembly_fragment:00094
                     fragment_chain:3
                     clone_end:T7
                     vector_side:right"
ORIGIN

  Query Match             8.5%;  Score 350.8;  DB 2;  Length 159603;
  Best Local Similarity   70.0%;  Pred. No. 3.4e-59;
  Matches  486;  Conservative    0;  Mismatches  207;  Indels    1;  Gaps    1;

Qy        209 AGAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACATA 268
              .| ||||||||||||   | | |||||||||||||||||||||||||| | | | | | |||

Db     121963 AAAAAAAAAAAAAAAAAATATATATATATATATATATATATATATATATATATATA 121904

Qy        269 TATATATACACATATATATACACATATATACATATATACACATATATACACATATATATA 328
              |||||||| | ||||||||| | ||||||| ||||||| | ||||||| | |||||||||
Db     121903 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 121844

Qy        329 CACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACA 388
              | ||| |||| || | ||||||||| | |||| |||| | ||||||| | ||||||| |
Db     121843 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 121784

Qy        389 CACATATATACACATATATACACACATATATACACATATATACACATATATACACATATA 448
              | ||||||| | ||||||| | | ||||||| | ||||||| | ||||||| | |||||
Db     121783 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 121724

Qy        449 CACATATATACACATATATACATATATACACATATATATA-ATATACACACATATATATA 507
              | ||||||| | ||||||| ||||||| | ||||||||| ||||| | | |||||||||
Db     121723 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 121664

Qy        508 CACATATATACACACATATATACACATATATACACATATATATACACATATATACACATA 567
              | ||||||| | | ||||||| | ||||||| | ||||||||| | ||||||| | |||
Db     121663 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 121604

Qy        568 TATACATATATACACATATATATACATATATACACATATATACATATATACACATATATA 627
              |||| ||||||| | |||||||| ||||||| | ||||||| ||||||| | |||||||
Db     121603 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 121544

Qy        628 CATATATACACACATATATACACATACATATACACACACATAGATATACATATATATACA 687
              ||||||| | | ||||||| | ||| ||||| | | | ||| ||||| |||||||| |
Db     121543 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 121484

Qy        688 CATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGT 747
              ||||||||  |||||||| ||||||||||           |||      |
Db     121483 TATATATATATATATATATATATATATATATATATATATATAAATATATATATATATATA 121424

Qy        748 TGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACA 807
              |   ||    |   ||| |  | | | |  | | |    | |||   | |  | |
Db     121423 TATATATATATATATATATATATACATATATATATATAAATATATATATATATATATAAA 121364

Qy        808 TTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA 867
              | |   |  | |||      | ||| |||||| | |||||| |  ||| | ||
Db     121363 TATATATATATATATATATATATATATATATATAAATATATAATAATAAATATTTATA 121304

Qy        868 AGGTAGTTAAGAAGGTAACACTATATATATATAT 901
              |  |  | | ||  ||  |   |||   | ||
Db     121303 ATATTATATATAATTTATAATATAATAATAAAAT 121270


RESULT 57
CR762498/c
LOCUS       CR762498              200607 bp    DNA     linear   HTG 10-OCT-2004
DEFINITION  Danio rerio clone DKEY-4J1, WORKING DRAFT SEQUENCE, 9 unordered
            pieces.
ACCESSION   CR762498
VERSION     CR762498.4  GI:52346306
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 200607)
  AUTHORS   McLaren,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 19, 2004 this sequence version replaced gi:52345308.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
```

```
          Center project name: zK4J1
          ---------- Summary Statistics
          Assembly program: XGAP4; version 4.5
          Chemistry: Dye-terminator; 100% of reads
          Consensus quality: 196926 bases at least Q40
          Consensus quality: 197654 bases at least Q30
          Consensus quality: 198332 bases at least Q20
          Insert size: 199807; sum-of-contigs
          Insert size: 210583; 10.0% error; agarose-fp
          Quality coverage: 10.81x in Q20 bases; sum-of-contigs Quality
          coverage: 10.33x in Q20 bases; agarose-fp
          ----------.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 9 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1    18795: contig of 18795 bp in length
          *    18796    18895: gap of 100 bp
          *    18896    73470: contig of 54575 bp in length
          *    73471    73570: gap of 100 bp
          *    73571    76405: contig of 2835 bp in length
          *    76406    76505: gap of 100 bp
          *    76506    89061: contig of 12556 bp in length
          *    89062    89161: gap of 100 bp
          *    89162   100153: contig of 10992 bp in length
          *   100154   100253: gap of 100 bp
          *   100254   157131: contig of 56878 bp in length
          *   157132   157231: gap of 100 bp
          *   157232   161046: contig of 3815 bp in length
          *   161047   161146: gap of 100 bp
          *   161147   167068: contig of 5922 bp in length
          *   167069   167168: gap of 100 bp
          *   167169   200607: contig of 33439 bp in length.
FEATURES             Location/Qualifiers
     source          1. .200607
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-4J1"
                     /clone_lib="DanioKey"
     misc_feature    1. .18795
                     /note="assembly_fragment:00545
                     fragment_chain:1"
     misc_feature    18896. .73470
                     /note="assembly_fragment:02336
                     fragment_chain:1"
     misc_feature    73571. .76405
                     /note="assembly_fragment:00024
                     fragment_chain:1"
     misc_feature    76506. .89061
                     /note="assembly_fragment:00349
                     fragment_chain:1"
     misc_feature    89162. .100153
                     /note="assembly_fragment:00174
                     fragment_chain:2"
     misc_feature    100254. .157131
                     /note="assembly_fragment:01357
                     fragment_chain:2"
     misc_feature    157232. .161046
                     /note="assembly_fragment:00051.0"
     misc_feature    161147. .167068
                     /note="assembly_fragment:00099"
     misc_feature    167169. .200607
                     /note="assembly_fragment:00747"
ORIGIN

  Query Match             8.5%;  Score 350.6;  DB 2;  Length 200607;
  Best Local Similarity   62.3%;  Pred. No. 3.6e-59;
  Matches  472;  Conservative    0;  Mismatches  285;  Indels    1;  Gaps    1;

Qy        197 AAATGTCATCTCAGAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACA 256
              ||||||  |    | ||||||||  |      | ||||||||||||||||||||||||  | |
Db      77043 AAATGTAAAAAAAAAAAAAAAAATATATATATATATATATATATATATATATATATA 76984

Qy        257 CACACACACATATATATATACACATATATATACACATATATACATATATACACATATATA 316
              | | | | |||||||||||| | ||||||||| | ||||||| ||||||| | |||||||
Db      76983 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 76924

Qy        317 CACATATATATACACATACATATGTACACATATATATACACATATGTATACACATATATA 376
              | ||||||||| | ||| |||| || | ||||||||| | |||| |||| | |||||||
Db      76923 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 76864

Qy        377 CACATATATACACACATATATACACATATATACACACATATATACACATATATACACATA 436
              | ||||||| | | ||||||| | ||||||| | | ||||||| | ||||||| | |||
Db      76863 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 76804

Qy        437 TATACACATATACACATATATACACATATATACATATATACACATATATATAATATACAC 496
              |||| | ||||| | ||||||| | ||||||| ||||||| | |||||||| ||||| |
Db      76803 TATATATATATATATATATATATATATATATATATATATATATATATATATAT-ATATATAT 76745

Qy        497 ACATATATATACACATATATACACACATATATACACATATATACACATATATATACACAT 556
              | ||||||||| | ||||||| | | ||||||| | ||||||| | ||||||||| | ||
Db      76744 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 76685

Qy        557 ATATACACATATATACATATATACACATATATATACATATATACACATATATACATATAT 616
              ||||| | ||||||| ||||||| | ||||||||| ||||||| | ||||||| ||||||
Db      76684 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 76625

Qy        617 ACACATATATACATATATACACACATATATACACATACATATACACACACATAGATATAC 676
              | | ||||||| ||||||| | | ||||||| | ||| ||||| | | | ||| |||||
Db      76624 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 76565

Qy        677 ATATATATACACATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAA 736
              ||||||||| | |||||||||  ||||||||| |||||||||||         |  |||
Db      76564 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAN 76505

Qy        737 GAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTG 796

Db      76504 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 76445

Qy        797 TGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAA 856
                                                     ||| ||| || | ||   |
Db      76444 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTATATATATATATATATATAT 76385

Qy        857 ACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATATATATATAGTGAAAGCAGTGTA 916
              |  ||| | ||    |  | |   |    | | | | | || | |  | ||   |
Db      76384 ATATATATATATATATATATATATTTTTTTTTTTTTTTNTCTTTACAATATATACAAATTT 76325

Qy        917 AACCTAAAGGATGGGCCAAGGATTTAAATGTTATAGAA 954
              |    | |          |   || |||  | |||| ||
Db      76324 ACAAAATATATATATATATATATATAATGGATATATAA 76287


RESULT 58
CR589875/c
LOCUS       CR589875              227441 bp    DNA     linear   HTG 19-JUL-2004
DEFINITION  Danio rerio clone DKEY-15K16, WORKING DRAFT SEQUENCE, 17 unordered
            pieces.
ACCESSION   CR589875
VERSION     CR589875.2  GI:50399773
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 227441)
  AUTHORS   Burton,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JUL-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
```

```
COMMENT     ------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            ------------- Project Information
            Center project name: zK15K16
            ------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 221759 bases at least Q40
            Consensus quality: 222952 bases at least Q30
            Consensus quality: 223650 bases at least Q20
            Insert size: 225841; sum-of-contigs
            Insert size: 225442; 2.6% error; agarose-fp
            Quality coverage: 7.98x in Q20 bases; sum-of-contigs Quality
            coverage: 8.44x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 17 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     7750: contig of 7750 bp in length
            *     7751     7850: gap of 100 bp
            *     7851    17808: contig of 9958 bp in length
            *    17809    17908: gap of 100 bp
            *    17909    26300: contig of 8392 bp in length
            *    26301    26400: gap of 100 bp
            *    26401    37193: contig of 10793 bp in length
            *    37194    37293: gap of 100 bp
            *    37294    42790: contig of 5497 bp in length
            *    42791    42890: gap of 100 bp
            *    42891    75037: contig of 32147 bp in length
            *    75038    75137: gap of 100 bp
            *    75138    77164: contig of 2027 bp in length
            *    77165    77264: gap of 100 bp
            *    77265    93266: contig of 16002 bp in length
            *    93267    93366: gap of 100 bp
            *    93367   109748: contig of 16382 bp in length
            *   109749   109848: gap of 100 bp
            *   109849   156408: contig of 46560 bp in length
            *   156409   156508: gap of 100 bp
            *   156509   168441: contig of 11933 bp in length
            *   168442   168541: gap of 100 bp
            *   168542   193830: contig of 25289 bp in length
            *   193831   193930: gap of 100 bp
            *   193931   212755: contig of 18825 bp in length
            *   212756   212855: gap of 100 bp
            *   212856   215964: contig of 3109 bp in length
            *   215965   216064: gap of 100 bp
            *   216065   218399: contig of 2335 bp in length
            *   218400   218499: gap of 100 bp
            *   218500   221549: contig of 3050 bp in length
            *   221550   221649: gap of 100 bp
            *   221650   227441: contig of 5792 bp in length.
FEATURES             Location/Qualifiers
     source          1. .227441
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-15K16"
                     /clone_lib="DanioKey"
     misc_feature    1. .7750
                     /note="assembly_fragment:00294
                     fragment_chain:1"
     misc_feature    7851. .17808
                     /note="assembly_fragment:00581
                     fragment_chain:1"
     misc_feature    17909. .26300
                     /note="assembly_fragment:00382
                     fragment_chain:1"
     misc_feature    26401. .37193
                     /note="assembly_fragment:00471
                     fragment_chain:1"
     misc_feature    37294. .42790
                     /note="assembly_fragment:00222
                     fragment_chain:1"
     misc_feature    42891. .75037
                     /note="assembly_fragment:02032
                     fragment_chain:1"
     misc_feature    75138. .77164
                     /note="assembly_fragment:00092
                     fragment_chain:1"
     misc_feature    77265. .93266
                     /note="assembly_fragment:01346
                     fragment_chain:1"
     misc_feature    93367. .109748
                     /note="assembly_fragment:00863
                     fragment_chain:1"
     misc_feature    109849. .156408
                     /note="assembly_fragment:02587
                     fragment_chain:1"
     misc_feature    156509. .168441
                     /note="assembly_fragment:00719
                     fragment_chain:1"
     misc_feature    168542. .193830
                     /note="assembly_fragment:01629
                     fragment_chain:1"
     misc_feature    193931. .212755
                     /note="assembly_fragment:01104
                     fragment_chain:2"
     misc_feature    212856. .215964
                     /note="assembly_fragment:00130
                     fragment_chain:2"
     misc_feature    216065. .218399
                     /note="assembly_fragment:00077"
     misc_feature    218500. .221549
                     /note="assembly_fragment:00107.0"
     misc_feature    221650. .227441
                     /note="assembly_fragment:00155.0"
ORIGIN

  Query Match             8.5%;  Score 350.4;  DB 2;  Length 227441;
  Best Local Similarity   64.4%;  Pred. No. 3.8e-59;
  Matches  466;  Conservative    0;  Mismatches  256;  Indels    2;  Gaps    2;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db      18405 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 18346

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db      18345 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 18286

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db      18285 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 18226

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db      18225 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 18166

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db      18165 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 18106

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db      18105 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 18046

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
```

```
Db      18045 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 17986

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| |||||||||| | |||||||||  ||||||
Db      17985 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 17926

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||||
Db      17925 TATATATATATATATAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 17866

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
                                                                      |||
Db      17865 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNAAA 17806

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTA 884
              ||        | ||| ||| || | ||    | |  ||| | || |  ||  ||    |  ||
Db      17805 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 17746

Qy        885 ACACTATATATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAA 944
                  ||||| |||||||| |  | |  || | |  |        || |||   | ||   |
Db      17745 TATATATATGTATATATATTCTATGATGTTTTAGTCATTTTTGTGTGCCTCTGTTTCTCA 17686

Qy        945 TGTT 948
              | ||
Db      17685 TATT 17682
```

```
RESULT 59
CR788315
LOCUS       CR788315             153947 bp    DNA     linear   HTG 24-SEP-2004
DEFINITION  Danio rerio clone DKEY-64E1, *** SEQUENCING IN PROGRESS ***, 14
            unordered pieces.
ACCESSION   CR788315
VERSION     CR788315.3  GI:52673623
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 153947)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 24, 2004 this sequence version replaced gi:52626672.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK64E1
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 150089 bases at least Q40
            Consensus quality: 150989 bases at least Q30
            Consensus quality: 151632 bases at least Q20
            Insert size: 152647; sum-of-contigs
            Insert size: 169881; 7.1% error; agarose-fp
            Quality coverage: 10.12x in Q20 bases; sum-of-contigs Quality
            coverage: 9.68x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 14 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     9129: contig of 9129 bp in length
            *    9130     9229: gap of 100 bp
            *    9230    14190: contig of 4961 bp in length
            *   14191    14290: gap of 100 bp
            *   14291    46311: contig of 32021 bp in length
            *   46312    46411: gap of 100 bp
            *   46412    54860: contig of 8449 bp in length
            *   54861    54960: gap of 100 bp
            *   54961    62361: contig of 7401 bp in length
            *   62362    62461: gap of 100 bp
            *   62462    70465: contig of 8004 bp in length
            *   70466    70565: gap of 100 bp
            *   70566    85672: contig of 15107 bp in length
            *   85673    85772: gap of 100 bp
            *   85773    90034: contig of 4262 bp in length
            *   90035    90134: gap of 100 bp
            *   90135   104444: contig of 14310 bp in length
            *  104445   104544: gap of 100 bp
            *  104545   111290: contig of 6746 bp in length
            *  111291   111390: gap of 100 bp
            *  111391   123791: contig of 12401 bp in length
            *  123792   123891: gap of 100 bp
            *  123892   125954: contig of 2063 bp in length
            *  125955   126054: gap of 100 bp
            *  126055   129586: contig of 3532 bp in length
            *  129587   129686: gap of 100 bp
            *  129687   153947: contig of 24261 bp in length.
FEATURES             Location/Qualifiers
     source          1. .153947
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-64E1"
                     /clone_lib="DanioKey"
     misc_feature    1. .9129
                     /note="assembly_fragment:00638
                     fragment_chain:1"
     misc_feature    9230. .14190
                     /note="assembly_fragment:00181
                     fragment_chain:1"
     misc_feature    14291. .46311
                     /note="assembly_fragment:01886
                     fragment_chain:1"
     misc_feature    46412. .54860
                     /note="assembly_fragment:00323
                     fragment_chain:1"
     misc_feature    54961. .62361
                     /note="assembly_fragment:00252
                     fragment_chain:1"
     misc_feature    62462. .70465
                     /note="assembly_fragment:00518
                     fragment_chain:1"
     misc_feature    70566. .85672
                     /note="assembly_fragment:00985
                     fragment_chain:1"
     misc_feature    85773. .90034
                     /note="assembly_fragment:00086
                     fragment_chain:1"
     misc_feature    90135. .104444
                     /note="assembly_fragment:00797
                     fragment_chain:1"
     misc_feature    104545. .111290
                     /note="assembly_fragment:00417
                     fragment_chain:2"
     misc_feature    111391. .123791
                     /note="assembly_fragment:01185
                     fragment_chain:2"
     misc_feature    123892. .125954
                     /note="assembly_fragment:00065"
     misc_feature    126055. .129586
                     /note="assembly_fragment:00133.0"
     misc_feature    129687. .153947
                     /note="assembly_fragment:01427"
```

```
ORIGIN

  Query Match             8.5%;  Score 350.2;  DB 2;  Length 153947;
  Best Local Similarity   64.4%;  Pred. No. 4.4e-59;
  Matches  450;  Conservative    0;  Mismatches  248;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db     123283 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 123342

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db     123343 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 123402

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db     123403 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 123462

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db     123463 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 123522

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db     123523 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 123582

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db     123583 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 123642

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | |||||||| ||||||| | ||||||| ||||||| | | |||||
Db     123643 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 123702

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db     123703 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 123762

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              |||  ||||||||||          |  ||
Db     123763 TATAAATATATATATATAAATATATATATNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 123822

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825

Db     123823 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 123882

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885
                      |   || || | || |||||  ||| | ||   || ||    | |||
Db     123883 NNNNNNNNNTTTTTTTTATATATATATTAAAAAATATATATATATTATTTTTATATATAA 123942

Qy        886 CACTATATATATATATAGTGAAAGCAGTGTAAACCTAAA 924
                  | ||||||||||||   | |    | || |  || |
Db     123943 ATAAAAATATATATATATATATATATATATATATATA 123981


RESULT 60
CR788302/c
LOCUS       CR788302              175303 bp    DNA     linear   HTG 24-SEP-2004
DEFINITION  Danio rerio clone DKEYP-38A10, *** SEQUENCING IN PROGRESS ***, 9
            unordered pieces.
ACCESSION   CR788302
VERSION     CR788302.3  GI:52673620
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 175303)
  AUTHORS   McLay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 24, 2004 this sequence version replaced gi:52420996.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zKp38A10
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 173044 bases at least Q40
            Consensus quality: 173386 bases at least Q30
            Consensus quality: 173663 bases at least Q20
            Insert size: 174503; sum-of-contigs
            Insert size: 174235; 1.3% error; agarose-fp
            Quality coverage: 8.55x in Q20 bases; sum-of-contigs Quality
            coverage: 8.62x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     6846: contig of 6846 bp in length
            *     6847     6946: gap of 100 bp
            *     6947    39633: contig of 32687 bp in length
            *    39634    39733: gap of 100 bp
            *    39734    44270: contig of 4537 bp in length
            *    44271    44370: gap of 100 bp
            *    44371    81545: contig of 37175 bp in length
            *    81546    81645: gap of 100 bp
            *    81646    84574: contig of 2929 bp in length
            *    84575    84674: gap of 100 bp
            *    84675   103084: contig of 18410 bp in length
            *   103085   103184: gap of 100 bp
            *   103185   122912: contig of 19728 bp in length
            *   122913   123012: gap of 100 bp
            *   123013   130331: contig of 7319 bp in length
            *   130332   130431: gap of 100 bp
            *   130432   175303: contig of 44872 bp in length.
FEATURES             Location/Qualifiers
     source          1. .175303
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEYP-38A10"
                     /clone_lib="DanioKeypilot"
     misc_feature    1. .6846
                     /note="assembly_fragment:00101
                     fragment_chain:1"
     misc_feature    6947. .39633
                     /note="assembly_fragment:01279
                     fragment_chain:1"
     misc_feature    39734. .44270
                     /note="assembly_fragment:00052
                     fragment_chain:1"
     misc_feature    44371. .81545
                     /note="assembly_fragment:00801
                     fragment_chain:1"
     misc_feature    81646. .84574
                     /note="assembly_fragment:00024
                     fragment_chain:1"
     misc_feature    84675. .103084
                     /note="assembly_fragment:00248
                     fragment_chain:2"
     misc_feature    103185. .122912
                     /note="assembly_fragment:00522
                     fragment_chain:2"
```

```
     misc_feature    123013. .130331
                     /note="assembly_fragment:00171.0"
     misc_feature    130432. .175303
                     /note="assembly_fragment:01761"
ORIGIN

  Query Match             8.5%;  Score 350.2;  DB 2;  Length 175303;
  Best Local Similarity   63.5%;  Pred. No. 4.3e-59;
  Matches  459;  Conservative    0;  Mismatches  263;  Indels    1;  Gaps    1;

Qy        202 TCATCTCAGAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACAC 261
              | | ||   ||||||||||      | ||||||||||||||||||||||||| | | | ||
Db     123573 TTACCTTTAAAAAAAAAAAAATATATATATATATATATATATATATATATATATATAT 123514

Qy        262 ACACATATATATATACACATATATATACACATATATACATATATACACATATATACACAT 321
              | | |||||||||| | |||||||| | ||||||| ||||||| | |||||||| | ||
Db     123513 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 123454

Qy        322 ATATATACACATACATATGTACACATATATATACACATATGTATACACATATATACACAT 381
              |||||| | ||| |||| || | |||||||| | |||| |||| | ||||||| | ||
Db     123453 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 123394

Qy        382 ATATACACACATATATACACATATATACACACATATATACACATATATACACATATATAC 441
              ||||| | | ||||||| | ||||||| | | |||||||| | ||||||| | |||||||
Db     123393 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 123334

Qy        442 ACATATACACATATATACACATATATACATATATACACATATATATAATATACACACATA 501
              | ||||| | ||||||| | ||||||| ||||||| | |||||||| ||||| | | |||
Db     123333 ATATATATATATATATATATATATATATATATATATATATATATATAT-ATATATATATATA 123275

Qy        502 TATATACACATATATACACACATATATACACATATATACACATATATATACACATATATA 561
              |||||| | ||||||| | | ||||||| | ||||||| | ||||||||| | |||||||
Db     123274 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 123215

Qy        562 CACATATATACATATATACACATATATATACATATATACACATATATACATATATACACA 621
              | ||||||| ||||||| | ||||||||| ||||||| | ||||||| ||||||| | |
Db     123214 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 123155

Qy        622 TATATACATATATACACACATATATACACATACATATACACACACATAGATATACATATA 681
              |||||| ||||||| | | ||||||| | ||| ||||| | | | ||| ||||| |||||
Db     123154 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 123095

Qy        682 TATACACATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGT 741
              |||| | |||||||||  |||||||| ||||||||||          |  |||  |
Db     123094 TATATATATATATATATATATATATATATATATATATATATATATATATATATATAAATA 123035

Qy        742 AGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGA 801
                |  |  ||    |   |
Db     123034 TATATATATATATATATAAAAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 122975

Qy        802 AGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTA 861

Db     122974 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 122915

Qy        862 TCTGTAAGGTAGTTAAGAAGGTAACACTATATATATATATAGTGAAAGCAGTGTAAACCT 921
              | |||  |  ||   |  ||   |||||||||||||   | |   | || |  |
Db     122914 NNTTTAATATTATTTTTTATATATATATATATATATATATATATATATATATATATAT 122855

Qy        922 AAA 924
              |||
Db     122854 AAA 122852


RESULT 61
CR753838
LOCUS       CR753838              204662 bp    DNA     linear   HTG 10-OCT-2004
DEFINITION  Danio rerio clone DKEY-189H5, WORKING DRAFT SEQUENCE, 10 unordered
            pieces.
ACCESSION   CR753838
VERSION     CR753838.3  GI:54021757
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 204662)
  AUTHORS   McLaren,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 10, 2004 this sequence version replaced gi:51592199.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK189H5
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 200213 bases at least Q40
            Consensus quality: 201277 bases at least Q30
            Consensus quality: 202286 bases at least Q20
            Insert size: 203762; sum-of-contigs
            Insert size: 192359; 2.6% error; agarose-fp
            Quality coverage: 8.93x in Q20 bases; sum-of-contigs Quality
            coverage: 10.24x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    54213: contig of 54213 bp in length
            *   54214    54313: gap of 100 bp
            *   54314    67044: contig of 12731 bp in length
            *   67045    67144: gap of 100 bp
            *   67145   118416: contig of 51272 bp in length
            *  118417   118516: gap of 100 bp
            *  118517   132097: contig of 13581 bp in length
            *  132098   132197: gap of 100 bp
            *  132198   135006: contig of 2809 bp in length
            *  135007   135106: gap of 100 bp
            *  135107   137155: contig of 2049 bp in length
            *  137156   137255: gap of 100 bp
            *  137256   139284: contig of 2029 bp in length
            *  139285   139384: gap of 100 bp
            *  139385   141755: contig of 2371 bp in length
            *  141756   141855: gap of 100 bp
            *  141856   146655: contig of 4800 bp in length
            *  146656   146755: gap of 100 bp
            *  146756   204662: contig of 57907 bp in length.
FEATURES             Location/Qualifiers
     source          1. .204662
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-189H5"
                     /clone_lib="DanioKey"
     misc_feature    1. .54213
                     /note="assembly_fragment:01901
                     fragment_chain:1"
     misc_feature    54314. .67044
                     /note="assembly_fragment:00280
                     fragment_chain:1"
     misc_feature    67145. .118416
                     /note="assembly_fragment:00441
                     fragment_chain:1"
     misc_feature    118517. .132097
                     /note="assembly_fragment:00148
```

```
                     fragment_chain:1"
     misc_feature    132198. .135006
                     /note="assembly_fragment:00046
                     fragment_chain:2"
     misc_feature    135107. .137155
                     /note="assembly_fragment:00042
                     fragment_chain:2"
     misc_feature    137256. .139284
                     /note="assembly_fragment:00024"
     misc_feature    139385. .141755
                     /note="assembly_fragment:00059"
     misc_feature    141856. .146655
                     /note="assembly_fragment:00081.0"
     misc_feature    146756. .204662
                     /note="assembly_fragment:01145"
ORIGIN

  Query Match             8.5%;  Score 350.2;  DB 2;  Length 204662;
  Best Local Similarity   68.7%;  Pred. No. 4.2e-59;
  Matches  497;  Conservative    0;  Mismatches  223;  Indels    3;  Gaps    1;

Qy        183 AGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATA 242
              | | ||||| |   |||    ||   | || ||||  |       | ||||||||||||||
Db     203771 ATGAAACAATAAAAAAAAACAATAAAAAAATAAAATTATATATATATATATATATATATATA 203830

Qy        243 TATATATATACACACACACACATATATATATACACATATATATACACATATATACATA 302
              |||||||||| | | | | | |||||||||||| | |||||||| | ||||||| |||
Db     203831 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 203890

Qy        303 TATACACATATATACACATATATATACACATACATATGTACACATATATATACACATATG 362
              |||| | ||||||| | ||||||||| | ||| |||| || | ||||||||| | ||||
Db     203891 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 203950

Qy        363 TATACACATATATACACATATATACACACATATATACACATATATACACACATATATACA 422
              |||| | ||||||| | ||||||| | | ||||||| | ||||||| | | |||||||  |
Db     203951 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 204010

Qy        423 CATATATACACATATATACACATATACACATATATACACATATATACATATATACACATA 482
              ||||||| | ||||||| | ||||| | ||||||| | ||||||| ||||||| | |||
Db     204011 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 204070

Qy        483 TATATA---ATATACACACATATATATACACATATATACACACATATATACACATATATA 539
              ||||||   ||||| | | ||||||||| | ||||||| | | ||||||| | |||||||
Db     204071 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 204130

Qy        540 CACATATATATACACATATATACACATATATACATATATACACATATATATACATATATA 599
              | ||||||||| | ||||||| | ||||||| ||||||| | ||||||||| |||||||
Db     204131 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 204190

Qy        600 CACATATATACATATATACACATATATACATATATACACACATATATACACATACATATA 659
              | |||||||| ||||||| | ||||||| ||||||| | | ||||||| | ||| |||||
Db     204191 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 204250

Qy        660 CACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATA 719
              | | | ||| ||||| ||||||||| | |||||||||  ||||||||| ||||||||||
Db     204251 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 204310

Qy        720 TGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGT 779
              |         | |||          | |   ||     |    ||| |   | | | |
Db     204311 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 204370

Qy        780 TTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATA 839
              |  | |    | |||    | |   | || |      |      | |||       | |||
Db     204371 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 204430

Qy        840 CATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATATATAT 899
              ||| || | ||  |||| ||  | |||  ||  ||      ||    |||| ||||||
Db     204431 TATATATATATATAAAAATATAAATATAATATATATATTTTAATATATATATAAATATAT 204490

Qy        900 ATA 902
              |||
Db     204491 ATA 204493
```

```
RESULT 62
CR626907
LOCUS       CR626907              218701 bp    DNA     linear   HTG 30-JUL-2004
DEFINITION  Danio rerio clone DKEY-52D15, WORKING DRAFT SEQUENCE, 7 unordered
            pieces.
ACCESSION   CR626907
VERSION     CR626907.2  GI:50871854
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 218701)
  AUTHORS   Burton,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUL-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 30, 2004 this sequence version replaced gi:50787983.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK52D15
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 216083 bases at least Q40
            Consensus quality: 216675 bases at least Q30
            Consensus quality: 216971 bases at least Q20
            Insert size: 218101; sum-of-contigs
            Insert size: 195570; 5.6% error; agarose-fp
            Quality coverage: 8.49x in Q20 bases; sum-of-contigs Quality
            coverage: 9.59x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     6174: contig of 6174 bp in length
            *     6175     6274: gap of 100 bp
            *     6275    10497: contig of 4223 bp in length
            *    10498    10597: gap of 100 bp
            *    10598    49931: contig of 39334 bp in length
            *    49932    50031: gap of 100 bp
            *    50032   118150: contig of 68119 bp in length
            *   118151   118250: gap of 100 bp
            *   118251   146821: contig of 28571 bp in length
            *   146822   146921: gap of 100 bp
            *   146922   186335: contig of 39414 bp in length
            *   186336   186435: gap of 100 bp
            *   186436   218701: contig of 32266 bp in length.
FEATURES             Location/Qualifiers
     source          1. .218701
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-52D15"
                     /clone_lib="DanioKey"
     misc_feature    1. .6174
                     /note="assembly_fragment:00016
                     fragment_chain:1"
     misc_feature    6275. .10497
                     /note="assembly_fragment:00055
                     fragment_chain:1"
```

```
     misc_feature    10598. .49931
                     /note="assembly_fragment:00497
                     fragment_chain:1"
     misc_feature    50032. .118150
                     /note="assembly_fragment:01902
                     fragment_chain:1"
     misc_feature    118251. .146821
                     /note="assembly_fragment:00934"
     misc_feature    146922. .186335
                     /note="assembly_fragment:01398"
     misc_feature    186436. .218701
                     /note="assembly_fragment:00096
                     clone_end:SP6
                     vector_side:right"
ORIGIN

  Query Match             8.5%;  Score 350.2;  DB 2;  Length 218701;
  Best Local Similarity   44.5%;  Pred. No. 4.2e-59;
  Matches  554;  Conservative    0;  Mismatches  689;  Indels    2;  Gaps    2;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db       5167 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 5226

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db       5227 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 5286

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db       5287 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 5346

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db       5347 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 5406

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || |||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db       5407 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 5466

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db       5467 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 5526

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db       5527 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 5586

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db       5587 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 5646

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||||         | |||          | |    ||    |    |||
Db       5647 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 5706

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
              |   | | | |   | | |     | |||     | |   | || |      |      | |
Db       5707 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 5766

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTA 884
              ||      | ||| ||| || | ||   | |  ||| |  | |  ||  |||  |    |
Db       5767 TATATATATATATATATATATATATATATATANATATATAAATATATATATAAANANANA 5826

Qy        885 ACACTATATATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAA 944
                |||||||| ||      ||  | |    |    |          |        ||
Db       5827 TATATATATATATNTAANNATNAAAAAATNNNTATNNNANAANNNNNNNANNNNNNNAA 5886

Qy        945 TGTTATAGAAGAATGGCTAAGATGCCAAAGCTCAGTGTATGTGGCAGAGGCATGGTGTAG 1004
              |  |             |      |  |                |      |    |
Db       5887 ANNNNTTTANNNNNNANNNNTNTNTTNNANNTTNNNNNNNNNNTAANNNNNTNNNTTTN 5946

Qy       1005 GGTGTGTCCAGGTTCATATATTGCATTAAGTGTGAGAACACCCTGGAGTATGAACCAAGA 1064
```

```
                              | ||     ||   |    |        | | |     |
Db       5947 TNNANNNAATAAAAANNNTTTTANNNNAATNNTNNTTATNNANNNNAATNTTNTNTANNN 6006

Qy       1065 AAATGCAAAAGCCAGAAGTGATGGAGGAAATGAGACACAATAATGAAGATATTGAGAGGA 1124
                |  |                    |   |   |                      | |
Db       6007 TNAAAAANNNNNNNNNNNNANNTNNNNNNTNNNANNNNANNNNNNNNNNNNNNNANAA 6066

Qy       1125 GGGTGTGGGCCTAGAGTGAAGCTTTTCGTGCCAGTACTTCTTTTGAAGGCCCAGTTCTCT 1184
                |                 ||||      | || || |          |
Db       6067 AATATTNNTTTNTTNNNNNNNTTTTTTNNNTAATTATTTTTNNNTTNNNNTAATNNNNNN 6126

Qy       1185 TCTCTCTCGGGGGCTCCTTCATCTCTCATAGAGTCCACAGCTTTTAAGGGCCAACACTTG 1244
                |                          |                 || ||
Db       6127 NATNANNNTNNTAANAAANAAANNNNNNNNNNNNNNNNNTNTTNAAANNNNNNNNNNNN 6186

Qy       1245 AGGTCAGCCTGGCTCTCTCATTTGAGCTGGATAGAACATTTTAGAGCACCATCTATTCTT 1304

Db       6187 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 6246

Qy       1305 CAAGAGGAAGTTTAAAAATAAAAGAACCTTGAAGAGGAAAAAATGTAGACATTCAATCTA 1364
                                          ||          |   |  | |      || |
Db       6247 NNNNNNNNNNNNNNNNNNNNNNNNNNNTTANNNNNNNTATTNTANATANNNNTNATTTT 6306

Qy       1365 ACCTTTTCATTTTACTAGCCAAAGCTAAATAGAATGCAGGTTACCTGTTTTTCAGCCAGG 1424
                |    ||          |
Db       6307 TNNTNNNNATNNNTTNNNNANANNNNNNNNNNNNNNNNNANNNNNNNNNNNNNNNNNN 6366

Qy       1425 CACCATCATTTCCTAATTGTTATAAAATTTATTATTATTGTTGTT 1469
                          |   | || |    ||| ||  |||      ||
Db       6367 NNNNNNNNNNNNNTTNNTTTTTTTTTTTTTTTTTNNTATNNNNNTT 6411
```

```
RESULT 63
CR769779
LOCUS       CR769779              219074 bp    DNA     linear   HTG 18-SEP-2004
DEFINITION  Danio rerio clone DKEY-18C12, *** SEQUENCING IN PROGRESS ***, 5
            unordered pieces.
ACCESSION   CR769779
VERSION     CR769779.3  GI:52345314
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 219074)
  AUTHORS   McLay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 18, 2004 this sequence version replaced gi:52214022.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK18C12
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 217273 bases at least Q40
            Consensus quality: 217586 bases at least Q30
            Consensus quality: 217870 bases at least Q20
            Insert size: 218674; sum-of-contigs
            Insert size: 217168; 4.7% error; agarose-fp
            Quality coverage: 8.00x in Q20 bases; sum-of-contigs Quality
            coverage: 8.05x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
```

```
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    31758: contig of 31758 bp in length
            *    31759    31858: gap of 100 bp
            *    31859    46986: contig of 15128 bp in length
            *    46987    47086: gap of 100 bp
            *    47087    62942: contig of 15856 bp in length
            *    62943    63042: gap of 100 bp
            *    63043   117588: contig of 54546 bp in length
            *   117589   117688: gap of 100 bp
            *   117689   219074: contig of 101386 bp in length.
FEATURES             Location/Qualifiers
     source          1. .219074
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-18C12"
                     /clone_lib="DanioKey"
     misc_feature    1. .31758
                     /note="assembly_fragment:00384
                     fragment_chain:1"
     misc_feature    31859. .46986
                     /note="assembly_fragment:00008
                     fragment_chain:1"
     misc_feature    47087. .62942
                     /note="assembly_fragment:00176.0"
     misc_feature    63043. .117588
                     /note="assembly_fragment:00812"
     misc_feature    117689. .219074
                     /note="assembly_fragment:01508"
ORIGIN

  Query Match             8.5%;  Score 350.2;  DB 2;  Length 219074;
  Best Local Similarity   67.3%;  Pred. No. 4.2e-59;
  Matches  524;  Conservative    0;  Mismatches  253;  Indels    2;  Gaps    2;

Qy        187 AACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATA 246
              | ||| |   |||     |     | ||||||||||    | | ||||||||||||||||
Db     218052 ACCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATATATATATATATATATA 218111

Qy        247 TATATACACACACACACATATATATATACACATATATATACACATATATACATATATA 306
              |||||| | | | | | | |||||||||||| | ||||||||| | ||||||| |||||||
Db     218112 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 218171

Qy        307 CACATATATACACATATATATACACATACATATGTACACATATATATACACATATGTATA 366
              | ||||||| | ||||||||| | ||| |||| || | ||||||||| | |||| ||||
Db     218172 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 218231

Qy        367 CACATATATACACATATATACACACATATATACACATATATACACACATATATACACATA 426
              | ||||||| | ||||||| | | ||||||| | ||||||| | | ||||||| | |||
Db     218232 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 218291

Qy        427 TATACACATATATACACATATACACATATATACACATATATACATATATACACATATATA 486
              |||| | |||||||| | ||||| | ||||||| | ||||||| ||||||| | |||||||
Db     218292 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 218351

Qy        487 TAATATACACACATATATATACACATATATACACACATATATACACATATATACACATAT 546
              | ||||| | | ||||||||| | ||||||| | | ||||||| | ||||||| | ||||
Db     218352 T-ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 218410

Qy        547 ATATACACATATATACACATATATACATATATACACATATATATACATATATACACATAT 606
              |||||| | ||||||| | ||||||| ||||||| | |||||||||| ||||||| | ||||
Db     218411 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 218470

Qy        607 ATACATATATACACATATATACATATATACACACATATATACACATACATATACACACAC 666
              ||| ||||||| | ||||||| ||||||| | | ||||||| | ||| ||||| | | |
Db     218471 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 218530

Qy        667 ATAGATATACATATATATACACATATATATACGTATATATATGTATATATATATGCTCCA 726
              .||| ||||| ||||||||| | |||||||||  ||||||||| ||||||||||
```

```
Db     218531 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 218590

Qy        727 GAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAG 786
              |  |||  |     |  |   ||    |   ||| |  | | | | |  | |
Db     218591 ATATATATATAAATATATATATATATATATATATATATATATATATATATATATATAT 218650

Qy        787 CAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAAT 846
              | |||    | |   | || |      |      | |||       | ||| ||| ||
Db     218651 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 218710

Qy        847 GTGTAATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTAACACTATATATATATATAGTG 905
              | |||  | |  ||| | || |  ||    | |  ||    || ||||| |||||
Db     218711 ATATAAATATATATATATATATATAAAATAATATATATATTTAAATATAAATATAAAT 218770

Qy        906 AAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTATAGAAGAATGGCTAA 964
              | |      |||   |   ||    ||  || ||||| ||  |  | |||   |||
Db     218771 ATAAATTATAAAAATATATATATATATTAATAATATAAATATTTAAATAAAATATATAA 218829


RESULT 64
CR749742
LOCUS       CR749742             194235 bp    DNA     linear   HTG 23-AUG-2004
DEFINITION  Danio rerio clone DKEYP-90H9, *** SEQUENCING IN PROGRESS ***, 9
            unordered pieces.
ACCESSION   CR749742
VERSION     CR749742.3  GI:51510860
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 194235)
  AUTHORS   McLay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-AUG-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 23, 2004 this sequence version replaced gi:51491327.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zKp90H9
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 190869 bases at least Q40
            Consensus quality: 191868 bases at least Q30
            Consensus quality: 192427 bases at least Q20
            Insert size: 193435; sum-of-contigs
            Insert size: 191728; 8.4% error; agarose-fp
            Quality coverage: 8.95x in Q20 bases; sum-of-contigs Quality
            coverage: 9.11x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2462: contig of 2462 bp in length
            *     2463     2562: gap of 100 bp
            *     2563    19346: contig of 16784 bp in length
            *    19347    19446: gap of 100 bp
            *    19447    64582: contig of 45136 bp in length
            *    64583    64682: gap of 100 bp
            *    64683    99973: contig of 35291 bp in length
            *    99974   100073: gap of 100 bp
```

```
           *   100074   123327: contig of 23254 bp in length
           *   123328   123427: gap of 100 bp
           *   123428   130685: contig of 7258 bp in length
           *   130686   130785: gap of 100 bp
           *   130786   142959: contig of 12174 bp in length
           *   142960   143059: gap of 100 bp
           *   143060   157214: contig of 14155 bp in length
           *   157215   157314: gap of 100 bp
           *   157315   194235: contig of 36921 bp in length.
FEATURES             Location/Qualifiers
     source          1. .194235
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEYP-90H9"
                     /clone_lib="DanioKeypilot"
     misc_feature    1. .2462
                     /note="assembly_fragment:00014
                     fragment_chain:1"
     misc_feature    2563. .19346
                     /note="assembly_fragment:00268
                     fragment_chain:1"
     misc_feature    19447. .64582
                     /note="assembly_fragment:02137
                     fragment_chain:1"
     misc_feature    64683. .99973
                     /note="assembly_fragment:01580
                     fragment_chain:1"
     misc_feature    100074. .123327
                     /note="assembly_fragment:00761
                     fragment_chain:1"
     misc_feature    123428. .130685
                     /note="assembly_fragment:00032
                     fragment_chain:1"
     misc_feature    130786. .142959
                     /note="assembly_fragment:00136.0"
     misc_feature    143060. .157214
                     /note="assembly_fragment:00500"
     misc_feature    157315. .194235
                     /note="assembly_fragment:01065"
ORIGIN

  Query Match             8.5%;  Score 350;  DB 2;  Length 194235;
  Best Local Similarity   69.2%;  Pred. No. 4.7e-59;
  Matches  504;  Conservative    0;  Mismatches  222;  Indels    2;  Gaps    2;

Qy        175 TCCAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATA 234
              | ||  |||   |   |||   | |   || ||   ||||||||||   |   ||||||
Db     122601 TTCATTCCATATATTCAGACTTAGAATCCAACTTGAAAAAAAAAAAAAAAAAAAATATATA 122660

Qy        235 TATATATATATATATATACACACACACACATATATATATACACATATATATACACATA 294
              |||||||||||||||| | | | | | |||||||||| | ||||||||| | |||
Db     122661 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 122720

Qy        295 TATACATATATACACATATATACACATATATATACACATACATATGTACACATATATATA 354
              |||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |||||||||
Db     122721 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 122780

Qy        355 CACATATGTATACACATATATACACATATATACACACATATATACACATATATACACACA 414
               | |||| |||| | ||||||| | ||||||| | | ||||||| | ||||||| | | |
Db     122781 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 122840

Qy        415 TATATACACATATATACACATATATACACATATACACATATATACACATATATACATATA 474
              |||||| | ||||||| | ||||||| | ||||| | ||||||| | ||||||| |||||
Db     122841 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 122900

Qy        475 TACACATATATATAATATACACACATATATATACACATATATACACACATATATACACAT 534
              || | |||||||| ||||| | | ||||||||| | ||||||| | | ||||||| | ||
Db     122901 TATATATATATAT-ATATATATATATATATATATATATATATATATATATATATATAT 122959

Qy        535 ATATACACATATATATACACATATATACACATATATACATATATACACATATATATACAT 594
              |||||| | ||||||||| | ||||||| | ||||||| ||||||| | ||||||||| ||
Db     122960 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 123019

Qy        595 ATATACACATATATACATATATACACATATATACATATATACACACATATATACACATAC 654
              |||||| | ||||||| ||||||| | ||||||| ||||||| | | ||||||| | |||
Db     123020 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 123079

Qy        655 ATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATAT 714
              ||||| | | | ||| ||||| ||||||||| | ||||||||  |||||||| ||||| 
Db     123080 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 123139

Qy        715 ATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAG 774
              ||||||          | |||          | |   ||    |   ||| |   | |
Db     123140 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 123199

Qy        775 AGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGG 834
              | |  | | |    | |||    | |   | || |     |      | |||
Db     123200 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 123259

Qy        835 AAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATA 894
              | ||| ||| || | ||   | |  ||| | | |  ||  ||  ||  |||   ||||||
Db     123260 ATATAAATATATATATATATATATATATAT-ANATATNTATAAATATAAATTTATATA 123318

Qy        895 TATATATA 902
              ||||||||
Db     123319 TATATATA 123326


RESULT 65
AL513530/c
LOCUS       AL513530               50612 bp    DNA     linear   PRI 15-MAY-2001
DEFINITION  Human DNA sequence from clone RP11-447M22 on chromosome 6, complete
            sequence.
ACCESSION   AL513530
VERSION     AL513530.8  GI:14133035
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 50612)
  AUTHORS   Sycamore,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On May 16, 2001 this sequence version replaced gi:13751553.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            RP11-447M22 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-447M22 It may be shorter because we sequence overlapping
```

```
          sections only once, except for a 100 base overlap.
          The true left end of clone RP4-708F5 is at 50513 in this sequence.
          The true right end of clone RP4-682J15 is at 100 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .50612
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP11-447M22"
                     /clone_lib="RPCI-11.2"
     repeat_region   1485. .1570
                     /note="MIR repeat: matches 88. .174 of consensus"
     repeat_region   1745. .1774
                     /note="15 copies 2 mer tt 86% conserved"
     repeat_region   3540. .3811
                     /note="MLT1E repeat: matches 265. .561 of consensus"
     repeat_region   3800. .4122
                     /note="MLT1D repeat: matches 1. .322 of consensus"
     repeat_region   4169. .4648
                     /note="L1M4 repeat: matches 3589. .4085 of consensus"
     repeat_region   5046. .5161
                     /note="58 copies 2 mer aa 59% conserved"
     repeat_region   6172. .6350
                     /note="MER5A repeat: matches 1. .189 of consensus"
     repeat_region   6352. .6435
                     /note="42 copies 2 mer ta 63% conserved"
     repeat_region   6783. .8591
                     /note="TIGGER2 repeat: matches 895. .2718 of consensus"
     repeat_region   8592. .8882
                     /note="AluSx repeat: matches 1. .291 of consensus"
     repeat_region   8883. .9301
                     /note="TIGGER2 repeat: matches 474. .895 of consensus"
     repeat_region   9302. .9598
                     /note="AluJb repeat: matches 1. .291 of consensus"
     repeat_region   9599. .10107
                     /note="TIGGER2 repeat: matches 4. .474 of consensus"
     repeat_region   10470. .10756
                     /note="AluSc repeat: matches 1. .288 of consensus"
     repeat_region   13276. .13732
                     /note="MLT1D repeat: matches 1. .505 of consensus"
     repeat_region   13898. .14119
                     /note="MLT1-INTERNAL repeat: matches 951. .1182 of
                     consensus"
     repeat_region   14554. .15103
                     /note="MLT1-INTERNAL repeat: matches 36. .592 of consensus"
     repeat_region   15130. .15428
                     /note="MLT2E repeat: matches 136. .397 of consensus"
     repeat_region   15401. .15553
                     /note="MLT1D repeat: matches 2. .154 of consensus"
     repeat_region   16010. .16185
                     /note="L2 repeat: matches 2191. .2372 of consensus"
     repeat_region   16302. .16343
                     /note="L1MB8 repeat: matches 6117. .6163 of consensus"
     repeat_region   16353. .16656
                     /note="AluSx repeat: matches 1. .304 of consensus"
     repeat_region   16841. .16908
                     /note="L2 repeat: matches 2641. .2709 of consensus"
     repeat_region   17903. .17997
                     /note="L2 repeat: matches 2575. .2680 of consensus"
     repeat_region   18264. .19239
                     /note="488 copies 2 mer ta 87% conserved"
     misc_feature    18959
                     /note="Tandem repeat. Forced join. Gap size estimated to
                     be approximately 250bp by restriction digest data."
     repeat_region   19240. .19524
                     /note="AluY repeat: matches 1. .285 of consensus"
     repeat_region   19735. .20045
                     /note="AluY repeat: matches 1. .308 of consensus"
     repeat_region   20344. .20626
                     /note="L2 repeat: matches 2002. .2294 of consensus"
     repeat_region   20849. .21066
                     /note="MIR repeat: matches 34. .244 of consensus"
     repeat_region   21538. .21606
                     /note="MER5A repeat: matches 38. .103 of consensus"
     repeat_region   23617. .23800
                     /note="MER5A repeat: matches 2. .186 of consensus"
     repeat_region   23953. .24005
                     /note="MADE1 repeat: matches 28. .80 of consensus"
     repeat_region   24271. .24453
                     /note="MER5A repeat: matches 2. .189 of consensus"
     repeat_region   26159. .26323
                     /note="MIR repeat: matches 78. .249 of consensus"
     repeat_region   26313. .26567
                     /note="L2 repeat: matches 397. .654 of consensus"
     repeat_region   26731. .26876
                     /note="L2 repeat: matches 879. .1028 of consensus"
     repeat_region   27304. .27596
                     /note="AluY repeat: matches 1. .292 of consensus"
     repeat_region   28752. .29042
                     /note="L1 repeat: matches 2111. .2413 of consensus"
     repeat_region   29124. .30099
                     /note="L1 repeat: matches 2504. .3534 of consensus"
     repeat_region   30174. .31422
                     /note="L1M4 repeat: matches 3279. .4581 of consensus"
     repeat_region   31424. .32086
                     /note="L1MC3 repeat: matches 7067. .7739 of consensus"
     repeat_region   32092. .32407
                     /note="AluSq repeat: matches 1. .306 of consensus"
     repeat_region   32417. .32806
                     /note="L1M4 repeat: matches 4603. .5002 of consensus"
     repeat_region   32818. .33181
                     /note="L1MA1 repeat: matches -1388. .-1025 of consensus"
     repeat_region   33182. .33493
                     /note="AluSp repeat: matches 1. .313 of consensus"
     repeat_region   33494. .34874
                     /note="L1MA1 repeat: matches -1025. .356 of consensus"
     repeat_region   34876. .34981
                     /note="L1P repeat: matches 3001. .3103 of consensus"
     repeat_region   34966. .35806
                     /note="L1PA2 repeat: matches 5275. .6144 of consensus"
     repeat_region   35825. .41028
                     /note="L1MA1 repeat: matches 347. .5615 of consensus"
     repeat_region   41029. .41334
                     /note="AluSq repeat: matches 6. .310 of consensus"
     repeat_region   41335. .41728
                     /note="L1MA1 repeat: matches 5615. .6005 of consensus"
     repeat_region   41727. .41963
                     /note="L1MC2 repeat: matches 5825. .6063 of consensus"
     repeat_region   41976. .42095
                     /note="60 copies 2 mer at 85% conserved"
     repeat_region   42097. .42330
                     /note="L1MC1 repeat: matches 6081. .6322 of consensus"
     repeat_region   42547. .42818
                     /note="L1MB7 repeat: matches 5778. .6033 of consensus"
     repeat_region   42819. .43022
                     /note="AluY repeat: matches 102. .303 of consensus"
     repeat_region   43023. .43157
                     /note="L1MB7 repeat: matches 6033. .6171 of consensus"
     repeat_region   43283. .43474
                     /note="MIR repeat: matches 21. .247 of consensus"
     repeat_region   45140. .45454
                     /note="AluY repeat: matches 1. .311 of consensus"
     repeat_region   45522. .45828
                     /note="AluSg repeat: matches 1. .307 of consensus"
     repeat_region   47590. .47773
                     /note="HAL1 repeat: matches 1309. .1491 of consensus"
     repeat_region   47984. .48257
                     /note="HAL1 repeat: matches 630. .931 of consensus"
     repeat_region   49161. .49464
                     /note="AluSq repeat: matches 1. .305 of consensus"
     repeat_region   50177. .50264
                     /note="22 copies 4 mer agag 70% conserved"
     repeat_region   50186. .50263
                     /note="39 copies 2 mer ga 74% conserved"
     repeat_region   50465. .50490
```

```
                     /note="13 copies 2 mer gt 96% conserved"
ORIGIN

  Query Match             8.5%;  Score 349.6;  DB 9;  Length 50612;
  Best Local Similarity   72.2%;  Pred. No. 6.8e-59;
  Matches  523;  Conservative    0;  Mismatches  194;  Indels    7;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC--AAAAT 58
              || |  |||||| ||||   |||| |||||||||||| ||||||||||||||||  |||||
Db      19451 AGATTGAGACCATCCTGGCTAACACGGTGAAACCCCGTCTCTACTAAAAATACAAAAAAT 19392

Qy         59 TAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGA 118
              |||   ||| |||||  ||  ||||||| |||||||||||  ||||||  ||||||||||
Db      19391 TAGCAGGGCGTGGTGGCACGTGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGA 19332

Qy        119 ATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCC 177
              ||| ||||||||||  || ||||| |||||||| ||| |||||||  || | ||||||||
Db      19331 ATCGCTTGAACCTG--GGAGGTGGGGGTTGCGGTGAGCCAAGATCGCGCTACTGCACTCC 19274

Qy        178 AGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATAT 237
              ||||  ||| || ||||| | |   | |||||| || | | | |      | |||||| |||
Db      19273 AGCCTGGGCGAC-AGAGCAAGACCCCGTCTCAAAATATATATATATTTATATATATTTAT 19215

Qy        238 ATATATATATATATACACACACACACATATATATATACACATATATATACACATATATAT 297
              |||||||| ||||||   | | | |   |||||||||| | |||||| || | || |||
Db      19214 ATATATATTTATATATTTATATATATTTATATATATTTATATATATTTATATATTTAT 19155

Qy        298 ACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACAC 357
              | |||| || | ||||||  | ||||||| || | |||  ||| || | || ||||||
Db      19154 ATATATTTATATATATATTTATATATATTTATATATATTTATATATATATTTATATATTT 19095

Qy        358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417
              |||| |||  | ||||||  | |||| || | | || |||| |  ||||||  | |  |||
Db      19094 ATATATATTTATATATATTTATATATTTATATATATTTATATATTTATATATATATTTAT 19035

Qy        418 ATACACATATATACACATATATACACATATACACATATATACACATATATACATATATAC 477
              |||   |||||||  |  |||||| |  |||| |  ||||||   |||| || ||||||
Db      19034 ATATTTATATATATATTTATATATATCTATATATTTATATATGTATATTTATATATATCT 18975

Qy        478 ACATATATATAATATACACACATATATATACACATATATACACACATATATACACATATA 537
              | |||| ||| ||||| | | |||| |||| | || |||| |   ||||||  | |||||
Db      18974 ATATATTTAT-ATATATATATATATTTATATATATTTATATATTTATATATTTATATATA 18916

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACATATATATACATATA 597
              |  | ||||||||  | |||| || | |||| || |||||| || ||||||||  |||||
Db      18915 TTTATATATATATTTATATATTTATATATATTTATATATATTCATATATATATTTATATA 18856

Qy        598 TACACATATATACATATATACACATATATACATATATACACACATATATACACATACATA 657
              |  |||||| |  ||||||  |||||||  ||||||  ||    |||  ||
Db      18855 TTTACATATGTGTTTATATATTTATATATATTTATATATTTACATATGTATTTATATATT 18796

Qy        658 TACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATA 717
              || | | |  || ||||  |||||| ||   |||||| ||  |||||| ||| ||| ||||
Db      18795 TATATATATTTATATATTTATATATGTATTTATATATTTATATATATTTATATATTTATA 18736

Qy        718 TATG 721
              ||||
Db      18735 TATG 18732


RESULT 66
AF438327_3
WPCOMMENT
  Sequence split into 4 fragments  LOCUS AF438327 Accession AF438327
     Fragment Name      Begin     End
     AF438327_0             1     110000
     AF438327_1        100001     210000
     AF438327_2        200001     310000
     AF438327_3        300001     381696
Continuation (4 of 4) of AF438327 from base 300001 (AF438327 Homo sapiens alpha 1 type X

  Query Match             8.5%;  Score 349.6;  DB 9;  Length 81696;
  Best Local Similarity   72.2%;  Pred. No. 6.4e-59;
  Matches  523;  Conservative    0;  Mismatches  194;  Indels    7;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC--AAAAT 58
              || |  |||||| ||||   |||| |||||||||||| ||||||||||||||||  |||||
Db      83851 AGATTGAGACCATCCTGGCTAACACGGTGAAACCCCGTCTCTACTAAAAATACAAAAAAT 83910

Qy         59 TAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGA 118
              |||   ||| |||||  ||  ||||||| |||||||||||  ||||||  ||||||||||
Db      84451 TAGCAGGGCGTGGTGGCACGTGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGA 84510

Qy        119 ATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCC 177
              ||| ||||||||||  || ||||| |||||||| ||| |||||||  || | ||||||||
Db      85051 ATCGCTTGAACCTG--GGAGGTGGGGGTTGCGGTGAGCCAAGATCGCGCTACTGCACTCC 85108

Qy        178 AGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATAT 237
              ||||  ||| || ||||| | |   | |||||| || | | | |      | |||||| |||
Db      85631 AGCCTGGGCGAC-AGAGCAAGACCCCGTCTCAAAATATATATATATTTATATATATTTAT 85689

Qy        238 ATATATATATATATACACACACACACATATATATATACACATATATATACACATATATAT 297
              |||||||| ||||||   | | | |   |||||||||| | |||||| || | || |||
Db      86221 ATATATATTTATATATTTATATATATTTATATATATTTATATATATTTATATATTTAT 86280

Qy        298 ACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACAC 357
              | |||| || | ||||||  | ||||||| || | |||  ||| || | || ||||||
Db      86821 ATATATTTATATATATATTTATATATATTTATATATATTTATATATATATTTATATATTT 86880

Qy        358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417
              |||| |||  | ||||||  | |||| || | | || |||| |  ||||||  | |  |||
Db      86881 ATATATATTTATATATATTTATATATTTATATATATTTATATTTATATATATATTTTAT 86940

Qy        418 ATACACATATATACACATATATACACATATACACATATATACACATATATACATATATAC 477
              |||   |||||||  |  |||||| |  |||| |  ||||||   |||| || ||||||
Db      87461 ATATTTATATATATATTTATATATATCTATATATTTATATATGTATATTTATATATATCT 87520

Qy        478 ACATATATATAATATACACACATATATATACACATATATACACACATATATACACATATA 537
              | |||| ||| ||||| | | |||| |||| | || |||| |   ||||||  | |||||
Db      87521 ATATATTTAT-ATATATATATATATTTATATATATTTATATATTTATATATTTATATATA 87579

Qy        538 TACACATATATATACACATATATACACATATATACATATATACACACATATATACACATATA 597
              |  | ||||||||  | |||| || | |||| || |||||| || ||||||||  |||||
Db      87580 TTTATATATATATTTATATATTTATATATATTTATATATATTCATATATATATTTATATA 87639

Qy        598 TACACATATATACATATATACACATATATACATATATACACACATATATACACATACATA 657
              |  |||||| |  ||||||  |||||||  ||||||  ||    |||  ||
Db      87640 TTTACATATGTGTTTATATATTTATATATATTTATATATTTACATATGTATTTATATATT 87699

Qy        658 TACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATA 717
              || | | |  || ||||  |||||| ||   |||||| ||  |||||| ||| ||| ||||
Db      87700 TATATATATTTATATATTTATATATGTATTTATATATTTATATATATTTATATATTTATA 87759

Qy        718 TATG 721
              ||||
Db      87760 TATG 87763


RESULT 67
AF438327_2
WPCOMMENT
  Sequence split into 4 fragments  LOCUS AF438327 Accession AF438327
     Fragment Name      Begin     End
     AF438327_0             1     110000
     AF438327_1        100001     210000
     AF438327_2        200001     310000
     AF438327_3        300001     381696
Continuation (3 of 4) of AF438327 from base 200001 (AF438327 Homo sapiens alpha 1 type X

  Query Match             8.5%;  Score 349.6;  DB 9;  Length 110000;
  Best Local Similarity   72.2%;  Pred. No. 6.1e-59;
  Matches  523;  Conservative    0;  Mismatches  194;  Indels    7;  Gaps    5;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC--AAAAT 58
              || |  |||||| ||||   |||| |||||||||||| ||||||||||||||||  |||||
Db     108385 AGATTGAGACCATCCTGGCTAACACGGTGAAACCCCGTCTCTACTAAAAATACAAAAAAT 108444
```

```
Qy         59 TAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGA 118
              |||   ||| |||||  ||  ||||||| ||||||||||  |||||| ||||||||||
Db     108445 TAGCAGGGCGTGGTGGCACGTGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGA 108504

Qy        119 ATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCC 177
              ||| ||||||||||  || ||||| |||||||| ||| |||||||  || | ||||||||
Db     108505 ATCGCTTGAACCTG--GGAGGTGGGGGTTGCGGTGAGCCAAGATCGCGCTACTGCACTCC 108562

Qy        178 AGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATAT 237
              ||||  ||| || ||||| | |    | ||||| || | | |       | |||||  |||
Db     108563 AGCCTGGGCGAC-AGAGCAAGACCCCGTCTCAAAATATATATATTTATATATATTTAT 108621

Qy        238 ATATATATATATATACACACACACACATATATATATACACATATATATACACATATAT 297
              |||||||| ||||||   | | | |   |||||||||  | |||||| || | || |||
Db     108622 ATATATATTTATATATTTATATATATTTATATATATATTTATATATATTTATATATTTAT 108681

Qy        298 ACATATATACACATATATACACATATATATACACATACATATGTACACATATATATACAC 357
              | |||| || | ||||||  | ||||||| || | |||  ||| || | || ||||||
Db     108682 ATATATTTATATATATATTTATATATATTTATATATATTTATATATATATTTATATATTT 108741

Qy        358 ATATGTATACACATATATACACATATATACACACATATATACACATATATACACACATAT 417
              |||| |||  | ||||||  | |||| || | | || |||| |  |||||| | |  |||
Db     108742 ATATATATTTATATATATTTATATATTTATATATATTTATATATTTATATATATATTTAT 108801

Qy        418 ATACACATATATACACATATATACACATATACACATATATACACATATATACATATATAC 477
              |||   |||||||  |  |||||| |  |||| |  ||||||   |||| || ||||||
Db     108802 ATATTTATATATATATTTATATATATCTATATATTTATATATGTATATTTATATATATCT 108861

Qy        478 ACATATATATAATATACACACATATATATACACATATATACACACATATATACACATATA 537
              | |||| ||| ||||| | | |||| |||| | || |||| |   ||||||  | |||||
Db     108862 ATATATTTAT-ATATATATATATATTTATATATATTTATATATTTATATATTTATATATA 108920

Qy        538 TACACATATATATACACATATATACACATATATACACATATATACACATATATATACATATA 597
              |  | ||||||||  | |||| || | |||| || |||||| || |||||||| |||||
Db     108921 TTTATATATATATTTATATATTTATATATATTTATATATATTCATATATATATTTATATA 108980

Qy        598 TACACATATATACACATATATACACATATATACACATATATACACACATATATACACATACATA 657
              |  ||||||  |   ||||||   |||||||  ||||||   |||||| ||   ||| ||
Db     108981 TTTACATATGTGTTTATATATTTATATATATTTATATATTTACATATGTATTTATATATT 109040

Qy        658 TACACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATA 717
              || | | |  || ||||  |||||| ||   |||||| ||  ||||| ||| ||| ||||
Db     109041 TATATATATTTATATATTTATATATGTATTTATATATTTATATATATTTATATATTTATA 109100

Qy        718 TATG 721
              ||||
Db     109101 TATG 109104


RESULT 68
CR759826/c
LOCUS       CR759826              157042 bp    DNA     linear   HTG 05-SEP-2004
DEFINITION  Danio rerio clone CH211-266K13, *** SEQUENCING IN PROGRESS ***, 10
            unordered pieces.
ACCESSION   CR759826
VERSION     CR759826.1  GI:51890341
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 157042)
  AUTHORS   McLay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     ---------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            ---------- Project Information
            Center project name: zC266K13
            ---------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 153933 bases at least Q40
            Consensus quality: 154487 bases at least Q30
            Consensus quality: 154866 bases at least Q20
            Insert size: 156142; sum-of-contigs
            Insert size: 159217; 0.8% error; agarose-fp
            Quality coverage: 6.01x in Q20 bases; sum-of-contigs Quality
            coverage: 5.90x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    44418: contig of 44418 bp in length
            *    44419    44518: gap of 100 bp
            *    44519    52984: contig of 8466 bp in length
            *    52985    53084: gap of 100 bp
            *    53085    65205: contig of 12121 bp in length
            *    65206    65305: gap of 100 bp
            *    65306    67571: contig of 2266 bp in length
            *    67572    67671: gap of 100 bp
            *    67672   110374: contig of 42703 bp in length
            *   110375   110474: gap of 100 bp
            *   110475   142025: contig of 31551 bp in length
            *   142026   142125: gap of 100 bp
            *   142126   146693: contig of 4568 bp in length
            *   146694   146793: gap of 100 bp
            *   146794   151219: contig of 4426 bp in length
            *   151220   151319: gap of 100 bp
            *   151320   154853: contig of 3534 bp in length
            *   154854   154953: gap of 100 bp
            *   154954   157042: contig of 2089 bp in length.
FEATURES             Location/Qualifiers
     source          1. .157042
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-266K13"
                     /clone_lib="CHORI-211"
     misc_feature    1. .44418
                     /note="assembly_fragment:00615
                     fragment_chain:1"
     misc_feature    44519. .52984
                     /note="assembly_fragment:00139
                     fragment_chain:1"
     misc_feature    53085. .65205
                     /note="assembly_fragment:00214
                     fragment_chain:1"
     misc_feature    65306. .67571
                     /note="assembly_fragment:00005
                     fragment_chain:1"
     misc_feature    67672. .110374
                     /note="assembly_fragment:01059
                     fragment_chain:1"
     misc_feature    110475. .142025
                     /note="assembly_fragment:00318
                     fragment_chain:2"
     misc_feature    142126. .146693
                     /note="assembly_fragment:00071
                     fragment_chain:2"
     misc_feature    146794. .151219
                     /note="assembly_fragment:00104"
     misc_feature    151320. .154853
                     /note="assembly_fragment:00042
```

```
                     fragment_chain:3"
     misc_feature    154954. .157042
                     /note="assembly_fragment:00020
                     fragment_chain:3
                     clone_end:SP6
                     vector_side:right"
ORIGIN

  Query Match             8.5%;  Score 349.6;  DB 2;  Length 157042;
  Best Local Similarity   63.1%;  Pred. No. 5.8e-59;
  Matches  462;  Conservative    0;  Mismatches  269;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db      68181 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 68122

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db      68121 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 68062

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db      68061 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 68002

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db      68001 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 67942

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db      67941 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 67882

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db      67881 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 67822

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db      67821 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 67762

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||  ||||| | | | ||| ||||| ||||||||| | |||||||||  || |||
Db      67761 TATATATTTATATATATATATATATATATATATATATATATATATATATATATATACATA 67702

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||||| ||||     | |
Db      67701 TATATATATATATATGATCCATCTCTACTCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 67642

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825

Db      67641 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 67582

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885
                    | | | || | ||   | |  ||| | ||   |  ||       ||
Db      67581 NNNNNNNNNTTTTAAATATATATATATATATATATATTTTTTTTTTTTTTATAT 67522

Qy        886 CACTATATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAAT 945
              |||||||||||||   | |    | || |  || |          |   || || ||
Db      67521 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 67462

Qy        946 GTTATAGAAGAA 957
              | | || ||
Db      67461 AAAAAAAAAAAA 67450


RESULT 69
CR589883/c
LOCUS       CR589883              180560 bp    DNA     linear   HTG 27-JUL-2004
DEFINITION  Danio rerio clone CH211-51N18, WORKING DRAFT SEQUENCE, 6 unordered
            pieces.
ACCESSION   CR589883
VERSION     CR589883.2  GI:50724901
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 180560)
  AUTHORS   Burton,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JUL-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 27, 2004 this sequence version replaced gi:50399732.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zC51N18
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 177647 bases at least Q40
            Consensus quality: 178065 bases at least Q30
            Consensus quality: 178456 bases at least Q20
            Insert size: 180060; sum-of-contigs
            Insert size: 184777; 2.8% error; agarose-fp
            Quality coverage: 8.72x in Q20 bases; sum-of-contigs Quality
            coverage: 8.60x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     7987: contig of 7987 bp in length
            *     7988     8087: gap of 100 bp
            *     8088    33232: contig of 25145 bp in length
            *    33233    33332: gap of 100 bp
            *    33333    70114: contig of 36782 bp in length
            *    70115    70214: gap of 100 bp
            *    70215   126158: contig of 55944 bp in length
            *   126159   126258: gap of 100 bp
            *   126259   162306: contig of 36048 bp in length
            *   162307   162406: gap of 100 bp
            *   162407   180560: contig of 18154 bp in length.
FEATURES             Location/Qualifiers
     source          1. .180560
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-51N18"
                     /clone_lib="CHORI-211"
     misc_feature    1. .7987
                     /note="assembly_fragment:00018
                     clone_end:SP6
                     vector_side:left"
     misc_feature    8088. .33232
                     /note="assembly_fragment:00362.0"
     misc_feature    33333. .70114
                     /note="assembly_fragment:01176
                     fragment_chain:1"
     misc_feature    70215. .126158
                     /note="assembly_fragment:01644
                     fragment_chain:1"
     misc_feature    126259. .162306
                     /note="assembly_fragment:00726
                     fragment_chain:1"
     misc_feature    162407. .180560
                     /note="assembly_fragment:00143
```

```
                    fragment_chain:1
                    clone_end:T7
                    vector_side:right"
ORIGIN

  Query Match             8.5%;  Score 349.6;  DB 2;  Length 180560;
  Best Local Similarity   43.8%;  Pred. No. 5.7e-59;
  Matches  560;  Conservative    0;  Mismatches  717;  Indels    2;  Gaps    2;

Qy        205 TCTCAGAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACA 264
              | | || | | | | |       | ||||||||||||||||||||||||| | | | | |
Db     127006 TTTAAGTATATATATATATATATATATATATATATATATATATATATATATATATA 126947

Qy        265 CATATATATATACACATATATATACACATATATACATATATACACATATATACACATATA 324
              |||||||||||| | ||||||||| | ||||||| ||||||| | ||||||| | |||||
Db     126946 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 126887

Qy        325 TATACACATACATATGTACACATATATATACACATATGTATACACATATATACACATATA 384
              |||| | ||| |||| || | ||||||||| | |||| |||| | ||||||| | |||||
Db     126886 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 126827

Qy        385 TACACACATATATACACATATATACACACATATATACACATATATACACATATATACACA 444
              || | | ||||||| | ||||||| | | ||||||| | ||||||| | ||||||| | |
Db     126826 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 126767

Qy        445 TATACACATATATACACATATATACATATATACACATATATATAATATACACACATATAT 504
              |||| | ||||||| | ||||||| ||||||| | |||||||| ||||| | | ||||||
Db     126766 TATATATATATATATATATATATATATATATATATATATATAT-ATATATATATATAT 126708

Qy        505 ATACACATATATACACACATATATACACATATATACACATATATATACACATATATACAC 564
              ||| | ||||||| | | ||||||| | ||||||| | ||||||||| | ||||||| |
Db     126707 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 126648

Qy        565 ATATATACATATATACACATATATATACATATATACACATATATACATATATACACATAT 624
              ||||||| ||||||| | ||||||||| ||||||| | ||||||| ||||||| | ||||
Db     126647 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 126588

Qy        625 ATACATATATACACACATATATACACATACATATACACACACATAGATATACATATATAT 684
              ||| ||||||| | | ||||||| | ||| ||||| | | | ||| ||||| |||||||||
Db     126587 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 126528

Qy        685 ACACATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGC 744
              | | |||||||||  |||||||| ||||||||||        |  |||
Db     126527 ATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 126468

Qy        745 AGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGG 804
              | |   ||    |    ||| |   | | | |   | | |    | |||    | | |
Db     126467 ATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 126408

Qy        805 ACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCT 864
              | || |     |      | |||       | ||| ||| || | ||   | |  ||| |
Db     126407 ATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 126348

Qy        865 GTA-AGGTAGTTAAGAAGGTAACACTATATATATATATAGTGAAAGCAGTGTAAACCTAA 923
               || |  ||  ||   |  ||    |||||||||||||   | |   | ||   ||
Db     126347 ATATATATATATATATATATATATATATATATATATATATATATATATATCTATATGAATAT 126288

Qy        924 AGGATGGGCCAAGGATTTAAATGTTATAGAAGAATGGCTAAGATGCCAAAGCTCAGTGTA 983
              |         |   || || || |
Db     126287 ATATATATATATATATATATATATATANNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 126228

Qy        984 TGTGGCAGAGGCATGGTGTAGGGTGTGTCCAGGTTCATATATTGCATTAAGTGTGAGAAC 1043

Db     126227 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 126168

Qy       1044 ACCCTGGAGTATGAACCAAGAAAATGCAAAAGCCAGAAGTGATGGAGGAAATGAGACACA 1103
                 ||  ||       |                 |
Db     126167 NNNNNNNNNTAAAAANNNNNNNANNATNNNNNNNNNNANNNNNNNNNNNTNNNNNNNTN 126108

Qy       1104 ATAATGAAGATATTGAGAGGAGGGTGTGGGCCTAGAGTGAAGCTTTTCGTGCCAGTACTT 1163
               |                                          ||
Db     126107 TTNNNTNNNNNNNNNNNNANNTNNNNNTNNNNNNNNNAAANNNNNNNNNNNNNNNNN 126048

Qy       1164 CTTTTGAAGGCCCAGTTCTCTTCTCTCTCGGGGGCTCCTTCATCTCTCATAGAGTCCACA 1223
                                    |
Db     126047 NNANNNNNNNNTANNNNNTNNANNTNNNTNNNNNNNNNNNNNNNNNANNNNNNNNNNN 125988

Qy       1224 GCTTTTAAGGGCCAACACTTGAGGTCAGCCTGGCTCTCTCATTTGAGCTGGATAGAACAT 1283

Db     125987 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 125928

Qy       1284 TTTAGAGCACCATCTATTCTTCAAGAGGAAGTTTAAAAATAAAAGAACCTTGAAGAGGAA 1343
                       | | || ||  | |  |  |  ||||    |     |
Db     125927 NNNNNNNTTNNNTTTTTTTTTNTATAANANATANTAAAAANNNNNANNNNNNANNNNANT 125868

Qy       1344 AAAATGTAGACATTCAATCTAACCTTTTCATTTTACTAGCCAAAGCTAAATAGAATGCAG 1403
                  |                   |                |      |   || ||      |
Db     125867 TNNNNNTTNNNNNNNNNNNNNNNTAAANNNNNNNNNAANNANNATTAAATTANTTAAAAT 125808

Qy       1404 GTTACCTGTTTTTCAGCCAGGCACCATCATTTCCTAATTGTTATAAAATTTATTATTATT 1463
                | |     ||||                |       || || |     ||    |  |
Db     125807 ATAANTNNNTTTTTTTNTNTNNNNNTNTNTNNNNNTTTTTTTNTNTNTNTTTNNNTNNNT 125748

Qy       1464 GTTGTTATTATTATTATTT 1482
                ||    ||    || | |
Db     125747 NTTNNNTTTNNNTTTNTNT 125729


RESULT 70
CR556714
LOCUS       CR556714               210384 bp    DNA     linear   HTG 12-JUL-2004
DEFINITION  Danio rerio clone DKEY-29P13, WORKING DRAFT SEQUENCE, 5 unordered
            pieces.
ACCESSION   CR556714
VERSION     CR556714.2  GI:50251122
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 210384)
  AUTHORS   Burton,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUL-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 12, 2004 this sequence version replaced gi:50234128.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK29P13
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 209301 bases at least Q40
            Consensus quality: 209711 bases at least Q30
            Consensus quality: 209847 bases at least Q20
            Insert size: 209984; sum-of-contigs
            Insert size: 209994; 2.1% error; agarose-fp
            Quality coverage: 10.01x in Q20 bases; sum-of-contigs Quality
            coverage: 10.13x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    19750: contig of 19750 bp in length
            *    19751    19850: gap of 100 bp
```

```
            *    19851    37399: contig of 17549 bp in length
            *    37400    37499: gap of 100 bp
            *    37500   112088: contig of 74589 bp in length
            *   112089   112188: gap of 100 bp
            *   112189   196376: contig of 84188 bp in length
            *   196377   196476: gap of 100 bp
            *   196477   210384: contig of 13908 bp in length.
FEATURES             Location/Qualifiers
     source          1. .210384
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-29P13"
                     /clone_lib="DanioKey"
     misc_feature    1. .19750
                     /note="assembly_fragment:00422
                     fragment_chain:1"
     misc_feature    19851. .37399
                     /note="assembly_fragment:00191
                     fragment_chain:1"
     misc_feature    37500. .112088
                     /note="assembly_fragment:00690
                     fragment_chain:2"
     misc_feature    112189. .196376
                     /note="assembly_fragment:01732
                     fragment_chain:2"
     misc_feature    196477. .210384
                     /note="assembly_fragment:00009.0"
ORIGIN

  Query Match             8.5%;  Score 349.6;  DB 2;  Length 210384;
  Best Local Similarity  68.8%;  Pred. No. 5.5e-59;
  Matches  509;  Conservative    0;  Mismatches  229;  Indels    2;  Gaps    2;

Qy        187 AACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATA 246
              ||  | |   |||     | |  |  |||| ||||||  | | ||||||||||||||||||
Db     111340 AAGTAAAAAAAAAGTAAAACAAAATAAAATAAAAAGTAAATATATATATATATATATATA 111399

Qy        247 TATATACACACACACACATATATATATACACATATATATACACATATATACATATATA 306
              |||||| | | | | | ||||||||||| | |||||||| | |||||| ||||||
Db     111400 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 111459

Qy        307 CACATATATACACATATATATACACATACATATGTACACATATATATACACATATGTATA 366
              | ||||||| | ||||||||| | ||| |||| || | ||||||||| | |||| ||||
Db     111460 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 111519

Qy        367 CACATATATACACATATATACACACATATATACACATATATACACACATATATACACATA 426
              | ||||||| | ||||||| | | ||||||| | ||||||| | | ||||||| | |||
Db     111520 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 111579

Qy        427 TATACACATATATACACATATACACATATATACACATATATACATATATACACATATATA 486
              |||| | ||||||| | ||||| | ||||||| | ||||||| ||||||| | |||||||
Db     111580 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 111639

Qy        487 TA-ATATACACACATATATATACACATATATACACACATATATACACATATATACACATA 545
              || ||||| | | ||||||||| | ||||||| | | ||||||| | ||||||| | |||
Db     111640 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 111699

Qy        546 TATATACACATATATACACATATATACATATATACACATATATATACATATATACACATA 605
              |||||| | ||||||| | ||||||| ||||||| | ||||||||| ||||||| | |||
Db     111700 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 111759

Qy        606 TATACATATATACACATATATACATATATACACACATATATACACATACATATACACACA 665
              |||| ||||||| | ||||||| ||||||| | | ||||||| | ||| ||||| | | |
Db     111760 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 111819

Qy        666 CATAGATATACATATATATACACATATATATACGTATATATATGTATATATATATGCTCC 725
              ||| ||||| ||||||||| | |||||||||  |||||||||  ||||||||||
Db     111820 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 111879

Qy        726 AGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACA 785
                  | |||          | |   ||     |    ||| |   | | | |  | | |
Db     111880 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 111939

Qy        786 GCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAA 845
              | |||    | |   | || |      |       | |||       | ||| ||| |
Db     111940 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 111999

Qy        846 TGTGTAATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTAACACTATATATATATAGT 904
              | | ||  | | |||| | || |  ||  ||   |  |||   ||||||||||||||
Db     112000 TATATATATATATATATATATATATATATATATATATAAATATATATATATATATA 112059

Qy        905 GAAAGCAGTGTAAACCTAAA 924
              | |  | | || |  || |
Db     112060 TATATAAATATATATATATA 112079


RESULT 71
CR790372
LOCUS       CR790372              299106 bp    DNA     linear   HTG 05-OCT-2004
DEFINITION  Danio rerio clone DKEY-70K13, *** SEQUENCING IN PROGRESS ***, 25
            unordered pieces.
ACCESSION   CR790372
VERSION     CR790372.4  GI:53828302
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 299106)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 5, 2004 this sequence version replaced gi:53755940.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK70K13
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 289331 bases at least Q40
            Consensus quality: 291856 bases at least Q30
            Consensus quality: 293478 bases at least Q20
            Insert size: 296706; sum-of-contigs
            Insert size: 225769; 2.9% error; agarose-fp
            Quality coverage: 6.22x in Q20 bases; sum-of-contigs Quality
            coverage: 9.16x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 25 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     3928: contig of 3928 bp in length
            *     3929     4028: gap of 100 bp
            *     4029     8417: contig of 4389 bp in length
            *     8418     8517: gap of 100 bp
            *     8518    12283: contig of 3766 bp in length
            *    12284    12383: gap of 100 bp
            *    12384    16946: contig of 4563 bp in length
            *    16947    17046: gap of 100 bp
            *    17047    19834: contig of 2788 bp in length
            *    19835    19934: gap of 100 bp
            *    19935    26863: contig of 6929 bp in length
            *    26864    26963: gap of 100 bp
            *    26964    29135: contig of 2172 bp in length
```

```
            *    29136    29235: gap of 100 bp
            *    29236    43722: contig of 14487 bp in length
            *    43723    43822: gap of 100 bp
            *    43823   109813: contig of 65991 bp in length
            *   109814   109913: gap of 100 bp
            *   109914   140749: contig of 30836 bp in length
            *   140750   140849: gap of 100 bp
            *   140850   156362: contig of 15513 bp in length
            *   156363   156462: gap of 100 bp
            *   156463   165292: contig of 8830 bp in length
            *   165293   165392: gap of 100 bp
            *   165393   243285: contig of 77893 bp in length
            *   243286   243385: gap of 100 bp
            *   243386   245562: contig of 2177 bp in length
            *   245563   245662: gap of 100 bp
            *   245663   250094: contig of 4432 bp in length
            *   250095   250194: gap of 100 bp
            *   250195   256405: contig of 6211 bp in length
            *   256406   256505: gap of 100 bp
            *   256506   258788: contig of 2283 bp in length
            *   258789   258888: gap of 100 bp
            *   258889   262780: contig of 3892 bp in length
            *   262781   262880: gap of 100 bp
            *   262881   265140: contig of 2260 bp in length
            *   265141   265240: gap of 100 bp
            *   265241   269635: contig of 4395 bp in length
            *   269636   269735: gap of 100 bp
            *   269736   272991: contig of 3256 bp in length
            *   272992   273091: gap of 100 bp
            *   273092   291385: contig of 18294 bp in length
            *   291386   291485: gap of 100 bp
            *   291486   293629: contig of 2144 bp in length
            *   293630   293729: gap of 100 bp
            *   293730   296461: contig of 2732 bp in length
            *   296462   296561: gap of 100 bp
            *   296562   299106: contig of 2545 bp in length.
FEATURES             Location/Qualifiers
     source          1. .299106
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-70K13"
                     /clone_lib="DanioKey"
     misc_feature    1. .3928
                     /note="assembly_fragment:00025
                     fragment_chain:1"
     misc_feature    4029. .8417
                     /note="assembly_fragment:00178
                     fragment_chain:1"
     misc_feature    8518. .12283
                     /note="assembly_fragment:00051
                     fragment_chain:1"
     misc_feature    12384. .16946
                     /note="assembly_fragment:00068
                     fragment_chain:1"
     misc_feature    17047. .19834
                     /note="assembly_fragment:03721
                     fragment_chain:1"
     misc_feature    19935. .26863
                     /note="assembly_fragment:00206
                     fragment_chain:1"
     misc_feature    26964. .29135
                     /note="assembly_fragment:03693
                     fragment_chain:1"
     misc_feature    29236. .43722
                     /note="assembly_fragment:00405
                     fragment_chain:2"
     misc_feature    43823. .109813
                     /note="assembly_fragment:01445
                     fragment_chain:2"
     misc_feature    109914. .140749
                     /note="assembly_fragment:01030
                     fragment_chain:2"
     misc_feature    140850. .156362
                     /note="assembly_fragment:00596
                     fragment_chain:3"
     misc_feature    156463. .165292
                     /note="assembly_fragment:00282
                     fragment_chain:3"
     misc_feature    165393. .243285
                     /note="assembly_fragment:02358
                     fragment_chain:3"
     misc_feature    243386. .245562
                     /note="assembly_fragment:03730
                     fragment_chain:4"
     misc_feature    245663. .250094
                     /note="assembly_fragment:00104
                     fragment_chain:4"
     misc_feature    250195. .256405
                     /note="assembly_fragment:00151
                     fragment_chain:4"
     misc_feature    256506. .258788
                     /note="assembly_fragment:00124
                     fragment_chain:5"
     misc_feature    258889. .262780
                     /note="assembly_fragment:00234
                     fragment_chain:5"
     misc_feature    262881. .265140
                     /note="assembly_fragment:03713
                     fragment_chain:6"
     misc_feature    265241. .269635
                     /note="assembly_fragment:00038
                     fragment_chain:6"
     misc_feature    269736. .272991
                     /note="assembly_fragment:00085"
     misc_feature    273092. .291385
                     /note="assembly_fragment:00801.0"
     misc_feature    291486. .293629
                     /note="assembly_fragment:03699"
     misc_feature    293730. .296461
                     /note="assembly_fragment:03706"
     misc_feature    296562. .299106
                     /note="assembly_fragment:03750"
ORIGIN

  Query Match             8.5%;  Score 349.6;  DB 2;  Length 299106;
  Best Local Similarity   60.4%;  Pred. No. 5.3e-59;
  Matches  492;  Conservative    0;  Mismatches  321;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db     242607 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 242666

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | |||||||||| | ||| |||| || | |
Db     242667 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 242726

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db     242727 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 242786

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db     242787 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 242846

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | |||||||||| ||||| | | |||||||||| | ||||||| | | |||
Db     242847 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 242906

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | |||||||||| | ||||||| | ||||||| ||||||| | |||
Db     242907 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 242966

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db     242967 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 243026
```

```
Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db     243027 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 243086

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||||          | |||          | |   ||     |    |||
Db     243087 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 243146

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
              |   | | | |  | | |    | |||    | |   | || |      |      | |
Db     243147 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 243206

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885
              ||       | ||| ||| || | |||  | |  ||| | ||  || ||||   |   ||
Db     243207 TATATATATATATATATATATATATAAATATATATATATTTATTGTTGTTATATANANAA 243266

Qy        886 CACTATATATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAAT 945
              |  | | |||||||||
Db     243267 AAAAAAAAATATATATATANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 243326

Qy        946 GTTATAGAAGAATGGCTAAGATGCCAAAGCTCAGTGTATGTGGCAGAGGCATGGTGTAGG 1005

Db     243327 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNC 243386

Qy       1006 GTGTGTCCAGGTTCATATATTGCATTAAGTGTGA 1039
                | |    ||| | |  || ||||||
Db     243387 TGAGCAAAGGCTGTGAATACTTCTGTACGTGTGA 243420


RESULT 72
CR790384/c
LOCUS       CR790384               182223 bp    DNA     linear   HTG 08-OCT-2004
DEFINITION  Danio rerio clone DKEY-208I6, *** SEQUENCING IN PROGRESS ***, 7
            unordered pieces.
ACCESSION   CR790384
VERSION     CR790384.3  GI:54019829
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 182223)
  AUTHORS   McLay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 9, 2004 this sequence version replaced gi:52694441.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK208I6
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 180226 bases at least Q40
            Consensus quality: 180685 bases at least Q30
            Consensus quality: 180956 bases at least Q20
            Insert size: 181623; sum-of-contigs
            Insert size: 177450; 3.6% error; agarose-fp
            Quality coverage: 17.50x in Q20 bases; sum-of-contigs Quality
            coverage: 18.50x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    27274: contig of 27274 bp in length
            *    27275    27374: gap of 100 bp
            *    27375    66205: contig of 38831 bp in length
            *    66206    66305: gap of 100 bp
            *    66306   114861: contig of 48556 bp in length
            *   114862   114961: gap of 100 bp
            *   114962   118489: contig of 3528 bp in length
            *   118490   118589: gap of 100 bp
            *   118590   147487: contig of 28898 bp in length
            *   147488   147587: gap of 100 bp
            *   147588   158792: contig of 11205 bp in length
            *   158793   158892: gap of 100 bp
            *   158893   182223: contig of 23331 bp in length.
FEATURES             Location/Qualifiers
     source          1. .182223
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-208I6"
                     /clone_lib="DanioKey"
     misc_feature    1. .27274
                     /note="assembly_fragment:01073
                     fragment_chain:1"
     misc_feature    27375. .66205
                     /note="assembly_fragment:01954
                     fragment_chain:1"
     misc_feature    66306. .114861
                     /note="assembly_fragment:03886
                     fragment_chain:1"
     misc_feature    114962. .118489
                     /note="assembly_fragment:00044
                     fragment_chain:2"
     misc_feature    118590. .147487
                     /note="assembly_fragment:02902
                     fragment_chain:2"
     misc_feature    147588. .158792
                     /note="assembly_fragment:00096.0"
     misc_feature    158893. .182223
                     /note="assembly_fragment:00385"
ORIGIN

  Query Match             8.5%;  Score 349.4;  DB 2;  Length 182223;
  Best Local Similarity   68.3%;  Pred. No. 6.2e-59;
  Matches  499;  Conservative    0;  Mismatches  231;  Indels    1;  Gaps    1;

Qy        199 ATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACA 258
              || | | |    |||||||||      | |||||||||||||||||||||||| | | |
Db      67159 ATTTAAGCAAGTAAAAAAAAAATATATATATATATATATATATATATATATATATATA 67100

Qy        259 CACACACATATATATATACACATATATATACACATATATACATATATACACATATATACA 318
              | | | |||||||||||| | ||||||||| | ||||||| ||||||| | ||||||| |
Db      67099 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 67040

Qy        319 CATATATATACACATACATATGTACACATATATATACACATATGTATACACATATATACA 378
              ||||||||| | ||| |||| || | ||||||||| | |||| |||| | ||||||| |
Db      67039 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 66980

Qy        379 CATATATACACACATATATACACATATATACACACATATATACACATATATACACATATA 438
              ||||||| | | ||||||| | ||||||| | | ||||||| | ||||||| | |||||
Db      66979 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 66920

Qy        439 TACACATATACACATATATACACATATATACATATATACACATATATATA-ATATACACA 497
              || | ||||| | ||||||| | ||||||| ||||||| | ||||||||| ||||| | |
Db      66919 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 66860

Qy        498 CATATATATACACATATATACACACATATATACACATATATACACATATATATACACATA 557
              ||||||||| | ||||||| | | ||||||| | ||||||| | ||||||||| | |||
Db      66859 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 66800

Qy        558 TATACACATATATACATATATACACATATATATACATATATACACATATATACATATATA 617
```

```
              |||| | ||||||| ||||||| | ||||||||| ||||||| | ||||||| |||||||
Db      66799 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 66740

Qy        618 CACATATATACATATATACACACATATATACACATACATATACACACACATAGATATACA 677
              | ||||||| ||||||| | | ||||||| | ||| ||||| | | | ||| ||||| |
Db      66739 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 66680

Qy        678 TATATATACACATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAG 737
              |||||||| | |||||||||  |||||||| ||||||||||          |   |||
Db      66679 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 66620

Qy        738 AGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGT 797
                  |   ||    | |    ||| |   ||| | |  |  | |    | |||
Db      66619 TATATATATAATATATATAATATATATATATATAAATAAATATATAAATATATATATATA 66560

Qy        798 GAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAA 857
              | |   | |||| | |  | ||    |  |        | ||| | |||      ||  ||
Db      66559 TATATATATATTTTAAATATATATATATAATATATAAATATATAAAAAATAAATATTTAA 66500

Qy        858 CCTATCTGTAAGGTAGTTAAGAAGGTAACACTATATATATATATAGTGAAAGCAGTGTAA 917
              ||   | |   |  ||| |  ||| |  || ||  ||||   | ||   |   |||
Db      66499 ATAATAATTTATAAATATAAATATATAATAAAATTTAATTATAATATAAATTAATAATAA 66440

Qy        918 ACCTAAAGGAT 928
              |  | ||  ||
Db      66439 ATATTAAATAT 66429


RESULT 73
CR790362
LOCUS       CR790362              223166 bp    DNA     linear   HTG 05-OCT-2004
DEFINITION  Danio rerio clone DKEY-219H14, *** SEQUENCING IN PROGRESS ***, 13
            unordered pieces.
ACCESSION   CR790362
VERSION     CR790362.3  GI:53828288
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 223166)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 5, 2004 this sequence version replaced gi:52626676.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK219H14
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 218162 bases at least Q40
            Consensus quality: 219035 bases at least Q30
            Consensus quality: 219848 bases at least Q20
            Insert size: 221966; sum-of-contigs
            Insert size: 231484; 3.1% error; agarose-fp
            Quality coverage: 9.07x in Q20 bases; sum-of-contigs Quality
            coverage: 9.03x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 13 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    45629: contig of 45629 bp in length
            *    45630    45729: gap of 100 bp
            *    45730    47743: contig of 2014 bp in length
            *    47744    47843: gap of 100 bp
            *    47844    71492: contig of 23649 bp in length
            *    71493    71592: gap of 100 bp
            *    71593    80196: contig of 8604 bp in length
            *    80197    80296: gap of 100 bp
            *    80297    94270: contig of 13974 bp in length
            *    94271    94370: gap of 100 bp
            *    94371   105308: contig of 10938 bp in length
            *   105309   105408: gap of 100 bp
            *   105409   113062: contig of 7654 bp in length
            *   113063   113162: gap of 100 bp
            *   113163   125638: contig of 12476 bp in length
            *   125639   125738: gap of 100 bp
            *   125739   129703: contig of 3965 bp in length
            *   129704   129803: gap of 100 bp
            *   129804   143077: contig of 13274 bp in length
            *   143078   143177: gap of 100 bp
            *   143178   148174: contig of 4997 bp in length
            *   148175   148274: gap of 100 bp
            *   148275   180179: contig of 31905 bp in length
            *   180180   180279: gap of 100 bp
            *   180280   223166: contig of 42887 bp in length.
FEATURES             Location/Qualifiers
     source          1. .223166
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-219H14"
                     /clone_lib="DanioKey"
     misc_feature    1. .45629
                     /note="assembly_fragment:02641
                     fragment_chain:1"
     misc_feature    45730. .47743
                     /note="assembly_fragment:00088
                     fragment_chain:1"
     misc_feature    47844. .71492
                     /note="assembly_fragment:01125
                     fragment_chain:1"
     misc_feature    71593. .80196
                     /note="assembly_fragment:00289
                     fragment_chain:1"
     misc_feature    80297. .94270
                     /note="assembly_fragment:00715
                     fragment_chain:1"
     misc_feature    94371. .105308
                     /note="assembly_fragment:00568
                     fragment_chain:2"
     misc_feature    105409. .113062
                     /note="assembly_fragment:00191
                     fragment_chain:2"
     misc_feature    113163. .125638
                     /note="assembly_fragment:00919
                     fragment_chain:2"
     misc_feature    125739. .129703
                     /note="assembly_fragment:00071
                     fragment_chain:3"
     misc_feature    129804. .143077
                     /note="assembly_fragment:00426
                     fragment_chain:3"
     misc_feature    143178. .148174
                     /note="assembly_fragment:00118.0"
     misc_feature    148275. .180179
                     /note="assembly_fragment:01481"
     misc_feature    180280. .223166
                     /note="assembly_fragment:01968"
ORIGIN

  Query Match             8.5%;  Score 349.4;  DB 2;  Length 223166;
  Best Local Similarity   79.2%;  Pred. No. 6e-59;
```

```
  Matches  427;  Conservative    0;  Mismatches  111;  Indels    1;  Gaps    1;

Qy        188 ACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATAT 247
              | || |  ||||| |  | | |||||||||| |||   ||| ||||||||||||||||||
Db     201702 AAAAAAAAGAAATTTTTTTTTAGAAAAAAATAAATAAATTATATATATATATATATATAT 201761

Qy        248 ATATACACACACACACATATATATATACACATATATATACACATATATACATATATAC 307
              ||||| | | | | | | |||||||||||| | ||||||||| | ||||||| |||||||
Db     201762 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 201821

Qy        308 ACATATATACACATATATATACACATACATATGTACACATATATATACACATATGTATAC 367
              | ||||||| | ||||||||| | ||| |||| || | ||||||||| | |||| ||||
Db     201822 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 201881

Qy        368 ACATATATACACATATATACACACATATATACACATATATACACACATATATACACATAT 427
              | ||||||| | ||||||| | | ||||||| | ||||||| | | ||||||| | ||||
Db     201882 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 201941

Qy        428 ATACACATATATACACATATACACATATATACACATATATACATATATACACATATATAT 487
              ||| | ||||||| | ||||| | ||||||| | ||||||| ||||||| | ||||||||
Db     201942 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 202001

Qy        488 A-ATATACACACATATATATACACATATATACACACATATATACACATATATACACATAT 546
              | ||||| | | ||||||||| | ||||||| | | ||||||| | ||||||| | ||||
Db     202002 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 202061

Qy        547 ATATACACATATATACACATATATACATATATACACATATATATACATATATACACATAT 606
              ||||| | ||||||| | ||||||| ||||||| | ||||||||| ||||||| | ||||
Db     202062 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 202121

Qy        607 ATACATATATACACATATATACATATATACACACATATATACACATACATATACACACAC 666
              ||| || |||| | ||||||| ||||||| | | ||||||| | ||| ||||| | | |
Db     202122 ATATATCTATATATATATATATATATATATATATATATATATATATATATATATATAT 202181

Qy        667 ATAGATATACATATATATACACATATATATACGTATATATATGTATATATATATGCTCC 725
              ||| ||||| ||||||||| | |||||||||  |||||||| ||||||||||  || |
Db     202182 ATATATATATATATATATATATATATATATATATATATATATATATATATATAACCTTC 202240


RESULT 74
CR759946
LOCUS       CR759946               232824 bp    DNA     linear   HTG 11-SEP-2004
DEFINITION  Danio rerio clone DKEY-79C1, *** SEQUENCING IN PROGRESS ***, 8
            unordered pieces.
ACCESSION   CR759946
VERSION     CR759946.2  GI:51988227
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 232824)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 11, 2004 this sequence version replaced gi:51965771.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK79C1
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 229932 bases at least Q40
            Consensus quality: 230704 bases at least Q30
            Consensus quality: 231171 bases at least Q20
            Insert size: 232124; sum-of-contigs
            Insert size: 219964; 3.3% error; agarose-fp
            Quality coverage: 8.73x in Q20 bases; sum-of-contigs Quality
            coverage: 9.26x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    33972: contig of 33972 bp in length
            *    33973    34072: gap of 100 bp
            *    34073    37589: contig of 3517 bp in length
            *    37590    37689: gap of 100 bp
            *    37690    45659: contig of 7970 bp in length
            *    45660    45759: gap of 100 bp
            *    45760    54370: contig of 8611 bp in length
            *    54371    54470: gap of 100 bp
            *    54471    87298: contig of 32828 bp in length
            *    87299    87398: gap of 100 bp
            *    87399   111650: contig of 24252 bp in length
            *   111651   111750: gap of 100 bp
            *   111751   170131: contig of 58381 bp in length
            *   170132   170231: gap of 100 bp
            *   170232   232824: contig of 62593 bp in length.
FEATURES             Location/Qualifiers
     source          1. .232824
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-79C1"
                     /clone_lib="DanioKey"
     misc_feature    1. .33972
                     /note="assembly_fragment:00912
                     fragment_chain:1"
     misc_feature    34073. .37589
                     /note="assembly_fragment:00025
                     fragment_chain:1"
     misc_feature    37690. .45659
                     /note="assembly_fragment:00053
                     fragment_chain:1"
     misc_feature    45760. .54370
                     /note="assembly_fragment:00127
                     fragment_chain:1"
     misc_feature    54471. .87298
                     /note="assembly_fragment:00512
                     fragment_chain:2"
     misc_feature    87399. .111650
                     /note="assembly_fragment:00209
                     fragment_chain:2"
     misc_feature    111751. .170131
                     /note="assembly_fragment:01393"
     misc_feature    170232. .232824
                     /note="assembly_fragment:02188.0"
ORIGIN

  Query Match             8.5%;  Score 349.4;  DB 2;  Length 232824;
  Best Local Similarity   59.8%;  Pred. No. 6e-59;
  Matches  502;  Conservative    0;  Mismatches  336;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | |||||||||||||||||||||||||| | | | | | | |||||||||||| | |||||||
Db     169438 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 169497

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db     169498 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 169557

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db     169558 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 169617
```

```
Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db     169618 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 169677

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db     169678 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 169737

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db     169738 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 169797

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db     169798 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 169857

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db     169858 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 169917

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||          | |||          | |   ||    |    |||
Db     169918 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 169977

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
              |   | | | |  | | |    | |||    | |   | || |      |      | |
Db     169978 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 170037

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885
              ||       | ||| ||| || | ||   | |  ||| | ||   ||  ||      ||
Db     170038 TATATATATATATATATATATATATATATATATATATATATATATAATATATATTTATAT 170097

Qy        886 CACTATATATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAAT 945
                |||||||||||||   | |     || |
Db     170098 ATATATATATATATATATATATATACATATATNNNNNNNNNNNNNNNNNNNNNNNNNN 170157

Qy        946 GTTATAGAAGAATGGCTAAGATGCCAAAGCTCAGTGTATGTGGCAGAGGCATGGTGTAGG 1005

Db     170158 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 170217

Qy       1006 GTGTGTCCAGGTTCATATATTGCATTAAGTGTGAGAACACCCTGGAGTATGAACCAAGA 1064
                      ||| |||   | |     | ||   |||   || |||  | | |
Db     170218 NNNNNNNNNNNNNNATACATTTTGTCACTAAAAACAATGTCCTCTGGTGTGACACTATA 170276
```

RESULT 75
CR318636
LOCUS CR318636 233296 bp DNA linear HTG 31-AUG-2004
DEFINITION Danio rerio clone DKEY-249N13, *** SEQUENCING IN PROGRESS ***, 10 unordered pieces.
ACCESSION CR318636
VERSION CR318636.3 GI:51861411
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 233296)
AUTHORS Sims,S.
TITLE Direct Submission
JOURNAL Submitted (30-AUG-2004) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT On Sep 2, 2004 this sequence version replaced gi:44909414.
---------- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfish-help@sanger.ac.uk
---------- Project Information
Center project name: zK249N13
---------- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 229955 bases at least Q40
Consensus quality: 230652 bases at least Q30
Consensus quality: 231088 bases at least Q20
Insert size: 232396; sum-of-contigs
Insert size: 229559; 5.6% error; agarose-fp
Quality coverage: 7.04x in Q20 bases; sum-of-contigs Quality coverage: 7.26x in Q20 bases; agarose-fp
----------.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 2974: contig of 2974 bp in length
* 2975 3074: gap of 100 bp
* 3075 13370: contig of 10296 bp in length
* 13371 13470: gap of 100 bp
* 13471 23811: contig of 10341 bp in length
* 23812 23911: gap of 100 bp
* 23912 37932: contig of 14021 bp in length
* 37933 38032: gap of 100 bp
* 38033 54166: contig of 16134 bp in length
* 54167 54266: gap of 100 bp
* 54267 102290: contig of 48024 bp in length
* 102291 102390: gap of 100 bp
* 102391 125681: contig of 23291 bp in length
* 125682 125781: gap of 100 bp
* 125782 142646: contig of 16865 bp in length
* 142647 142746: gap of 100 bp
* 142747 165218: contig of 22472 bp in length
* 165219 165318: gap of 100 bp
* 165319 233296: contig of 67978 bp in length.

FEATURES Location/Qualifiers
source 1. .233296
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="DKEY-249N13"
/clone_lib="DanioKey"
misc_feature 1. .2974
/note="assembly_fragment:00024
fragment_chain:1"
misc_feature 3075. .13370
/note="assembly_fragment:00163
fragment_chain:1"
misc_feature 13471. .23811
/note="assembly_fragment:00045
fragment_chain:1"
misc_feature 23912. .37932
/note="assembly_fragment:00283
fragment_chain:2"
misc_feature 38033. .54166
/note="assembly_fragment:00602
fragment_chain:2"
misc_feature 54267. .102290
/note="assembly_fragment:01245
fragment_chain:2"
misc_feature 102391. .125681
/note="assembly_fragment:01003
fragment_chain:3"
misc_feature 125782. .142646
/note="assembly_fragment:00432
fragment_chain:3"
misc_feature 142747. .165218
/note="assembly_fragment:00787
fragment_chain:3"
misc_feature 165319. .233296

```
                     /note="assembly_fragment:01715.0"
ORIGIN

  Query Match             8.5%;  Score 349.4;  DB 2;  Length 233296;
  Best Local Similarity   69.7%;  Pred. No. 6e-59;
  Matches  487;  Conservative    0;  Mismatches  211;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||| | | | | | | |||||||||||| | |||||||
Db     164502 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 164561

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | | |
Db     164562 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 164621

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db     164622 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 164681

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db     164682 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 164741

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db     164742 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 164801

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db     164802 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 164861

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db     164862 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 164921

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db     164922 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 164981

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||||          | |||          | |   ||    |   |||
Db     164982 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 165041

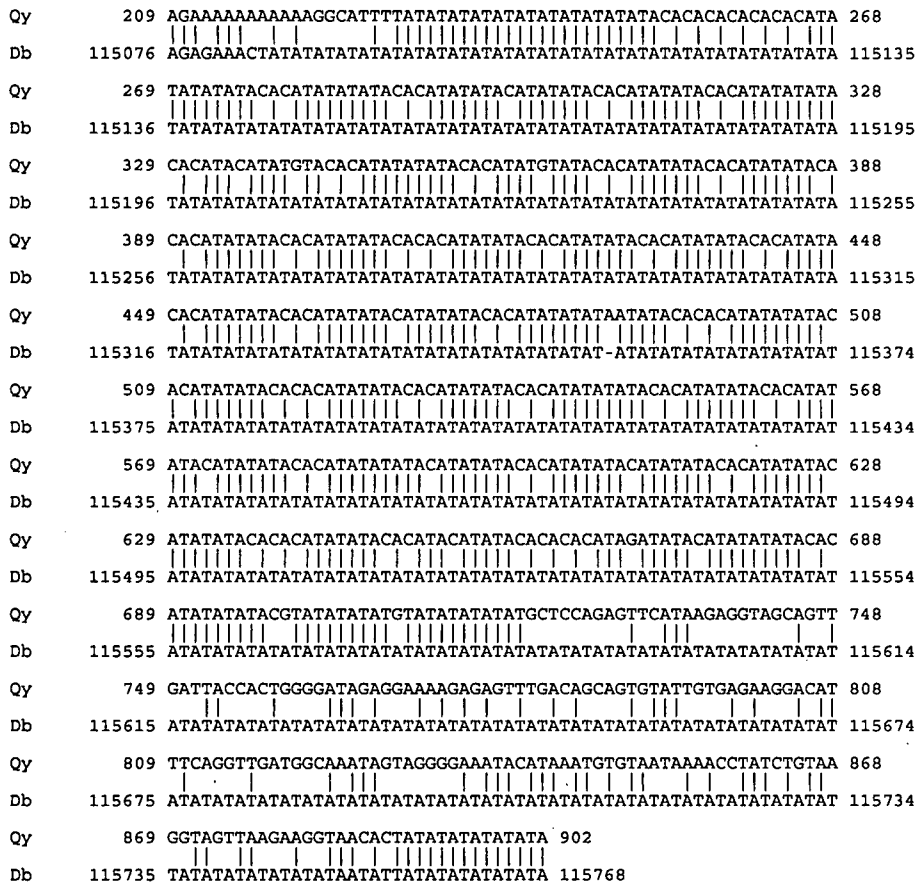
Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
              |   | | | |  |  | |    | |||    | |   | || |     |      | |
Db     165042 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 165101

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885
              |       | ||| ||| || | ||   | |  ||| | ||  ||  |||  |  |||
Db     165102 AATATATATATATATATATATATATATATATATATATATATATAATATATAAATATATAA 165161

Qy        886 CACTATATATATATATAGTGAAAGCAGTGTAAACCTAAA 924
              | ||||  ||||||||   | |  | | || |  ||||
Db     165162 AAATATAATTATATATATATATATAAATATATATATAAA 165200


RESULT 76
CR631122
LOCUS       CR631122              162977 bp    DNA     linear   HTG 13-AUG-2004
DEFINITION  Danio rerio clone DKEYP-36H11, *** SEQUENCING IN PROGRESS ***, 7
            unordered pieces.
ACCESSION   CR631122
VERSION     CR631122.2  GI:51241630
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 162977)
  AUTHORS   McLay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-AUG-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 13, 2004 this sequence version replaced gi:51127401.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zKp36H11
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 159934 bases at least Q40
            Consensus quality: 160525 bases at least Q30
            Consensus quality: 161089 bases at least Q20
            Insert size: 162377; sum-of-contigs
            Insert size: 159156; 13.8% error; agarose-fp
            Quality coverage: 7.39x in Q20 bases; sum-of-contigs Quality
            coverage: 7.54x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    11615: contig of 11615 bp in length
            *   11616    11715: gap of 100 bp
            *   11716    30875: contig of 19160 bp in length
            *   30876    30975: gap of 100 bp
            *   30976    47841: contig of 16866 bp in length
            *   47842    47941: gap of 100 bp
            *   47942   115881: contig of 67940 bp in length
            *  115882   115981: gap of 100 bp
            *  115982   134454: contig of 18473 bp in length
            *  134455   134554: gap of 100 bp
            *  134555   137147: contig of 2593 bp in length
            *  137148   137247: gap of 100 bp
            *  137248   162977: contig of 25730 bp in length.
FEATURES             Location/Qualifiers
     source          1. .162977
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEYP-36H11"
                     /clone_lib="DanioKeypilot"
     misc_feature    1. .11615
                     /note="assembly_fragment:00030
                     fragment_chain:1"
     misc_feature    11716. .30875
                     /note="assembly_fragment:00862
                     fragment_chain:1"
     misc_feature    30976. .47841
                     /note="assembly_fragment:00397
                     fragment_chain:1"
     misc_feature    47942. .115881
                     /note="assembly_fragment:01121
                     fragment_chain:1"
     misc_feature    115982. .134454
                     /note="assembly_fragment:00177
                     fragment_chain:1"
     misc_feature    134555. .137147
                     /note="assembly_fragment:00008"
     misc_feature    137248. .162977
                     /note="assembly_fragment:00618.0"
ORIGIN

  Query Match             8.5%;  Score 349.2;  DB 2;  Length 162977;
  Best Local Similarity   69.9%;  Pred. No. 6.9e-59;
  Matches  485;  Conservative    0;  Mismatches  208;  Indels    1;  Gaps    1;
```



```
Qy        209 AGAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACATA 268
Db     115076 AGAGAAACTATATATATATATATATATATATATATATATATATATATATATATATA 115135

Qy        269 TATATATACACATATATATACACATATATACATATATACACATATATACACATATATATA 328
Db     115136 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 115195

Qy        329 CACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACA 388
Db     115196 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 115255

Qy        389 CACATATATACACATATATACACACATATATACACATATATACACATATATACACATATA 448
Db     115256 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 115315

Qy        449 CACATATATACACATATATACATATATACACATATATATAATATACACACATATATATAC 508
Db     115316 TATATATATATATATATATATATATATATATATATATAT-ATATATATATATATATAT 115374

Qy        509 ACATATATACACACATATATACACATATATACACATATATATACACATATATACACATAT 568
Db     115375 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 115434

Qy        569 ATACATATATACACATATATATACATATATACACATATATACATATATACACATATATAC 628
Db     115435 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 115494

Qy        629 ATATATACACACATATATACACATACATATACACACACATAGATATACATATATATACAC 688
Db     115495 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 115554

Qy        689 ATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTT 748
Db     115555 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 115614

Qy        749 GATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACAT 808
Db     115615 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 115674

Qy        809 TTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAA 868
Db     115675 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 115734

Qy        869 GGTAGTTAAGAAGGTAACACTATATATATATATA 902
Db     115735 TATATATATATATAATATTATATATATATATA 115768
```

```
RESULT 77
AC021590/c
LOCUS       AC021590               171427 bp    DNA     linear   PRI 01-JAN-2002
DEFINITION  Homo sapiens chromosome 8, clone RP11-143H8, complete sequence.
ACCESSION   AC021590
VERSION     AC021590.10  GI:17432882
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 171427)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 8, clone RP11-143H8
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 171427)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 171427)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JAN-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Dec 10, 2001 this sequence version replaced gi:17402804.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L5777
                Center clone name: 143_H_8
            --------------.
FEATURES             Location/Qualifiers
     source          1. .171427
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="RP11-143H8"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   complement(2020. .2137)
                     /rpt_family="L2"
     repeat_region   complement(2160. .2295)
                     /rpt_family="L2"
     repeat_region   complement(2349. .2479)
                     /rpt_family="FLAM_C"
     repeat_region   complement(2483. .2835)
                     /rpt_family="L2"
     repeat_region   complement(3975. .4107)
                     /rpt_family="FLAM_C"
     repeat_region   complement(4639. .4747)
                     /rpt_family="L2"
```

```
repeat_region   complement(5060. .5878)
                /rpt_family="L1PA16"
repeat_region   complement(6034. .6096)
                /rpt_family="MIR"
repeat_region   6248. .6555
                /rpt_family="AluJb"
repeat_region   6716. .7017
                /rpt_family="L2"
repeat_region   7035. .7070
                /rpt_family="(CA)n"
repeat_region   7333. .7572
                /rpt_family="MIR"
repeat_region   7627. .7903
                /rpt_family="L2"
repeat_region   8277. .8484
                /rpt_family="AluJo"
repeat_region   8718. .8877
                /rpt_family="L3"
repeat_region   complement(9247. .9483)
                /rpt_family="MER46A"
repeat_region   10206. .10263
                /rpt_family="AT_rich"
repeat_region   10421. .10493
                /rpt_family="AT_rich"
repeat_region   complement(10891. .11060)
                /rpt_family="L2"
repeat_region   11256. .11305
                /rpt_family="AT_rich"
repeat_region   11809. .12090
                /rpt_family="AluSx"
repeat_region   12091. .12189
                /rpt_family="(TAGA)n"
repeat_region   12406. .12512
                /rpt_family="FRAM"
repeat_region   12513. .12539
                /rpt_family="AT_rich"
repeat_region   12566. .12755
                /rpt_family="MER61A"
repeat_region   12883. .18519
                /rpt_family="L1PA6"
repeat_region   18929. .18994
                /rpt_family="MIR"
repeat_region   20160. .20182
                /rpt_family="(TGGA)n"
repeat_region   complement(20466. .20667)
                /rpt_family="MIR"
repeat_region   20671. .20850
                /rpt_family="FAM"
repeat_region   20880. .20909
                /rpt_family="AT_rich"
repeat_region   complement(20910. .21037)
                /rpt_family="FRAM"
repeat_region   21121. .21419
                /rpt_family="AluJo"
repeat_region   complement(21424. .21508)
                /rpt_family="L1ME"
repeat_region   21934. .22046
                /rpt_family="MIR3"
repeat_region   23344. .23388
                /rpt_family="AT_rich"
repeat_region   23449. .23483
                /rpt_family="(TA)n"
repeat_region   complement(23484. .23779)
                /rpt_family="AluSx"
repeat_region   24093. .24190
                /rpt_family="L2"
repeat_region   24985. .25019
                /rpt_family="L1MA1"
repeat_region   25020. .25319
                /rpt_family="AluSx"
repeat_region   25320. .25557
                /rpt_family="L1MA1"
repeat_region   25558. .25581
                /rpt_family="AT_rich"
repeat_region   25777. .25802
                /rpt_family="AT_rich"
repeat_region   complement(26663. .26756)
                /rpt_family="L1ME4A"
repeat_region   26764. .26817
                /rpt_family="(TAAAAA)n"
repeat_region   27050. .27082
                /rpt_family="AT_rich"
repeat_region   complement(27455. .27626)
                /rpt_family="L1ME4A"
repeat_region   30048. .30108
                /rpt_family="AluJ/FRAM"
repeat_region   30249. .30269
                /rpt_family="AT_rich"
repeat_region   complement(30270. .30568)
                /rpt_family="AluSq"
repeat_region   31382. .31690
                /rpt_family="AluSx"
repeat_region   complement(32052. .32214)
                /rpt_family="L2"
repeat_region   32452. .32743
                /rpt_family="AluSc"
repeat_region   32839. .32969
                /rpt_family="L2"
repeat_region   complement(33055. .33190)
                /rpt_family="FLAM_C"
repeat_region   complement(33266. .33548)
                /rpt_family="L2"
repeat_region   33600. .33795
                /rpt_family="L2"
repeat_region   33939. .34117
                /rpt_family="L1MDa"

Query Match             8.5%;  Score 349.2;  DB 9;  Length 171427;
Best Local Similarity   65.2%;  Pred. No. 6.8e-59;
Matches  609;  Conservative    0;  Mismatches  313;  Indels   12;  Gaps    6;

Qy          1 AGTTCAAGACCAGCCTGAGCAACATGGTGAAACCCCATCTCTACTAAAAATAC--AAAAT 58
              || ||||||||| ||||    |||| |||||||||||| ||||||||||||||  |||||
Db      74773 AGATCAAGACCATCCTGGCTAACACGGTGAAACCCCGTCTCTACTAAAAATACAAAAAT 74714

Qy         59 TAGACAGGCATGGTGATACACGCCTGTAATCCCAGCTACTTCGGAGGCCGAGGCAGGAGA 118
              |||   |||||||||     ||||||| |||||||||| |||||| ||||||||||
Db      74713 TAGCAGGGCATGGTGGCGGGCGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGA 74654

Qy        119 ATCACTTGAACCTGCTGGGGGTGGAGGTTGCGGGGAG-CAAGATCATGCCATTGCACTCC 177
              ||  |  |||  | | || || |||| |||| | ||| | ||||   |||| ||||||||
Db      74653 ATGGCGGGAA--TCCGGGAGGCGGAGCTTGCAGTGAGCCGAGATAGCGCCACTGCACTCC 74596

Qy        178 AGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAGGCATTT-----TATA 232
              ||||  ||  |||       | | |       | ||||||||  ||  |||  |     || |
Db      74595 AGCCTGGGGGACAGAGCGAGACTCTGTCTCAAAAAAAAAAAGAAAACATATAAATATAAA 74536

Qy        233 TATATATATATATATATATACACACACACACATATATATATACACATATATATACACA 292
              ||| |||||| ||||||||| |||   | | |||| |||||||   | |||||  ||| |||
Db      74535 TATGTATATAAATATATATATACATTTATATACATTTATATATTTATATATAAATATACA 74476

Qy        293 TATATACATATATACACATATATACACATATATATACACATACATATGTACACATATATA 352
              ||||||  | ||||   ||||||    ||||||  || | ||   |||  ||   ||||||
Db      74475 TATATATGTGTATATTTATATATGTTTATATATTTATATATGTTTATATATTTATATATG 74416

Qy        353 TACACATATGTATACACATATATACACATATATACACACATATATACACATATATACACA 412
              |  | |||| |||| | ||||  | | || | || | |  |||||| | ||||||| | |
Db      74415 TTTATATATTTATATATATATTAATATATGTTTATATATTTATATATAAATATATATAAA 74356

Qy        413 CATATATACACATATATACACATATATACACATATACACATATATACACATATATACATA 472
              ||| |||| | | ||||| | ||||||| | | ||  | ||||| | | ||| ||| |||
Db      74355 CATTTATATATAAATATATAAATATATATAAACATTTATATATAAATATATAAATATATA 74296

Qy        473 TATACACATATATAT-AATATACACACATATATATACACATATATACACACATATATACA 531
              || |||  ||||||| |||||| | | ||||||| |||  |||||| | | ||||| | |
Db      74295 TAAACATTTATATATAAATATATAAATATATATAAACATTTATATATAAATATATAAATA 74236
```

```
Qy        532 CATATATACACATATATATACACATATATACACATATATACATATATACACATATATATA 591
              ||||| |||  |||||||| | ||||||| |||| |||| ||| ||| | | |||||||
Db      74235 TATATAAACATTTATATATAAATATATATAAACATTTATATATAAATATATAAATATATA 74176

Qy        592 CATATATACACATATATACATATATACACATATATACATATATACACACATATATACACA 651
              | | ||  | ||||| | ||||| | | ||| | |  |||||| | | ||||| | | |
Db      74175 TAAACATTTATATATAAATATATAAATATATAAACATTTATATATAAATATATAAATATA 74116

Qy        652 TACATATACACACACATAGATATACATATATATACACATATATATACGTATATATATGTA 711
              || | ||  | | | | | ||||| ||||||| |||  |||||||  ||||||||   |
Db      74115 TAAACATTTATATATAAATATATAAATATATAAACATTTATATATAAATATATATAAACA 74056

Qy        712 TATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAA 771
              | |||||||         || |||         |  | | |    |    | | | |
Db      74055 TTTATATATAAATATATATTTATATATAAATATATATAAACATTTATATATAAATATATA 73996

Qy        772 AAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAG 831
              || | |  || | |    ||||    | |   | ||||     |      | |||
Db      73995 AATATATATTTATATATACGTATATATAAATATATATTTATATATACGTATATATAAATA 73936

Qy        832 GGGAAATACATAAATGTGTAATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTAACACTA 890
              | ||  ||||| | ||  || |  ||| | || | |||   |   |  |||   ||
Db      73935 TATATATGTATAAACATATATAAATATATATTTATATATGTATAAACATATATAAATATA 73876

Qy        891 TATATATATATAGTGAAAGCAGTGTAAACCTAAA 924
              ||| |||||||    |||    | ||||  || |
Db      73875 TATTTATATATGTATAAACATATATAAATATATA 73842


RESULT 78
CR759788/c
LOCUS       CR759788              195311 bp    DNA     linear   HTG 07-SEP-2004
DEFINITION  Danio rerio clone CH211-153F13, *** SEQUENCING IN PROGRESS ***, 11
            unordered pieces.
ACCESSION   CR759788
VERSION     CR759788.3  GI:51950464
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 195311)
  AUTHORS   McLay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 9, 2004 this sequence version replaced gi:51890035.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zC153F13
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 191941 bases at least Q40
            Consensus quality: 192830 bases at least Q30
            Consensus quality: 193638 bases at least Q20
            Insert size: 194311; sum-of-contigs
            Insert size: 184045; 2.4% error; agarose-fp
            Quality coverage: 6.41x in Q20 bases; sum-of-contigs Quality
            coverage: 6.89x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     4825: contig of 4825 bp in length
            *     4826     4925: gap of 100 bp
            *     4926    14966: contig of 10041 bp in length
            *    14967    15066: gap of 100 bp
            *    15067    17900: contig of 2834 bp in length
            *    17901    18000: gap of 100 bp
            *    18001    22764: contig of 4764 bp in length
            *    22765    22864: gap of 100 bp
            *    22865    40985: contig of 18121 bp in length
            *    40986    41085: gap of 100 bp
            *    41086    65693: contig of 24608 bp in length
            *    65694    65793: gap of 100 bp
            *    65794   118569: contig of 52776 bp in length
            *   118570   118669: gap of 100 bp
            *   118670   131148: contig of 12479 bp in length
            *   131149   131248: gap of 100 bp
            *   131249   172432: contig of 41184 bp in length
            *   172433   172532: gap of 100 bp
            *   172533   178335: contig of 5803 bp in length
            *   178336   178435: gap of 100 bp
            *   178436   195311: contig of 16876 bp in length.
FEATURES             Location/Qualifiers
     source          1. .195311
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-153F13"
                     /clone_lib="CHORI-211"
     misc_feature    1. .4825
                     /note="assembly_fragment:00162
                     fragment_chain:1
                     clone_end:SP6
                     vector_side:left"
     misc_feature    4926. .14966
                     /note="assembly_fragment:00231
                     fragment_chain:1"
     misc_feature    15067. .17900
                     /note="assembly_fragment:00038
                     fragment_chain:1"
     misc_feature    18001. .22764
                     /note="assembly_fragment:00112
                     fragment_chain:1"
     misc_feature    22865. .40985
                     /note="assembly_fragment:00460
                     fragment_chain:2"
     misc_feature    41086. .65693
                     /note="assembly_fragment:00820
                     fragment_chain:2"
     misc_feature    65794. .118569
                     /note="assembly_fragment:01583
                     fragment_chain:2"
     misc_feature    118670. .131148
                     /note="assembly_fragment:00325
                     fragment_chain:2"
     misc_feature    131249. .172432
                     /note="assembly_fragment:01060
                     fragment_chain:2"
     misc_feature    172533. .178335
                     /note="assembly_fragment:00062
                     fragment_chain:2"
     misc_feature    178436. .195311
                     /note="assembly_fragment:00627
                     fragment_chain:2"
ORIGIN

  Query Match             8.5%;  Score 349.2;  DB 2;  Length 195311;
  Best Local Similarity   63.3%;  Pred. No. 6.7e-59;
  Matches  475;  Conservative    0;  Mismatches  273;  Indels    2;  Gaps    2;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
```

```
Db     178944 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 178885

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db     178884 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 178825

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db     178824 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 178765

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db     178764 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 178705

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db     178704 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 178645

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db     178644 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 178585

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db     178584 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 178525

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | ||||||||  ||||||
Db     178524 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 178465

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||||          |  ||
Db     178464 TATATATATATATATATATATATATATNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 178405

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825

Db     178404 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 178345

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTA 884
                  | ||| ||| || | ||   | |  ||| | || |  ||  ||    |  ||
Db     178344 NNNNNNNNNATATATATATATATATATATATATATATATATATATATATATATATATATA 178285

Qy        885 ACACTATATATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAA 944
                ||||||||||||||   | |    | || |  ||      | |   ||    ||
Db     178284 TATATATATATATATATATATATATATATATATATATTCATAGATCTGTGGCGAGAAT 178225

Qy        945 TGTTATAGAAGAATGGCTAAGATGCCAAAG 974
              ||  | | |||   |  |||   |  ||||
Db     178224 TGCAAGAAAAGTTAGAATAACTGGTTAAAG 178195


RESULT 79
CR762470
LOCUS       CR762470             199230 bp    DNA     linear   HTG 16-SEP-2004
DEFINITION  Danio rerio clone DKEY-12N16, *** SEQUENCING IN PROGRESS ***, 15
            unordered pieces.
ACCESSION   CR762470
VERSION     CR762470.2  GI:52213994
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 199230)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 16, 2004 this sequence version replaced gi:52000316.
            ---------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            ---------- Project Information
            Center project name: zK12N16
            ---------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 192957 bases at least Q40
            Consensus quality: 193980 bases at least Q30
            Consensus quality: 194701 bases at least Q20
            Insert size: 197830; sum-of-contigs
            Insert size: 204603; 6.7% error; agarose-fp
            Quality coverage: 9.51x in Q20 bases; sum-of-contigs Quality
            coverage: 9.54x in Q20 bases; agarose-fp
            ----------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 15 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    22618: contig of 22618 bp in length
            *    22619    22718: gap of 100 bp
            *    22719    28510: contig of 5792 bp in length
            *    28511    28610: gap of 100 bp
            *    28611    61395: contig of 32785 bp in length
            *    61396    61495: gap of 100 bp
            *    61496    65559: contig of 4064 bp in length
            *    65560    65659: gap of 100 bp
            *    65660    72529: contig of 6870 bp in length
            *    72530    72629: gap of 100 bp
            *    72630    92239: contig of 19610 bp in length
            *    92240    92339: gap of 100 bp
            *    92340    98985: contig of 6646 bp in length
            *    98986    99085: gap of 100 bp
            *    99086   146980: contig of 47895 bp in length
            *   146981   147080: gap of 100 bp
            *   147081   152421: contig of 5341 bp in length
            *   152422   152521: gap of 100 bp
            *   152522   157138: contig of 4617 bp in length
            *   157139   157238: gap of 100 bp
            *   157239   167850: contig of 10612 bp in length
            *   167851   167950: gap of 100 bp
            *   167951   170933: contig of 2983 bp in length
            *   170934   171033: gap of 100 bp
            *   171034   186729: contig of 15696 bp in length
            *   186730   186829: gap of 100 bp
            *   186830   195213: contig of 8384 bp in length
            *   195214   195313: gap of 100 bp
            *   195314   199230: contig of 3917 bp in length.
FEATURES             Location/Qualifiers
     source          1. .199230
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-12N16"
                     /clone_lib="DanioKey"
     misc_feature    1. .22618
                     /note="assembly_fragment:01219
                     fragment_chain:1"
     misc_feature    22719. .28510
                     /note="assembly_fragment:00267
                     fragment_chain:1"
     misc_feature    28611. .61395
                     /note="assembly_fragment:01636
                     fragment_chain:1"
     misc_feature    61496. .65559
                     /note="assembly_fragment:00116
                     fragment_chain:1"
```

```
     misc_feature    65660. .72529
                     /note="assembly_fragment:00202
                     fragment_chain:1"
     misc_feature    72630. .92239
                     /note="assembly_fragment:00921
                     fragment_chain:2"
     misc_feature    92340. .98985
                     /note="assembly_fragment:00329
                     fragment_chain:2"
     misc_feature    99086. .146980
                     /note="assembly_fragment:02193
                     fragment_chain:2"
     misc_feature    147081. .152421
                     /note="assembly_fragment:00157
                     fragment_chain:3"
     misc_feature    152522. .157138
                     /note="assembly_fragment:00407
                     fragment_chain:3"
     misc_feature    157239. .167850
                     /note="assembly_fragment:00593
                     fragment_chain:4"
     misc_feature    167951. .170933
                     /note="assembly_fragment:00070
                     fragment_chain:4"
     misc_feature    171034. .186729
                     /note="assembly_fragment:00727
                     fragment_chain:5"
     misc_feature    186830. .195213
                     /note="assembly_fragment:00491
                     fragment_chain:5"
     misc_feature    195314. .199230
                     /note="assembly_fragment:00081.0"
ORIGIN

  Query Match             8.5%;  Score 349.2;  DB 2;  Length 199230;
  Best Local Similarity   68.5%;  Pred. No. 6.7e-59;
  Matches  511;  Conservative    0;  Mismatches  233;  Indels    2;  Gaps    2;

Qy        181 CCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTTTATATATATATA 240
              |||| ||| || |   |||  | |  | |  ||| ||| ||   | |  ||||||||||
Db      26837 CCAGACAAAAAAAAAAAAAAATTAAATAAATAAATAAATAAATAAATAAATATATATATA 26896

Qy        241 TATATATATATACACACACACACATATATATATACACATATATATACACATATATACA 300
              |||||||||||| | | | | | ||||||||||| | |||||||| | ||||||| |
Db      26897 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 26956

Qy        301 TATATACACATATATACACATATATATACACATACATATGTACACATATATATACACATA 360
              |||||| | ||||||| | |||||||| | ||| |||| || | |||||||| | |||
Db      26957 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 27016

Qy        361 TGTATACACATATATACACATATATACACACATATATACACATATATACACACATATATA 420
              | |||| | ||||||| | ||||||| | | ||||||| | ||||||| | | |||||||
Db      27017 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 27076

Qy        421 CACATATATACACATATATACACATATACACATATATACACATATATACATATATACACA 480
              | ||||||| | ||||||| | ||||| | ||||||| | ||||||| ||||||| | |
Db      27077 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 27136

Qy        481 TATATATA-ATATACACACATATATATACACATATATACACACATATATACACATATATA 539
              |||||||| ||||| | | ||||||||| | ||||||| | | ||||||| | |||||||
Db      27137 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 27196

Qy        540 CACATATATATACACATATATACACATATATACATATATACACATATATATACATATATA 599
              | ||||||||| | ||||||| | ||||||| ||||||| | ||||||||| |||||||
Db      27197 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 27256

Qy        600 CACATATATACATATATACACATATATACATATATACACACATATATACACATACATATA 659
              | ||||||| ||||||| | ||||||| ||||||| | | ||||||| | ||| |||||
Db      27257 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 27316

Qy        660 CACACACATAGATATACATATATATACACATATATATACGTATATATATGTATATATATA 719
              | | | ||| ||||| ||||||||| | |||||||||  ||||||||| ||||||||||
Db      27317 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 27376

Qy        720 TGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGT 779
              |          | |||          | |   ||     |    ||| |   | | | |
Db      27377 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 27436

Qy        780 TTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATA 839
              | | |    | |||    | |   | || |      |        | |||        | |||
Db      27437 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 27496

Qy        840 CATAAATGTGTAATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTAACACTATATATATA 898
              |||| || | ||   | |  ||| | || |  ||  ||   |  ||    ||||||||||
Db      27497 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 27556

Qy        899 TATAGTGAAAGCAGTGTAAACCTAAA 924
              ||||   | |    | || |  || |
Db      27557 TATATATATATATATATATATATA 27582


RESULT 80
AC098061/c
LOCUS       AC098061               247615 bp    DNA     linear   HTG 10-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-115G17, WORKING DRAFT SEQUENCE, 2
            unordered pieces.
ACCESSION   AC098061
VERSION     AC098061.7  GI:30522584
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 247615)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
```

```
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 247615)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-OCT-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 247615)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 10, 2003 this sequence version replaced gi:23096179.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GGEZ
                Center clone name: CH230-115G17
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 219440 bases at least Q40
                Consensus quality: 221621 bases at least Q30
                Consensus quality: 223241 bases at least Q20
                Estimated insert size: 234299; sum-of-contigs estimation
                Quality coverage: 8x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     9721: contig of 9721 bp in length
            *    9722     9821: gap of unknown length
            *    9822   247615: contig of 237794 bp in length.
FEATURES              Location/Qualifiers
     source           1. .247615
                      /organism="Rattus norvegicus"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:10116"
                      /clone="CH230-115G17"
ORIGIN

  Query Match             8.5%;  Score 349.2;  DB 2;  Length 247615;
  Best Local Similarity   67.1%;  Pred. No. 6.5e-59;
  Matches  523;  Conservative    0;  Mismatches  254;  Indels    2;  Gaps    2;

Qy        227 TTTATATATATATATATATATATATACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||||| | | | | | |||||||||| | |||||||
Db       9080 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 9021

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | |||||||||| | ||| |||| || | |
Db       9020 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 8961

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db       8960 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 8901

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db       8900 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 8841

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | |||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db       8840 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 8781

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db       8780 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 8721

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db       8720 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 8661

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db       8660 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 8601

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||          | |||          | |  |  ||   |  |||
Db       8600 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 8541

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
              |  | | | |  | | |  | |||  | |  | || |  |  |  | |
Db       8540 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 8481

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTA 884
              ||       | ||| ||| || | ||   | |  ||| | || |  ||  ||   |  ||
Db       8480 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 8421

Qy        885 ACACTATATATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAA 944
              |||||||||||||||  | ||  | || |  | | ||      ||  || | ||  | |
Db       8420 TATATATATATATATATAGANAGAGATATATATATAGATATGAATATATAAAAATGTAGA 8361

Qy        945 TGTTATAGAAGAATGGCTAAGATGCCAAAGCTCAGTGTATGTGGCAGAGGCATGGTGTA 1003
              | | | || |  || | | |  || ||  |   |      || || | || | | |
Db       8360 GATAAAAAAACAGAGGTTTAAAGGGGAAGGCATGGATCAATGACAAGGGGAAAGGGGGA 8302


RESULT 81
CR759820/c
LOCUS       CR759820              162435 bp    DNA     linear   HTG 10-SEP-2004
DEFINITION  Danio rerio clone DKEY-208C7, *** SEQUENCING IN PROGRESS ***, 7
            unordered pieces.
ACCESSION   CR759820
VERSION     CR759820.2  GI:51987786
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 162435)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
```

```
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 10, 2004 this sequence version replaced gi:51890335.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK208C7
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 159742 bases at least Q40
            Consensus quality: 160382 bases at least Q30
            Consensus quality: 160932 bases at least Q20
            Insert size: 161835; sum-of-contigs
            Insert size: 166033; 3.0% error; agarose-fp
            Quality coverage: 10.96x in Q20 bases; sum-of-contigs Quality
            coverage: 10.78x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     9773: contig of 9773 bp in length
            *     9774     9873: gap of 100 bp
            *     9874    59608: contig of 49735 bp in length
            *    59609    59708: gap of 100 bp
            *    59709    64998: contig of 5290 bp in length
            *    64999    65098: gap of 100 bp
            *    65099    70537: contig of 5439 bp in length
            *    70538    70637: gap of 100 bp
            *    70638    90121: contig of 19484 bp in length
            *    90122    90221: gap of 100 bp
            *    90222    92310: contig of 2089 bp in length
            *    92311    92410: gap of 100 bp
            *    92411   162435: contig of 70025 bp in length.
FEATURES            Location/Qualifiers
     source         1. .162435
                    /organism="Danio rerio"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:7955"
                    /clone="DKEY-208C7"
                    /clone_lib="DanioKey"
     misc_feature   1. .9773
                    /note="assembly_fragment:00141
                    fragment_chain:1"
     misc_feature   9874. .59608
                    /note="assembly_fragment:00514
                    fragment_chain:1"
     misc_feature   59709. .64998
                    /note="assembly_fragment:00049
                    fragment_chain:1"
     misc_feature   65099. .70537
                    /note="assembly_fragment:00086
                    fragment_chain:2"
     misc_feature   70638. .90121
                    /note="assembly_fragment:00272
                    fragment_chain:2"
     misc_feature   90222. .92310
                    /note="assembly_fragment:00018.0"
     misc_feature   92411. .162435
                    /note="assembly_fragment:01362"
ORIGIN

  Query Match             8.5%;  Score 349;  DB 2;  Length 162435;
  Best Local Similarity   68.7%;  Pred. No. 7.5e-59;
  Matches  509;  Conservative    0;  Mismatches  230;  Indels    2;  Gaps    2;
```

```
Qy        209 AGAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACATA 268
              |||||||||||    | | |||||||||||||||||||||||| | | | | | |||
Db        850 AAAAAAAAAAAAAAAAAATATATATATATATATATATATATATATATATATATATA 791

Qy        269 TATATATACACATATATATACACATATATACATATATACACATATATACACATATATATA 328
              |||||||| | ||||||||| | ||||||| ||||||| | ||||||| | |||||||||
Db        790 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 731

Qy        329 CACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACA 388
              | ||| |||| || | ||||||||| | |||| |||| | ||||||| | ||||||| |
Db        730 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 671

Qy        389 CACATATATACACATATATACACACATATATACACATATATACACATATATACACATATA 448
              | ||||||| | ||||||| | | ||||||| | ||||||| | ||||||| | |||||
Db        670 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 611

Qy        449 CACATATATACACATATATACATATATACACATATATATA-ATATACACACATATATATA 507
              | ||||||| | ||||||| ||||||| | ||||||||| ||||| | | |||||||||
Db        610 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 551

Qy        508 CACATATATACACACATATATACACATATATACACATATATATACACATATATACACATA 567
              | ||||||| | | ||||||| | ||||||| | ||||||||| | ||||||| | |||
Db        550 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 491

Qy        568 TATACATATATACACATATATATACATATATACACATATATACATATATACACATATATA 627
              |||| ||||||| | ||||||||| ||||||| | ||||||| ||||||| | |||||||
Db        490 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 431

Qy        628 CATATATACACACATATATACACATACATATACACACACATAGATATACATATATATACA 687
              ||||||| | | ||||||| | ||| ||||| | | | ||| ||||| |||||||||
Db        430 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 371

Qy        688 CATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGT 747
              |||||||||  |||||||||| ||||||||||          |  |||         |
Db        370 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 311

Qy        748 TGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACA 807
              |   ||    |    ||| |   | | | |  | | |    | |||    | |   | |
Db        310 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 251

Qy        808 TTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA 867
              | |    |    | |||       | ||| ||| || | ||   | |  ||| | ||
Db        250 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 191

Qy        868 -AGGTAGTTAAGAAGGTAACACTATATATATATATAGTGAAAGCAGTGTAAACCTAAAGG 926
               |  ||  ||   |  ||    |||||||||||||   | |    | || |  || |
Db        190 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 131

Qy        927 ATGGGCCAAGGATTTAAATGT 947
                |  |  ||||| || |
Db        130 TATAGAGATATATTTATATAT 110


RESULT 82
CR788258
LOCUS       CR788258              165992 bp    DNA     linear   HTG 25-SEP-2004
DEFINITION  Danio rerio clone DKEY-182D21, *** SEQUENCING IN PROGRESS ***, 10
            unordered pieces.
ACCESSION   CR788258
VERSION     CR788258.3  GI:52694485
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 165992)
  AUTHORS   McLay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
```

```
COMMENT     On Sep 25, 2004 this sequence version replaced gi:52353217.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK182D21
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 162106 bases at least Q40
            Consensus quality: 163025 bases at least Q30
            Consensus quality: 163727 bases at least Q20
            Insert size: 165092; sum-of-contigs
            Insert size: 165748; 0.8% error; agarose-fp
            Quality coverage: 7.00x in Q20 bases; sum-of-contigs Quality
            coverage: 7.22x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    12528: contig of 12528 bp in length
            *    12529    12628: gap of 100 bp
            *    12629    31404: contig of 18776 bp in length
            *    31405    31504: gap of 100 bp
            *    31505    37666: contig of 6162 bp in length
            *    37667    37766: gap of 100 bp
            *    37767    62052: contig of 24286 bp in length
            *    62053    62152: gap of 100 bp
            *    62153    79002: contig of 16850 bp in length
            *    79003    79102: gap of 100 bp
            *    79103   113175: contig of 34073 bp in length
            *   113176   113275: gap of 100 bp
            *   113276   122334: contig of 9059 bp in length
            *   122335   122434: gap of 100 bp
            *   122435   137396: contig of 14962 bp in length
            *   137397   137496: gap of 100 bp
            *   137497   147411: contig of 9915 bp in length
            *   147412   147511: gap of 100 bp
            *   147512   165992: contig of 18481 bp in length.
FEATURES             Location/Qualifiers
     source          1. .165992
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-182D21"
                     /clone_lib="DanioKey"
     misc_feature    1. .12528
                     /note="assembly_fragment:00331
                     fragment_chain:1"
     misc_feature    12629. .31404
                     /note="assembly_fragment:00639
                     fragment_chain:1"
     misc_feature    31505. .37666
                     /note="assembly_fragment:00044
                     fragment_chain:1"
     misc_feature    37767. .62052
                     /note="assembly_fragment:01292
                     fragment_chain:2"
     misc_feature    62153. .79002
                     /note="assembly_fragment:00843
                     fragment_chain:2"
     misc_feature    79103. .113175
                     /note="assembly_fragment:01612
                     fragment_chain:2"
     misc_feature    113276. .122334
                     /note="assembly_fragment:00114
                     fragment_chain:3"
     misc_feature    122435. .137396
                     /note="assembly_fragment:00465
                     fragment_chain:3"
     misc_feature    137497. .147411
                     /note="assembly_fragment:00219"
     misc_feature    147512. .165992
                     /note="assembly_fragment:01055.0"
ORIGIN

  Query Match             8.5%;  Score 349;  DB 2;  Length 165992;
  Best Local Similarity   63.5%;  Pred. No. 7.5e-59;
  Matches  455;  Conservative    0;  Mismatches  261;  Indels    1;  Gaps    1;

Qy        209 AGAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACACATA 268
              |||||||||||      |||||||||||||||||||||||| | | | | | |||
Db      11982 AAAAAAAAAAAAAAAATATATATATATATATATATATATATATATATATATATATA 12041

Qy        269 TATATATACACATATATATACACATATATACATATATACACATATATACACATATATATA 328
              |||||||| | ||||||||| | ||||||| ||||||| | ||||||| | |||||||||
Db      12042 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 12101

Qy        329 CACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACA 388
              | ||| |||| || | ||||||||| | |||| |||| | ||||||| | ||||||| |
Db      12102 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 12161

Qy        389 CACATATATACACATATATACACACATATATACACATATATACACATATATACACATATA 448
              | ||||||| | ||||||| | | ||||||| | ||||||| | ||||||| | |||||
Db      12162 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 12221

Qy        449 CACATATATACACATATATACATATATACACATATATATA-ATATACACACATATATATA 507
              | ||||||| | ||||||| ||||||| | ||||||||| ||||| | | |||||||||
Db      12222 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 12281

Qy        508 CACATATATACACACATATATACACATATATACACATATATATACACATATATACACATA 567
              | ||||||| | | ||||||| | ||||||| | ||||||||| | ||||||| | |||
Db      12282 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 12341

Qy        568 TATACATATATACACATATATATACATATATACACATATATACATATATACACATATATA 627
              |||| ||||||| | ||||||||| ||||||| | ||||||| ||||||| | |||||||
Db      12342 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 12401

Qy        628 CATATATACACACATATATACACATACATATACACACACATAGATATACATATATATACA 687
              ||||||| | | ||||||| | ||| ||||| | | | ||| ||||| |||||||||
Db      12402 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 12461

Qy        688 CATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGT 747
              |||||||||  |||||||||| ||||||||||          |  |||         |
Db      12462 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 12521

Qy        748 TGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACA 807
              |   ||
Db      12522 TATATATNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 12581

Qy        808 TTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA 867
                                                          |  |    ||
Db      12582 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTATATTTATATTA 12641

Qy        868 AGGTAGTTAAGAAGGTAACACTATATATATATAGTGAAAGCAGTGTAAACCTAAA 924
              |  ||  ||      |||   |||||||||||||   | |    | || |  || |
Db      12642 ATATAAATATTTTTTTAATTATATATATATATATATATATATATATATATATA 12698


RESULT 83
CR769782
LOCUS       CR769782              248767 bp    DNA     linear   HTG 27-SEP-2004
DEFINITION  Danio rerio clone DKEY-41C6, *** SEQUENCING IN PROGRESS ***, 16
            unordered pieces.
ACCESSION   CR769782
VERSION     CR769782.4  GI:52748693
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
```

```
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
          Cypriniformes; Cyprinidae; Danio.
REFERENCE 1  (bases 1 to 248767)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
          Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
          zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT   On Sep 27, 2004 this sequence version replaced gi:52694483.
          -------------- Genome Center
          Center: Wellcome Trust Sanger Institute
          Center code: SC
          Web site: http://www.sanger.ac.uk
          Contact: zfish-help@sanger.ac.uk
          -------------- Project Information
          Center project name: zK41C6
          -------------- Summary Statistics
          Assembly program: XGAP4; version 4.5
          Chemistry: Dye-terminator; 100% of reads
          Consensus quality: 243682 bases at least Q40
          Consensus quality: 244936 bases at least Q30
          Consensus quality: 245625 bases at least Q20
          Insert size: 247267; sum-of-contigs
          Insert size: 236837; 3.1% error; agarose-fp
          Quality coverage: 8.86x in Q20 bases; sum-of-contigs Quality
          coverage: 9.25x in Q20 bases; agarose-fp
          --------------.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 16 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *      1    10635: contig of 10635 bp in length
          *  10636    10735: gap of 100 bp
          *  10736    54062: contig of 43327 bp in length
          *  54063    54162: gap of 100 bp
          *  54163    65940: contig of 11778 bp in length
          *  65941    66040: gap of 100 bp
          *  66041   115833: contig of 49793 bp in length
          * 115834   115933: gap of 100 bp
          * 115934   161746: contig of 45813 bp in length
          * 161747   161846: gap of 100 bp
          * 161847   167273: contig of 5427 bp in length
          * 167274   167373: gap of 100 bp
          * 167374   174818: contig of 7445 bp in length
          * 174819   174918: gap of 100 bp
          * 174919   184757: contig of 9839 bp in length
          * 184758   184857: gap of 100 bp
          * 184858   186977: contig of 2120 bp in length
          * 186978   187077: gap of 100 bp
          * 187078   202458: contig of 15381 bp in length
          * 202459   202558: gap of 100 bp
          * 202559   212782: contig of 10224 bp in length
          * 212783   212882: gap of 100 bp
          * 212883   215321: contig of 2439 bp in length
          * 215322   215421: gap of 100 bp
          * 215422   220884: contig of 5463 bp in length
          * 220885   220984: gap of 100 bp
          * 220985   237401: contig of 16417 bp in length
          * 237402   237501: gap of 100 bp
          * 237502   241277: contig of 3776 bp in length
          * 241278   241377: gap of 100 bp
          * 241378   248767: contig of 7390 bp in length.
FEATURES             Location/Qualifiers
     source          1. .248767
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-41C6"
                     /clone_lib="DanioKey"
     misc_feature    1. .10635
                     /note="assembly_fragment:00860
                     fragment_chain:1"
     misc_feature    10736. .54062
                     /note="assembly_fragment:02081
                     fragment_chain:1"
     misc_feature    54163. .65940
                     /note="assembly_fragment:00547
                     fragment_chain:1"
     misc_feature    66041. .115833
                     /note="assembly_fragment:02702
                     fragment_chain:1"
     misc_feature    115934. .161746
                     /note="assembly_fragment:01463
                     fragment_chain:1"
     misc_feature    161847. .167273
                     /note="assembly_fragment:00149
                     fragment_chain:1"
     misc_feature    167374. .174818
                     /note="assembly_fragment:00212
                     fragment_chain:1"
     misc_feature    174919. .184757
                     /note="assembly_fragment:00417
                     fragment_chain:2"
     misc_feature    184858. .186977
                     /note="assembly_fragment:00013
                     fragment_chain:2"
     misc_feature    187078. .202458
                     /note="assembly_fragment:01020
                     fragment_chain:2"
     misc_feature    202559. .212782
                     /note="assembly_fragment:00703
                     fragment_chain:2"
     misc_feature    212883. .215321
                     /note="assembly_fragment:00030
                     fragment_chain:2"
     misc_feature    215422. .220884
                     /note="assembly_fragment:00089
                     fragment_chain:3"
     misc_feature    220985. .237401
                     /note="assembly_fragment:01224
                     fragment_chain:3"
     misc_feature    237502. .241277
                     /note="assembly_fragment:00048.0"
     misc_feature    241378. .248767
                     /note="assembly_fragment:00309"
ORIGIN

  Query Match             8.5%;  Score 349;  DB 2;  Length 248767;
  Best Local Similarity   68.3%;  Pred. No. 7.1e-59;
  Matches  498;  Conservative    0;  Mismatches  230;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db      53249 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 53308

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db      53309 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 53368

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db      53369 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 53428

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db      53429 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 53488

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db      53489 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 53548
```

```
Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db      53549 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 53608

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db      53609 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 53668

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db      53669 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 53728

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||||          | |||          | |   ||    |   |||
Db      53729 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 53788

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
              |   | | | |  |  | |    | |||    | |   | || |     |       | |
Db      53789 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 53848

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885
              ||       | ||| ||| ||   ||   | |  ||| | ||   ||  ||   |  ||
Db      53849 TATATATATATATATATATATACATATATATATATATATATAATAAATATATATATAT 53908

Qy        886 CACTATATATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAAT 945
                |||||||||||||   | |  | | |||   ||||  ||      |    |  |||
Db      53909 ATATATATATATATATATATATATAAATATAATAATAAATAATATAAATATTTATATAAT 53968

Qy        946 GTTATAGAA 954
              |||   ||
Db      53969 ATTAATTAA 53977
```

```
RESULT 84
CR812467
LOCUS       CR812467               326441 bp    DNA     linear   HTG 27-SEP-2004
DEFINITION  Danio rerio clone CH211-134M19, *** SEQUENCING IN PROGRESS ***, 23
            unordered pieces.
ACCESSION   CR812467
VERSION     CR812467.2  GI:52748666
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 326441)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 27, 2004 this sequence version replaced gi:52694423.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zC134M19
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 317403 bases at least Q40
            Consensus quality: 319890 bases at least Q30
            Consensus quality: 321680 bases at least Q20
            Insert size: 324241; sum-of-contigs
            Insert size: 167531; 3.4% error; agarose-fp
            Quality coverage: 6.43x in Q20 bases; sum-of-contigs Quality
            coverage: 12.99x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 23 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     9379: contig of 9379 bp in length
            *     9380     9479: gap of 100 bp
            *     9480    17719: contig of 8240 bp in length
            *    17720    17819: gap of 100 bp
            *    17820    22219: contig of 4400 bp in length
            *    22220    22319: gap of 100 bp
            *    22320    33493: contig of 11174 bp in length
            *    33494    33593: gap of 100 bp
            *    33594    57232: contig of 23639 bp in length
            *    57233    57332: gap of 100 bp
            *    57333    63423: contig of 6091 bp in length
            *    63424    63523: gap of 100 bp
            *    63524    67014: contig of 3491 bp in length
            *    67015    67114: gap of 100 bp
            *    67115   106755: contig of 39641 bp in length
            *   106756   106855: gap of 100 bp
            *   106856   121891: contig of 15036 bp in length
            *   121892   121991: gap of 100 bp
            *   121992   135062: contig of 13071 bp in length
            *   135063   135162: gap of 100 bp
            *   135163   153032: contig of 17870 bp in length
            *   153033   153132: gap of 100 bp
            *   153133   190703: contig of 37571 bp in length
            *   190704   190803: gap of 100 bp
            *   190804   193044: contig of 2241 bp in length
            *   193045   193144: gap of 100 bp
            *   193145   198079: contig of 4935 bp in length
            *   198080   198179: gap of 100 bp
            *   198180   207026: contig of 8847 bp in length
            *   207027   207126: gap of 100 bp
            *   207127   214185: contig of 7059 bp in length
            *   214186   214285: gap of 100 bp
            *   214286   235774: contig of 21489 bp in length
            *   235775   235874: gap of 100 bp
            *   235875   268606: contig of 32732 bp in length
            *   268607   268706: gap of 100 bp
            *   268707   270870: contig of 2164 bp in length
            *   270871   270970: gap of 100 bp
            *   270971   278926: contig of 7956 bp in length
            *   278927   279026: gap of 100 bp
            *   279027   294890: contig of 15864 bp in length
            *   294891   294990: gap of 100 bp
            *   294991   302499: contig of 7509 bp in length
            *   302500   302599: gap of 100 bp
            *   302600   326441: contig of 23842 bp in length.
FEATURES             Location/Qualifiers
     source          1. .326441
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-134M19"
                     /clone_lib="CHORI-211"
     misc_feature    1. .9379
                     /note="assembly_fragment:00388
                     fragment_chain:1"
     misc_feature    9480. .17719
                     /note="assembly_fragment:00332
                     fragment_chain:1"
     misc_feature    17820. .22219
                     /note="assembly_fragment:00211
                     fragment_chain:1"
     misc_feature    22320. .33493
                     /note="assembly_fragment:00785
                     fragment_chain:1"
     misc_feature    33594. .57232
                     /note="assembly_fragment:01360
                     fragment_chain:1"
```

```
     misc_feature    57333. .63423
                     /note="assembly_fragment:00160
                     fragment_chain:1"
     misc_feature    63524. .67014
                     /note="assembly_fragment:00138
                     fragment_chain:1"
     misc_feature    67115. .106755
                     /note="assembly_fragment:02559
                     fragment_chain:2"
     misc_feature    106856. .121891
                     /note="assembly_fragment:00914
                     fragment_chain:2"
     misc_feature    121992. .135062
                     /note="assembly_fragment:00663
                     fragment_chain:2"
     misc_feature    135163. .153032
                     /note="assembly_fragment:00542
                     fragment_chain:2"
     misc_feature    153133. .190703
                     /note="assembly_fragment:01930
                     fragment_chain:3"
     misc_feature    190804. .193044
                     /note="assembly_fragment:00089
                     fragment_chain:3"
     misc_feature    193145. .198079
                     /note="assembly_fragment:00185
                     fragment_chain:3"
     misc_feature    198180. .207026
                     /note="assembly_fragment:00288
                     fragment_chain:4"
     misc_feature    207127. .214185
                     /note="assembly_fragment:00248
                     fragment_chain:4"
     misc_feature    214286. .235774
                     /note="assembly_fragment:01110
                     fragment_chain:5"
     misc_feature    235875. .268606
                     /note="assembly_fragment:02953
                     fragment_chain:5"
     misc_feature    268707. .270870
                     /note="assembly_fragment:00100"
     misc_feature    270971. .278926
                     /note="assembly_fragment:00463"
     misc_feature    279027. .294890
                     /note="assembly_fragment:01641.0"
     misc_feature    294991. .302499
                     /note="assembly_fragment:01641.1"
     misc_feature    302600. .326441
                     /note="assembly_fragment:02243"
ORIGIN

  Query Match             8.5%;  Score 349;  DB 2;  Length 326441;
  Best Local Similarity   64.1%;  Pred. No. 6.8e-59;
  Matches  467;  Conservative    0;  Mismatches  260;  Indels    2;  Gaps    2;

Qy        198 AATGTCATCTCAGAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACAC 257
              ||| | || |     | | | | |       | ||||||||||||||||||||||| | |
Db     235247 AATTTTATATATATATATATATATATATATATATATATATATATATATATATATATAT 235306

Qy        258 ACACACACATATATATATACACATATATATACACATATATACATATATACACATATATAC 317
              | | | | |||||||||| | ||||||||| | ||||||| ||||||| | |||||||
Db     235307 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 235366

Qy        318 ACATATATATACACATACATATGTACACATATATATACACATATGTATACACATATATAC 377
              | ||||||||| | ||| |||| || | ||||||||| | |||| |||| | |||||||
Db     235367 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 235426

Qy        378 ACATATATACACACATATATACACATATATACACACATATATACACATATATACACATAT 437
              | |||||||| | | ||||||| | ||||||| | | ||||||| | ||||||| | ||||
Db     235427 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 235486

Qy        438 ATACACATATACACATATATACACATATATACATATATACACATATATATA-ATATACAC 496
                 ||| | ||||| | ||||||| | ||||||| ||||||| | ||||||||| ||||| |

Db     235487 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 235546

Qy        497 ACATATATATACACATATATACACACATATATACACATATATACACATATATATACACAT 556
              | ||||||||| | ||||||| | | ||||||| | ||||||| | ||||||||| | ||
Db     235547 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 235606

Qy        557 ATATACACATATATACATATATACACATATATATACATATATACACATATATACATATAT 616
              ||||| | ||||||| ||||||| | |||||||||| ||||||| | ||||||| ||||||
Db     235607 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 235666

Qy        617 ACACATATATACATATATACACACATATATACACATACATATACACACACATAGATATAC 676
              | | ||||||| ||||||| | | ||||||| | ||| ||||| | | | |||  ||||
Db     235667 ATATATATATATATATATATATATATATATATATATATATATATATATATATATCTATAT 235726

Qy        677 ATATATATACACATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAA 736
              ||||||||| | ||||||||| |||||||||| ||||||||||
Db     235727 ATATATATATATATATATATATATGTATATATATATATATATATATATNNNNNNNNNNNN 235786

Qy        737 GAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTG 796

Db     235787 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 235846

Qy        797 TGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAA 856
                                                 |||       | ||| ||| || | || | |
Db     235847 NNNNNNNNNNNNNNNNNNNNNNNNNNNNATATATATATATATATATATATATTTAT 235906

Qy        857 ACCTATCTGTA-AGGTAGTTAAGAAGGTAACACTATATATATATATAGTGAAAGCAGTGT 915
              |  ||| | || |  ||  ||   |  ||    ||||||||||||||   | |    | |
Db     235907 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 235966

Qy        916 AAACCTAAA 924
              | |  || |
Db     235967 ATATATATA 235975


RESULT 85
CR627490
LOCUS       CR627490              216831 bp    DNA     linear   HTG 09-AUG-2004
DEFINITION  Danio rerio clone DKEY-52H23, WORKING DRAFT SEQUENCE, 15 unordered
            pieces.
ACCESSION   CR627490
VERSION     CR627490.3  GI:51090209
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 216831)
  AUTHORS   Burton,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-AUG-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 9, 2004 this sequence version replaced gi:51036472.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK52H23
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 210790 bases at least Q40
            Consensus quality: 211954 bases at least Q30
            Consensus quality: 213128 bases at least Q20
            Insert size: 215431; sum-of-contigs
            Insert size: 201083; 6.6% error; agarose-fp
            Quality coverage: 9.53x in Q20 bases; sum-of-contigs Quality
            coverage: 10.53x in Q20 bases; agarose-fp
            --------------.
```

```
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 15 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    10834: contig of 10834 bp in length
            *    10835    10934: gap of 100 bp
            *    10935    15833: contig of 4899 bp in length
            *    15834    15933: gap of 100 bp
            *    15934    43416: contig of 27483 bp in length
            *    43417    43516: gap of 100 bp
            *    43517    50746: contig of 7230 bp in length
            *    50747    50846: gap of 100 bp
            *    50847    56683: contig of 5837 bp in length
            *    56684    56783: gap of 100 bp
            *    56784    61053: contig of 4270 bp in length
            *    61054    61153: gap of 100 bp
            *    61154    65291: contig of 4138 bp in length
            *    65292    65391: gap of 100 bp
            *    65392    93832: contig of 28441 bp in length
            *    93833    93932: gap of 100 bp
            *    93933   118737: contig of 24805 bp in length
            *   118738   118837: gap of 100 bp
            *   118838   122857: contig of 4020 bp in length
            *   122858   122957: gap of 100 bp
            *   122958   175495: contig of 52538 bp in length
            *   175496   175595: gap of 100 bp
            *   175596   203060: contig of 27465 bp in length
            *   203061   203160: gap of 100 bp
            *   203161   206216: contig of 3056 bp in length
            *   206217   206316: gap of 100 bp
            *   206317   208899: contig of 2583 bp in length
            *   208900   208999: gap of 100 bp
            *   209000   216831: contig of 7832 bp in length.
FEATURES             Location/Qualifiers
     source          1. .216831
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-52H23"
                     /clone_lib="DanioKey"
     misc_feature    1. .10834
                     /note="assembly_fragment:00683
                     fragment_chain:1"
     misc_feature    10935. .15833
                     /note="assembly_fragment:00240
                     fragment_chain:1"
     misc_feature    15934. .43416
                     /note="assembly_fragment:01139
                     fragment_chain:1"
     misc_feature    43517. .50746
                     /note="assembly_fragment:00404
                     fragment_chain:1"
     misc_feature    50847. .56683
                     /note="assembly_fragment:00315
                     fragment_chain:1"
     misc_feature    56784. .61053
                     /note="assembly_fragment:00165
                     fragment_chain:1"
     misc_feature    61154. .65291
                     /note="assembly_fragment:00123
                     fragment_chain:1"
     misc_feature    65392. .93832
                     /note="assembly_fragment:02008
                     fragment_chain:1"
     misc_feature    93933. .118737
                     /note="assembly_fragment:00887
                     fragment_chain:1"
     misc_feature    118838. .122857
                     /note="assembly_fragment:00091
                     fragment_chain:1"
     misc_feature    122958. .175495
                     /note="assembly_fragment:02459
                     fragment_chain:2"
     misc_feature    175596. .203060
                     /note="assembly_fragment:01567
                     fragment_chain:2"
     misc_feature    203161. .206216
                     /note="assembly_fragment:00050"
     misc_feature    206317. .208899
                     /note="assembly_fragment:00069"
     misc_feature    209000. .216831
                     /note="assembly_fragment:00508.0"
ORIGIN

  Query Match             8.5%;  Score 348.8;  DB 2;  Length 216831;
  Best Local Similarity   64.1%;  Pred. No. 7.9e-59;
  Matches  448;  Conservative    0;  Mismatches  250;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              |||||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db      93324 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 93383

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| |||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db      93384 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 93443

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db      93444 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 93503

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db      93504 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 93563

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db      93564 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 93623

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db      93624 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 93683

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db      93684 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 93743

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||| ||
Db      93744 TATATATATATATATATATATATATATATATATATATATATANATATATATATATTTA 93803

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||          | ||
Db      93804 TATATATATATATATATATATATATATATNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 93863

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825

Db      93864 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 93923

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885
                    | ||| | |||| | |       | |||      |  ||  ||       ||
Db      93924 NNNNNNNNATATATAAAAATATATTTATTTATATATAATATATATATATTTTTAT 93983

Qy        886 CACTATATATATATATAGTGAAAGCAGTGTAAACCTAAA 924
                |||||||||||||| | | |    | || |  || |
Db      93984 ATATATATATATATATATTTATATATATATANANATATA 94022


RESULT 86
CR450804/c
LOCUS       CR450804              220888 bp    DNA     linear   HTG 08-OCT-2004
DEFINITION  Danio rerio clone CH211-275D22, *** SEQUENCING IN PROGRESS ***, 28
            unordered pieces.
```

```
ACCESSION   CR450804
VERSION     CR450804.4  GI:54019855
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 220888)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 9, 2004 this sequence version replaced gi:53828284.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zC275D22
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 206726 bases at least Q40
            Consensus quality: 211339 bases at least Q30
            Consensus quality: 214628 bases at least Q20
            Insert size: 218188; sum-of-contigs
            Insert size: 163802; 4.5% error; agarose-fp
            Quality coverage: 7.02x in Q20 bases; sum-of-contigs Quality
            coverage: 11.27x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 28 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    25610: contig of 25610 bp in length
            *    25611    25710: gap of 100 bp
            *    25711    32841: contig of 7131 bp in length
            *    32842    32941: gap of 100 bp
            *    32942    47962: contig of 15021 bp in length
            *    47963    48062: gap of 100 bp
            *    48063    99124: contig of 51062 bp in length
            *    99125    99224: gap of 100 bp
            *    99225   126254: contig of 27030 bp in length
            *   126255   126354: gap of 100 bp
            *   126355   132148: contig of 5794 bp in length
            *   132149   132248: gap of 100 bp
            *   132249   135973: contig of 3725 bp in length
            *   135974   136073: gap of 100 bp
            *   136074   160340: contig of 24267 bp in length
            *   160341   160440: gap of 100 bp
            *   160441   163288: contig of 2848 bp in length
            *   163289   163388: gap of 100 bp
            *   163389   165705: contig of 2317 bp in length
            *   165706   165805: gap of 100 bp
            *   165806   168126: contig of 2321 bp in length
            *   168127   168226: gap of 100 bp
            *   168227   173446: contig of 5220 bp in length
            *   173447   173546: gap of 100 bp
            *   173547   175654: contig of 2108 bp in length
            *   175655   175754: gap of 100 bp
            *   175755   178489: contig of 2735 bp in length
            *   178490   178589: gap of 100 bp
            *   178590   181085: contig of 2496 bp in length
            *   181086   181185: gap of 100 bp
            *   181186   185198: contig of 4013 bp in length
            *   185199   185298: gap of 100 bp
            *   185299   188726: contig of 3428 bp in length
            *   188727   188826: gap of 100 bp
            *   188827   191247: contig of 2421 bp in length
            *   191248   191347: gap of 100 bp
            *   191348   194660: contig of 3313 bp in length
            *   194661   194760: gap of 100 bp
            *   194761   197060: contig of 2300 bp in length
            *   197061   197160: gap of 100 bp
            *   197161   199305: contig of 2145 bp in length
            *   199306   199405: gap of 100 bp
            *   199406   202170: contig of 2765 bp in length
            *   202171   202270: gap of 100 bp
            *   202271   204970: contig of 2700 bp in length
            *   204971   205070: gap of 100 bp
            *   205071   207333: contig of 2263 bp in length
            *   207334   207433: gap of 100 bp
            *   207434   210161: contig of 2728 bp in length
            *   210162   210261: gap of 100 bp
            *   210262   213219: contig of 2958 bp in length
            *   213220   213319: gap of 100 bp
            *   213320   217663: contig of 4344 bp in length
            *   217664   217763: gap of 100 bp
            *   217764   220888: contig of 3125 bp in length.
FEATURES             Location/Qualifiers
     source          1. .220888
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-275D22"
                     /clone_lib="CHORI-211"
     misc_feature    1. .25610
                     /note="assembly_fragment:01114
                     fragment_chain:1
                     clone_end:T7
                     vector_side:left"
     misc_feature    25711. .32841
                     /note="assembly_fragment:00426
                     fragment_chain:1"
     misc_feature    32942. .47962
                     /note="assembly_fragment:00502
                     fragment_chain:1"
     misc_feature    48063. .99124
                     /note="assembly_fragment:02105
                     fragment_chain:1"
     misc_feature    99225. .126254
                     /note="assembly_fragment:01573
                     fragment_chain:1"
     misc_feature    126355. .132148
                     /note="assembly_fragment:00373
                     fragment_chain:1"
     misc_feature    132249. .135973
                     /note="assembly_fragment:00352
                     fragment_chain:1"
     misc_feature    136074. .160340
                     /note="assembly_fragment:00794
                     fragment_chain:1
                     clone_end:SP6
                     vector_side:right"
     misc_feature    160441. .163288
                     /note="assembly_fragment:00075
                     fragment_chain:2"
     misc_feature    163389. .165705
                     /note="assembly_fragment:00116
                     fragment_chain:2"
     misc_feature    165806. .168126
                     /note="assembly_fragment:00155
                     fragment_chain:3"
     misc_feature    168227. .173446
                     /note="assembly_fragment:00298
                     fragment_chain:3"
     misc_feature    173547. .175654
                     /note="assembly_fragment:00176
                     fragment_chain:4"
```

```
     misc_feature    175755. .178489
                     /note="assembly_fragment:00200
                     fragment_chain:4"
     misc_feature    178590. .181085
                     /note="assembly_fragment:00192
                     fragment_chain:5"
     misc_feature    181186. .185198
                     /note="assembly_fragment:00276
                     fragment_chain:5"
     misc_feature    185299. .188726
                     /note="assembly_fragment:00244
                     fragment_chain:6"
     misc_feature    188827. .191247
                     /note="assembly_fragment:00254
                     fragment_chain:6"
     misc_feature    191348. .194660
                     /note="assembly_fragment:00311
                     fragment_chain:7"
     misc_feature    194761. .197060
                     /note="assembly_fragment:00225
                     fragment_chain:7"
     misc_feature    197161. .199305
                     /note="assembly_fragment:00131"
     misc_feature    199406. .202170
                     /note="assembly_fragment:00143"
     misc_feature    202271. .204970
                     /note="assembly_fragment:00162"
     misc_feature    205071. .207333
                     /note="assembly_fragment:00169"
     misc_feature    207434. .210161
                     /note="assembly_fragment:00184"
     misc_feature    210262. .213219
                     /note="assembly_fragment:00265"
     misc_feature    213320. .217663
                     /note="assembly_fragment:00324"
     misc_feature    217764. .220888
                     /note="assembly_fragment:00338"
ORIGIN

  Query Match             8.5%;  Score 348.8;  DB 2;  Length 220888;
  Best Local Similarity   65.1%;  Pred. No. 7.9e-59;
  Matches  456;  Conservative    0;  Mismatches  242;  Indels    2;  Gaps    2;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db      26206 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 26147

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | |||||||| | ||||||||| | ||| |||| || | |
Db      26146 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 26087

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db      26086 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 26027

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | |||||||| | |||||||| | ||||||| | ||||| | ||||||| | |||||
Db      26026 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 25967

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db      25966 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 25907

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db      25906 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 25847

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db      25846 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 25787

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db      25786 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 25727

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||
Db      25726 TATATATATATATAAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 25667

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
                                                                        |
Db      25666 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNATTA 25607

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTA 884
              . || ||    | ||| ||| || | ||   | |  ||| | || |  ||  |||  |  ||
Db      25606 TAATATATTATATATATATATTTATATATATATATATATATATATATATAAATATATA 25547

Qy        885 ACACTATATATATATATAGTGAAAGCAGTGTAAACCTAAA 924
                  |||||||||||||  | |    | || |  ||||
Db      25546 TATATATATATATATATATATATATATATATATATATA 25507


RESULT 87
CR812957/c
LOCUS       CR812957              224175 bp    DNA     linear   HTG 28-SEP-2004
DEFINITION  Danio rerio clone DKEY-41L5, *** SEQUENCING IN PROGRESS ***, 13
            unordered pieces.
ACCESSION   CR812957
VERSION     CR812957.2  GI:52839447
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 224175)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 28, 2004 this sequence version replaced gi:52749045.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK41L5
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 219065 bases at least Q40
            Consensus quality: 220415 bases at least Q30
            Consensus quality: 221446 bases at least Q20
            Insert size: 222975; sum-of-contigs
            Insert size: 218290; 5.1% error; agarose-fp
            Quality coverage: 6.26x in Q20 bases; sum-of-contigs Quality
            coverage: 6.49x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 13 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2025: contig of 2025 bp in length
            *     2026     2125: gap of 100 bp
            *     2126    41693: contig of 39568 bp in length
            *    41694    41793: gap of 100 bp
            *    41794    50195: contig of 8402 bp in length
            *    50196    50295: gap of 100 bp
            *    50296    54932: contig of 4637 bp in length
            *    54933    55032: gap of 100 bp
```

```
            *    55033    75269: contig of 20237 bp in length
            *    75270    75369: gap of 100 bp
            *    75370   105478: contig of 30109 bp in length
            *   105479   105578: gap of 100 bp
            *   105579   128261: contig of 22683 bp in length
            *   128262   128361: gap of 100 bp
            *   128362   130955: contig of 2594 bp in length
            *   130956   131055: gap of 100 bp
            *   131056   177561: contig of 46506 bp in length
            *   177562   177661: gap of 100 bp
            *   177662   185583: contig of 7922 bp in length
            *   185584   185683: gap of 100 bp
            *   185684   195396: contig of 9713 bp in length
            *   195397   195496: gap of 100 bp
            *   195497   208546: contig of 13050 bp in length
            *   208547   208646: gap of 100 bp
            *   208647   224175: contig of 15529 bp in length.
FEATURES             Location/Qualifiers
     source          1. .224175
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-41L5"
                     /clone_lib="DanioKey"
     misc_feature    1. .2025
                     /note="assembly_fragment:00011
                     fragment_chain:1"
     misc_feature    2126. .41693
                     /note="assembly_fragment:01456
                     fragment_chain:1"
     misc_feature    41794. .50195
                     /note="assembly_fragment:00090
                     fragment_chain:1"
     misc_feature    50296. .54932
                     /note="assembly_fragment:00059
                     fragment_chain:1"
     misc_feature    55033. .75269
                     /note="assembly_fragment:00674
                     fragment_chain:1"
     misc_feature    75370. .105478
                     /note="assembly_fragment:01132
                     fragment_chain:2"
     misc_feature    105579. .128261
                     /note="assembly_fragment:00900
                     fragment_chain:2"
     misc_feature    128362. .130955
                     /note="assembly_fragment:00024
                     fragment_chain:2"
     misc_feature    131056. .177561
                     /note="assembly_fragment:01832
                     fragment_chain:2"
     misc_feature    177662. .185583
                     /note="assembly_fragment:00258
                     fragment_chain:2"
     misc_feature    185684. .195396
                     /note="assembly_fragment:00172
                     fragment_chain:3"
     misc_feature    195497. .208546
                     /note="assembly_fragment:00383
                     fragment_chain:3"
     misc_feature    208647. .224175
                     /note="assembly_fragment:00528"
ORIGIN

  Query Match           8.5%;  Score 348.8;  DB 2;  Length 224175;
  Best Local Similarity   60.6%;  Pred. No. 7.9e-59;
  Matches  507;  Conservative    0;  Mismatches  327;  Indels    2;  Gaps    2;

Qy        191 AGAGCGAAATGTCATCTCAGAAAAAAAAAAAGGCATTTTATATATATATATATATATATA 250
              | | |  ||   |||   | ||||||||||||   |  ||||||||||||||||||||
Db       2838 ATAACACAAAACCATAAAAAAAAAAAAAAAAAAAAAAAAATATATATATATATATATATA 2779

Qy        251 TACACACACACACACATATATATATACACATATATATACACATATATACATATATACACA 310
```

```
              || | | | | | | |||||||||||| | ||||||||| | ||||||| ||||||| | |
Db       2778 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 2719

Qy        311 TATATACACATATATATACACATACATATGTACACATATATATACACATATGTATACACA 370
              |||||| | ||||||||| | ||| |||| || | ||||||||| | |||| |||| | |
Db       2718 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 2659

Qy        371 TATATACACATATATACACACATATATACACATATATACACACATATATACACATATATA 430
              |||||| | ||||||| | | ||||||| | ||||||| | | ||||||| | |||||||
Db       2658 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 2599

Qy        431 CACATATATACACATATACACATATATACACATATATACATATATACACATATATATAAT 490
              | ||||||| | ||||| | ||||||| | ||||||| ||||||| | |||||||| ||
Db       2598 TATATATATATATATATATATATATATATATATATATATATATATATATATATATAT-AT 2540

Qy        491 ATACACACATATATATACACATATATACACACATATATACACATATATACACATATATAT 550
              ||| | | ||||||||| | ||||||| | | ||||||| | ||||||| | ||||||||
Db       2539 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 2480

Qy        551 ACACATATATACACATATATACATATATACACATATATATACATATATACACATATATAC 610
              | | |||||||| | ||||||| ||||||| | ||||||||| ||||||| | |||||||
Db       2479 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 2420

Qy        611 ATATATACACATATATACATATATACACACATATATACACATACATATACACACACATAG 670
              ||||||| | ||||||| ||||||| | | ||||||| | ||| ||||| | | | |||
Db       2419 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 2360

Qy        671 ATATACATATATATACACATATATATACGTATATATATGTATATATATATGCTCCAGAGT 730
              ||||| ||||||||| | |||||||||  |||||||| ||||||||||          |
Db       2359 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 2300

Qy        731 TCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGT 790
                |||          |  |   ||     |   ||| |   | | | |  | | |    |
Db       2299 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 2240

Qy        791 GTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGT 850
              |||   | |   | || |     |       | |||        | ||| ||| || | |
Db       2239 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 2180

Qy        851 AATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTAACACTATATATATATATAGTGAAAG 909
              |   | |  ||| | || |  ||  ||  |  ||    ||||||||||||||
Db       2179 ATATATATATATATATATATATATATATATATATATATATATATATATATATATNNNNNN 2120

Qy        910 CAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTATAGAAGAATGGCTAAGATGC 969

Db       2119 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 2060

Qy        970 CAAAGCTCAGTGTATGTGGCAGAGGCATGGTGTAGGGTGTGTCCAGGTTCATATAT 1025
                                                  | | |  |  | ||||||
Db       2059 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNATATATATATATATATATAT 2004
```

```
RESULT 88
AC094041/c
LOCUS       AC094041               236551 bp    DNA     linear   HTG 10-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-18C6, WORKING DRAFT SEQUENCE, 4
            unordered pieces.
ACCESSION   AC094041
VERSION     AC094041.8  GI:30520885
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 236551)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
```

Chacko,J., Chavez;D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 236551)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (14-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 236551)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (10-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On May 10, 2003 this sequence version replaced gi:24942723.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

---------- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
Center project name: GFNH
Center clone name: CH230-18C6
---------- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 222835 bases at least Q40
Consensus quality: 225635 bases at least Q30
Consensus quality: 228117 bases at least Q20
Estimated insert size: 236184; sum-of-contigs estimation
Quality coverage: 9x in Q20 bases; sum-of-contigs estimation
----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 230804: contig of 230804 bp in length
* 230805 230904: gap of unknown length
* 230905 232373: contig of 1469 bp in length
* 232374 232473: gap of unknown length
* 232474 233531: contig of 1058 bp in length
* 233532 233631: gap of unknown length
* 233632 236551: contig of 2920 bp in length.

FEATURES Location/Qualifiers
source 1. .236551
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-18C6"

ORIGIN

Query Match 8.5%; Score 348.8; DB 2; Length 236551;
Best Local Similarity 69.6%; Pred. No. 7.8e-59;
Matches 487; Conservative 0; Mismatches 212; Indels 1; Gaps 1;

```
Qy       227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
             | ||||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db    234988 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 234929

Qy       287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
             || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db    234928 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 234869

Qy       347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
             |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db    234868 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 234809

Qy       407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
             || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db    234808 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 234749

Qy       467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
             || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db    234748 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 234689

Qy       526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
             |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db    234688 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 234629

Qy       586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
             |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db    234628 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 234569

Qy       646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
             || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db    234568 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 234509
```

```
Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||||          |  |||          |  |   ||    |    |||
Db     234508 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 234449

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
              |   | | | |  |  | |    | |||    | |   | || |     |      | |
Db     234448 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 234389

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885
              ||        | ||| ||| || | ||   | |  ||| | ||      |  | | | ||
Db     234388 TATATATATATATATATATATATATATATATATATATATATATATATATATATGATAT 234329

Qy        886 CACTATATATATATATAGTGAAAGCAGTGTAAACCTAAAG 925
              ||||||||||||| | | |||  | | || |  || ||
Db     234328 ATATATATATATATAAAATTAAAAAATTATATATATATAG 234289
```

RESULT 89
BX322798/c
LOCUS BX322798 109868 bp DNA linear VRT 24-JUN-2004
DEFINITION Zebrafish DNA sequence from clone CH211-231F6 in linkage group 24, complete sequence.
ACCESSION BX322798
VERSION BX322798.6 GI:49203424
KEYWORDS HTG.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 109868)
AUTHORS Heath,P.
TITLE Direct Submission
JOURNAL Submitted (24-JUN-2004) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT On Jun 24, 2004 this sequence version replaced gi:49065270.
---------- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfish-help@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep Clone-derived Zebrafish pUC subclones occasionally display inconsistency over the length of mononucleotide A/T runs and conserved TA repeats. Where this is found the longest good quality representation will be submitted.
Repeat names beginning 'Dr' were identified by the Recon repeat discovery system (Zhirong Bao and Sean Eddy, submitted), and those beginning 'drr' were identified by Rick Waterman (Stephen Johnson lab, WashU). For further information see http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml CH211-231F6 is from a CHORI-211 BAC library
VECTOR: pTARBAC2.1.

FEATURES Location/Qualifiers
source 1. .109868
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="CH211-231F6"
/clone_lib="CHORI-211"
ORIGIN

Query Match 8.5%; Score 348.6; DB 5; Length 109868;
Best Local Similarity 76.5%; Pred. No. 9.6e-59;
Matches 440; Conservative 0; Mismatches 134; Indels 1; Gaps 1;

```
Qy        147 TGCGGGGAGCAAGATCATGCCATTGCACTCCAGCCCAGGCAACAAGAGCGAAATGTCATC 206
              || | |||   ||  ||      ||||||  ||   |   ||   | |   || | |
Db      78151 TGTGTGGAAACAGTCCAAAAACATGCACTATAGGTGAATGAATGGGTTCTTTATTTGACT 78092

Qy        207 TCAGAAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACACA 266
              | | | ||| | |   || ||||| ||| ||||| |||||||||| | | | | |||
Db      78091 ATATATATAAATATATATATATTTATACATACATATACATATATATATATATATATATACA 78032

Qy        267 TATATATATACACATATATATACACATATATACATATATACACATATATACACATATATA 326
              |||||||||||| ||||||||| | ||||| | ||||| | | |||||||||| |||||||
Db      78031 TATATATATACATATATATATATATATACATATATACATATATATATACATATATATA 77972

Qy        327 TACACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATA 386
              | |||||||||| |||| |||||||||| |||| |||||| ||||||| | |||||||
Db      77971 CATACATACATATATACATATATATATACATATATATATACATATATATATATATATATA 77912

Qy        387 CACACATATATACACATATATACACACATATATACACATATATACACATATATACACATA 446
              || ||||||| | | ||||||||| | ||||||||| | ||||| | ||||| | | |||
Db      77911 CATACATATACATATATATATACATATATATATACATACATATATATATATACATATATA 77852

Qy        447 TACACATATATACACATATATACATATATACACATATATATA-ATATACACACATATATA 505
              || ||||| ||| |||||||||| ||| ||| | ||||| ||| ||||||| | |||||||
Db      77851 TATACATACATATACATATATATATATACATATATATATACATACATATACATATATATATA 77792

Qy        506 TACACATATATACACACATATATACACATATATACACATATATATACACATATATACACA 565
              | | |||||||||| ||||||| | | |||||||||| |||||||||| ||||| ||| |||
Db      77791 CATATATATATACATACATATACATATATATATACATATATATATATACATACATATACA 77732

Qy        566 TATATACATATATACACATATATATACATATATACACATATATACATATATACACATATA 625
              |||||| ||| ||||| ||||| |||||||||| | | ||||||| ||| ||||| |||||
Db      77731 TATATATATACATACATATATACATACATATACATATATATATATATACATACATATATA 77672

Qy        626 TACATATATACACACATATATACACATACATATACACACACATAGATATACATATATATA 685
              | ||||||| ||| | ||||| ||| |  |||| | | | ||| ||||| |||||||||
Db      77671 CATATATATATACATACATATATACACATGTATATATATATATATATATATATATATA 77612

Qy        686 CACATATATATACGTATATATATGTATATATATAT 720
              |||||| |||||  |||||||||| ||||||||||
Db      77611 CACATACATATATATATATATATATATATATAT 77577
```

RESULT 90
CR392338/c
LOCUS CR392338 132545 bp DNA linear HTG 26-AUG-2004
DEFINITION Danio rerio clone DKEY-262M9, *** SEQUENCING IN PROGRESS ***, 9 unordered pieces.
ACCESSION CR392338
VERSION CR392338.3 GI:51571696
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 132545)
AUTHORS Sims,S.
TITLE Direct Submission
JOURNAL Submitted (24-AUG-2004) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

```
COMMENT     On Aug 26, 2004 this sequence version replaced gi:51490965.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK262M9
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 130047 bases at least Q40
            Consensus quality: 130566 bases at least Q30
            Consensus quality: 130976 bases at least Q20
            Insert size: 131745; sum-of-contigs
            Insert size: 144213; 1.4% error; agarose-fp
            Quality coverage: 7.43x in Q20 bases; sum-of-contigs Quality
            coverage: 6.98x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2980: contig of 2980 bp in length
            *     2981     3080: gap of 100 bp
            *     3081     7869: contig of 4789 bp in length
            *     7870     7969: gap of 100 bp
            *     7970    25024: contig of 17055 bp in length
            *    25025    25124: gap of 100 bp
            *    25125    34501: contig of 9377 bp in length
            *    34502    34601: gap of 100 bp
            *    34602    42063: contig of 7462 bp in length
            *    42064    42163: gap of 100 bp
            *    42164    45417: contig of 3254 bp in length
            *    45418    45517: gap of 100 bp
            *    45518    62732: contig of 17215 bp in length
            *    62733    62832: gap of 100 bp
            *    62833    89992: contig of 27160 bp in length
            *    89993    90092: gap of 100 bp
            *    90093   132545: contig of 42453 bp in length.
FEATURES            Location/Qualifiers
     source         1. .132545
                    /organism="Danio rerio"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:7955"
                    /clone="DKEY-262M9"
                    /clone_lib="DanioKey"
     misc_feature   1. .2980
                    /note="assembly_fragment:00025
                    fragment_chain:1"
     misc_feature   3081. .7869
                    /note="assembly_fragment:00103
                    fragment_chain:1"
     misc_feature   7970. .25024
                    /note="assembly_fragment:00458
                    fragment_chain:1"
     misc_feature   25125. .34501
                    /note="assembly_fragment:00156
                    fragment_chain:1"
     misc_feature   34602. .42063
                    /note="assembly_fragment:00280
                    fragment_chain:1"
     misc_feature   42164. .45417
                    /note="assembly_fragment:00074
                    fragment_chain:2"
     misc_feature   45518. .62732
                    /note="assembly_fragment:00711
                    fragment_chain:2"
     misc_feature   62833. .89992
                    /note="assembly_fragment:00974
                    fragment_chain:2"
     misc_feature   90093. .132545
                    /note="assembly_fragment:01385
                    fragment_chain:2"
ORIGIN

  Query Match             8.5%;  Score 348.6;  DB 2;  Length 132545;
  Best Local Similarity   62.2%;  Pred. No. 9.3e-59;
  Matches  472;  Conservative    0;  Mismatches  284;  Indels    3;  Gaps    1;

Qy        169 TTGCACTCCAGCCCAGGCAACAAGAGCGAAATGTCATCTCAGAAAAAAAAAAAAGGCATTT 228
              ||  || |   | |   ||  | |     ||||| |   ||| |||| || |    | ||
Db      63457 TTTTACACACCCTGAAAAAAGTAAAAACCAATGTAAATACAGCAAAACAATATTTTAATT 63398

Qy        229 TATATATATATATATATATATATACACACACACACACATATATATATACACATATATATA 288
              ||||||||||||||||||||||| | | | | | | |||||||||||| | |||||||||
Db      63397 AATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 63338

Qy        289 CACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATA 348
              | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |||
Db      63337 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 63278

Qy        349 TATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATA 408
              |||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||||
Db      63277 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 63218

Qy        409 CACACATATATACACATATATACACATATATACACATATACACATATATACACATATATA 468
              | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||||
Db      63217 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 63158

Qy        469 CATATATACACATATATATA---ATATACACACATATATATACACATATATACACACATA 525
              ||||||| | |||||||||    ||||| | | ||||||||| | ||||||| | | |||
Db      63157 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 63098

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db      63097 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 63038

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db      63037 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 62978

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db      62977 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 62918

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||          | |||          | |   ||    |    |||
Db      62917 TATATATATATATATATATATATATATATATATATATATATATATATATATATAAATA 62858

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
              |  || | | |  | | |    |
Db      62857 AAATAATATATATATATATATATNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 62798

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885

Db      62797 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 62738

Qy        886 CACTATATATATATAGTGAAAGCAGTGTAAACCTAAA 924
                |||||||||||   | |    | || | ||| |
Db      62737 NNNNNTATATATATATATATATATATATATATCTATA 62699


RESULT 91
AC004966
LOCUS       AC004966              134084 bp    DNA     linear   PRI 03-OCT-2003
DEFINITION  Homo sapiens PAC clone RP5-1107K15 from 7, complete sequence.
ACCESSION   AC004966
VERSION     AC004966.2  GI:5306307
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 134084)
AUTHORS Sulston,J.E. and Wilson,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
PUBMED 9847074
REFERENCE 2 (bases 1 to 134084)
AUTHORS Smith-Craig,Renee., Strowmatt,C. and Wohldmann,P.
TITLE The sequence of Homo sapiens PAC clone RP5-1107K15
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 134084)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (12-JUN-1998) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 134084)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (01-JUL-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 5 (bases 1 to 134084)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (23-OCT-1999) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 6 (bases 1 to 134084)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-1999) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 7 (bases 1 to 134084)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (26-APR-2003) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 8 (bases 1 to 134084)
AUTHORS Wilson,R.
TITLE Direct Submission
JOURNAL Submitted (03-OCT-2003) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On Jul 1, 1999 this sequence version replaced gi:3907492.
---------- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: sapiens@watson.wustl.edu
---------- Summary Statistics
Center project name: H_DJ1107K15
----------.

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see http://www.nhgri.nih.gov/DIR/GTB/CHR7 , send mailto:egreen@nhgri.nih.gov , or see http://genome.wustl.edu

SOURCE INFORMATION:
This clone was derived from human PAC library RPCI-5, prepared by Pieter de Jong and coworkers at http://www.chori.org using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor.
The clone may be obtained either from Genome Systems, Inc. (http://www.genomesystems.com) or Research Genetics, Inc. (http://www.resgen.com); or from Pieter de Jong.
VECTOR: pCYPAC2

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP5-942I16, 200 bp overlap.
Actual start of this clone is at base position 197 of RP5-1107K15
actual end is at base position 134084 of RP5-1107K15.

FEATURES Location/Qualifiers
source 1. .134084
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="7"
/map="7"
/clone="RP5-1107K15"
/clone_lib="RPCI-5"
repeat_region 84. .925
/rpt_family="L1"
repeat_region 1199. .1495
/rpt_family="Alu"
repeat_region 1692. .2013
/rpt_family="Alu"
repeat_region 2023. .2340
/rpt_family="Alu"
repeat_region 3225. .3438
/rpt_family="CR1"
repeat_region 3459. .3768
/rpt_family="Alu"
repeat_region 4040. .4346
/rpt_family="Alu"
repeat_region 4474. .4594
/rpt_family="MIR"
repeat_region 4854. .5166
/rpt_family="Alu"
repeat_region 5172. .5337
/rpt_family="Alu"
repeat_region 5338. .5387
/rpt_family="(TTA)n"
repeat_region 5388. .5507
/rpt_family="Alu"
repeat_region 5523. .5800
/rpt_family="Alu"
repeat_region 6624. .6922
/rpt_family="Alu"
repeat_region 7057. .7353
/rpt_family="Alu"
repeat_region 7360. .7642
/rpt_family="Alu"
repeat_region 8314. .8428
/rpt_family="L2"
repeat_region 8464. .8762
/rpt_family="Alu"
repeat_region 8767. .9078
/rpt_family="Alu"
repeat_region 9642. .9950
/rpt_family="Alu"
repeat_region 9959. .10274
/rpt_family="Alu"
repeat_region 10789. .11096
/rpt_family="Alu"
repeat_region 11354. .11489
/rpt_family="Alu"

```
     repeat_region   11526. .11806
                     /rpt_family="Alu"
     repeat_region   11810. .12113
                     /rpt_family="Alu"
     repeat_region   12388. .12679
                     /rpt_family="Alu"
     repeat_region   13254. .13297
                     /rpt_family="(TG)n"
     repeat_region   13298. .13445
                     /rpt_family="Alu"
     repeat_region   13448. .13774
                     /rpt_family="Alu"
     repeat_region   13776. .14066
                     /rpt_family="Alu"
     repeat_region   14091. .14478
                     /rpt_family="ERVL"
     repeat_region   14550. .14612
                     /rpt_family="Alu"
     repeat_region   14618. .14897
                     /rpt_family="Alu"
     repeat_region   14936. .15029
                     /rpt_family="MER1_type"
     repeat_region   15078. .15244
                     /rpt_family="L2"
     repeat_region   15248. .15469
                     /rpt_family="MaLR"
     repeat_region   15485. .15561
                     /rpt_family="MaLR"
     repeat_region   15700. .15955
                     /rpt_family="Alu"
     repeat_region   15997. .16096
                     /rpt_family="L1"
     repeat_region   16097. .16124
                     /rpt_family="(TTTA)n"
     repeat_region   16135. .16422
                     /rpt_family="Alu"
     repeat_region   16877. .17035
                     /rpt_family="Alu"
     repeat_region   17036. .17345
                     /rpt_family="Alu"
     repeat_region   17530. .17671
                     /rpt_family="MIR"
     repeat_region   18281. .18563
                     /rpt_family="Alu"
     repeat_region   18564. .18609
                     /rpt_family="(CA)n"
     repeat_region   18711. .18888

  Query Match             8.5%;  Score 348.6;  DB 9;  Length 134084;
  Best Local Similarity   83.3%;  Pred. No. 9.3e-59;
  Matches  409;  Conservative    0;  Mismatches   79;  Indels    3;  Gaps    1;

Qy        230 ATATATATATATATATATATATACACACACACACATATATATATACACATATATATAC 289
              |||||||||| | |||||||||| | | | | ||| | ||||| ||||| | | ||||||||
Db     108308 ATATATATACACATATATATACATATATATACATATATATACATATATATACATATATAC 108367

Qy        290 ACATATATACATATATACACATATATACACATATATATACACATACATATGTACACATAT 349
              |||||||||| ||||| ||| ||||| ||| ||||||| ||||| ||||  | | ||||
Db     108368 ACATATATACACATATATACACATATATACACATATATATACATATATATACATATATAT 108427

Qy        350 ATATACACATATGTATACACATATATACACATATATACACACATATATACACATATATAC 409
              | ||| | |||  |||| ||||| ||| ||||||||| ||| ||||||||| ||||||||
Db     108428 ACATATATATACATATATACATACATATACATATATATACATATATATACATATATATAC 108487

Qy        410 ACACATATATACACATATATACACATATATACACATATACACATATATACACATATATAC 469
              ||| ||||| | | ||||| ||| ||||| ||| ||||||||||||||||| ||||||||
Db     108488 ACATATATACATATATATACACATATATACACATATATACACATATATACATATATATAC 108547

Qy        470 ATATATACACATATATATA---ATATACACACATATATATACACATATATACACACATAT 526
              ||||||| ||||||||||    ||||||| | ||| | ||| | | ||||| | ||| ||
Db     108548 ATATATATACATATATATACATATATACATATATACACATATATACATATATATACACAT 108607

Qy        527 ATACACATATATACACATATATATACACATATATACACATATATACATATATACACATAT 586
              |||| |||||||| ||| |||||||| |||||||| |||||||| ||||| ||||||
Db     108608 ATATACATATATATACACATATATACATATATATACATATATATACACATATATACATAT 108667

Qy        587 ATATACATATATACACATATATACATATATACACATATATACATATATACACACATATAT 646
              |||||||||||||| |||||||||||||| |||||||||||||||| ||| ||||||
Db     108668 ATATACATATATACATATATATACATATATATACATATATACATATATATACATATATAT 108727

Qy        647 ACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATAT 706
              ||| ||| | ||| | | ||||| ||||||||||||||||| ||||||| || |||||||
Db     108728 ACATATATACATATATATACATATATATACATATATATACATATATATACACATATATAT 108787

Qy        707 ATGTATATATA 717
              |   |||||||
Db     108788 ACACATATATA 108798


RESULT 92
CR735106/c
LOCUS       CR735106              170355 bp    DNA     linear   HTG 19-AUG-2004
DEFINITION  Danio rerio clone CH211-232N15, WORKING DRAFT SEQUENCE, 13
            unordered pieces.
ACCESSION   CR735106
VERSION     CR735106.2  GI:51476923
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 170355)
  AUTHORS   Burton,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-AUG-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 20, 2004 this sequence version replaced gi:51241603.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zC232N15
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 167417 bases at least Q40
            Consensus quality: 168324 bases at least Q30
            Consensus quality: 168699 bases at least Q20
            Insert size: 169155; sum-of-contigs
            Insert size: 178984; 3.2% error; agarose-fp
            Quality coverage: 6.27x in Q20 bases; sum-of-contigs Quality .
            coverage: 6.05x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 13 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    11635: contig of 11635 bp in length
            *    11636    11735: gap of 100 bp
            *    11736    22534: contig of 10799 bp in length
            *    22535    22634: gap of 100 bp
            *    22635    25737: contig of 3103 bp in length
            *    25738    25837: gap of 100 bp
            *    25838    30700: contig of 4863 bp in length
            *    30701    30800: gap of 100 bp
            *    30801    47115: contig of 16315 bp in length
            *    47116    47215: gap of 100 bp
            *    47216    67157: contig of 19942 bp in length
```

```
            *    67158    67257: gap of 100 bp
            *    67258    93368: contig of 26111 bp in length
            *    93369    93468: gap of 100 bp
            *    93469   103340: contig of 9872 bp in length
            *   103341   103440: gap of 100 bp
            *   103441   113531: contig of 10091 bp in length
            *   113532   113631: gap of 100 bp
            *   113632   125308: contig of 11677 bp in length
            *   125309   125408: gap of 100 bp
            *   125409   138161: contig of 12753 bp in length
            *   138162   138261: gap of 100 bp
            *   138262   143042: contig of 4781 bp in length
            *   143043   143142: gap of 100 bp
            *   143143   170355: contig of 27213 bp in length.
FEATURES             Location/Qualifiers
     source          1. .170355
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-232N15"
                     /clone_lib="CHORI-211"
     misc_feature    1. .11635
                     /note="assembly_fragment:00706
                     fragment_chain:1"
     misc_feature    11736. .22534
                     /note="assembly_fragment:00576
                     fragment_chain:1"
     misc_feature    22635. .25737
                     /note="assembly_fragment:00028
                     fragment_chain:1"
     misc_feature    25838. .30700
                     /note="assembly_fragment:00056
                     fragment_chain:1"
     misc_feature    30801. .47115
                     /note="assembly_fragment:00862
                     fragment_chain:1"
     misc_feature    47216. .67157
                     /note="assembly_fragment:01028
                     fragment_chain:1"
     misc_feature    67258. .93368
                     /note="assembly_fragment:01537
                     fragment_chain:1"
     misc_feature    93469. .103340
                     /note="assembly_fragment:00351
                     fragment_chain:2"
     misc_feature    103441. .113531
                     /note="assembly_fragment:00156
                     fragment_chain:2"
     misc_feature    113632. .125308
                     /note="assembly_fragment:00462
                     fragment_chain:2"
     misc_feature    125409. .138161
                     /note="assembly_fragment:00246
                     fragment_chain:2"
     misc_feature    138262. .143042
                     /note="assembly_fragment:00100.0"
     misc_feature    143143. .170355
                     /note="assembly_fragment:01234"
ORIGIN

  Query Match             8.5%;  Score 348.6;  DB 2;  Length 170355;
  Best Local Similarity   65.6%;  Pred. No. 9e-59;
  Matches  451;  Conservative    0;  Mismatches  234;  Indels    2;  Gaps    2;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||| | | | | | | |||||||||||| | |||||||
Db     114143 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 114084

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db     114083 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 114024

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              ||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db     114023 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 113964

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db     113963 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 113904

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || |||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db     113903 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 113844

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db     113843 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 113784

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db     113783 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 113724

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db     113723 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 113664

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| |||||||||||  |  ||       ||
Db     113663 TATATATATATATATATTGGAGTCAGATGAATNNNNNNNNNNNNNNNNNNNNNNNNNN 113604

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825

Db     113603 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 113544

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTA 884
                      || ||| || | ||   | |  ||| | || |  ||  ||   |  ||
Db     113543 NNNNNNNNNNNTATATATATATATATATATATATATATATATATATATATATATATATA 113484

Qy        885 ACACTATATATATATATAGTGAAAGCA 911
                    |||||||||||||| || | ||
Db     113483 TATATATATATATATATAGGGAGATCA 113457


RESULT 93
CR762394
LOCUS       CR762394              233491 bp    DNA     linear   HTG 13-SEP-2004
DEFINITION  Danio rerio clone DKEY-108C6, *** SEQUENCING IN PROGRESS ***, 11
            unordered pieces.
ACCESSION   CR762394
VERSION     CR762394.3  GI:52000387
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 233491)
  AUTHORS   McLay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 13, 2004 this sequence version replaced gi:51988247.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK108C6
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 229390 bases at least Q40
            Consensus quality: 230228 bases at least Q30
            Consensus quality: 230919 bases at least Q20
```

```
        Insert size: 232491; sum-of-contigs
        Insert size: 232003; 3.4% error; agarose-fp
        Quality coverage: 6.99x in Q20 bases; sum-of-contigs Quality
        coverage: 7.05x in Q20 bases; agarose-fp
        --------------.
        * NOTE: This is a 'working draft' sequence. It currently
        * consists of 11 contigs. The true order of the pieces
        * is not known and their order in this sequence record is
        * arbitrary. Gaps between the contigs are represented as
        * runs of N, but the exact sizes of the gaps are unknown.
        * This record will be updated with the finished sequence
        * as soon as it is available and the accession number will
        * be preserved.
        *       1     2489: contig of 2489 bp in length
        *    2490     2589: gap of 100 bp
        *    2590    44861: contig of 42272 bp in length
        *   44862    44961: gap of 100 bp
        *   44962    83642: contig of 38681 bp in length
        *   83643    83742: gap of 100 bp
        *   83743   120587: contig of 36845 bp in length
        *  120588   120687: gap of 100 bp
        *  120688   129304: contig of 8617 bp in length
        *  129305   129404: gap of 100 bp
        *  129405   132761: contig of 3357 bp in length
        *  132762   132861: gap of 100 bp
        *  132862   147991: contig of 15130 bp in length
        *  147992   148091: gap of 100 bp
        *  148092   152438: contig of 4347 bp in length
        *  152439   152538: gap of 100 bp
        *  152539   199358: contig of 46820 bp in length
        *  199359   199458: gap of 100 bp
        *  199459   209632: contig of 10174 bp in length
        *  209633   209732: gap of 100 bp
        *  209733   233491: contig of 23759 bp in length.
FEATURES             Location/Qualifiers
     source          1. .233491
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-108C6"
                     /clone_lib="DanioKey"
     misc_feature    1. .2489
                     /note="assembly_fragment:00019
                     fragment_chain:1"
     misc_feature    2590. .44861
                     /note="assembly_fragment:01734
                     fragment_chain:1"
     misc_feature    44962. .83642
                     /note="assembly_fragment:01269
                     fragment_chain:1"
     misc_feature    83743. .120587
                     /note="assembly_fragment:00842
                     fragment_chain:1"
     misc_feature    120688. .129304
                     /note="assembly_fragment:00086
                     fragment_chain:1"
     misc_feature    129405. .132761
                     /note="assembly_fragment:00035
                     fragment_chain:1"
     misc_feature    132862. .147991
                     /note="assembly_fragment:00295
                     fragment_chain:1"
     misc_feature    148092. .152438
                     /note="assembly_fragment:00060
                     fragment_chain:2"
     misc_feature    152539. .199358
                     /note="assembly_fragment:02231
                     fragment_chain:2"
     misc_feature    199459. .209632
                     /note="assembly_fragment:00184
                     fragment_chain:2"
     misc_feature    209733. .233491
                     /note="assembly_fragment:00499.0"
```

```
ORIGIN

  Query Match             8.5%;  Score 348.6;  DB 2;  Length 233491;
  Best Local Similarity   64.2%;  Pred. No. 8.5e-59;
  Matches  449;  Conservative    0;  Mismatches  249;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db      83133 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 83192

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db      83193 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 83252

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db      83253 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 83312

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db      83313 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 83372

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db      83373 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 83432

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db      83433 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 83492

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db      83493 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 83552

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| |||||||||| | |||||||||  ||||||
Db      83553 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 83612

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||          |  |||
Db      83613 TATATATATATATATATATATATATATATANNNNNNNNNNNNNNNNNNNNNNNNNNNNN 83672

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825

Db      83673 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 83732

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885
                    || ||| || |    || |   ||| | ||      || | |   ||
Db      83733 NNNNNNNNNNTTTATATATATATAATTTATATATTATATATATTTATATTTATATTATAT 83792

Qy        886 CACTATATATATATATAGTGAAAGCAGTGTAAACCTAAA 924
                | |||||||||||||   | |     | || |  || |
Db      83793 ATATTTATATATATATATATATATATATATATATATA 83831


RESULT 94
CR548629
LOCUS       CR548629              131550 bp    DNA     linear   HTG 06-JUL-2004
DEFINITION  Danio rerio clone DKEY-173G4, WORKING DRAFT SEQUENCE, 12 unordered
            pieces.
ACCESSION   CR548629
VERSION     CR548629.2  GI:50057711
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 131550)
  AUTHORS   Burton,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUL-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
```

```
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 8, 2004 this sequence version replaced gi:49658992.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK173G4
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 127208 bases at least Q40
            Consensus quality: 128066 bases at least Q30
            Consensus quality: 128642 bases at least Q20
            Insert size: 130450; sum-of-contigs
            Insert size: 145332; 1.1% error; agarose-fp
            Quality coverage: 7.95x in Q20 bases; sum-of-contigs Quality
            coverage: 7.21x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 12 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    18381: contig of 18381 bp in length
            *    18382    18481: gap of 100 bp
            *    18482    25435: contig of 6954 bp in length
            *    25436    25535: gap of 100 bp
            *    25536    29512: contig of 3977 bp in length
            *    29513    29612: gap of 100 bp
            *    29613    40983: contig of 11371 bp in length
            *    40984    41083: gap of 100 bp
            *    41084    56393: contig of 15310 bp in length
            *    56394    56493: gap of 100 bp
            *    56494    70939: contig of 14446 bp in length
            *    70940    71039: gap of 100 bp
            *    71040    80278: contig of 9239 bp in length
            *    80279    80378: gap of 100 bp
            *    80379    87641: contig of 7263 bp in length
            *    87642    87741: gap of 100 bp
            *    87742   103997: contig of 16256 bp in length
            *   103998   104097: gap of 100 bp
            *   104098   106761: contig of 2664 bp in length
            *   106762   106861: gap of 100 bp
            *   106862   110734: contig of 3873 bp in length
            *   110735   110834: gap of 100 bp
            *   110835   131550: contig of 20716 bp in length.
FEATURES             Location/Qualifiers
     source          1. .131550
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-173G4"
                     /clone_lib="DanioKey"
     misc_feature    1. .18381
                     /note="assembly_fragment:01412
                     fragment_chain:1"
     misc_feature    18482. .25435
                     /note="assembly_fragment:00193
                     fragment_chain:1"
     misc_feature    25536. .29512
                     /note="assembly_fragment:00013
                     fragment_chain:1"
     misc_feature    29613. .40983
                     /note="assembly_fragment:00264
                     fragment_chain:1"
     misc_feature    41084. .56393
                     /note="assembly_fragment:00930
                     fragment_chain:1"
     misc_feature    56494. .70939
                     /note="assembly_fragment:00480
                     fragment_chain:1"
     misc_feature    71040. .80278
                     /note="assembly_fragment:00367
                     fragment_chain:2"
     misc_feature    80379. .87641
                     /note="assembly_fragment:00063
                     fragment_chain:2"
     misc_feature    87742. .103997
                     /note="assembly_fragment:00696
                     fragment_chain:2"
     misc_feature    104098. .106761
                     /note="assembly_fragment:00032"
     misc_feature    106862. .110734
                     /note="assembly_fragment:00122.0"
     misc_feature    110835. .131550
                     /note="assembly_fragment:01164"
ORIGIN

  Query Match             8.5%;  Score 348.4;  DB 2;  Length 131550;
  Best Local Similarity   69.5%;  Pred. No. 1e-58;
  Matches  500;  Conservative    0;  Mismatches  217;  Indels    2;  Gaps    2;

Qy        208 CAGAAAAAAAAAAAGGCATTTTATATATATATATATATATATACACACACACACACAT 267
              || |||   || ||     |||||||||||||||||||||||||||||||||||||||
Db      86675 CAAAAAGGCAATAAATACATTTATATATATATATATATATATACACACACACACACAC 86734

Qy        268 ATATATATACACATATATATACACATATATACATATATACACATATATACACATATATAT 327
              | | | ||||| ||||||||| | ||||||| ||||||| | ||||||| | ||||||||
Db      86735 ACACACATACATATATATATATATATATATATATATATATATATATATATATATATAT 86794

Qy        328 ACACATACATATGTACACATATATATACACATATGTATACACATATATACACATATATAC 387
              | | ||| |||| || | |||||||||| | |||| |||| | ||||||| | |||||||
Db      86795 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 86854

Qy        388 ACACATATATACACATATATACACACATATATACACATATATACACATATATACACATAT 447
              | | ||||||| | ||||||| | | ||||||| | ||||||| | ||||||| | ||||
Db      86855 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 86914

Qy        448 ACACATATATACACATATATACATATATACACATATATATA-ATATACACACATATATAT 506
              | | ||||||| | ||||||| ||||||| | ||||||||| ||||| | | ||||||||
Db      86915 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 86974

Qy        507 ACACATATATACACACATATATACACATATATACACATATATATACACATATATACACAT 566
              | | ||||||| | | ||||||| | ||||||| | ||||||||| | ||||||| | ||
Db      86975 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 87034

Qy        567 ATATACATATATACACATATATATACATATATACACATATATACATATATACACATATAT 626
              ||||| ||||||| | ||||||||| ||||||| | ||||||| ||||||| | ||||||
Db      87035 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 87094

Qy        627 ACATATATACACACATATATACACATACATATACACACACATAGATATACATATATATAC 686
              | ||||||| | | ||||||| | ||| ||||| | | | ||| ||||| |||||||||
Db      87095 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 87154

Qy        687 ACATATATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAG 746
              | |||||||||  |||||||| ||||||||||        |  |||          |
Db      87155 ATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 87214

Qy        747 TTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGAC 806
              |   ||    |   ||| |   | | | |  | | |    | |||    | |   |
Db      87215 ATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 87274

Qy        807 ATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGT 866
              || |     |     | |||       | ||| ||| || | ||   | |  ||| | |
Db      87275 ATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 87334

Qy        867 A-AGGTAGTTAAGAAGGTAACACTATATATATATATAGTGAAAGCAGTGTAAACCTAAA 924
              | |  ||  ||   |  ||    |||||||||||||   | |    | || |  || |
Db      87335 ATATATATATATATATATATATATATATATATATATATATATATATATATATATNTATA 87393
```

```
RESULT 95
CR388149/c
LOCUS       CR388149             158085 bp    DNA     linear   HTG 19-AUG-2004
DEFINITION  Danio rerio clone DKEY-223J15, WORKING DRAFT SEQUENCE, 6 unordered
            pieces.
ACCESSION   CR388149
VERSION     CR388149.3  GI:51470665
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 158085)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-AUG-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 20, 2004 this sequence version replaced gi:46408652.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK223J15
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 156562 bases at least Q40
            Consensus quality: 156889 bases at least Q30
            Consensus quality: 157151 bases at least Q20
            Insert size: 157585; sum-of-contigs
            Insert size: 159679; 5.7% error; agarose-fp
            Quality coverage: 8.57x in Q20 bases; sum-of-contigs Quality
            coverage: 8.73x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    64210: contig of 64210 bp in length
            *    64211    64310: gap of 100 bp
            *    64311    82166: contig of 17856 bp in length
            *    82167    82266: gap of 100 bp
            *    82267    87201: contig of 4935 bp in length
            *    87202    87301: gap of 100 bp
            *    87302   120999: contig of 33698 bp in length
            *   121000   121099: gap of 100 bp
            *   121100   131190: contig of 10091 bp in length
            *   131191   131290: gap of 100 bp
            *   131291   158085: contig of 26795 bp in length.
FEATURES             Location/Qualifiers
     source          1. .158085
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-223J15"
                     /clone_lib="DanioKey"
     misc_feature    1. .64210
                     /note="assembly_fragment:01629
                     fragment_chain:1"
     misc_feature    64311. .82166
                     /note="assembly_fragment:00258
                     fragment_chain:1"
     misc_feature    82267. .87201
                     /note="assembly_fragment:00044
                     fragment_chain:1"
     misc_feature    87302. .120999
                     /note="assembly_fragment:00990
                     fragment_chain:1"
     misc_feature    121100. .131190
                     /note="assembly_fragment:00092
                     fragment_chain:1"
     misc_feature    131291. .158085
                     /note="assembly_fragment:00487.0"
ORIGIN

  Query Match             8.5%;  Score 348.4;  DB 2;  Length 158085;
  Best Local Similarity   67.9%;  Pred. No. 9.9e-59;
  Matches  515;  Conservative    0;  Mismatches  241;  Indels    2;  Gaps    2;

Qy        227 TTTATATATATATATATATATATATACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db      65123 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 65064

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | ||||||||| | ||| |||| || | |
Db      65063 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 65004

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db      65003 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 64944

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db      64943 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 64884

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db      64883 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 64824

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db      64823 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 64764

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db      64763 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 64704

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db      64703 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 64644

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||||          | |||          |  |   ||    |    |||
Db      64643 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 64584

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
              |  |||| | |  |  | |  | |||    | |   | || |     |      | |
Db      64583 TATAAAAATATATATATATATATATATATATATATATATATATATATATATATATATA 64524

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885
              ||      | ||| ||| || | ||  | |  | | | ||   ||  |||  |  ||
Db      64523 TATATATATATATATATATATATATATATATATATTTATATATAATAAATAAATAAAAAA 64464

Qy        886 CACTATATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAAT 945
              |||||||||||||  |   |  | | || |  ||||  |       ||  || ||| |
Db      64463 ATATATATATATATTATTTTATTA-TATATATATAAAAAAAAATAAAAATATATAATT 64405

Qy        946 GTTATAGAAGAATGGCTAAGATGCCAAAGCTCAGTGTA 983
              | | | |  |||   |||  |    | |  | | | ||
Db      64404 TTAAAATATAAATAATTAAAAAAATATATTTAATTATA 64367


RESULT 96
CR774185/c
LOCUS       CR774185             167023 bp    DNA     linear   HTG 01-OCT-2004
DEFINITION  Danio rerio clone CH211-237H18, WORKING DRAFT SEQUENCE, 11
            unordered pieces.
ACCESSION   CR774185
```

```
VERSION     CR774185.2  GI:52346261
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 167023)
  AUTHORS   McLaren,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 19, 2004 this sequence version replaced gi:52213970.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zC237H18
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 164787 bases at least Q40
            Consensus quality: 165392 bases at least Q30
            Consensus quality: 165717 bases at least Q20
            Insert size: 166023; sum-of-contigs
            Insert size: 188931; 10.3% error; agarose-fp
            Quality coverage: 9.96x in Q20 bases; sum-of-contigs Quality
            coverage: 8.81x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    21904: contig of 21904 bp in length
            *    21905    22004: gap of 100 bp
            *    22005    25287: contig of 3283 bp in length
            *    25288    25387: gap of 100 bp
            *    25388    69373: contig of 43986 bp in length
            *    69374    69473: gap of 100 bp
            *    69474    71725: contig of 2252 bp in length
            *    71726    71825: gap of 100 bp
            *    71826    74171: contig of 2346 bp in length
            *    74172    74271: gap of 100 bp
            *    74272    81210: contig of 6939 bp in length
            *    81211    81310: gap of 100 bp
            *    81311    91830: contig of 10520 bp in length
            *    91831    91930: gap of 100 bp
            *    91931    99897: contig of 7967 bp in length
            *    99898    99997: gap of 100 bp
            *    99998   111895: contig of 11898 bp in length
            *   111896   111995: gap of 100 bp
            *   111996   149605: contig of 37610 bp in length
            *   149606   149705: gap of 100 bp
            *   149706   167023: contig of 17318 bp in length.
FEATURES             Location/Qualifiers
     source          1. .167023
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-237H18"
                     /clone_lib="CHORI-211"
     misc_feature    1. .21904
                     /note="assembly_fragment:00892
                     clone_end:SP6
                     vector_side:left"
     misc_feature    22005. .25287
                     /note="assembly_fragment:00055
                     fragment_chain:1"
     misc_feature    25388. .69373
                     /note="assembly_fragment:01905
                     fragment_chain:1"
     misc_feature    69474. .71725
                     /note="assembly_fragment:00023"
     misc_feature    71826. .74171
                     /note="assembly_fragment:00031"
     misc_feature    74272. .81210
                     /note="assembly_fragment:00083
                     fragment_chain:2"
     misc_feature    81311. .91830
                     /note="assembly_fragment:00416
                     fragment_chain:2"
     misc_feature    91931. .99897
                     /note="assembly_fragment:00157
                     fragment_chain:2"
     misc_feature    99998. .111895
                     /note="assembly_fragment:00260
                     fragment_chain:2"
     misc_feature    111996. .149605
                     /note="assembly_fragment:01303
                     fragment_chain:2"
     misc_feature    149706. .167023
                     /note="assembly_fragment:00585
                     fragment_chain:2
                     clone_end:T7
                     vector_side:right"
ORIGIN

  Query Match             8.5%;  Score 348.4;  DB 2;  Length 167023;
  Best Local Similarity   64.6%;  Pred. No. 9.8e-59;
  Matches  461;  Conservative    0;  Mismatches  251;  Indels    2;  Gaps    2;

Qy        212 AAAAAAAAAAGGCATTTTATATATATATATATATATATATACACACACACACACATATAT 271
              ||  |  ||||   ||| ||||||||||||||||||||||||| | | | | | | ||||||
Db     100501 AATATTAAAAAAAATTATATATATATATATATATATATATATATATATATATATATATAT 100442

Qy        272 ATATACACATATATATACACATATATACATATATACACATATATACACATATATATACAC 331
              ||||| | |||||||||| | ||||||| ||||||| | ||||||| | |||||||| |
Db     100441 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 100382

Qy        332 ATACATATGTACACATATATATACACATATGTATACACATATATACACATATATACACAC 391
              ||| |||| || | ||||||||| | |||| |||| | ||||||| | ||||||| | |
Db     100381 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 100322

Qy        392 ATATATACACATATATACACACATATATACACATATATACACATATATACACATATACAC 451
              ||||||| | ||||||| | | ||||||| | ||||||| | ||||||| | ||||| |
Db     100321 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 100262

Qy        452 ATATATACACATATATACATATATACACATATATATAATATACACACATATATATACACA 511
              ||||||| | ||||||| ||||||| | |||||||| ||||| | | ||||||||| | |
Db     100261 ATATATATATATATATATATATATATATATATATAT-ATATATATATATATATATATATA 100203

Qy        512 TATATACACACATATATACACATATATACACATATATATACACATATATACACATATATA 571
              |||||| | | ||||||| | ||||||| | ||||||||| | ||||||| | |||||||
Db     100202 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 100143

Qy        572 CATATATACACATATATATACATATATACACATATATACATATATACACATATATACATA 631
              ||||||| | ||||||||| ||||||| | ||||||| ||||||| | ||||||| |||
Db     100142 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 100083

Qy        632 TATACACACATATATACACATACATATACACACACATAGATATACATATATATACACATA 691
              |||| | | ||||||| | ||| ||||| | | | ||| ||||| ||||||||| | |||
Db     100082 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 100023

Qy        692 TATATACGTATATATATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGAT 751
              ||||||  |||||||||| |||||||
Db     100022 TATATATATATATATATATATATATNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 99963

Qy        752 TACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTC 811
```

```
Db      99962 NNNN
```

```
            pieces.
ACCESSION   CR749743
VERSION     CR749743.4  GI:54021791
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 171424)
  AUTHORS   McLaren,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 10, 2004 this sequence version replaced gi:52213800.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK97M3
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 165755 bases at least Q40
            Consensus quality: 166566 bases at least Q30
            Consensus quality: 167672 bases at least Q20
            Insert size: 170624; sum-of-contigs
            Insert size: 189327; 3.5% error; agarose-fp
            Quality coverage: 8.39x in Q20 bases; sum-of-contigs Quality
            coverage: 7.61x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    13261: contig of 13261 bp in length
            *    13262    13361: gap of 100 bp
            *    13362    19777: contig of 6416 bp in length
            *    19778    19877: gap of 100 bp
            *    19878    33915: contig of 14038 bp in length
            *    33916    34015: gap of 100 bp
            *    34016    61771: contig of 27756 bp in length
            *    61772    61871: gap of 100 bp
            *    61872    94572: contig of 32701 bp in length
            *    94573    94672: gap of 100 bp
            *    94673   131525: contig of 36853 bp in length
            *   131526   131625: gap of 100 bp
            *   131626   133888: contig of 2263 bp in length
            *   133889   133988: gap of 100 bp
            *   133989   157725: contig of 23737 bp in length
            *   157726   157825: gap of 100 bp
            *   157826   171424: contig of 13599 bp in length.
FEATURES             Location/Qualifiers
     source          1. .171424
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-97M3"
                     /clone_lib="DanioKey"
     misc_feature    1. .13261
                     /note="assembly_fragment:00444
                     fragment_chain:1"
     misc_feature    13362. .19777
                     /note="assembly_fragment:00033
                     fragment_chain:1"
     misc_feature    19878. .33915
                     /note="assembly_fragment:00098
                     fragment_chain:1"
     misc_feature    34016. .61771
                     /note="assembly_fragment:00628
                     fragment_chain:2"
     misc_feature    61872. .94572
                     /note="assembly_fragment:01012
                     fragment_chain:2"
     misc_feature    94673. .131525
                     /note="assembly_fragment:01459
                     fragment_chain:2"
     misc_feature    131626. .133888
                     /note="assembly_fragment:00008
                     fragment_chain:3"
     misc_feature    133989. .157725
                     /note="assembly_fragment:01969
                     fragment_chain:3"
     misc_feature    157826. .171424
                     /note="assembly_fragment:00264.0"
ORIGIN

  Query Match             8.5%;  Score 348.4;  DB 2;  Length 171424;
  Best Local Similarity   67.6%;  Pred. No. 9.8e-59;
  Matches  495;  Conservative    0;  Mismatches  236;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db     134802 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 134743

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | |||||||||| | ||| |||| || | |
Db     134742 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 134683

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              ||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db     134682 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 134623

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db     134622 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 134563

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | |||||||| ||||| | | |||||||| | ||||||| | | |||
Db     134562 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 134503

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db     134502 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 134443

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              |||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | ||||||
Db     134442 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 134383

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db     134382 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 134323

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||         | |||          | |  ||   |    |||
Db     134322 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 134263

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
              |  | | | |  | | |   | |||   | |  | || |    | |   | |
Db     134262 TATATATATATATATATATATATATATATATATATATATATATATATATAAATATATA 134203

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885
              ||  |   | ||| ||| || | ||    |      ||     |  |  ||  |  ||
Db     134202 TATAAATATATATATATATATATTTATATATTTTAATAATATAANNAAAAAAATATATAN 134143

Qy        886 CACTATATATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAAT 945
              |||||||||||||| | |||   | || | |    ||    ||  |   | |
Db     134142 ATATATATATATATATATTTAAATATATATATATTTTTTTNATNTTAAAANNANAAANAN 134083
```

```
Qy        946 GTTATAGAAGAA 957
              | | || ||
Db     134082 AAAAAATAATAA 134071


RESULT 99
CR759841
LOCUS       CR759841              178975 bp    DNA     linear   HTG 08-SEP-2004
DEFINITION  Danio rerio clone DKEY-192G7, *** SEQUENCING IN PROGRESS ***, 6
            unordered pieces.
ACCESSION   CR759841
VERSION     CR759841.2  GI:51965207
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 178975)
  AUTHORS   Burton,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 9, 2004 this sequence version replaced gi:51890356.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK192G7
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 176373 bases at least Q40
            Consensus quality: 176964 bases at least Q30
            Consensus quality: 177367 bases at least Q20
            Insert size: 178475; sum-of-contigs
            Insert size: 178296; 1.0% error; agarose-fp
            Quality coverage: 11.15x in Q20 bases; sum-of-contigs Quality
            coverage: 11.48x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     3686: contig of 3686 bp in length
            *     3687     3786: gap of 100 bp
            *     3787    62427: contig of 58641 bp in length
            *    62428    62527: gap of 100 bp
            *    62528    84899: contig of 22372 bp in length
            *    84900    84999: gap of 100 bp
            *    85000   136604: contig of 51605 bp in length
            *   136605   136704: gap of 100 bp
            *   136705   158073: contig of 21369 bp in length
            *   158074   158173: gap of 100 bp
            *   158174   178975: contig of 20802 bp in length.
FEATURES             Location/Qualifiers
     source          1. .178975
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-192G7"
                     /clone_lib="DanioKey"
     misc_feature    1. .3686
                     /note="assembly_fragment:00051
                     fragment_chain:1"
     misc_feature    3787. .62427
                     /note="assembly_fragment:01908
                     fragment_chain:1"
     misc_feature    62528. .84899
                     /note="assembly_fragment:00400
                     fragment_chain:1"
     misc_feature    85000. .136604
                     /note="assembly_fragment:01093
                     fragment_chain:2"
     misc_feature    136705. .158073
                     /note="assembly_fragment:00722
                     fragment_chain:2"
     misc_feature    158174. .178975
                     /note="assembly_fragment:00096
                     fragment_chain:2"
ORIGIN

  Query Match             8.5%;  Score 348.4;  DB 2;  Length 178975;
  Best Local Similarity   60.4%;  Pred. No. 9.7e-59;
  Matches  492;  Conservative    0;  Mismatches  321;  Indels    1;  Gaps    1;

Qy        227 TTTATATATATATATATATATATATACACACACACACATATATATATACACATATATA 286
              | ||||||||||||||||||||||||| | | | | | |||||||||||| | |||||||
Db     135912 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 135971

Qy        287 TACACATATATACATATATACACATATATACACATATATATACACATACATATGTACACA 346
              || | ||||||| ||||||| | ||||||| | |||||||||| | ||| |||| || | |
Db     135972 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 136031

Qy        347 TATATATACACATATGTATACACATATATACACATATATACACACATATATACACATATA 406
              |||||||| | |||| |||| | ||||||| | ||||||| | | ||||||| | |||||
Db     136032 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 136091

Qy        407 TACACACATATATACACATATATACACATATATACACATATACACATATATACACATATA 466
              || | | ||||||| | ||||||| | ||||||| | ||||| | ||||||| | |||||
Db     136092 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 136151

Qy        467 TACATATATACACATATATATA-ATATACACACATATATATACACATATATACACACATA 525
              || ||||||| | ||||||||| ||||| | | ||||||||| | ||||||| | | |||
Db     136152 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 136211

Qy        526 TATACACATATATACACATATATATACACATATATACACATATATACATATATACACATA 585
              |||| | ||||||| | ||||||||| | ||||||| | ||||||| ||||||| | |||
Db     136212 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 136271

Qy        586 TATATACATATATACACATATATACATATATACACATATATACATATATACACACATATA 645
              ||||| ||||||| | ||||||| ||||||| | ||||||| ||||||| | | |||||
Db     136272 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 136331

Qy        646 TACACATACATATACACACACATAGATATACATATATATACACATATATATACGTATATA 705
              || | ||| ||||| | | | ||| ||||| ||||||||| | |||||||||  ||||||
Db     136332 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 136391

Qy        706 TATGTATATATATATGCTCCAGAGTTCATAAGAGGTAGCAGTTGATTACCACTGGGGATA 765
              ||| ||||||||||||          | |||          |  |   ||    |   |||
Db     136392 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 136451

Qy        766 GAGGAAAAGAGAGTTTGACAGCAGTGTATTGTGAGAAGGACATTTCAGGTTGATGGCAAA 825
              |   | | | |  | |    | |||    | |  | || |      |    | ||
Db     136452 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 136511

Qy        826 TAGTAGGGGAAATACATAAATGTGTAATAAAACCTATCTGTAAGGTAGTTAAGAAGGTAA 885
              ||      | ||| ||| || | ||   | |  ||| | ||      |  |  |  ||
Db     136512 TATATATATATATATATATATATATATATATATATATATATATATATATATAATATATAT 136571

Qy        886 CACTATATATATATAGTGAAAGCAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAAT 945
              ||||||||||||||   | |    | || |
Db     136572 ATATATATATATATATATATATATATATATANNNNNNNNNNNNNNNNNNNNNNNNNN 136631

Qy        946 GTTATAGAAGAATGGCTAAGATGCCAAAGCTCAGTGTATGTGGCAGAGGCATGGTGTAGG 1005

Db     136632 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 136691

Qy       1006 GTGTGTCCAGGTTCATATATTGCATTAAGTGTGA 1039
```

```
                          | | ||  ||||| | | |
Db     136692 NNNNNNNNNNNNNTTTTTTTTTTATTAATTATTA 136725


RESULT 100
CR792414
LOCUS       CR792414              179213 bp    DNA     linear   HTG 25-SEP-2004
DEFINITION  Danio rerio clone CH211-95O10, *** SEQUENCING IN PROGRESS ***, 9
            unordered pieces.
ACCESSION   CR792414
VERSION     CR792414.2  GI:52694443
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 179213)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 25, 2004 this sequence version replaced gi:52626621.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zC95O10
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 176437 bases at least Q40
            Consensus quality: 176944 bases at least Q30
            Consensus quality: 177295 bases at least Q20
            Insert size: 178413; sum-of-contigs
            Insert size: 195379; 9.0% error; agarose-fp
            Quality coverage: 8.87x in Q20 bases; sum-of-contigs Quality
            coverage: 8.31x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     4542: contig of 4542 bp in length
            *     4543     4642: gap of 100 bp
            *     4643    41427: contig of 36785 bp in length
            *    41428    41527: gap of 100 bp
            *    41528    50062: contig of 8535 bp in length
            *    50063    50162: gap of 100 bp
            *    50163    59923: contig of 9761 bp in length
            *    59924    60023: gap of 100 bp
            *    60024   103384: contig of 43361 bp in length
            *   103385   103484: gap of 100 bp
            *   103485   121057: contig of 17573 bp in length
            *   121058   121157: gap of 100 bp
            *   121158   148007: contig of 26850 bp in length
            *   148008   148107: gap of 100 bp
            *   148108   169047: contig of 20940 bp in length
            *   169048   169147: gap of 100 bp
            *   169148   179213: contig of 10066 bp in length.
FEATURES             Location/Qualifiers
     source          1. .179213
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-95O10"
                     /clone_lib="CHORI-211"
     misc_feature    1. .4542
                     /note="assembly_fragment:00045
                     fragment_chain:1
                     clone_end:T7
                     vector_side:left"
     misc_feature    4643. .41427
                     /note="assembly_fragment:01381
                     fragment_chain:1"
     misc_feature    41528. .50062
                     /note="assembly_fragment:00108
                     fragment_chain:1"
     misc_feature    50163. .59923
                     /note="assembly_fragment:00205
                     fragment_chain:1"
     misc_feature    60024. .103384
                     /note="assembly_fragment:01959"
     misc_feature    103485. .121057
                     /note="assembly_fragment:00463
                     fragment_chain:2"
     misc_feature    121158. .148007
                     /note="assembly_fragment:00984
                     fragment_chain:2"
     misc_feature    148108. .169047
                     /note="assembly_fragment:00719
                     fragment_chain:2"
     misc_feature    169148. .179213
                     /note="assembly_fragment:00316
                     fragment_chain:2
                     clone_end:SP6
                     vector_side:right"
ORIGIN

  Query Match             8.5%;  Score 348.4;  DB 2;  Length 179213;
  Best Local Similarity   66.8%;  Pred. No. 9.7e-59;
  Matches  524;  Conservative    0;  Mismatches  258;  Indels    2;  Gaps    2;

Qy        191 AGAGCGAAATGTCATCTCAGAAAAAAAAAAAGGCATTTTATATATATATATATATATATA 250
              |||  |  ||  ||  |   ||  |||||| |   |  |||||||||||||||||||||
Db     120224 AGACAGGCATTTCGACAGTGATTAAAAAACAACAACATTATATATATATATATATATATA 120283

Qy        251 TACACACACACACACATATATATATACACATATATATACACATATATACATATATACACA 310
              || | | | | | ||||||||||| | ||||||||| | ||||||| ||||||| | |
Db     120284 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 120343

Qy        311 TATATACACATATATATACACATACATATGTACACATATATATACACATATGTATACACA 370
              |||||| | |||||||||| | ||| |||| || | |||||||||| | |||| |||| | |
Db     120344 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 120403

Qy        371 TATATACACATATATACACACATATATACACATATATACACACATATATACACATATATA 430
              |||||| | ||||||| | | ||||||| | ||||||| | | ||||||| | ||||||||
Db     120404 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 120463

Qy        431 CACATATATACACATATACACATATATACACATATATACATATATACACATATATATAAT 490
              | ||||||| | ||||| | ||||||| | ||||||| ||||||| | |||||||| ||
Db     120464 TATATATATATATATATATATATATATATATATATATATATATATATATATATATAT-AT 120522

Qy        491 ATACACACATATATATACACATATATACACACATATATACACATATATACACATATATAT 550
              ||| | | |||||||||| | ||||||| | | ||||||| | ||||||| | ||||||||
Db     120523 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 120582

Qy        551 ACACATATATACACATATATACATATATACACATATATATACATATATACACATATATAC 610
              | | ||||||| | ||||||| ||||||| | ||||||||| ||||||| | |||||||
Db     120583 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 120642

Qy        611 ATATATACACATATATACATATATACACACATATATACACATACATATACACACACATAG 670
              ||||||| | ||||||| ||||||| | | ||||||| | ||| ||||| | | | |||
Db     120643 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 120702

Qy        671 ATATACATATATATACACATATATATACGTATATATATGTATATATATATGCTCCAGAGT 730
              ||||| ||||||||| | |||||||||  |||||||||| ||||||||||         |
Db     120703 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 120762
```

```
Qy        731 TCATAAGAGGTAGCAGTTGATTACCACTGGGGATAGAGGAAAAGAGAGTTTGACAGCAGT 790
              |||           |  |   ||    |    ||| |   | | | |  |  | |     |
Db     120763 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 120822

Qy        791 GTATTGTGAGAAGGACATTTCAGGTTGATGGCAAATAGTAGGGGAAATACATAAATGTGT 850
              |||    | |   | || |     |       | |||       | ||| ||| || | |
Db     120823 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 120882

Qy        851 AATAAAACCTATCTGTA-AGGTAGTTAAGAAGGTAACACTATATATATATATAGTGAAAG 909
              |   | |  ||| | || |  ||  ||   |  ||    |||||||||||||   |||
Db     120883 ATATATATATATATATATATATATATATATATATATATATATATATATATATATAAAT 120942

Qy        910 CAGTGTAAACCTAAAGGATGGGCCAAGGATTTAAATGTTATAGAAGAATGGCTAAGATGC 969
              | || |  || |   |      |   || || || |     || | |   ||  |
Db     120943 ATATATATATATATATATTTATATATATATATATATATATAAAAANTAAATATAAATA 121002

Qy        970 CAAA 973
              |||
Db     121003 NAAA 121006
```

Search completed: April 21, 2005, 17:01:58
Job time : 11574.3 secs

This Page Blank (uspto)

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 20, 2005, 22:52:20 ; Search time 2182.35 Seconds
(without alignments)
19011.657 Million cell updates/sec

Title: US-09-939-209A-3_COPY_11056_12145
Perfect score: 1090
Sequence: 1 tgctttccgtgtggtggtgc.....taactctaggatctcaggtg 1090

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_htc:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		129.6	11.9	719	9	CE498118	CE498118 tigr-gss-
2	c	121.6	11.2	733	9	CE251959	CE251959 tigr-gss-
3		100.4	9.2	527	9	CE406579	CE406579 tigr-gss-
4		96.4	8.8	825	5	BX924658	BX924658 BX924658
5		87.4	8.0	798	5	BX914754	BX914754 BX914754
6		68.4	6.3	1225	9	CNS0161D	AL106171 Drosophil
7	c	66	6.1	1101	9	CNS00EVL	AL069706 Drosophil
8	c	65.2	6.0	1896	9	CG753083	CG753083 P048-1-C0
9		63	5.8	1101	9	CNS00EVL	AL069706 Drosophil
10		61.2	5.6	1101	9	CNS00FMC	AL070972 Drosophil
11	c	60.8	5.6	1101	9	CNS003DQ	AL064580 Drosophil
12		60.4	5.5	1101	9	CNS00YWL	AL096927 Drosophil
13		60.2	5.5	910	9	CNS01G8P	AL142826 Anopheles
14		59.8	5.5	1101	9	CNS0039G	AL063921 Drosophil
15		59.8	5.5	1101	9	CNS00EO7	AL069440 Drosophil
16	c	59.2	5.4	895	9	CNS06FSV	AL396821 T7 end of
17	c	59	5.4	1626	7	CF238805	CF238805 AGENCOURT
18		58.8	5.4	836	9	CNS01100	AL099642 Drosophil
19		58.6	5.4	772	4	BM617963	BM617963 170006871
20	c	58.2	5.3	1101	9	CNS004ZW	AL055440 Drosophil
21	c	57.8	5.3	843	9	CNS00CS1	AL059666 Drosophil
22	c	57.8	5.3	1101	9	CNS0039G	AL063921 Drosophil
23		57.8	5.3	1579	9	AG311129	AG311129 Mus muscu
24		57.6	5.3	886	8	BH177277	BH177277 008_L_22-
25		57.6	5.3	886	9	CNS07JUX	AL614235 T3 end of
26	c	57.4	5.3	762	8	BH673700	BH673700 BOHUT26TF
27		57.4	5.3	1101	9	CNS00LT2	AL078714 Drosophil
28		57.2	5.2	1092	9	CNS020K7	AL175696 Tetraodon
29	c	56.6	5.2	935	9	CNS00KOI	AL078237 Drosophil
30	c	56.6	5.2	1254	3	CR720510	CR720510 Tetraodon
31	c	56.6	5.2	1313	7	CK997149	CK997149 ip15c02.b
32	c	56.4	5.2	979	9	CNS0161W	AL106190 Drosophil
33	c	56	5.1	1259	9	AG360787	AG360787 Mus muscu
34		55.8	5.1	967	9	CNS00KTZ	AL078030 Drosophil
35		55.8	5.1	1101	9	CNS00K1F	AL077002 Drosophil
36		55.8	5.1	1132	8	BZ695089	BZ695089 SP__Ba005
37	c	55.6	5.1	1043	9	CNS0145P	AL103735 Drosophil
38		55.6	5.1	1335	9	CL646802	CL646802 CH213-123
39	c	55.6	5.1	1807	9	CL515401	CL515401 SAIL_89_H
40		55.4	5.1	928	9	CNS00DKY	AL071865 Drosophil
41		55.4	5.1	1592	9	CG750135	CG750135 P044-3-D0
42	c	55.2	5.1	690	8	BH988719	BH988719 oee37h07.
43		55.2	5.1	1201	9	CNS015ZZ	AL106121 Drosophil
44		55.2	5.1	1626	7	CF238805	CF238805 AGENCOURT
45		55	5.0	996	9	CNS00FUH	AL071063 Drosophil

ALIGNMENTS

RESULT 1
CE498118
LOCUS CE498118 719 bp DNA linear GSS 28-SEP-2003
DEFINITION tigr-gss-dog-17000327316840 Dog Library Canis familiaris genomic, genomic survey sequence.
ACCESSION CE498118
VERSION CE498118.1 GI:36814899
KEYWORDS GSS.
SOURCE Canis familiaris (dog)
ORGANISM Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE 1 (bases 1 to 719)
AUTHORS Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K., Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and Venter,J.C.
TITLE The dog genome: survey sequencing and comparative analysis
JOURNAL Science 301 (5641), 1898-1903 (2003)
MEDLINE 22875432
PUBMED 14512627
COMMENT Contact: Kirkness EF
The Institute for Genomic Research
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-0200
Fax: 301-838-0208
Email: ekirknes@tigr.org
Class: shotgun.
FEATURES Location/Qualifiers
source 1. .719
/organism="Canis familiaris"
/mol_type="genomic DNA"
/strain="Standard Poodle"
/db_xref="taxon:9615"
/clone_lib="Dog Library"
/note="Site_1: BstXI; Libraries were prepared from peripheral blood"
ORIGIN

Query Match 11.9%; Score 129.6; DB 9; Length 719;
Best Local Similarity 63.7%; Pred. No. 7e-18;
Matches 321; Conservative 0; Mismatches 129; Indels 54; Gaps 6;

```
Qy        565 TAAAATGAAGCAGGAGGGTCTGGCAGAGAGTATCTATGAAATAAGTTCCTCTGACCTTCA 624
              | |||   |  ||||||||||| ||  ||||   ||| |   ||| || |||| || ||
Db          5 TTAAAGTGAAGCGGAGGGTCTGGTAGGAAGTACTCATGGAGGGAGTGCCCCTGAGCTCCA 64
```

```
Qy        625 CGCTTAATTTTCTGAATGGAGTGGAGCAAATTACTTCAAGCTTCACTTAACTTGCATATG 684
Db         65 TCATGACTATCCCGAGCTGAGTGGAGCAAATTAGTCTGAGCTTCAGTTAACTCATATATA 124

Qy        685 AAATGAACCGTACAAAAATACAA---------------------------------------- 707
Db        125 AAATTAACCATACAAAAATAAATAAAATAAAATAAAATAAAATAAACCATACAAAATATA 184

Qy        708 -GAGTGTCAGGAGAAAGTTATGCTCTGGTAAATATTTTGCAAAACAGATAAAAGATAATA 766
Db        185 GGAGTGTCAGGAAACCGTTATGCTCTGGTAAGTATTGCAGGAGACAGATGAATGAT---- 240

Qy        767 CTAGAGCTCTGTCCTCAAAGAGTTAAGCAGCTAATCTAAGGAGGTAAACTCTATGTCAGC 826
Db        241 -TGCACAGATCTGCCCAAGGAGCTAAGAAGCTAAGCTGGGGAGGTAACCTGTATATCAGC 299

Qy        827 AGGATGAACTGCTCTTCCCTTTCCTCCTCAATAAATTGCAAATCATCTAGTCCAACATCT 886
Db        300 AAAGTGAATGGCTCTTCCC--TCTTCACTGGTCAACTGCAAATCAACTAGTCTAGCACGT 357

Qy        887 T------TACCACCAGTGCCTGAGGCTCCAGAGGAGCCATTGCCTTCTCAAGGTCACATA 940
Db        358 TCCCCCCTAGCATCAGTGCCTGAGGCTGCAG-GAAGCCCTGAACTGCTCAAGGTCAAATA 416

Qy        941 GGTGGTGGGTGAGTTAGGACCA--AATCTAGAATTCCTGACTCCAGTAACTTCTGAAGTC 998
Db        417 GGCAGTGGATGGCCAGTGCACATCAAATCCGGTGTCATGACTCCAGTGACCTCTAAAGTC 476

Qy        999 ATTTTGTTTTTTATTTTTATGGTT 1022
Db        477 ATGCTGGCTTTTATTTTTATGGTT 500
```

```
RESULT 2
CE251959/c
LOCUS       CE251959                733 bp    DNA     linear   GSS 26-SEP-2003
DEFINITION  tigr-gss-dog-17000335916971 Dog Library Canis familiaris genomic,
            genomic survey sequence.
ACCESSION   CE251959
VERSION     CE251959.1  GI:35949172
KEYWORDS    GSS.
SOURCE      Canis familiaris (dog)
  ORGANISM  Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE   1  (bases 1 to 733)
  AUTHORS   Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
            Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
            Venter,J.C.
  TITLE     The dog genome: survey sequencing and comparative analysis
  JOURNAL   Science 301 (5641), 1898-1903 (2003)
  MEDLINE   22875432
   PUBMED   14512627
COMMENT     Contact: Kirkness EF
            The Institute for Genomic Research
            Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
            Rockville, MD 20850, USA
            Tel: 301-838-0200
            Fax: 301-838-0208
            Email: ekirknes@tigr.org
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .733
                     /organism="Canis familiaris"
                     /mol_type="genomic DNA"
                     /strain="Standard Poodle"
                     /db_xref="taxon:9615"
                     /clone_lib="Dog Library"
                     /note="Site_1: BstXI; Libraries were prepared from
                     peripheral blood"
ORIGIN

  Query Match             11.2%;  Score 121.6;  DB 9;  Length 733;
  Best Local Similarity   65.7%;  Pred. No. 4e-16;
  Matches  259;  Conservative    0;  Mismatches  119;  Indels   16;  Gaps    5;
```

```
Qy        637 TGAATGGAGTGGAGCAAATTACTTCAAGCTTCACTTAACTTGCATATGAAATGAACCGTA 696
Db        700 TAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAATAACCCAT 641

Qy        697 CAAAAATACAAGAGTGTCAGGAGAAAGTTATGCTCTGGTAAATATTTTGCAAAACAGATA 756
Db        640 ACAAAATATAGGAGTGTCAGGAAACCGTTATGCTCTGGTAAGTATTGCAGGAGACAGATG 581

Qy        757 AAAGATAATACTAGAGCTCTGTCCTCAAAGAGTTAAGCAGCTAATCTAAGGAGGTAAACT 816
Db        580 AATGAT-----TGCACAGATCTGCCCAAGGAGCTAAGAAGCTAAGCTGGGGAGGTAACCT 526

Qy        817 CTATGTCAGCAGGATGAACTGCTCTTCCCTTTCCTCCTCAATAAATTGCAAATCATCTAG 876
Db        525 GTATATCAGCAAAGTGAATGGCTCTTCCC--TCTTCACTGGTCAACTGCAAATCAACTAG 468

Qy        877 TCCAACATCTT------TACCACCAGTGCCTGAGGCTCCAGAGGAGCCATTGCCTTCTCA 930
Db        467 TCTAGCACGTTCCCCCCTAGCATCAGTGCCTGAGGCTGCAG-GAAGCCCTGAACTGCTCA 409

Qy        931 AGGTCACATAGGTGGTGGGTGAGTTAGGACCA--AATCTAGAATTCCTGACTCCAGTAAC 988
Db        408 AGGTCAAATAGGCAGTGGATGGCCAGTGCACATCAAATCCGGTGTCATGACTCCAGTGAC 349

Qy        989 TTCTGAAGTCATTTTGTTTTTTATTTTTATGGTT 1022
Db        348 CTCTAAAGTCATGCTGGCTTTTATTTTTATGGTT 315
```

```
RESULT 3
CE406579
LOCUS       CE406579                . 527 bp    DNA     linear   GSS 27-SEP-2003
DEFINITION  tigr-gss-dog-17000335605351 Dog Library Canis familiaris genomic,
            genomic survey sequence.
ACCESSION   CE406579
VERSION     CE406579.1  GI:36659296
KEYWORDS    GSS.
SOURCE      Canis familiaris (dog)
  ORGANISM  Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE   1  (bases 1 to 527)
  AUTHORS   Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
            Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
            Venter,J.C.
  TITLE     The dog genome: survey sequencing and comparative analysis
  JOURNAL   Science 301 (5641), 1898-1903 (2003)
  MEDLINE   22875432
   PUBMED   14512627
COMMENT     Contact: Kirkness EF
            The Institute for Genomic Research
            Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
            Rockville, MD 20850, USA
            Tel: 301-838-0200
            Fax: 301-838-0208
            Email: ekirknes@tigr.org
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .527
                     /organism="Canis familiaris"
                     /mol_type="genomic DNA"
                     /strain="Standard Poodle"
                     /db_xref="taxon:9615"
                     /clone_lib="Dog Library"
                     /note="Site_1: BstXI; Libraries were prepared from
                     peripheral blood"
ORIGIN

  Query Match             9.2%;  Score 100.4;  DB 9;  Length 527;
  Best Local Similarity   69.0%;  Pred. No. 1.8e-11;
```

```
  Matches  229;  Conservative    0;  Mismatches   86;  Indels   17;  Gaps    6;

Qy        699 AAAATACAAGAGTGTCAGGAGAAAGTTATGCTCTGGTAAATATTTTGCAAAACAGATAAA 758
              |||||| | ||||||| ||| |  ||||||||||||||| ||||     | |||||| ||
Db         11 AAAATATAGGAGTGTC-GGAAACCGTTATGCTCTGGTAAGTATTGCAGGAGACAGATGAA 69

Qy        759 AGATAATACTAGAGCTCTGTCCTCAAAGAGTTAAGCAGCTAATCTAAGGAGGTAAACTCT 818
              |||      |  |    | | | ||| ||| |||| |||||| ||  |||||||| || |
Db         70 TGAT-----TGCACAGATCTGCCCAAGGAGCTAAGAAGCTAAGCTGGGGAGGTAACCTGT 124

Qy        819 ATGTCAGCAGGATGAACTGCTCTTCCCTTTCCTCCTCAATAAATTGCAAATCATCTAGTC 878
              || ||||||   ||||  |||||||||  || ||     | || ||||||||| ||||||
Db        125 ATATCAGCAAAGTGAATGGCTCTTCCC--TCTTCACTGGTCAACTGCAAATCAACTAGTC 182

Qy        879 CAACATCTT------TACCACCAGTGCCTGAGGCTCCAGAGGAGCCATTGCCTTCTCAAG 932
              | ||  ||       || || ||||||||||||||| ||| | |||| |    || ||||||
Db        183 TAGCACGTTCCCCCCTAGCATCAGTGCCTGAGGCTGCAG-GAAGCCCTGAACTGCTCAAG 241

Qy        933 GTCACATAGGTGGTGGGTGAGTTAGGACCA--AATCTAGAATTCCTGACTCCAGTAACTT 990
              |||| |||||  |||| ||      |  ||  ||    |   || |||||||||| || |
Db        242 GTCAAATAGGCAGTGGATGGCCAGTGCACATCAAATCCGGTGTCATGACTCCAGTGACCT 301

Qy        991 CTGAAGTCATTTTGTTTTTTATTTTTATGGTT 1022
              || |||||||  ||  ||||||||||||||||
Db        302 CTAAAGTCATGCTGGCTTTTATTTTTATGGTT 333


RESULT 4
BX924658
LOCUS       BX924658                825 bp    mRNA    linear   EST 07-MAY-2004
DEFINITION  BX924658 Sus Scrofa library (scan) Sus scrofa cDNA clone
            scan0022d.p.16 5prim, mRNA sequence.
ACCESSION   BX924658
VERSION     BX924658.1  GI:41141506
KEYWORDS    EST.
SOURCE      Sus scrofa (pig)
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 825)
  AUTHORS   Bonnet,A., Tosser-Klopp,G., Benne,F., Cabau,C., Villeger,S.,
            Soares,M., Bonaldo,F. and Hatey,F.
  TITLE     A Pig Normalised Multi-Tissue cDNA Library
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Tosser-Klopp G
            Genetique Animale
            Institut National de la Recherche Agronomique
            Chemin de Borde-Rouge - Auzeville BP27, 31326 Castanet-Tolosan
            cedex, FRANCE
            Tel: 33 (0) 5.61.28.51.14
            Fax: 33 (0) 5.61.28.53.08
            Email: tosser@toulouse.inra.fr
            Sequence cleaned of vector, adaptator and repetitions. Contact us
            at sigenasupport@jouy.inra.fr to obtain the chromatogram of this
            sequence.
            Plate: 0022  row: p  column: 16.
FEATURES             Location/Qualifiers
     source          1. .825
                     /organism="Sus scrofa"
                     /mol_type="mRNA"
                     /db_xref="taxon:9823"
                     /clone="scan0022d.p.16"
                     /tissue_type="mixed"
                     /dev_stage="from embryos to adults"
                     /clone_lib="Sus Scrofa library (scan)"
                     /note="tissues: adipose tissue, brain, kidney, liver,
                     muscle, ovary, testis,  heart, hypothalamus, pancreas,
                     skin, spleen, thymus, placenta, pituitary  gland, seminal
                     vesicle, small intestine, uterus, adrenals, bulbo  uretral
                     gland, cerebral trunk, epididymis, female gonad,
                     gall-bladder,  hippocampus, large intestine, male gonad,
                     melanocytes, stomach, udder"
ORIGIN

  Query Match             8.8%;  Score 96.4;  DB 5;  Length 825;
  Best Local Similarity   76.6%;  Pred. No. 1.4e-10;
  Matches  170;  Conservative    0;  Mismatches   46;  Indels    6;  Gaps    4;

Qy          1 TGCTTTCCGTGTGGTGGTGCAGTAACTTTCACTCTCTATGTTCTTATAAGCAAATGTTAC 60
              |||| || ||||| || ||||||| ||||     |   ||||||||| ||||||||| |||
Db        590 TGCTGTCTGTGTGATGCTGCAGTATCTTT--TGCCTCATGTTCTTAAAAGCAAATGCTAC 647

Qy         61 AATGAGATATGAGTTTTAAAGCCAGATCTTCCTTATCTCTCTGCCCCAT-CTCTAGTTCT 119
              ||| | |||  ||||| |||    | || |   || ||| |||||| |   ||||||||||
Db        648 AATAAAATA--AGTTTCAAAAGTATATATCTATTCTCTTTCTGCCTCTGCCTCTAGTTCT 705

Qy        120 TGAAGTGTCTCATATGAGT-TTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGT 178
              |||| ||||  || |||||  |||||||||| |||||||||  |||||||||| ||  |
Db        706 TGAAATGTCATATGTGAGTGGTGGTTGAGAAGTATTGATCACAACAAATCAGTGAAGGGC 765

Qy        179 TTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATAT 220
              | |||| || |||||||||| |||||||||||||||| ||| |
Db        766 TCTGTAAAAAATCTCATCTTTAAGACATTGTTTTGTCAATGT 807


RESULT 5
BX914754
LOCUS       BX914754                798 bp    mRNA    linear   EST 07-MAY-2004
DEFINITION  BX914754 Sus Scrofa library (scan) Sus scrofa cDNA clone
            scan0028d.k.07 5prim, mRNA sequence.
ACCESSION   BX914754
VERSION     BX914754.1  GI:41131533
KEYWORDS    EST.
SOURCE      Sus scrofa (pig)
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 798)
  AUTHORS   Bonnet,A., Tosser-Klopp,G., Benne,F., Cabau,C., Villeger,S.,
            Soares,M., Bonaldo,F. and Hatey,F.
  TITLE     A Pig Normalised Multi-Tissue cDNA Library
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Tosser-Klopp G
            Genetique Animale
            Institut National de la Recherche Agronomique
            Chemin de Borde-Rouge - Auzeville BP27, 31326 Castanet-Tolosan
            cedex, FRANCE
            Tel: 33 (0) 5.61.28.51.14
            Fax: 33 (0) 5.61.28.53.08
            Email: tosser@toulouse.inra.fr
            Sequence cleaned of vector, adaptator and repetitions. Contact us
            at sigenasupport@jouy.inra.fr to obtain the chromatogram of this
            sequence.
            Plate: 0028  row: k  column: 7.
FEATURES             Location/Qualifiers
     source          1. .798
                     /organism="Sus scrofa"
                     /mol_type="mRNA"
                     /db_xref="taxon:9823"
                     /clone="scan0028d.k.07"
                     /tissue_type="mixed"
                     /dev_stage="from embryos to adults"
                     /clone_lib="Sus Scrofa library (scan)"
                     /note="tissues: adipose tissue, brain, kidney, liver,
                     muscle, ovary, testis,  heart, hypothalamus, pancreas,
                     skin, spleen, thymus, placenta, pituitary  gland, seminal
                     vesicle, small intestine, uterus, adrenals, bulbo  uretral
                     gland, cerebral trunk, epididymis, female gonad,
                     gall-bladder,  hippocampus, large intestine, male gonad,
                     melanocytes, stomach, udder"
ORIGIN

  Query Match             8.0%;  Score 87.4;  DB 5;  Length 798;
  Best Local Similarity   75.6%;  Pred. No. 1.3e-08;
  Matches  161;  Conservative    0;  Mismatches   46;  Indels    6;  Gaps    4;
```

```
Qy          1 TGCTTTCCGTGTGGTGGTGCAGTAACTTTCACTCTCTATGTTCTTATAAGCAAATGTTAC 60
              |||| || ||||| || ||||||| ||||    |   |||||||| |||||||| |||
Db        590 TGCTGTCTGTGTGATGCTGCAGTATCTTT--TGCCTCATGTTCTTAAAAGCAAATGCTAC 647

Qy         61 AATGAGATATGAGTTTTAAAGCCAGATCTTCCTTATCTCTCTGCCCCAT-CTCTAGTTCT 119
              ||| | |||  ||||| |||   | || |   || ||| |||||| |   ||||||||||
Db        648 AATAAAATA--AGTTTCAAAAGTATATATCTATTCTCTTTCTGCCTCTGCCTCTAGTTCT 705

Qy        120 TGAAGTGTCTCATATGAGT-TTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGT 178
              |||| ||||  || |||||  |||||||||| |||||||||  |||||||||| ||  |
Db        706 TGAAATGTCATATGTGAGTGGTGGTTGAGAAGTATTGATCACAACAAATCAGTGAAGGGC 765

Qy        179 TTTGTAGAAGATCTCATCTTAAAGACATTGTTT 211
              | |||| || |||||||||| |||||||  |||
Db        766 TCTGTAAAAAATCTCATCTTTAAGACATGTTTT 798


RESULT 6
CNS0161D
LOCUS       CNS0161D               1225 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN15C18 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL106171
VERSION     AL106171.1  GI:5620504
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1225)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1225
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN15C18"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : SP6"
ORIGIN

  Query Match             6.3%;  Score 68.4;  DB 9;  Length 1225;
  Best Local Similarity   28.3%;  Pred. No. 0.00021;
  Matches  143;  Conservative  122;  Mismatches  240;  Indels    0;  Gaps    0;

Qy         23 TAACTTTCACTCTCTATGTTCTTATAAGCAAATGTTACAATGAGATATGAGTTTTAAAGC 82
              | : : |:| |   |:| || ||  |  :   : |:  ||  |  ||       :  : :
Db        600 TTWTYATYAATAAWTWTTTTWTTTAAYYMCCKYTTWTAAAAAAATTAACCTCCMYYTHYY 659

Qy         83 CAGATCTTCCTTATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGG 142
               : |:::|| ||:| ::: | |: | :: |     | |     | : :  | |   |||
Db        660 AWTAYMWTCNTTMTWHYYTTCCYTCCYYCCCTCCCCCTCTCYTTYTYTTTYTHTCTTTTT 719

Qy        143 TTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAG 202
              ||    :|:|:   ::  :::||: :|  :  : :::  :: ::   ::  :  : :|:
Db        720 TTTTYTHAHAWCCKYYYCHMMAAMAMAYAKCCKBKKKNNKWCKMKMKHMAADAAKAMAMN 779

Qy        203 ACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTA 262
              : |:: |||: : |:   |     || |  |*:  ||:| | :  |:: | |  |  | :
Db        780 WAAHHWTTTYTWAAHAYNAAAAATTTAAAATCKCAAAMAMAAMAAAMMAAAAACCCMCCV 839

Qy        263 TTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCAT 322
                  ::::::::::::::   :|:|:::: | ||| ||    |   | | ||| |  ::
Db        840 CCKKKKDKKKKKKKKDDKDAAKRARADDRKAAAAAAAAATATAAAAAAAAAAAAAAAWW 899

Qy        323 ATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCA 382
              :  |:  ::|  : || |:  || : :  :| |    | |  | | :| :|  :     |
Db        900 WATTYTYYYAYAWAAAYAWATAAWWAWTTWTTAAAWWAAAAAAAAAAWAWWTTTYAAAAAA 959

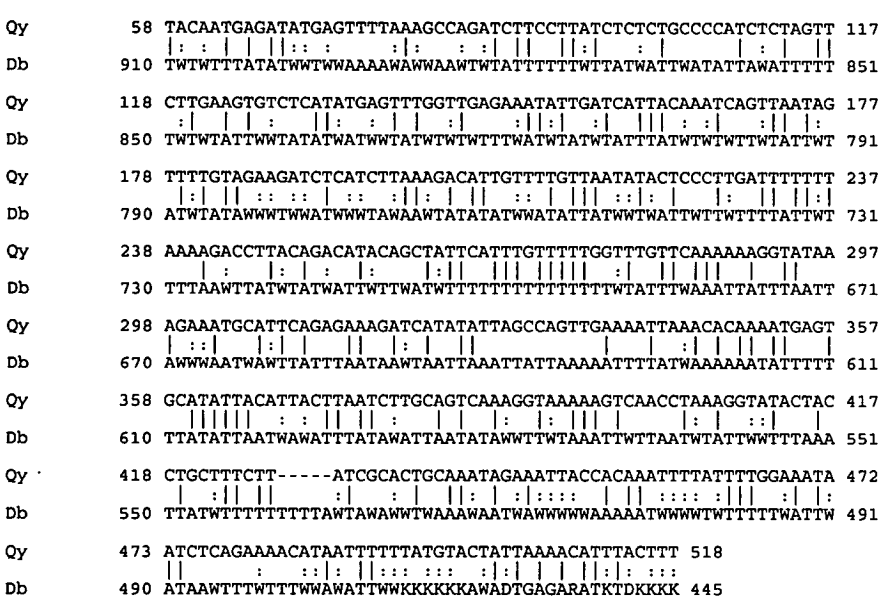
Qy        383 GTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATA 442
              |  :|  |||:   |     |  |  : |  |:  |  ||| |: |     |    :||
Db        960 WTWTWAAATAAWTTATTTTTTTTTAWWWTTTATWWATAATTTATWTTTWTTTTTTTTWTA 1019

Qy        443 GAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTA 502
              ||:  |  | ||  | | ||:|  :::|: : |    | | ||:  :|||||| | ::|
Db       1020 AAAWAAAAAAAAAATATATTTTWTWAWWWTWTWATTTTTATATATWTAYTTTTATATTMWA 1079

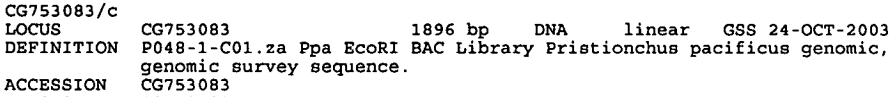
Qy        503 TTAAAACATTTACTTTCCAAATATT 527
              : ::|  :||:  |:|    :::|:
Db       1080 WAWWATWWTTWTWTWTTWTTWWWTW 1104


RESULT 7
CNS00EVL/c
LOCUS       CNS00EVL               1101 bp    DNA     linear   GSS 04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC:
            BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL069706
VERSION     AL069706.1  GI:4949849
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR29B23"
                     /clone_lib="RPCI-98"
                     /note="end : T7"
ORIGIN

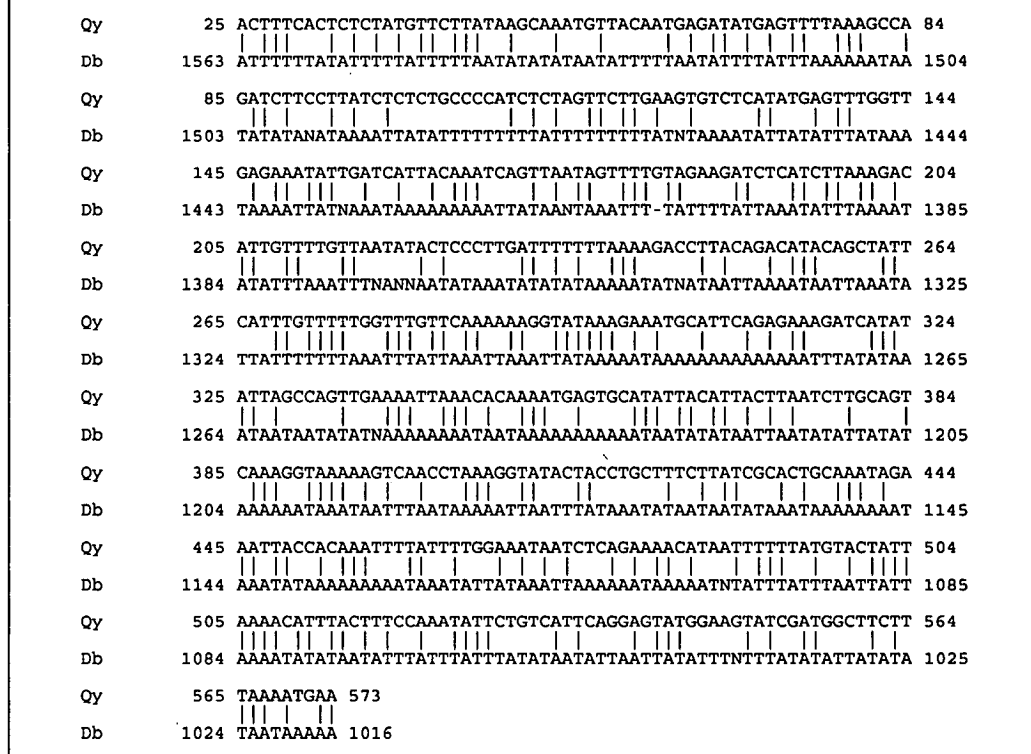
  Query Match             6.1%;  Score 66;  DB 9;  Length 1101;
  Best Local Similarity   33.3%;  Pred. No. 0.0007;
  Matches  155;  Conservative   92;  Mismatches  214;  Indels    5;  Gaps    1;
```



RESULT 8
CG753083/c
LOCUS CG753083 1896 bp DNA linear GSS 24-OCT-2003
DEFINITION P048-1-C01.za Ppa EcoRI BAC Library Pristionchus pacificus genomic, genomic survey sequence.
ACCESSION CG753083
VERSION CG753083.1 GI:37977199
KEYWORDS GSS.
SOURCE Pristionchus pacificus
ORGANISM Pristionchus pacificus
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida; Neodiplogasteridae; Pristionchus.
REFERENCE 1 (bases 1 to 1896)
AUTHORS Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K., Buntjer,J., van der Meulen,M. and Sommer,R.J.
TITLE An integrated physical and genetic map of the nematode Pristionchus pacificus
JOURNAL Mol. Genet. Genomics 269 (5), 715-722 (2003)
MEDLINE 22835951
PUBMED 12884007
COMMENT Contact: Sommer RJ
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498
Email: ralf.sommer@tuebingen.mpg.de
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .1896
/organism="Pristionchus pacificus"
/mol_type="genomic DNA"
/strain="California"
/db_xref="taxon:54126"
/clone_lib="Ppa EcoRI BAC Library"
/note="The library was generated by a partial digest of the genomic DNA with EcoRI and cloning into the BAC vector."
ORIGIN



Query Match 6.0%; Score 65.2; DB 9; Length 1896;
Best Local Similarity 45.4%; Pred. No. 0.0011;
Matches 249; Conservative 0; Mismatches 299; Indels 1; Gaps 1;



RESULT 9
CNS00EVL
LOCUS CNS00EVL 1101 bp DNA linear GSS 04-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC: BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL069706
VERSION AL069706.1 GI:4949849
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,

NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

```
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR29B23"
                     /clone_lib="RPCI-98"
                     /note="end : T7"
ORIGIN
```

```
  Query Match             5.8%;  Score 63;  DB 9;  Length 1101;
  Best Local Similarity   32.5%;  Pred. No. 0.0032;
  Matches  194;  Conservative  102;  Mismatches  297;  Indels    4;  Gaps    1;

Qy        131 ATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGAT 190
              |::: | ||| :: : :| |:|| :  :|:  : |  |    ||:|   ||  | :: ||
Db        502 AWWWWATTTTTWWWWWTWATTWTTTWAWWTWTAWTAAAAAAAAAWATAATTTAAAWWAAT 561

Qy        191 CTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACA 250
               :  |  :||    :   :| | |||||:||     :| |||  | |||||  |   |
Db        562 AWATTAAWAATTTAWAAWWTATATTAATWTATAAATWTWATTAATATAAAAAAATATTTT 621

Qy        251 GACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCA 310
               |||  |  | || ||   ||  |  :|| || || ||:    |::: || |  ||
Db        622 TTWATAAAATTTTTAATAATTTAATTAWTTATTAAATAAWTWATTWWWTAATTAAATAAT 681

Qy        311 GAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTA 370
              |||  |: | | |  |   |  :  :||: | :   :| |       |:| || | :|
Db        682 TTWAAATAWAAAAAAAAAAAAAAAWATWAAWAATWATAWATAAWTTAAAAWAATAAAAWA 741

Qy        371 CTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATC 430
               : | : ::  | |  ::    | |: |  |  :| :  :::  ||  |    | : |:
Db        742 AWAATWAWWATAATATWWATATATAWTTWTAWWWATWWAWWWTATAWATAWAATAWAAWA 801

Qy        431 GCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATT 490
               : |  | |:| |:||:|  : :|: | :: |:    |::|||     :|:| | |||:
Db        802 WAWATAAATAWATAWATWAAAWAWAWATAWWATWATATAWWAAT----AWAWAAAAAATW 857

Qy        491 TTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGTCATTCAGGAGTATGGAAGT 550
              |  |||  | :|| |:|| |  ||  :|  : :| || : : ||  |  : :||  || :
Db        858 TAATATWAATWATAAWAAAAAATAWAWTTWWTWTTTTWWAWWATATAAAWAWATAWAAAW 917

Qy        551 ATCGATGGCTTCTTTAAAATGAAGCAGGAGGGTCTGGCAGAGAGTATCTATGAAATAAGT 610
              |   |         | |:|::  :  |  :   |           :|| :|| ::|: :
Db        918 AAAAAAAAAAAAATAAWAWWWTWATATTWTTATTAAAWTWTATWWATTWATTWWAWTWTA 977

Qy        611 TCCTCTGACCTTCACGCTTAATTTTCTGAATGGAGTGGAGCAAATTACTTCAAGCTTCAC 670
              |  | : |  |  |   :: |:| : |  ||  | |  :  | |||   : ||     |
Db        978 TAWTTWTATWTATATATWWTAWTAWATATATTTATTAAWWTATATTTTAWAAAWTAATAT 1037

Qy        671 TTAACTTGCATATGAAATGAACCGTACAAAAATACAAGAGTGTCAGGAGAAAGTTAT 727
               |:|  :    :  | |: ::   ||    | ||| :  |     :  | |:|  :|:
Db       1038 ATWATAWWWTAWATATAWAWWAATTAWTTATATATWTAAWATAWWAAAAAWAAAWAW 1094
```

```
RESULT 10
CNS00FMC
LOCUS       CNS00FMC                1101 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC:
            BACR31N17 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL070972
VERSION     AL070972.1  GI:4950813
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR31N17"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN
```

```
  Query Match             5.6%;  Score 61.2;  DB 9;  Length 1101;
  Best Local Similarity   38.8%;  Pred. No. 0.008;
  Matches  119;  Conservative   46;  Mismatches  142;  Indels    0;  Gaps    0;

Qy        231 TTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAA 290
              |||||||: ||    :|| | | | | |  |||  |  | | | :  |:| |   : |:|
Db        773 TTTTTTTTWTAAATTTWTATATATAAAAAAAATATAAAAATWTATAWTWTWTTTATTWTAWA 832

Qy        291 GGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAA 350
               |:  ::  |||  ||| | : :|   |  |||| || : | :: ||:|||:|     |
Db        833 TTTWATWWATAATTTATTWAAVAWATATTATTATAATAABAAAWWTAAWATTWANATATA 892

Qy        351 AATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGT 410
              |:   |  :|||    : | :|| | |    |     |   |:| |  : :   : ||  |
Db        893 TAWATATAWWTATATTTWAATWTATTTTATWAAATTATTWTWATATTWAWTTWWTAAATT 952

Qy        411 ATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAA 470
              | ||  |  | | || |   :  |  | |:: ||: ||  | ||::  :: |:|  |||
Db        953 TAAATAAATTWTATATTTTTTWTTTWATAWWWAAWATAWAAWAAWWAAWWATWTAAAAA 1012

Qy        471 TAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTG 530
              ||: | : |||  | ||::|::||| :|   ||   : |::|  :|      ||:  |
Db       1013 AAAWATAWAAAATAAAAAWWTWWTATAWATATTTTTTWAAWWTTTWTATATTTTAWAWTA 1072

Qy        531 TCATTCA 537
               :  || |
Db       1073 WTTTTAA 1079
```

```
RESULT 11
CNS003DQ/c
LOCUS       CNS003DQ                1101 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR08I09 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL064580
VERSION     AL064580.1  GI:4941932
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
```

```
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR08I09"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN

  Query Match             5.6%;  Score 60.8;  DB 9;  Length 1101;
  Best Local Similarity   39.5%;  Pred. No. 0.0098;
  Matches  197;  Conservative   37;  Mismatches  265;  Indels    0;  Gaps    0;

Qy          27 TTTCACTCTCTATGTTCTTATAAGCAAATGTTACAATGAGATATGAGTTTTAAAGCCAGA 86
               ||| ||| | | : : :|  ::|  ||:   || :|  :  : : | ||||:|     |
Db         573 TTTTACTTTTTTYYYCYTCNYWAYAAAWAAATATWAATMACWTWAAATTTTWATTTAATT 514

Qy          87 TCTTCCTTATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGA 146
               |     ||     | |     | |  : || ||        |  |   | |||  |: |
Db         513 TAAATATTTAAATTTTTATTTTTATTCWATTTTTTTTCCCCWTTTTTATATTTTAATWWA 454

Qy         147 GAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACAT 206
               |:||| |  |||| |::  :    | :  :|  | |: |  | |  ||| | | ||
Db         453 ATTAWATTWAAAATTAWAWWAAWTAATAAWAAAWTAAAAAWAAAATAAAATTATAAAAAT 394

Qy         207 TGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCA 266
               ||||   |||| |     |:  ||| | ||||| |   || | || :  :||| :
Db         393 ATTTTTAAAAATAAATATWTTWTTTTTATATAAAAWAATATATAAATATTAWAAWATTWW 334

Qy         267 TTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATAT 326
               :| | |||  | | |    : :|  | |  :   ||   |      ::     |: :||
Db         333 AWTATATTTTTTATATATTTWTWATTTTTTTWTTTATATTTAWTWTTWWTATAWAWTWAT 274

Qy         327 TAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCA 386
               ||       | | ||: ||| | ||||  |     | | || |  | | ||  ||  | |
Db         273 TAAATTAATTACAAWAAAAAATAAAAAAAAAAAAAAAAATATAAAAATAAAATTTAAAATTT 214

Qy         387 AAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAA 446
               ||   | |||   ||   |||  | |  |          |    |||  | |||
Db         213 AAAAAATAAAAAAAAAAAAAAAAAAAAAACATAAAAAAAAAAAAAAAAAAAAAAAAAAA 154

Qy         447 TTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAA 506
               |    :| ||||  ||:  |||| |    | ||||  | ||     ::   |  |  ||
Db         153 AAAATTAWATTTTTCGTTYTTAAATTAAAAAAAAAAAAAAAAAAAAAAWWAAAAAAAAAA 94

Qy         507 AACATTTACTTTCCAAATA 525
               || |   | | |  ||| |
Db          93 AAAAAAAAATATAAAAAAA 75


RESULT 12
CNS00YWL
LOCUS       CNS00YWL                1101 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN01G13 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL096927
VERSION     AL096927.1  GI:5608538
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN01G13"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : SP6"
ORIGIN

  Query Match             5.5%;  Score 60.4;  DB 9;  Length 1101;
  Best Local Similarity   35.7%;  Pred. No. 0.012;
  Matches  131;  Conservative   69;  Mismatches  158;  Indels    9;  Gaps    1;

Qy         165 AATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTC 224
               ||  :: :|| :   |:|| | :  :  :  | :|:|::: :::::: :::  |   :|
Db         734 AARARRGDAAGRACNTKGTTGGRTGKAAMGGCMYAMAVVSSBBKKKKGKKKGGTKRTVTG 793

Qy         225 CCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTC 284
               :||| |||||| :||| :    | | |:||  |  ||   | |       :   |: |
Db         794 KMTTGTTTTTTTTAWAAATWAAAAAAATAYATNWAWATAAAAAATAWAARAWNWATWWTAT 853

Qy         285 AAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAA 344
               ||||::   | | | ||    :|  | | |:| |: |:||| :|   |  | |:||  ||
Db         854 AAAAWWAAAAAATATAATAARWTAAAAAAAWAAAWAAWATAAWATGTATATWAWAAAAAA 913

Qy         345 ACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCT 404
               : |  ||:: | |          |:|  |||  :   |   :||   :|||| : |
Db         914 WTATTAAWWAAWT---------AWWAKATAAAAWAAAAAARWAAWTAWAAAATWWATAWA 964

Qy         405 AAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTT 464
               |:  |||:         ||  :||     | |   | || || :||  : :|  | :: :
Db         965 TAWTATATWTATATATATTATWTAATATADTWATANTAWAATKTAWDWAWATATAWWAWA 1024

Qy         465 TGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAAT 524
               :  |||||:| | |  | : |||: :::||:| :| ||::|  | || ||  ::  ||:|
Db        1025 WAWAAATAWTATTAATATWAATAWNWWWTTWTAWAWTAKWATTATATATAWNWWKWAAWT 1084

Qy         525 ATTCTGT 531
               | | |
Db        1085 TATNTAT 1091
```

```
RESULT 13
CNS01G8P
LOCUS       CNS01G8P                910 bp    DNA     linear   GSS 01-JUN-2001
DEFINITION  Anopheles gambiae GSS SP6 end of clone 06E14 of NotreDame1 library
            from strain PEST of Anopheles gambiae (African malaria mosquito),
            genomic survey sequence.
ACCESSION   AL142826
VERSION     AL142826.1  GI:7000944
KEYWORDS    GSS.
SOURCE      Anopheles gambiae (African malaria mosquito)
  ORGANISM  Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
            Anopheles.
REFERENCE   1  (bases 1 to 910)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
REFERENCE   2  (bases 1 to 910)
  AUTHORS   Roth,C.W., Brey,P.T., Ke,Z., Collins,F.H. and Weissenbach,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-FEB-2000) BBMI, Institut Pasteur, 25, rue du Dr.
            Roux, Paris 75015, France
COMMENT     This clone is from an A. gambiae BAC library provided by F.H.
            Collins and sequenced by Genoscope in collaboration with the
            Laboratory of Biochem. and Biol. Molec. of Insects, Institut
            Pasteur.
FEATURES             Location/Qualifiers
     source          1. .910
                     /organism="Anopheles gambiae"
                     /mol_type="genomic DNA"
                     /strain="PEST"
                     /db_xref="taxon:7165"
                     /clone="06E14"
                     /clone_lib="NotreDame1"
                     /note="end : SP6"
ORIGIN

  Query Match             5.5%;  Score 60.2;  DB 9;  Length 910;
  Best Local Similarity   41.4%;  Pred. No. 0.013;
  Matches  207;  Conservative   36;  Mismatches  249;  Indels    8;  Gaps    2;

Qy         27 TTTCACTCTCTATGTTCTTATAAGCAAATGTTACAATGAGATATGAGTTTTAAAGCCAGA 86
Db        411 TNTATTTTTTTAWATTTATATNWAAATWAATTATAATAAAATATATTTTTTAAAAATTAA 470

Qy         87 TCTTCCTTATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGA 146
Db        471 TCWTTTTTAATTAAATAAWTATTNNAAAAAAWAAAAAWWTTTTWAWTATAAWATTTTWAA 530

Qy        147 GAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACAT 206
Db        531 AAAATATATATWTTTTTAAAWAATWTTA---ATATAAAAAAAATTTAATATWTWAATTAA 587

Qy        207 TGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCA 266
Db        588 AAAWWAAATTTTTTTATTATTAWTAATTWAANAAAATAATAAAAAAAATAAAAAAAAAAN 647

Qy        267 TTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATAT 326
Db        648 TTATWWWTTTAATTATATAAWAAAAAATTTTATAATAAWTTTATAAAAAAWATTTWAATW 707

Qy        327 TAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCA 386
Db        708 TACTTATAWATAAAAAAAAATATTWAAAAAWAAAAATATAATAATAAAATAANAAAAAAA 767

Qy        387 AAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAA 446
Db        768 AATANAAATTNWTAAAAAAATTWAWNANAA-----TTAYTWAACAAWNTNTWTATAAAAA 822

Qy        447 TTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAA 506
Db        823 TATNYWTAAAATAAWWTTTAATWTTNATTTTAATATCATTATTTTTATAATTTATATTAA 882

Qy        507 AACATTTACTTTCCAAATAT 526
Db        883 AATAATTATATCTAAATAWT 902


RESULT 14
CNS0039G
LOCUS       CNS0039G                1101 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL063921
VERSION     AL063921.1  GI:4941778
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR08K10"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN

  Query Match             5.5%;  Score 59.8;  DB 9;  Length 1101;
  Best Local Similarity   17.2%;  Pred. No. 0.016;
  Matches  120;  Conservative  284;  Mismatches  292;  Indels    3;  Gaps    1;

Qy          9 GTGTGGTGGTGCAGTAACTTTCACTCTCTATGTTCTTATAAGCAAATGTTACAATGAGAT 68
Db        385 GGGRGGGGGGMATATAWAWWWWWTTTTTTTTAWAAAWAAAATAATTWWAAWAWAAAAAT 444

Qy         69 ATGAGTTTTAAAGCCAGATCTTCCTTATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTC 128
Db        445 TWWAAAAWAAAAWAWTAWWTTTAWTWAAAAAAAAAAAATTWTTTTTTTTWTTTAWTTWATA 504

Qy        129 TCATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAG 188
Db        505 WWTTWWWTTAAAWAAAAAAAAAAAWAAAWAAAWATAAATWTWWTWWTTYTTWAAWATAAA 564

Qy        189 ATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTA 248
```

```
Db        565 MCMAAWYYHTYTTYTYHYYTYWTYTMTWHYHTMYTHAWAHTTWYHWYHTYAMWHWMTWHW 624

Qy        249 CAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATT 308

Db        625 TMWAWWHWTTYTAAYYYYYTCMYYYHYMHWHHAHAHAAAWWTTHTWWTHAYHWATYHYYY 684

Qy        309 CAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACAT 368

Db        685 YMYC---AMMCMCTHTCHHCYYYYHHYTAHHTHTHHWYAHYYMWYYMWAYYWMYCTACTY 741

Qy        369 TACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTA 428

Db        742 HYHHHHHYHWAYHTTWYAWAHAMWMWHHAHYAAAAAWAAWATTHHYHHTTHYMHHTYMYH 801

Qy        429 TCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAA 488

Db        802 YYMYTCCYMCTYHCWHYYHTAYTCWTWTHHWMWTWTHWYHHTWWHHTTTHWAWWHTHTWC 861

Qy        489 TTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGTCATTCAGGAGTATGGAA 548

Db        862 WWWWHATTWTWATHCWACMTMHWHHWMHMHHHHMACHAHHHTHMCMCHHHMHCTCHHHHT 921

Qy        549 GTATCGATGGCTTCTTTAAAATGAAGCAGGAGGGTCTGGCAGAGAGTATCTATGAAATAA 608

Db        922 MYHMTCHWMWHMHWHWHHWWMATWMTTMTTMMMCCMMHHHCHMYHMMHMYMYCCHYYCT 981

Qy        609 GTTCCTCTGACCTTCACGCTTAATTTTCTGAATGGAGTGGAGCAAATTACTTCAAGCTTC 668

Db        982 CHTHATTHYHYMCTCYHYCTWHTYWTAYWWAWTAHAMTTATWWWWMHWWAHWATWWWWWW 1041

Qy        669 ACTTAACTTGCATATGAAATGAACCGTACAAAAATACAA 707

Db       1042 WWATAWWACTCHHWTWYHTHCTWYYHHTYHMWWAWWMAW 1080
```

RESULT 15
CNS00EO7
LOCUS CNS00EO7 1101 bp DNA linear GSS 04-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC: BACR29P01 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL069440
VERSION AL069440.1 GI:4949583
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR29P01"
/clone_lib="RPCI-98"
/note="end : TET3"
ORIGIN

```
  Query Match              5.5%;  Score 59.8;  DB 9;  Length 1101;
  Best Local Similarity   35.6%;  Pred. No. 0.016;
  Matches  134;  Conservative   59;  Mismatches  183;  Indels    0;  Gaps    0;

Qy        151 TATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTT 210

Db        564 TATWWWATATAAAAAWAATTTAATTTTTWATTTTATATATTWAAWGWWAAAAAAAWWTAT 623

Qy        211 TTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTG 270

Db        624 AAWHTWTAAATTTATAWTTWAWATTTTAAATAWATTTATTTAAWTTAWWTWTTTAAWAWT 683

Qy        271 TTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGC 330

Db        684 TTAAWTATTAAARWWAATTWAAAAWAAAAWTTTWAATAAWAAATAAAAATAAATTTTTTT 743

Qy        331 CAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGG 390

Db        744 TATTTAAWWAWTTAAAATTAWTWATWAAWAWTTWTAWAAAAAAAAATNWTATAAWATTWWK 803

Qy        391 TAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTAC 450

Db        804 AATAATWTATAWWTTTAWATATWATRTAAATTWTAATWYTTTTTWTTTAWTTTAAWATTT 863

Qy        451 CACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACA 510

Db        864 TTTTAWWTAAWAWTGTAATWAAATTAWWTTAAAAAWAATGWTTTWTAAATTWWGTTAWTT 923

Qy        511 TTTACTTTCCAAATAT 526

Db        924 TTTATTTTWAAATTGT 939
```

RESULT 16
CNS06FSV/c
LOCUS CNS06FSV 895 bp DNA linear GSS 17-JUN-2001
DEFINITION T7 end of clone AR0AA032H05 of library AR0AA from strain CBS 732 of Zygosaccharomyces rouxii, genomic survey sequence.
ACCESSION AL396821
VERSION AL396821.1 GI:12149222
KEYWORDS GSS.
SOURCE Zygosaccharomyces rouxii
ORGANISM Zygosaccharomyces rouxii
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Zygosaccharomyces.
REFERENCE 1 (bases 1 to 895)
AUTHORS Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G., Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S., de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B., Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S., Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M., Wincker,P. and Weissenbach,J.
TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies
JOURNAL FEBS Lett. 487 (1), 3-12 (2000)
MEDLINE 20584711
PUBMED 11152876
REFERENCE 2 (bases 1 to 895)
AUTHORS de Montigny,J., Straub,M., Potier,S., Tekaia,F., Dujon,B., Wincker,P., Artiguenave,F. and Souciet,J.
TITLE Genomic exploration of the hemiascomycetous yeasts: 8. Zygosaccharomyces rouxii
JOURNAL FEBS Lett. 487 (1), 52-55 (2000)
MEDLINE 20584718
PUBMED 11152883
REFERENCE 3 (bases 1 to 895)

```
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT     This GSS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var.  hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES             Location/Qualifiers
     source          1. .895
                     /organism="Zygosaccharomyces rouxii"
                     /mol_type="genomic DNA"
                     /strain="CBS 732"
                     /db_xref="taxon:4956"
                     /clone="AR0AA032H05"
                     /clone_lib="AR0AA"
                     /note="end : T7"
     tRNA            <19. .>93
                     /product="tRNA-Thr"
                     /note="putative mitochondrial tRNA_thr gene"
                     /evidence=not_experimental
ORIGIN

  Query Match             5.4%;  Score 59.2;  DB 9;  Length 895;
  Best Local Similarity   42.1%;  Pred. No. 0.022;
  Matches  198;  Conservative   19;  Mismatches  252;  Indels    1;  Gaps    1;

Qy         75 TTTAAAGCCAGATCTTCCTTATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATAT 134
              | || |   |  | ||  || | | |      | | |  || ||    | | |  | |
Db        887 TATATATATATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 828

Qy        135 GAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCA 194
              ||:  ||     | ||  |  |    ||  | ||  ||  | | |   |  | | |
Db        827 TTTTTWTTTTTTTTTTTTTTWTTTTTNNATAAATATTTTTTATATATATTTTATTTTTTA 768

Qy        195 TCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACA 254
              | |        : ||||  |       |    ||  ||| | || :| |  ||| | | |
Db        767 TATANTTATATAWTTTTTTNTNTATATTTTTTTTTTTTTATATATWANAAWTTATATATA 708

Qy        255 TACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGA 314
              || |     |   |   |::|   |   | || |    |||  | ||||  ||  | |
Db        707 TATATNATATATATATATWWTATNTAAATWAATATNTATATTTATAAATATATATATATT 648

Qy        315 AAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTA 374
              |   | |||||||     :| :|| ||| : | | | |   |  :|||:   |:| ||
Db        647 ATANTTATATATTTATATAWTTWAATTTATWWAWATANTATNTNWWTATWNTNTWATTTT 588

Qy        375 ATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCAC 434
              || |     | ||       | | : |   | ||  |||| :|  |   :: || |
Db        587 ATWTATATATAAATATATTATATWTATANTTAATATATATWATTTTAAWWTTTNTTATTT 528

Qy        435 TGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTT 494
              |   :||  | ||||     ||| |  ||:|  | |||   | | |  | ||| || |||
Db        527 TAATWAT-TATATTATATATAATATATTTWTTWATATATATTTATATNATATATTTATTT 469

Qy        495 ATGTACTATTAAAACATTTACTTTCCAAATATTCTGTCATTCAGGAGTAT 544
              ||  : | ||: |   ||   ||    |||| |   |   | |  | |||
Db        468 ATNAWATNTTWTATATTTATATTATATAATAATTATTATATAATTATTAT 419


RESULT 17
CF238805/c
LOCUS       CF238805               1626 bp    mRNA    linear   EST 05-AUG-2003
DEFINITION  AGENCOURT_15099447 NICHD_XGC_Emb6 Xenopus tropicalis cDNA clone
            IMAGE:6995950 5', mRNA sequence.
ACCESSION   CF238805
VERSION     CF238805.1  GI:33442013
KEYWORDS    EST.
SOURCE      Xenopus tropicalis (western clawed frog)
  ORGANISM  Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 1626)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
            National Cancer Institute / NIH
            Bldg. 31 Rm10A07 Bethesda, MD 20892
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Robert M. Grainger
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM14680  row: g  column: 21
            High quality sequence start: 71
            High quality sequence stop: 316.
FEATURES             Location/Qualifiers
     source          1. .1626
                     /organism="Xenopus tropicalis"
                     /mol_type="mRNA"
                     /db_xref="taxon:8364"
                     /clone="IMAGE:6995950"
                     /tissue_type="neurula"
                     /dev_stage="embryo, stages 14-19"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NICHD_XGC_Emb6"
                     /note="Vector: pCMV-SPORT6.1; Site_1: NotI; Site_2: EcoRV;
                     Cloned unidirectionally. Primer: Oligo dT. Average insert
                     size 2.1 kb. Constructed by Invitrogen. Note: This is a
                     Xenopus Gene Collection (XGC) library."
ORIGIN

  Query Match             5.4%;  Score 59;  DB 7;  Length 1626;
  Best Local Similarity   42.2%;  Pred. No. 0.025;
  Matches  218;  Conservative    0;  Mismatches  291;  Indels    8;  Gaps    1;

Qy         27 TTTCACTCTCTATGTTCTTATAAGCAAATGTTACAATGAGATATGAGTTTTAAAGCCAGA 86
              |||       | | || ||| ||      | ||    |  | |   | |
Db        845 TTTTTANTNNTTTTTTTTAAAAAATTTTTTTTTTTTTTTTTTTTTTNNTTTTNNTTT 786

Qy         87 TCTTCCTTATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGA 146
              | ||  || | | | |      | | || ||    ||   | |  | |   |||  ||
Db        785 TTTTTTTTTTTTTTTTTTTTTTTTTTTTAATTNNATAAAATTTTNTTTTTTTTTTTTTTT 726

Qy        147 GAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACAT 206
              | ||  | | |   ||| | |   |  |||| |   | || |  | ||       |
Db        725 TTTTTTTTTTTACTTTTAATTATTATTTTTTTTTTTTTNANATTTNTTNTTTTTTTTTT 666

Qy        207 TGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCA 266
              | |||| |   | |  |   ||  |||||||        |   | | |||    ||||
Db        665 TTTTTTTTATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTANNANANATAATTATATTAT 606

Qy        267 TTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATAT 326
              ||| |||||  ||| ||           ||| ||||  |  | |       |  | | |
Db        605 TTTTTTTTTTTTTTTTTTTTNNNNNANNNNAAAAAAAATATTTATAAATTTTTTTTTTTTT 546

Qy        327 T--------AGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCT 378
              |                 ||||  |||   |||  |   |   | ||   ||  ||
Db        545 TTTTTNTNNNNNNNNNAAAAAAAAAATTAAAATTTTTTTTTTTTTTTTTTTTNNNNA 486

Qy        379 TGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCA 438
              |   |||  |||||   ||  |||  |||  |   |  ||| |||    |   |
Db        485 NANAAAAAAAAAAAAAAAANAAAAAAAAAATTTATNTTTTTTTTTTTTTTAANAAAANAAA 426
```

```
Qy        439 AATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGT 498
              ||   ||   |  | |||  || ||||  | |||     | | |     |||||||  |
Db        425 AAANAAANAAANAAAAAAAATTTTTTTATATATATCTCTATATATATATTTTTTTTTN 366

Qy        499 ACTATTAAAACATTTACTTTCCAAATATTCTGTCATT 535
              | |  | | |  ||  |       | || | |  ||
Db        365 TNTNTATATATANATATNTNTTTTTTTTTTTTNTNTTT 329


RESULT 18
CNS01100
LOCUS       CNS01100                836 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN05N18 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL099642
VERSION     AL099642.1  GI:5611253
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 836)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .836
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN05N18"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : T7"
ORIGIN

  Query Match             5.4%;  Score 58.8;  DB 9;  Length 836;
  Best Local Similarity   44.2%;  Pred. No. 0.026;
  Matches  168;  Conservative   15;  Mismatches  197;  Indels    0;  Gaps    0;

Qy        148 AAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATT 207
              ||| |||  |  |: | |||  | ||  ||     | |    | || | ||  |  |     ||
Db         17 AAAAATTTAAWAWAATAAAAAATTTTTTAAAAATATTTTAAATTTTATTATTTAATTTTT 76

Qy        208 GTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCAT 267
              | || || | ||| |    | | ||| || |||      | | :   | |   |:| |
Db         77 TTATTTTTTAAATATTTTAATAAATTTATTTAAATTTTATTTAAWTTAATAATAAWTAAA 136

Qy        268 TTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATT 327
                |  |  |||  |  |||||     || |    |    |:: | |    | || |:|| ||
Db        137 AAAATAATATTTTAATTTAAAAATTATTATATTTAAAAAWWTAAATTTATATAAWATTTT 196

Qy        328 AGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAA 387
                |    || |  ||| | ||||  | |  || |    | |        ||          |
Db        197 TTTAAAAAAAATAAAAAAWAAAAAAAAATTTAATTTATTTTAAAAATTAATTAATTGATA 256

Qy        388 AGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAAT 447
              |  ||| ||   ||  |  |  | | |  |   |   ||    |     || | ||||
Db        257 AATTAATAAAATAAAATTTAATAAAAAAAWATTTATATATAGAAAAAAAATAAAAAAAT 316

Qy        448 TACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAA 507
              ||  | ||||||::|| |  ||| ||  :   |||| | : |    :|  :|  |:|||| |
Db        317 TAAWAWAAATTTWWTTATTTAAAAAAAWWATWAAAAAAAWTTAAAAWAAAWAAAAWTAAA 376

Qy        508 ACATTTACTTTCCAAATATT 527
                ::||   ||  ||| | |
Db        377 TTWWTTTAATTATAAAAAAT 396


RESULT 19
BM617963
LOCUS       BM617963                772 bp    mRNA    linear   EST 25-FEB-2002
DEFINITION  17000687160179 A.Gam.ad.cDNA.blood1 Anopheles gambiae cDNA clone
            19600449700993 5', mRNA sequence.
ACCESSION   BM617963
VERSION     BM617963.1  GI:18916173
KEYWORDS    EST.
SOURCE      Anopheles gambiae (African malaria mosquito)
  ORGANISM  Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
            Anopheles.
REFERENCE   1  (bases 1 to 772)
  AUTHORS   Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L.,
            Charlab,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
  TITLE     Celera Anopheles gambiae EST project
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Holt R.A.
            Celera Genomics
            45 w. Gude Dr., Rockville, MD 20850, USA
            Tel: 2404533151
            Fax: 2404534580
            Email: HoltRA@celera.com
            Plate: NU01004AA1  row: B  column: 19
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .772
                     /organism="Anopheles gambiae"
                     /mol_type="mRNA"
                     /strain="RSP-ST (Reduced susc. to Permethrin - std.
                     chromosome)"
                     /db_xref="taxon:7165"
                     /clone="19600449700993"
                     /dev_stage="Adult"
                     /lab_host="DH10b"
                     /clone_lib="A.Gam.ad.cDNA.blood1"
                     /note="Vector: pSport1; Site_1: Sal1; Site_2: Not1; Whole
                     adult mosquitoes (mixed sex) frozen on liquid nitrogen 24
                     hours  after human blood feeding. cDNA inserts >500 bp
                     cloned directionally  into pSport 1. Not 1 site is 3'.
                     Clones available through the Malaria  Research and
                     Reference Reagent Resource Center (www.malaria.mr4.org)"
ORIGIN

  Query Match             5.4%;  Score 58.6;  DB 4;  Length 772;
  Best Local Similarity   46.0%;  Pred. No. 0.029;
  Matches  199;  Conservative    0;  Mismatches  234;  Indels    0;  Gaps    0;

Qy        141 GGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAA 200
              | ||   |||||  | |     |   ||| | ||        | |  |||||  | ||  ||
Db        173 GTTTATTAAATAATAAATTATTATAATAATTTTTCTCGAATTTTAAAGATTTAATTATAT 232

Qy        201 AGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGC 260
              | | |   | ||||  |   |||   || | | | | |  |||  | | |     |  |
Db        233 ATAAAAGTTATTGTGGAATAACTATTTTAAATATGTAATTAGAAATGAAATGTTAATCGT 292

Qy        261 TATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATC 320
              | |  | |   | | | |||  |   |||    | |||    |||  |    |  |
Db        293 TTTAAAATATATCTAGTTTTTTAATAAATGAATTTAATTTAGCTTATTTATTTTATTAAG 352

Qy        321 ATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTG 380
              ||  ||    | || | |||||||  | |  |  | |  |    |       ||||  |
Db        353 TTAATTTTTTAATTTAATAATTAAATAAAGTAATATTTTAAGGGATGAGCTTTAAAATAA 412
```

```
Qy        381 CAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAA 440
              | |  |   |   |        ||||   |||  | ||      |        |  | |
Db        413 AATTTTATATTTTTTATAATTTTTAAATAAATATAAGCTTAAAAATAGCTATTATTAATA 472

Qy        441 TAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTAC 500
              |    |||  |   ||||| | |   |||| || | |   ||  ||| || ||   ||
Db        473 AATTTGTTATAATTTATTTTTTATAAAAAATTATTTAATTTAAATTAAATTATTTATTAA 532

Qy        501 TATTAAAACATTTACTTTCCAAATATTCTGTCATTCAGGAGTATGGAAGTATCGATGGCT 560
              |||| |||  |||| |    |||  ||| |||| | | ||  |     || ||  |    |
Db        533 AATTTAAATTTTTAATAATAAAAATTTCAGTCAATTATGATAAAATTAGCATATAAAATT 592

Qy        561 TCTTTAAAATGAA 573
              | | | ||| |||
Db        593 TATATTAAAGGAA 605
```

RESULT 20
CNS004ZW/c
LOCUS CNS004ZW 1101 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC # BACR11E08 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL055440
VERSION AL055440.1 GI:4932241
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR11E08"
/clone_lib="RPCI-98"
/note="end : T7"
ORIGIN

Query Match 5.3%; Score 58.2; DB 9; Length 1101;
Best Local Similarity 38.7%; Pred. No. 0.037;
Matches 144; Conservative 50; Mismatches 175; Indels 3; Gaps 1;

```
Qy        197 TTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATA 256
              ||::| |: || | |  |  ||: | :   || |:| ||||: || |   :| | | |:
Db       1064 TTWMAAAMTTTATATATTATATWAAAWWATTTWAWTATTTTWTAACAAAAYAWAAATAWT 1005

Qy        257 CAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAA 316
                ||:| : |   |     |:: :| |::     |::| | |:||  :   | :  ::
```

```
Db       1004 TTTTTAWTAWATATATAAATWTWWAWTTAWWTTTTATWWATATAWATATWAATAWWATHW 945

Qy        317 GATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAAT 376
              |  ||| :  |   |    :| || : | | |   | : |   |  |: | || | ||
Db        944 TAAWATAAMAAAAWTAWAAAWATATAWTAAAWATTTTAWTTATTTKATWTAATATTAAAA 885

Qy        377 CTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTG 436
              ||  | | |    ||||      |   || |  |||| ||  |: |||  :||     |
Db        884 ATTATATTWATGAATAAATTTAAATAATATAANTATANTATTTKATTTAAWATATTBTTT 825

Qy        437 CAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTA- 495
              :  ||| | |  |    ||| |    |    ::| :|: |   :| | ||| :| ::|:
Db        824 STTATATATAAAATATAAAAATAANATAAWWWWAAWAWATANWWATATATATWTAWWTWA 765

Qy        496 --TGTACTATTAAAACATTTACTTTCCAAATATTCTGTCATTCAGGAGTATGGAAGTATC 553
                | ||   :||||: |||:  : |  |:||:|| | | :|| |  | |:|  :| | |
Db        764 ATTTTAWATWTAAAWWATTWKAWATATAWATWTTTTTTAMTTTANAAATMTAAWAMTGTK 705

Qy        554 GATGGCTTCTTT 565
              ||   || :||
Db        704 TATATATTTHTT 693
```

RESULT 21
CNS00CS1/c
LOCUS CNS00CS1 843 bp DNA linear GSS 04-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR26H19 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL059666
VERSION AL059666.1 GI:4947129
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 843)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1. .843
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR26H19"
/clone_lib="RPCI-98"
/note="end : TET3"
ORIGIN

Query Match 5.3%; Score 57.8; DB 9; Length 843;
Best Local Similarity 33.5%; Pred. No. 0.043;
Matches 121; Conservative 78; Mismatches 161; Indels 1; Gaps 1;

```
Qy        150 ATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGT 209
```

```
Db        843 ATWAAWWAAATATAWAWATAAWATATATWTWTTTTTATAATTTTDWTTWWDAATAWWTRT 784

Qy        210 TTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTT 269

Db        783 ATWATTTWTTTATTTTAWTWTTTTTTWWTATTATTAATTTWTATATWTWTATWTATATWT 724

Qy        270 GTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAG 329

Db        723 TWAATTWWTWATATTAWATATATATAWWATAAAWTTAATTATTWWAWWWWWTATTTATWT 664

Qy        330 CCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAG 389

Db        663 TTATWTTWATAWTTWTTWWTAWTAAWAWTATTAWAATAWTATAWAATTAAWTTTTWTRTA 604

Qy        390 GTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTA 449

Db        603 AAAAAWWWAAAAAATWATATAAAAAWWWATAMAHTWTTATMAAATWWAWAATAATAHWAA 544

Qy        450 CCACAAATTTTATTTTGGAAAT-AATCTCAGAAAACATAATTTTTTATGTACTATTAAAA 508

Db        543 MAATAAAAWWWAATWWATWATTWAAAMAAABAATAYWAWSAAWWAWATAHAWCAYWAAAA 484

Qy        509 C 509

Db        483 C 483
```

RESULT 22
CNS0039G/c
LOCUS CNS0039G 1101 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL063921
VERSION AL063921.1 GI:4941778
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR08K10"
/clone_lib="RPCI-98"
/note="end : TET3"
ORIGIN

```
Query Match              5.3%;  Score 57.8;  DB 9;  Length 1101;
Best Local Similarity   18.6%;  Pred. No. 0.045;
Matches   91;  Conservative  205;  Mismatches  191;  Indels    2;  Gaps    1;

Qy         39 TGTTCTTATAAGCAAATGTTACAATGAGATATGAGTTTTAAAGCCAGATCTTCCTTATCT 98

Db        954 KKKAAKAAKWATKWWDDWDWDKDWKWDGAKDRKADDDDGAGDKDDDGKGKDADDDTDGTK 895

Qy         99 CTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATC 158

Db        894 DDDDKDKWDDWDKAKGTWGDATWAWAATDWWWWGWADADWWTWDAAADDWWADDRWDAWA 835

Qy        159 ATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAAT 218

Db        834 WKWDDAWAWGARTADRRDWGDRAGKRGGARKRRDRKRADDKRDAADDRDDAATWTTWTTT 775

Qy        219 ATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGT 278

Db        774 TTRDTDDWKWKTDTWTRWAADRTWDRDDDDDRDRAGTAGRKWRRTWKRRWKRRDTRWDDA 715

Qy        279 TTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATAT--ATTAGCCAGTTG 336

Db        714 DADDTARDDRRRRGDDGADAGKGKKTGRKRRRRDRATWDRTDAWWADAAWWTTTDTDTDD 655

Qy        337 AAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAA 396

Db        654 WDKRDRRRKGARRRRRTTARAAWDWWTWKAWDWAKWDWKTRADRWDRWAADTWTDARKAD 595

Qy        397 GTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAA 456

Db        594 RDWAKARAWRARRDRARAARADRRWTTKGKTTTATWTTWAARAAWWAWWAWATTTATWTT 535

Qy        457 TTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACT 516

Db        534 TWTTTWTTTTTTTTTTTTWTTTAAWWWAAWWTATWAAWTAAAWAAAAAAAAWAATTTTTTTT 475

Qy        517 TTCCAAATA 525

Db        474 TTTTWAWTA 466
```

RESULT 23
AG311129
LOCUS AG311129 1579 bp DNA linear GSS 02-JUN-2004
DEFINITION Mus musculus molossinus DNA, clone:MSMg01-093M11.T7, genomic survey sequence.
ACCESSION AG311129
VERSION AG311129.1 GI:47884083
KEYWORDS GSS.
SOURCE Mus musculus molossinus
ORGANISM Mus musculus molossinus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
TITLE BAC end Sequences of Library MSMg01
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1579)
AUTHORS Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the mouse BAC library MSMg01. For BAC library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
Tsukuba Institude, Bio Resource Center,
The Institute of Physical and Chemical Research (RIKEN) 3-1-1 Koyadai, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, fax: 81-298-36-9199
e-mail: abe@rtc.riken.jp
PRIMERS

```
            Sequencing : T7
            LIBRARY
            Vector     : pBACe3.6
            R.Site 1   : EcoRI
            R.Site 2   : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .1579
                     /organism="Mus musculus molossinus"
                     /mol_type="genomic DNA"
                     /sub_species="molossinus"
                     /db_xref="taxon:57486"
                     /clone="MSMg01-093M11.T7"
                     /sex="male"
                     /tissue_type="mixture of kidney and spleen"
                     /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

  Query Match             5.3%;  Score 57.8;  DB 9;  Length 1579;
  Best Local Similarity   48.4%;  Pred. No. 0.047;
  Matches  245;  Conservative    0;  Mismatches  258;  Indels    3;  Gaps    3;

Qy         23 TAACTTTCACTCTCTATGTTCTTATAAGCAAATGTTACAATGAGATATGAGTTTTAAAGC 82
              || |||     | | | | ||    |||      | |||       | |    |  | |||  |
Db        854 TATCTTATTTTTTATTTATTTAATTAAATCTCTATTATTTATCGTTTATATATATAATTC 913

Qy         83 CAGATCTTCCTTATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGG 142
                 ||  ||  |||| | | |       | | ||  | ||  ||| | | ||||   |||
Db        914 ATTATNTTATTTATATTTATATATTTTATATATATTTTTCAGTATATTATATATTTTTCA 973

Qy        143 TTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAG 202
              |      || ||  |||||  |  |   || ||| ||    |||  |  | ||  |||
Db        974 TACCCTTATCTTTTTCATTTTATTTTTATTCATATTTAATAAGATCAGTTAATACTAATT 1033

Qy        203 ACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTA 262
              |   |  ||  || | |||       | |  |||| |     |  ||  | | |||||
Db       1034 ATTATAATTCATTTA-ATATCATTATTAAATTTTAATTTTATTTTTTATATATACATTAT 1092

Qy        263 TTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCAT-TCAGAGAAAGATCA 321
               ||| || |||||  |||  | ||  |   ||| ||   ||  || | |     || || |
Db       1093 ATCAGTTTTTTTTTCTTTCCTTAAGTATTTTATTAACTTATATATATTATTATAATATTA 1152

Qy        322 TATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGC 381
              |   |||       |  | |||  |  | | |   |  ||  ||  ||||||| |||
Db       1153 TTCTTTAATTTACTATATATTTTAATTATATTTTATATATTATAGCTTACTTATTCTATT 1212

Qy        382 AGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCA-AA 440
              |  | |  |||  |     |  ||     ||   |  |  | | ||||   | |  |   |
Db       1213 ACATATATCTAATTATAATAATTATTTTTAATTTACATTTATATTATTATATTAAATTA 1272

Qy        441 TAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTAC 500
              |  | ||||      | | ||||| |    | | |||| |  | |||  ||| ||| |
Db       1273 TCTATATTATATATATTATTATTATATGATTTATCTTAATATACAGTATTATTTTTAATA 1332

Qy        501 TATTAAAACATTTACTTTCCAAATAT 526
              || ||||    ||||||   | || |
Db       1333 TACTAAAGATATTACTTAATATATTT 1358


RESULT 24
BH177277
LOCUS       BH177277                886 bp    DNA     linear   GSS 19-OCT-2001
DEFINITION  008_L_22-rev SmBAC1 Schistosoma mansoni genomic clone 008L22 5',
            genomic survey sequence.
ACCESSION   BH177277
VERSION     BH177277.1  GI:16275873
KEYWORDS    GSS.
SOURCE      Schistosoma mansoni
  ORGANISM  Schistosoma mansoni
            Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
            Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.
REFERENCE   1  (bases 1 to 886)
  AUTHORS   Le Paslier,M.C., Pierce,R.J., Merlin,F., Hirai,H., Wu,W.,
            Williams,D.L., Johnston,D., LoVerde,P.T. and Le Paslier,D.
  TITLE     Construction and characterization of a Schistosoma mansoni
            bacterial artificial chromosome library
  JOURNAL   Genomics 65 (2), 87-94 (2000)
  MEDLINE   20247247
   PUBMED   10783255
COMMENT     Other_GSSs: 008_L_22-21
            Contact: Pierce RJ
            INSERM U 167
            Institut Pasteur de Lille
            1 rue du Professeur A. Calmette, 59019-Lille, France
            Tel: (33) (0)3 20877783
            Fax: (33) (0)3 20877888
            Email: Raymond.Pierce@pasteur-lille.fr
            CNS sequencing ID=DG0AA008DF11BP1
            Plate: 008  row: L  column: 22
            Seq primer: M13 reverse primer
            Class: BAC ends
            High quality sequence stop: 886.
FEATURES             Location/Qualifiers
     source          1. .886
                     /organism="Schistosoma mansoni"
                     /mol_type="genomic DNA"
                     /strain="Puerto-Rican"
                     /db_xref="taxon:6183"
                     /clone="008L22"
                     /sex="mixed"
                     /dev_stage="cercariae"
                     /lab_host="Biomphalaria glabrata"
                     /clone_lib="SmBAC1"
                     /note="Vector: pBeloBAC 11; Site_1: Hind III; Partially
                     Hind III digested and size-selected S. mansoni  cercarial
                     DNA was ligated into Hind III digested pBeloBAC 11 vector
                     and used to transform E. coli DH10B. The complete library
                     contains 23808 clones from 4 independent
                     sizing-ligation-transformations.  Average insert size
                     ranges from 70-127 kb and genome coverage is 7.9-fold."
ORIGIN

  Query Match             5.3%;  Score 57.6;  DB 8;  Length 886;
  Best Local Similarity   44.0%;  Pred. No. 0.048;
  Matches  165;  Conservative   26;  Mismatches  176;  Indels    8;  Gaps    1;

Qy        138 TTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCT 197
              |||  |  | :||: ||   | |: | ||:| |  |:|    ||| |: |  ||   |: |
Db        516 TTTTTTMTAAMAAWTTTATTTAWAAAAAWTAAWATWAATAATTTTTWAATTATAATAWAT 575

Qy        198 TAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATAC 257
              |: |   ||| :  |  | |: |: |    : | ||||||||||         | | |||
Db        576 TWTAATTATTTWAATTATTAWTTWATWAAAWTAATTTTTTAAAA--------AAAATAW 627

Qy        258 AGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAG 317
                |||  | :| || ||  |:| :|  |||||  | | ||   ||  |:| | | |
Db        628 TTTTATAAAAWTTTTATTAATWTAWTTTAAAAAAATTTTAATTTATAAAWTTAAAAATTT 687

Qy        318 ATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATC 377
              |  | : |||    | || |||||||| | ||||     |  | |  | :| |      |
Db        688 TTTTTTWTTTAATAATTTAAAAATTAAAAATAAAAATTTTAAAAAAAAAWTAATAATNTA 747

Qy        378 TTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGC 437
              :      |:|  |||||:   ||   ||:  : || ||  |  |      |    |
Db        748 AWTTTTATAWAATTAAAAWAAWAARAAAAWTAWTTAATAAATTTTAAAAAAAAAAAAAAT 807

Qy        438 AAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATG 497
              |:| |    |:|  | |||     |       |||||||   | |||| | : ||||| |
Db        808 AWAAAATTTTWAAAAWAAAAAAAAAAAAAAAAAAATAATAAAAAAAAAAAAWTTTTTTAAAA 867

Qy        498 TACTATTAAAACATT 512
              |  | |   | |||
Db        868 AAAAAATTTTAAATT 882


RESULT 25
CNS07JUX
```

```
LOCUS       CNS07JUX                886 bp    DNA     linear   GSS 30-NOV-2001
DEFINITION  T3 end of clone 008DF11 of library SmBAC1 from strain Puerto-Rican
            of Schistosoma mansoni, genomic survey sequence.
ACCESSION   AL614235
VERSION     AL614235.1  GI:16027459
KEYWORDS    GSS.
SOURCE      Schistosoma mansoni
  ORGANISM  Schistosoma mansoni
            Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
            Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.
REFERENCE   1  (bases 1 to 886)
  AUTHORS   Le Paslier,M.C., Pierce,R.J., Merlin,F., Hirai,H., Wu,W.,
            Williams,D.L., Johnston,D., LoVerde,P.T. and Le Paslier,D.
  TITLE     Construction and characterization of a Schistosoma mansoni
            bacterial artificial chromosome library
  JOURNAL   Genomics 65 (2), 87-94 (2000)
  MEDLINE   20247247
   PUBMED   10783255
REFERENCE   2  (bases 1 to 886)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-OCT-2001) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Partially Hind III digested and size-selected S. mansoni cercarial
            DNA was ligated into Hind III digested pBeloBAC 11 vector and used
            to transform E. coli DH10B. The complete library contains 23808
            clones from 4 independent sizing-ligation-transformations. Average
            insert size ranges from 70-127 kb and genome coverage is 7.9-fold.
FEATURES             Location/Qualifiers
     source          1. .886
                     /organism="Schistosoma mansoni"
                     /mol_type="genomic DNA"
                     /strain="Puerto-Rican"
                     /db_xref="taxon:6183"
                     /clone="008DF11"
                     /clone_lib="SmBAC1"
                     /note="end : T3"
ORIGIN

  Query Match             5.3%;  Score 57.6;  DB 9;  Length 886;
  Best Local Similarity   44.0%;  Pred. No. 0.048;
  Matches  165;  Conservative   26;  Mismatches  176;  Indels    8;  Gaps    1;

Qy        138 TTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCT 197
              |||   |   | :||: ||  | |: | ||:| |  |:|    ||| |: |  ||   |: |
Db        516 TTTTTTMTAAMAAWTTTATTTAWAAAAAWTAAWATWAATAATTTTTWAATTATAATAWAT 575

Qy        198 TAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATAC 257
              |: |   ||| :  |  | |: |: |    : | ||||||||||        | | |||
Db        576 TWTAATTATTTWAATTATTAWTTWATWAAAWTAATTTTTTAAAA--------AAAAATAW 627

Qy        258 AGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAG 317
                |||  | :| || ||  |:| :|  |||||  | | ||   ||  |:| | | |
Db        628 TTTTATAAAAWTTTTATTAATWTAWTTTAAAAAAATTTTAATTTATAAAWTTAAAAATTT 687

Qy        318 ATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATC 377
              |  | : |||   | || |||||||||| | ||||    |  | |  | :| |     |
Db        688 TTTTTTWTTTAATAATTTAAAAATTAAAAATAAAAATTTTAAAAAAAAAWTAATAATNTA 747

Qy        378 TTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGC 437
               :     |:|  |||||:   ||   ||:  : || ||  |  |     |    |
Db        748 AWTTTTATAWAATTAAAAWAAWAARAAAAWTAWTTAATAAATTTTAAAAAAAAAAAAAT 807

Qy        438 AAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATG 497
              |:| |    |:|  | |||    |       |||||||   | ||||  | : ||||| |
Db        808 AWAAAATTTTWAAAAWAAAAAAAAAAAAAAAAAATAATAAAAAAAAAAAAWTTTTTTAAAA 867

Qy        498 TACTATTAAAACATT 512
              |  | |   | |||
Db        868 AAAAAATTTTAAATT 882
```

```
RESULT 26
BH673700/c
LOCUS       BH673700                762 bp    DNA     linear   GSS 19-FEB-2002
DEFINITION  BOHUT26TF BO_2_3_KB Brassica oleracea genomic clone BOHUT26,
            genomic survey sequence.
ACCESSION   BH673700
VERSION     BH673700.1  GI:18741461
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 762)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .762
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOHUT26"
                     /clone_lib="BO_2_3_KB"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
ORIGIN

  Query Match             5.3%;  Score 57.4;  DB 8;  Length 762;
  Best Local Similarity   49.2%;  Pred. No. 0.053;
  Matches  151;  Conservative    0;  Mismatches  156;  Indels    0;  Gaps    0;

Qy        212 TGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGT 271
              | | | || | |  |||    | | | ||     |   || | |||  ||  | || |
Db        593 TTTAATTACAATATCTTTTAATATCTTAATCTTTTATTTGAAACACAAATAAACTTTCCT 534

Qy        272 TTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCC 331
              | |    || |   ||||    |  ||| |  |   | ||    |   || ||    |
Db        533 TATAAAATTCTAACAAAATCTTTAAAAAATCTATTTGCAATTTATTTTCGTAATAAATTA 474

Qy        332 AGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGT 391
              | || ||   | |  |   | |  | |  | | |||| || |  ||  | | ||| ||
Db        473 ATTTAAATTCTGATTATCACTAAAATATAATTAAAATTAATTCAGTTTTTATTTAAAAGT 414

Qy        392 AAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACC 451
              ||||  |||| || |    | |     |   |  |||   ||||  ||  ||||||
Db        413 TAAAATTCAAACTCACATAAAATATTTTAAATATATATTTTACTGATAACCGAAATTTAA 354

Qy        452 ACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACAT 511
              |  ||||   |||||||||||||  |      || ||| ||| |||        || | | ||
Db        353 ATTAATTGCTTTTTGGAAATAAAATTTTTTAAAATATTTTGTTAGAATTGTTTTAGAAAT 294

Qy        512 TTACTTT 518
              | | ||
Db        293 ATTCATT 287
```

```
RESULT 27
CNS00LT2
LOCUS       CNS00LT2                1101 bp    DNA     linear   GSS 14-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC:
            BACR48P19 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
```

```
ACCESSION   AL078714
VERSION     AL078714.1  GI:5102004
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR48P19"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN

  Query Match             5.3%;  Score 57.4;  DB 9;  Length 1101;
  Best Local Similarity   29.3%;  Pred. No. 0.055;
  Matches  132;  Conservative  101;  Mismatches  218;  Indels    0;  Gaps    0;

Qy        285 AAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAA 344
              ||||||    | ||| |||    |    | | ||| |  | | |  |    |     ||||  ||
Db        416 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAANAA 475

Qy        345 ACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCT 404
              | | | ||  | |  . | ||  :::| || :| ||: |:: :|| :       :  :  :
Db        476 AAAAANAAAAAATTWTTTTTTTWKKATTTTWTTTTDAADDKWAAWWDTTTTTAWTTWTTW 535

Qy        405 AAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTT 464
              |    |||  |   |: ||:  | |      : || || :||::      |||:|| |:|
Db        536 ATTAATATTTTTTTTKWTTWKATKTDWTKKAKTAAKTAAWAAWWTTTTTTAATWTTTTKT 595

Qy        465 TGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAAT 524
              :  |  |:|: : | : |   :  :|||||:| |  :| | |      ||:  : |      |
Db        596 WTTAKKTWAWKWWAWDTATTTWTTWTTTTTWTTTTTWAATTTTTWTTTWTAWATTTTTTT 655

Qy        525 ATTCTGTCATTCAGGAGTATGGAAGTATCGATGGCTTCTTTAAAATGAAGCAGGAGGGTC 584
              || : :  ||      |  |   |  :|   |:  ::  ::  : |  |:: :
Db        656 TTTAKTKTTTTTTTTTTATTAAATAAAAAAWTTTDTWAAAWWTTTKKKKKKAAAADKWKD 715

Qy        585 TGGCAGAGAGTATCTATGAAATAAGTTCCTCTGACCTTCACGCTTAATTTTCTGAATGGA 644
              :: :||:  | | : :: ||:|| ::    :|   :    : ::  :::: ::  :||:
Db        716 AKKWDGAKKATKTKKKDKKAAWAAADKKDRKKGGKKKGKGKKGKKGGKKKKKKKKGKGGW 775

Qy        645 GTGGAGCAAATTACTTCAAGCTTCACTTAACTTGCATATGAAATGAACCGTACAAAAATA 704
              |::| : | :::| ::   : :: | ||   ::|  : ::|| :: |  :|    ::| |
Db        776 GKKGKKKAGDDDAKDKTKKKKKKAATTTTKKKKGKKKKKKAAKKKAADRTKKTKWDAAA 835

Qy        705 CAAGAGTGTCAGGAGAAAGTTATGCTCTGGT 735
              || |:: :  |: :   ::: |: : ::::
```

```
Db        836 AAAAAKKTKDKGKKKKKTKKKTTKKKKKKKK 866


RESULT 28
CNS020K7
LOCUS       CNS020K7                1092 bp    DNA     linear   GSS 01-SEP-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            222L11 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL175696
VERSION     AL175696.1  GI:7813753
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE   1
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
  MEDLINE   20296633
   PUBMED   10835645
REFERENCE   2
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
            Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
            Saurin,W., Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
  MEDLINE   20359837
   PUBMED   10899143
REFERENCE   3  (bases 1 to 1092)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1092
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="222L11"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG222CF06LP1~end : T7"
ORIGIN

  Query Match             5.2%;  Score 57.2;  DB 9;  Length 1092;
  Best Local Similarity   38.1%;  Pred. No. 0.061;
  Matches  128;  Conservative   53;  Mismatches  153;  Indels    2;  Gaps    1;

Qy        238 AAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAA 297
              :||: :  :::   | | | |  | ||  ||| |||||  |||    ||||||   | ||
Db        663 WAAWWWWWWWWTTAAAAAAAAAATTTTTTTTTTTTTTTTTTTTAAAAAAAAAAAAAAAAA 722

Qy        298 AGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGT 357
              : :|||  | : | | ||: |: |:::|:|       || |::::::::| | :|||: : :
Db        723 WTWAATWAAAWWAAAAAAWAAWAAWWWAWTTWTTTTTTWAWWWWWWWAAAAWAAAWWWTW 782

Qy        358 GCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTAC 417
                |: ||  :::|  |:|: ||  | |  ||   ||| | |  |  ||:    | | :|
Db        783 TTAWTTTTAWWWA--TWAWWTTWTATTTTAAAWAAAANAWTTTAWTTAWTAAAAAAWWAA 840

Qy        418 CTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTC 477
```

```
Db        841 ATAATTWTTTAAAAWAWTTTAWAAATTWWAATWTAAAAAAAAAAAATTAAAAAAWAAWTWW 900

Qy        478 AGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGTCATTCA 537

Db        901 TTTTTTTTWTWAWWTWTAAATWTTTTTTTTTTTTWWTTTTTWTAAWAAAAAAAAAAAAA 960

Qy        538 GGAGTATGGAAGTATCGATGGCTTCTTTAAAATGAA 573

Db        961 TATAAAWNAAAAWWTTAATTTATTTTTAAAAAAAAA 996


RESULT 29
CNS00KOI/c
LOCUS       CNS00KOI                935 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC:
            BACR18E12 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL078237
VERSION     AL078237.1  GI:4957411
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 935)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES              Location/Qualifiers
     source           1. .935
                      /organism="Drosophila melanogaster"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:7227"
                      /clone="BACR18E12"
                      /clone_lib="RPCI-98"
                      /note="end : TET3"
ORIGIN

  Query Match             5.2%;  Score 56.6;  DB 9;  Length 935;
  Best Local Similarity   46.4%;  Pred. No. 0.081;
  Matches  176;  Conservative   14;  Mismatches  180;  Indels    9;  Gaps    1;

Qy        157 TCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTA 216

Db        415 TAATDAAAAATAATTAGATAATGATAAATAAAATTTATTHWTATGTTGAAAWWTTWTTCA 356

Qy        217 ATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTG 276

Db        355 TAAAAATATAAAATGTTTTTACATTCTGAAAAATGATATAATTATAATGATRAATTAMWT 296

Qy        277 GTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTG 336

Db        295 TTTTATTCTAATTARATATAWATAAWTATATAAAATTATATTTTTTATAATTTTTAAATT 236

Qy        337 AAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAA 396

Db        235 TAAAATAAATATATAAATAGATATTTWTTAAATTTTTA---------TATAAAWGAAAAA 185

Qy        397 GTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAA 456

Db        184 ATTTTAATAAATATATATAAATTTAAATATTATAGAAATTGAATGAATTTGWAAAWAAWA 125

Qy        457 TTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACT 516

Db        124 ATTTAATATTTGATTATTTTGTTAAGTATTTWTTTTTATTATATTTGTAAAAAATTTTAT 65

Qy        517 TTCCAAATATTCTGTCATT 535

Db         64 ATGGTTTGATTATTTTATT 46


RESULT 30
CR720510/c
LOCUS       CR720510                1254 bp    mRNA    linear   HTC 19-AUG-2004
DEFINITION  Tetraodon nigroviridis full-length cDNA.
ACCESSION   CR720510
VERSION     CR720510.1  GI:51218761
KEYWORDS    HTC; cDNA; full-length; Tetraodon nigroviridis.
SOURCE      Tetraodon nigroviridis
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 1254)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-AUG-2004) Genoscope - Centre National de Sequencage -
            : 2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
            (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT     The sequences are based on single pass reads.
            More information available at
            http://www.genoscope.cns.fr/tetraodon.
FEATURES              Location/Qualifiers
     source           1. .1254
                      /organism="Tetraodon nigroviridis"
                      /mol_type="mRNA"
                      /db_xref="taxon:99883"
                      /tissue_type="Kidney"
ORIGIN

  Query Match             5.2%;  Score 56.6;  DB 3;  Length 1254;
  Best Local Similarity   46.4%;  Pred. No. 0.083;
  Matches  185;  Conservative    0;  Mismatches  214;  Indels    0;  Gaps    0;

Qy        172 TAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGAT 231

Db       1250 TTATTATTATACTCATAATATTATATATTATTAATTATAATATATATAATAAACAATA 1191

Qy        232 TTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAG 291

Db       1190 TAATTATAAATAAATAAAAAAAAATAATTAAAATAAATATATTTTATATAATATATAAA 1131

Qy        292 GTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAA 351

Db       1130 TTAAAAAAAAATATTATATTATATAATTATTTTAATATTAATTTAAAAAAAAAATATAAA 1071

Qy        352 ATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTA 411

Db       1070 AATAAAAAAAAAATAAAAAATTTATATTATATAAAAAAAAAAAAAAAAAAAAAAATAAAA 1011

Qy        412 TACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAAT 471

Db       1010 TAATATATAATATTTTTTATAATAAAAATTATAATATAATATATTTATTTTATTATTATT 951

Qy        472 AATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGT 531
```

```
Db        950 TTTAATTTTTTATATTTATTTTTATTATTATTTATATTATATAATAATAACATATTTATA 891

Qy        532 CATTCAGGAGTATGGAAGTATCGATGGCTTCTTTAAAAT 570
              ||  |    ||   || |||  ||   ||  |  | ||
Db        890 TATATATATATAATAAATTATATATTATTTTATATATAT 852


RESULT 31
CK997149/c
LOCUS       CK997149               1313 bp    mRNA    linear   EST 26-MAR-2004
DEFINITION  ip15c02.b1 Brain - Cerebellum Library (DOGEST8) Canis familiaris
            cDNA clone ip15c02, mRNA sequence.
ACCESSION   CK997149
VERSION     CK997149.1  GI:45760829
KEYWORDS    EST.
SOURCE      Canis familiaris (dog)
  ORGANISM  Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE   1  (bases 1 to 1313)
  AUTHORS   Balija,V., Nascimento,L.U. and McCombie,W.R.
  TITLE     ESTs from Canis familiaris cerebellum (dog)
  JOURNAL   Unpublished (2004)
COMMENT     Contact: W. Richard McCombie
            Lita Annenberg Hazen Genome Sequencing Center
            Cold Spring Harbor Laboratory
            PO Box 100, Cold Spring Harbor, NY 11724, USA
            Tel: 516 367 8884
            Fax: 516 367 8874
            Email: mccombie@cshl.org
            Plate: ip15  row: c  column: 02
            High quality sequence stop: 1313.
FEATURES             Location/Qualifiers
     source          1. .1313
                     /organism="Canis familiaris"
                     /mol_type="mRNA"
                     /db_xref="taxon:9615"
                     /clone="ip15c02"
                     /sex="Unknown"
                     /tissue_type="Cerebellum"
                     /dev_stage="3 month old normal canine"
                     /lab_host="XL10 Gold"
                     /clone_lib="Brain - Cerebellum Library (DOGEST8)"
                     /note="Organ: Brain; Vector: pBluescript II SK; Site_1:
                     EcoRI; Site_2: XhoI; Library constructed using pBluescript
                     XR kit from Stratagene. Cloned cDNA was size selected
                     between 1-3 kb. Mark Haskins VMD, PhD, Pathology and
                     Medical Genetics, School of Veterinary Medicine,
                     University of Pennsylvania, 3800 Spruce Street,
                     Philadelphia, PA 19104-6051"
ORIGIN

  Query Match             5.2%;  Score 56.6;  DB 7;  Length 1313;
  Best Local Similarity   45.7%;  Pred. No. 0.084;
  Matches  302;  Conservative    0;  Mismatches  350;  Indels    9;  Gaps    3;

Qy        111 TCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAG 170
              | |  || || || | | | |  |   |||  |    ||||| | ||  | | || | |
Db        945 TATTTTTATTAAATTATTTAATTATTTTTATTAATTAAATAATTATTTTAAAAATTTAT 886

Qy        171 TTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGA 230
              |  | | |   |  | || ||   ||   ||    | | |||  | | ||||        |
Db        885 TATAAAATAATTAAAAAAATTAAATTAAAATATAAATATTTATTAATTATATAATTAATA 826

Qy        231 TT-TTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAA 289
              || ||||| | |    ||      |||    ||||  ||| ||||   ||| || |||||
Db        825 TTATTTTTTATATTTTTTTATTTAATAATTTTATTATTTTTTTTTATATTTATTTAAAAA 766

Qy        290 AGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACA 349
              || | || |  | | |   |  || ||| | ||  | |  |||  |  |   |
Db        765 TAATAAATAAAATTTTAATAATTAATTTATAATAAAATATAAAATATAAATATTTAATTA 706

Qy        350 AAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGG 409
              | || |     |||||   || |  |   |  | | |||  |    | | ||  |  |
Db        705 ATATAAAAAATTATTATTATATTATAATATTAAATAAAAATTTTTTATTTAAANTTTAAA 646

Qy        410 TATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAA 469
              ||       |  ||| |||       |    ||   |||||  | | ||||| |||  |
Db        645 AATTTATTTTTTTTTATTAAANANTTTTTTATTATAATTA-TATATATTTAATTTAAAT 587

Qy        470 ATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCT 529
              | |             ||  ||| ||||  | | |||  | |||      |  || | |  |
Db        586 AANAAAATTTTTTTTATTTTTTATTATTAATTTTTATTATATTATTAATAAAATTTTAAT 527

Qy        530 GTCATTCAGGAGTATGGAAGTA-------TCGATGGCTTCTTTAAAATGAAGCAGGAGGG 582
              ||  |  | | |  || ||       |  ||    | | |  | | ||
Db        526 TAAATATAATAATTTTTAAATATTAAAATTATATTATATATATTTATTAAATTTTTTTAT 467

Qy        583 TCTGGCAGAGAGTATCTATGAAATAAGTTCCTCTGACCTTCACGCTTAATTTTCTGAATG 642
              | |   | | | |||  || |||||| ||     |        |    ||  | | |   |
Db        466 TTTAATAAATAATATTAATTAAATAATTTNTATTATTAAAAAAAAATATATATTTTTTTT 407

Qy        643 GAGTGGAGCAAATTACTTCAAGCTTCACTTAACTTGCATATGAAATGAACCGTACAAAAA 702
              | |  |  ||||||  |  | |  |  | ||||  |        || | |     | || ||
Db        406 TAATTTATTAAATTATATATAAATAAAATTAAAATAATATAAAATTAATTTTAATAATAA 347

Qy        703 TACAAGAGTGTCAGGAGAAAGTTATGCTCTGGTAAATATTTTGCAAAACAGATAAAAGAT 762
              || || | |   |    || ||||  | |   | | |||     |||   ||| |   |
Db        346 TAAAATATTTAAATTTATAATTTATTNTATTTAATAAATTAAAATAAATTAATATATATT 287

Qy        763 A 763
              |
Db        286 A 286


RESULT 32
CNS0161W/c
LOCUS       CNS0161W               979 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN15B17 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL106190
VERSION     AL106190.1  GI:5620578
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 979)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .979
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN15B17"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : SP6"
ORIGIN

  Query Match             5.2%;  Score 56.4;  DB 9;  Length 979;
  Best Local Similarity   33.7%;  Pred. No. 0.09;
```

```
  Matches  142;  Conservative   79;  Mismatches  194;  Indels    6;  Gaps    1;

Qy        231 TTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAA 290
              ||||||||:|| :  ||    : :||       ||  ||| | |::  ::  ||    | :
Db        979 TTTTTTTAWAATWWTTTTTTTHTHTATTWTCTTTTYTTTTTATYYTTWYNTTTTTTTATH 920

Qy        291 GGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAA 350
              :|||| |||   | | : | : : :| ||:| |||   :      ::::||| :::|
Db        919 TAWATAAAAAAAAAAAATAWWAWWTWWWTWATWTTTTAWATWHAATCTWWKHAAATHMHA 860

Qy        351 AATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGT 410
              |   :   : |    ||  :|:|| |   |    |    :||||   :|  |||:  |
Db        859 TTTNTNWWATYGTMTWTTTTTWTWATWTMTWATMNTATWAAWAAAAMAWHAAATAAWATT 800

Qy        411 ATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAA 470
              ||: |   |  ||:| | |   |:   |:|: ::||: |  | ::   ::  ||:  |
Db        799 ATMTTTTDTCTTTYCSTTTTTTASCTWARAWBDRAADAANAAWRRSNAWDKATTBATAKT 740

Qy        471 TAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTG 530
              : |      :|| ::|:|| :::::| :| :|||| :|   ||:       || |::| |
Db        739 AWTT------WAAAWDAWTTADKWWTAWAWWATTANMARKATTWRRAAAWAAKTWWTDTW 686

Qy        531 TCATTCAGGAGTATGGAAGTATCGATGGCTTCTTTAAAATGAAGCAGGAGGGTCTGGCAG 590
              | ||| |  : :|    | :|   ||  ::   : || :| ||  || :  :: | |
Db        685 TAATTAATADADAAAWTACWAAAWATTTMHATAKAAAGWTTAAAAAGAWMADWATAGAGA 626

Qy        591 AGAGTATCTATGAAATAAGTTCCTCTGACCTTCACGCTTAATTTTCTGAATGGAGTGGAG 650
              | : | |  |  ||:::| |  | : ||  ||::   :  : |:: | |  |      |
Db        625 AAWATTTWAAAMAAMWWAATAACAMAGAAWTTBWTTTBABDNTKDTTTATAGMNNNNGNC 566

Qy        651 C 651
              |
Db        565 C 565
```

```
RESULT 33
AG360787/c
LOCUS       AG360787               1259 bp    DNA     linear   GSS 03-JUN-2004
DEFINITION  Mus musculus molossinus DNA, clone:MSMg01-161F20.T7, genomic survey
            sequence.
ACCESSION   AG360787
VERSION     AG360787.1  GI:47971992
KEYWORDS    GSS.
SOURCE      Mus musculus molossinus
  ORGANISM  Mus musculus molossinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     BAC end Sequences of Library MSMg01
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1259)
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the mouse BAC library MSMg01. For BAC
            library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
            Tsukuba Institude, Bio Resource Center,
            The Institute of Physical and Chemical Research (RIKEN) 3-1-1
            Koyadai, Tsukuba, 305-0074 Japan
            phone: 81-298-36-9189, fax: 81-298-36-9199
            e-mail: abe@rtc.riken.jp
            PRIMERS
            Sequencing : T7
            LIBRARY
            Vector     : pBACe3.6
            R.Site 1   : EcoRI
            R.Site 2   : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .1259
                     /organism="Mus musculus molossinus"
                     /mol_type="genomic DNA"
                     /sub_species="molossinus"
                     /db_xref="taxon:57486"
                     /clone="MSMg01-161F20.T7"
                     /sex="male"
                     /tissue_type="mixture of kidney and spleen"
                     /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN
```

```
  Query Match             5.1%;  Score 56;  DB 9;  Length 1259;
  Best Local Similarity   47.0%;  Pred. No. 0.11;
  Matches  173;  Conservative    0;  Mismatches  195;  Indels    0;  Gaps    0;

Qy        143 TTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAG 202
              || | ||||||  || |  |   |  |  || |   |  | || |  | |   |||
Db       1018 TTTAAAAATATAAATAAAAAATTAAAAAAAAAAAAAAATAAAAAAAAAAATTAAAAAAAAA 959

Qy        203 ACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTA 262
              | |        | || | |      |  ||       |||| | |  | | | ||| |   |
Db        958 AAAAAAAAAAATAAAAAAAAAAAAATAAATAAAAAAAAAAATAACAAAAAAAAATAAAATAA 899

Qy        263 TTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCAT 322
              |  | ||  | |          ||||||  || ||| |||   | | |  ||| |  ||
Db        898 TAAAATTAAATATATAAAAAAAAAAAAAAAAATAAAAAAAAAAATAATAAATAAAAAAAAAT 839

Qy        323 ATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCA 382
              | | ||   ||  | ||| | |||  | |  | | |  |  |||       |
Db        838 AAAATAAAAAAAAAAAATAAAAAATAAAAATATAATAAAAAAAAAAAAAAAATAAAAAAAAA 779

Qy        383 GTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATA 442
              |||   |||||   ||   |||  || | |          | ||   |    ||| |
Db        778 AAAAAAAAAAAAAATATAAAAAAAAAAATAAAAAAAAAAATAAAATAATACAATATAAAAAA 719

Qy        443 GAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTA 502
              ||| || || | | ||  |  |  ||||||   || |||| ||| || |  ||  |  |
Db        718 AAAAATAACAAATAATTATTAATACAAATAAAAACAAAAAAAATACTTATAAATAAAACA 659

Qy        503 TTAAAACA 510
              || ||||
Db        658 ATATAACA 651
```

```
RESULT 34
CNS00KTZ
LOCUS       CNS00KTZ                967 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC:
            BACR18K20 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL078030
VERSION     AL078030.1  GI:4957608
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 967)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
```

EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES Location/Qualifiers
source 1. .967
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR18K20"
/clone_lib="RPCI-98"
/note="end : TET3"

ORIGIN

Query Match 5.1%; Score 55.8; DB 9; Length 967;
Best Local Similarity 32.2%; Pred. No. 0.12;
Matches 141; Conservative 82; Mismatches 215; Indels 0; Gaps 0;

```
Qy          51 CAAATGTTACAATGAGATATGAGTTTTAAAGCCAGATCTTCCTTATCTCTCTGCCCCATC 110
               |||     ||:|  | |  | |     :|| ::| |: ||:|    : | |  ||::|::
Db         489 CAAYMNCCACMACMACAATTTAMMAMMMAAMMMAAAYTTTHCCMCCMCCCCCCCCMHAHH 548

Qy         111 TCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAG 170
               || | || ||        | || |  : :|         | ||       |:||  | |  |:
Db         549 TCCATTTTTTYCCCCTSCCCAAAACHCYTACCACCCAATATTCACCCCWTTTAATAAAAS 608

Qy         171 TTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGA 230
               :|| |           || ::: :  ||| || | ||  ||||   ||| ||
Db         609 YTATTTYCCCCMACMAAAMWMCMCCCTTTAAMARATAATTTTCCCAATTTAAAAAAAACC 668

Qy         231 TTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAA 290
               ||||   |       ||: :   :|::| :  |:|  || | :|: :  |: :   ||||
Db         669 TTTTSAAATTKCTTTTTMWWTTTWTMMATYCCTKCTATTTTNWTWTKSATKAKATKAAAA 728

Qy         291 GGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAA 350
                 ||::| ||||:  :: : : :||:|   | | :: :  |:: |  : |::|  :
Db         729 AAAATDRAAAAATKRTKYTMAWADAAKAMAGTTTGHWTKKKAKKGGKGRTTKRAGRMWGS 788

Qy         351 AATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGT 410
               |      |:| : |::::   |:|| |   | |:||:          | ||   ||:
Db         789 CTTTCCCCCMTTWGAHWKDNDMTWATTTSTRATTHAAMCAATTTTTCTAAAAGAAAWAWG 848

Qy         411 ATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAA 470
                :| |  :: : || | |:    |:::: |:  |: |:|: | |||   : :|:|    :
Db         849 TKATTTTHWTYCTTTTAADWTTYCYSVRTADTTARTTDAHAAAAAAAAAAWTYTYTTTTTW 908

Qy         471 TAATCTCAGAAAACATAA 488
               | || | :  ::: |: |
Db         909 TKATTTTWTTWDRKAKTA 926
```

RESULT 35
CNS00K1F
LOCUS CNS00K1F 1101 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC: BACR39N15 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL077002
VERSION AL077002.1 GI:4956580
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR39N15"
/clone_lib="RPCI-98"
/note="end : TET3"

ORIGIN

Query Match 5.1%; Score 55.8; DB 9; Length 1101;
Best Local Similarity 47.6%; Pred. No. 0.12;
Matches 150; Conservative 2; Mismatches 163; Indels 0; Gaps 0;

```
Qy         194 ATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGAC 253
               || ||||  | ||  |||| ||| | | |   || | | ||| | | |  || |
Db          54 ATATTAATTATATATTTTTTTTATTTTTTTTTATTTAAATATTTAATATTATTTTTTATTA 113

Qy         254 ATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAG 313
               || |   ||| |  | ||||   ||  |  |  ||  |||  | |||||| |||   |
Db         114 ATTTAATNATTKAAATTTTTTNANATTTATATATTAAAATATTTAAAAATGAATTGGTAN 173

Qy         314 AAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTT 373
               || || ||  | |||       |   |||||   | | | | |   ||| |  ||| |:
Db         174 TAATATTATTCAATAGTATAAAAATKTTTAAATTKAGATTAATTTTGTATAATTTTATTK 233

Qy         374 AATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCA 433
               |  ||: ||| |||||    | || |   |  || |    |   |    | || | | |
Db         234 TAGATTKTAGTTAAAGGGGGTATGTTAGTTTTTAGATTGTATGTTTWGGATATTGTGGGA 293

Qy         434 CTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTT 493
               || |       ||    |   ||| ||    |  | |  | ||||    ||| |
Db         294 ATGAATTGTTTTATAGTTATGGTTTTGGTTAAATATTGAGTAAAATAAAAATATATTATA 353

Qy         494 TATGTACTATTAAAA 508
               | |  |  |||||
Db         354 ATTTTKGTTATAAAA 368
```

RESULT 36
BZ695089
LOCUS BZ695089 1132 bp DNA linear GSS 02-JUL-2003
DEFINITION SP__Ba0057N22.f SP__Ba Sorghum propinquum genomic clone SP__Ba0057N22 5', genomic survey sequence.
ACCESSION BZ695089
VERSION BZ695089.1 GI:28387830
KEYWORDS GSS.
SOURCE Sorghum propinquum
ORGANISM Sorghum propinquum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 1132)
AUTHORS Wing,R., Yu,Y., Kim,H.R., Collura,K., Pries,G., Currie,J., Soderlund,C. and Hatfield,J.
TITLE Sequencing of Sorghum BAC ends.
http://genome.arizona.edu/stc/sorghum
JOURNAL Unpublished (2003)

```
COMMENT     Contact: Rod Wing
            Arizona Genomics Institute
            University of Arizona
            Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
            85721-0088, USA
            Tel: 520 626 3967
            Fax: 520 621 9288
            Email: http://genome.arizona.edu
            PCR PRimers
            FORWARD: atc agc ggc cgc gat cc
            BACKWARD: gta aaa cga cgg cca gtg
            Plate: 0057  row: N  column: 22
            Seq primer: atc agc ggc cgc gat cc
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .1132
                     /organism="Sorghum propinquum"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:132711"
                     /clone="SP__Ba0057N22"
                     /clone_lib="SP__Ba"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII; Paterson lab BAC library (HindIII)"
ORIGIN

  Query Match             5.1%;  Score 55.8;  DB 8;  Length 1132;
  Best Local Similarity   43.4%;  Pred. No. 0.12;
  Matches  189;  Conservative    0;  Mismatches  246;  Indels    0;  Gaps    0;

Qy        199 AAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACA 258
              ||| |  || || | | || |||        |  |||||  |  |  |      | ||  |
Db        436 AAAAAATTTNTTATNTNAANATAGGTNTAATAANTTTTTTTATTAATTATTTAAAATTNA 495

Qy        259 GCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGA 318
                |  |||   | |  ||  |  ||||   ||||   |  || |    || |
Db        496 TNAAAAAATTNTATATNTTTTATATATTAAAATAATTAAATTTTTANTTTAAATATAANA 555

Qy        319 TCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCT 378
              ||  |||    | || ||   | |  | || ||   |  |  ||   | | |  || |
Db        556 NAATTAATTTAAAAATTTAATNNTNATTAAAATATATATTAAATTTNATTAAATATATTT 615

Qy        379 TGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCA 438
              | | | |||   ||||    |   | ||  |  | |  ||      || ||||     |  |
Db        616 TATAATTAAAAAAAAAATTATATTTTTAAANTTAAATATTAATATTTTTATATATATTAA 675

Qy        439 AATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGT 498
              || |||| |    | | |||| | |  |      | |      | || ||  ||||| |
Db        676 TTTATAAATAAAATTATAATTTAATATTNATTANTTATTTTTTTATATTATNATTTATTT 735

Qy        499 ACTATTAAAACATTTACTTTCCAAATATTCTGTCATTCAGGAGTATGGAAGTATCGATGG 558
              | ||   | |   | | |  |   |||   | ||     | |||  |   |   ||
Db        736 TTTTTTTTTATAANAATTATTAATTAATTTAATTATATTAAAATATTTATTAAAAAATAA 795

Qy        559 CTTCTTTAAAATGAAGCAGGAGGGTCTGGCAGAGAGTATCTATGAAATAAGTTCCTCTGA 618
              |  | ||||   |   |      | |  | | |  || ||   ||| |  |    | |
Db        796 TTATTATAAATAAATATAATTTTATTTTANATANAAAATATAATTAATTAAATTAATTTA 855

Qy        619 CCTTCACGCTTAATT 633
                ||     ||||||
Db        856 TATTTTTAATTAATT 870


RESULT 37
CNS0145P/c
LOCUS       CNS0145P               1043 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN11G11 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL103735
VERSION     AL103735.1  GI:5615346
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1043)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1043
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN11G11"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : T7"
ORIGIN

  Query Match             5.1%;  Score 55.6;  DB 9;  Length 1043;
  Best Local Similarity   32.3%;  Pred. No. 0.14;
  Matches  193;  Conservative   79;  Mismatches  325;  Indels    0;  Gaps    0;

Qy        170 GTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTG 229
              |||  |  :::::: :    | |  | ||         | |:|| |:. :| |  |   |:|
Db        968 GTTTTTTTKKKKKKKKTTTTTTTTTTTTTTTTTTTTATATWTTTTDTRTKTTWTDATTWG 909

Qy        230 ATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAA 289
              ||||: || : :   ||       |    | ||  ||| |||||  ||| ||  : :
Db        908 KTTTTKGTATRTDTKTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTWTWTWT 849

Qy        290 AGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACA 349
              |  |::| | | :|  :   | ::||| |  |  |:||:   : :: :  :::::: :
Db        848 ATWTWWATATATWTARRGGRAADDAAAAAAAAATTWTTWTWWWWWWWWTTWWWWWWWWT 789

Qy        350 AAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGG 409
                :   |  | |  | |  |   ||      |   |||     |   : ||  || |
Db        788 TTTWTTTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAANTATNTWAAATATATAAA 729

Qy        410 TATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAA 469
              | :|  :     | |       |  |||||   ||:|  : ::||  |||  :   ||
Db        728 TTWAAAWATAWATATAAATATAWATAAAAAATATTTATWAWAWAWWATAATATAWAWAAA 669

Qy        470 ATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCT 529
              | :|: : : |||:    ||||:::|: ||  : :|::: |   | : |  |:|||| |
Db        668 AAWAWAWAWAAAAWATATATTTWWWAWATAAAWAWAWWWAAAAAAAWATATAWATATTWT 609

Qy        530 GTCATTCAGGAGTATGGAAGTATCGATGGCTTCTTTAAAATGAAGCAGGAGGGTCTGGCA 589
              | ||  |  |  ||  |  ||:   |    | | :| |::  |   |
Db        608 TTWATATATAAAAATATATATAKTATTTTTATATAWATAWWTATWTATATAWAAAAAAAG 549

Qy        590 GAGAGTATCTATGAAATAAGTTCCTCTGACCTTCACGCTTAATTTTCTGAATGGAGTGGA 649
              |||  |   | :| | ||         :| |  | | |         |
Db        548 AAGAAAAAAAAAARAGAGAAAAAAAAAASAACCCCCCCCCNNNNCCCCCCCCCCCCCMCCC 489

Qy        650 GCAAATTACTTCAAGCTTCACTTAACTTGCATATGAAATGAACCGTACAAAAATACAAGA 709
              || |             |       |   |: :  |||  || :  |:||| | |:|| |
Db        488 NCANANANANANNNNNGTNNGNACGCAMCCMNVCAAAAAAAAAAMAAAMAAACACAMAAAA 429

Qy        710 GTGTCAGGAGAAAGTTATGCTCTGGTAAATATTTTGCAAAACAGATAAAAGATAATA 766
                :  | |||    |  : :    ||| || |    ||| | |::||| : :|:|
Db        428 AAAMAMAMAAAAAAAMAMAVAMAAANAAAAATATRAATAAAAAAAWRAAAAWAWAWA 372
```

```
RESULT 38
CL646802
LOCUS       CL646802               1335 bp    DNA     linear   GSS 06-JUL-2004
DEFINITION  CH213-123P18.SP6 CH213 Gasterosteus aculeatus genomic clone
            CH213-123P18 3', genomic survey sequence.
ACCESSION   CL646802
VERSION     CL646802.1  GI:49666226
KEYWORDS    GSS.
SOURCE      Gasterosteus aculeatus (three spined stickleback)
  ORGANISM  Gasterosteus aculeatus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Gasterosteiformes;
            Gasterosteidae; Gasterosteus.
REFERENCE   1  (bases 1 to 1335)
  AUTHORS   Kingsley,D., Grimwood,J., Dickson,M., Schmutz,J. and Myers,R.M.
  TITLE     Expressed sequence tags from Gasterosteus aculeatus
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Grimwood, Jane
            Stanford Human Genome Center
            Stanford University School of Medicine
            975 S California Avenue, Palo Alto, CA 94304, USA
            Tel: 650 320 5917
            Fax: 650 320 5801
            Email: jane@shgc.stanford.edu
            Plate: 123
            Class: BAC ends
            High quality sequence start: 26
            High quality sequence stop: 686.
FEATURES             Location/Qualifiers
     source          1. .1335
                     /organism="Gasterosteus aculeatus"
                     /mol_type="genomic DNA"
                     /strain="Salmon River"
                     /db_xref="taxon:69293"
                     /clone="CH213-123P18"
                     /sex="Mixed"
                     /cell_type="Blood"
                     /clone_lib="CH213"
                     /note="Vector: pTARBAC2.1; Site_1: EcoRI; The sequence of
                     the clone was established as a mapping and sequencing
                     collaboration at the Stanford Genome Evolution Center,
                     funded by the NIH Centers of Excellence in Genomic Science
                     (CEGS) initiative (http://cegs.stanford.edu). The clone
                     was isolated from the BAC library CHORI-213 built by
                     Pieter deJong in collaboration with the Stanford Genome
                     Evolution Center (http://www.chori.org/bacpac/). Clones
                     may be purchased from BACPAC Resources
                     (http://www.chori.org/bacpac/ordering_information.htm)."
ORIGIN

  Query Match             5.1%;  Score 55.6;  DB 9;  Length 1335;
  Best Local Similarity   40.9%;  Pred. No. 0.14;
  Matches  190;  Conservative    0;  Mismatches  275;  Indels    0;  Gaps    0;

Qy        109 TCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAATC 168
              | | ||  | || |  | | |      |  |  ||  | | ||||  ||  ||| | |
Db        461 TATATATATATTTATATATATATATATATATATGTATATATATATATATATTTATATTTA 520

Qy        169 AGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTT 228
                |  |   | ||     | || |  |  || |    |  || | ||  | ||     ||
Db        521 TTTNTAATATATTTATATATATATTTTTATATAATTTTATTTATTTTTTTTTAAATATTT 580

Qy        229 GATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAA 288
              ||||  || ||| |  ||        |         |  ||| |||||  |||  |  | |
Db        581 TTTTTTATTTAAATATTTTTATATATTTATATATATTTTTTTTTTTTTTTTTTAATTTATA 640

Qy        289 AAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACAC 348
                |  |||| | | ||  ||  | | | | ||       ||   |      | ||  |
Db        641 TATATATATATATATATATATATATATATATATATATATANANANANATATATATANNTN 700

Qy        349 AAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAG 408
                |      |     ||         | | ||      ||   ||  || | |     |||
```

```
Db        701 NANNNTNATTNTNNTTNNNNNNNNNANTTTTTTTTNAAATTTTATTAATTTTTTTAAAAA 760

Qy        409 GTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGA 468
                  |   | ||   |||  |   |||| |||        ||||  ||||   |
Db        761 TATATNTTNTTNTTTATATATATAAAAANAAATATAAANATTTNTTTTTTTTTTTTTTAA 820

Qy        469 AATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTC 528
              |||  | |       | | || |  || |    ||     ||  |||| ||   || | |
Db        821 AATTTTATANNTTTAAANAAATNNTTTTTNTATANNNTAAATTTTATTTATAAATTTTAA 880

Qy        529 TGTCATTCAGGAGTATGGAAGTATCGATGGCTTCTTTAAAATGAA 573
              |      |      |       | |  |    ||  | |||||  ||
Db        881 TNNANNTANNNTATTAATTTTTTTTTAATTTTTTATNAAAATNAA 925


RESULT 39
CL515401/c
LOCUS       CL515401               1807 bp    DNA     linear   GSS 01-APR-2004
DEFINITION  SAIL_89_H04.v1 SAIL Collection Arabidopsis thaliana genomic clone
            SAIL_89_H04.v1, genomic survey sequence.
ACCESSION   CL515401
VERSION     CL515401.1  GI:46012721
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1807)
  AUTHORS   Sessions,A., Burke,E., Presting,G., Aux,G., McElver,J., Patton,D.,
            Dietrich,B., Ho,P., Bacwaden,J., Ko,C., Clarke,J.D., Cotton,D.,
            Bullis,D., Snell,J., Miguel,T., Hutchison,D., Kimmerly,B.,
            Mitzel,T., Katagiri,F., Glazebrook,J., Law,M. and Goff,S.A.
  TITLE     A high-throughput Arabidopsis reverse genetics system
  JOURNAL   Plant Cell 14 (12), 2985-2994 (2002)
  MEDLINE   22356987
   PUBMED   12468722
COMMENT     Contact: Sessions A
            Applied Trait Genetics
            Syngenta Biotechnology Inc.
            3054 Cornwallis Rd., Research Triangle Park, NC 27709, USA
            Email: allen.sessions@syngenta.com
            ABRC Stock Number CS804259; T-DNA left border flanking sequences of
            Syngenta Arabidopsis Insertion Library (SAIL) lines are available
            through the Arabidopsis Biological Resource Center (ABRC).
            Sequences represent a pool of amplified genomic regions and not
            single contiguous sequences.
            Class: TDNA tagged.
FEATURES             Location/Qualifiers
     source          1. .1807
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /ecotype="Columbia"
                     /db_xref="taxon:3702"
                     /clone="SAIL_89_H04.v1"
                     /clone_lib="SAIL Collection"
                     /note="T-DNA left border sequences were isolated using a
                     modified TAIL-PCR strategy"
ORIGIN

  Query Match             5.1%;  Score 55.6;  DB 9;  Length 1807;
  Best Local Similarity   41.8%;  Pred. No. 0.14;
  Matches  142;  Conservative    0;  Mismatches  198;  Indels    0;  Gaps    0;

Qy        165 AATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTC 224
              ||  | | || | | |    | |  |||| ||  | || |  |    |    |    | || |
Db        395 AAATAATAAAAAATATATAATATAATCTTATAATTAATAANTATNTAATTNTTTTTAATA 336

Qy        225 CCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTC 284
                  |||| |   | |  |||   |  || |  || | |  | |||  |    |  ||
Db        335 TTATATATTTTATTTTATAATTTAANTATTTATATATANTTANNTATTTATATATAATTN 276

Qy        285 AAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAA 344
```

```
Db        275 NATAAAAATATTTATTTATATATTNATTATATTATTATATANTNNNNNNNNTNNNTATAT 216

Qy        345 ACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCT 404
Db        215 ANNTTATTTATTTTNNTAATATATTANTTAATTTATTATTTTATATTATATTATATNNAT 156

Qy        405 AAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTT 464
Db        155 ATTTTTNNATTATNTTTTTATATATTATTTATATTTATATAATANNNNNNTTTNTTATTT 96

Qy        465 TGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATT 504
Db         95 TNNNTTATNTTNNANNTAANNAAATTATTTNTNNTTNNTT 56


RESULT 40
CNS00DKY
LOCUS       CNS00DKY                928 bp    DNA     linear   GSS 04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR27A24 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL071865
VERSION     AL071865.1  GI:4948170
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 928)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .928
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR27A24"
                     /clone_lib="RPCI-98"
                     /note="end : T7"
ORIGIN

  Query Match             5.1%;  Score 55.4;  DB 9;  Length 928;
  Best Local Similarity   33.6%;  Pred. No. 0.15;
  Matches  120;  Conservative   75;  Mismatches  158;  Indels    4;  Gaps    1;

Qy         35 TCTATGTTCTTATAAGCAAATGTTACAATGAGATATGAGTTTTAAAGCCAGATCTTCCTT 94
Db        478 TWTATATTATATAAWATAAATATGWTAWTAAATTTTTTATKAWWAAAAAAWAATWTTTTT 537

Qy         95 ATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATATT 154
Db        538 WWTTTTTWAWAAAWTATTTWTTTTTTTTWTTTWAWTTTAATTWTATWTTAWTTAAAWW 597

Qy        155 GATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGT 214
Db        598 TAWWATTTTATAWTWAATAWWAAATATWTW----ATATAATATWTWTTWTWTAWTWTTAT 653

Qy        215 TAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTT 274
Db        654 ATAWAAAWAAAAWAAAAAAWAATAAAAAATTTTAWAAWTAAAAWAAAAATTTAWTTTTTT 713

Qy        275 TGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGT 334
Db        714 TTTTTTTTWTATWTWAWATAWAAATATWWWWWWATWTWDGKNWNNNAWWWWWWWWWAWW 773

Qy        335 TGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGT 391
Db        774 WWAAAWWWWWWWWWWWAAAAAAAAAAAAAWAAAAAAAWDDDDDDDWDDDWWKAAKKKK 830


RESULT 41
CG750135
LOCUS       CG750135                1592 bp    DNA     linear   GSS 24-OCT-2003
DEFINITION  P044-3-D09.za Ppa EcoRI BAC Library Pristionchus pacificus genomic,
            genomic survey sequence.
ACCESSION   CG750135
VERSION     CG750135.1  GI:37971278
KEYWORDS    GSS.
SOURCE      Pristionchus pacificus
  ORGANISM  Pristionchus pacificus
            Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
            Neodiplogasteridae; Pristionchus.
REFERENCE   1  (bases 1 to 1592)
  AUTHORS   Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K.,
            Buntjer,J., van der Meulen,M. and Sommer,R.J.
  TITLE     An integrated physical and genetic map of the nematode Pristionchus
            pacificus
  JOURNAL   Mol. Genet. Genomics 269 (5), 715-722 (2003)
  MEDLINE   22835951
   PUBMED   12884007
COMMENT     Contact: Sommer RJ
            Evolutionary Biology
            Max-Planck-Institute for Developmental Biology
            Spemannstr. 37-39, Tuebingen D-72076, Germany
            Tel: 00497071601371
            Fax: 00497071601498
            Email: ralf.sommer@tuebingen.mpg.de
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .1592
                     /organism="Pristionchus pacificus"
                     /mol_type="genomic DNA"
                     /strain="California"
                     /db_xref="taxon:54126"
                     /clone_lib="Ppa EcoRI BAC Library"
                     /note="The library was generated by a partial digest of
                     the genomic DNA with EcoRI and cloning into the BAC
                     vector."
ORIGIN

  Query Match             5.1%;  Score 55.4;  DB 9;  Length 1592;
  Best Local Similarity   46.5%;  Pred. No. 0.16;
  Matches  190;  Conservative    0;  Mismatches  218;  Indels    1;  Gaps    1;

Qy        119 TTGAAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGT 178
Db        663 TTATAATTATTATTATAATATTTTTTAATAATTTTTTTTTATATTATATATATTTATTAT 722

Qy        179 TTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTA 238
Db        723 TATTAATAATATTTTATATAATAATTAATATAATTATTATATANTAAATAAATATTTATA 782

Qy        239 AAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAA 298
Db        783 ATTTTTATNATTTATATATA-TAATATATATTTAATTAANTTTTTANATAATAATTTTATT 841
```

```
Qy        299 GAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTG 358
Db        842 AATTTTTATTAAAATAATTTTNATATTATAANATTTATTAAATTTTTATTATTAATATTT 901

Qy        359 CATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACC 418
Db        902 TATANATTATAATTTTANAAATATAATAAATATAATAATANAAAATTATTTTATATTTAA 961

Qy        419 TGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCA 478
Db        962 TTATATATTTNNATTTTATTATTATTATTTTTTTTTAAATATTATATTATTTATATTATA 1021

Qy        479 GAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATT 527
Db       1022 TTTATATTTATTTTTAATATATNATTTAAATATATAAAAATTAAATATT 1070


RESULT 42
BH988719/c
LOCUS       BH988719                690 bp    DNA     linear   GSS 07-OCT-2002
DEFINITION  oee37h07.g1 B.oleracea002 Brassica oleracea genomic, genomic survey
            sequence.
ACCESSION   BH988719
VERSION     BH988719.1  GI:23525111
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 690)
  AUTHORS   Delehaunty,K., Fewell,G., Fulton,L., McCombie,W.R., Miner,T.,
            Nash,W., Rabinowicz,P.D. and Wilson,R.K.
  TITLE     Whole genome shotgun reads from Brassica oleracea
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Richard K. Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Plate: oee37  row: h  column: 07
            Seq primer: -28RPpOT reverse
            Class: shotgun
            High quality sequence start: 90
            High quality sequence stop: 551.
FEATURES             Location/Qualifiers
     source          1. .690
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3712"
                     /clone_lib="B.oleracea002"
                     /note="Vector: pOTw13; Whole genome shotgun library from
                     flowering buds. DNA was purified  from a crude nuclear
                     prep using Brassica oleracea TO1000DH3 buds  provided by
                     Thosmas Osborn at the University of Wisconsin. Genomic
                     DNA was provided by Pablo Rabinowicz (CSHL) and the
                     shotgun library  prepared at Washington University Genome
                     Sequencing Center."
ORIGIN

  Query Match             5.1%;  Score 55.2;  DB 8;  Length 690;
  Best Local Similarity   48.7%;  Pred. No. 0.16;
  Matches  150;  Conservative    0;  Mismatches  158;  Indels    0;  Gaps    0;

Qy        207 TGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCA 266
Db        543 TGTTATTCAAAAATCTACTTTTTTTTCTTTTAACTGAAGAAATATATATAGAGATTTATA 484

Qy        267 TTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATAT 326
Db        483 TTTGTTTTCTTAAATTTAAAAAAAGTTATGAAAGAAATTATTTTTGGATAAACTGTAAAT 424

Qy        327 TAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCA 386
```

```
Db        423 ACTTTAAGATAAATATAATTATCAAAATCTTTAGAATGTTTTCTTTTAATTTATGCTTAA 364

Qy        387 AAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAA 446
Db        363 TTATATATTAGTAAATTTAGTTATATATTAAATTACCAAAGATTTTAGCTATTCTAATTT 304

Qy        447 TTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAA 506
Db        303 TTAACTTAAAGTTCAAGCAAAAAAAAAACTCTCTGAAAATAATAGTATAAATAGAATAAA 244

Qy        507 AACATTTA 514
Db        243 ATGATATA 236


RESULT 43
CNS015ZZ
LOCUS       CNS015ZZ                1201 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN15K05 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL106121
VERSION     AL106121.1  GI:5620267
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1201
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN15K05"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : T7"
ORIGIN

  Query Match             5.1%;  Score 55.2;  DB 9;  Length 1201;
  Best Local Similarity   31.1%;  Pred. No. 0.17;
  Matches  143;  Conservative  102;  Mismatches  214;  Indels    1;  Gaps    1;

Qy         34 CTCTATGTTCTTATAAGCAAATGTTACAATGAGATATGAGTTTTAAAGCCAGATCTTCCT 93
Db        736 CCCTAMATTYTTAYTTTTAATTAWTTCHAAMAAAAAAAAMCHHWAHHAAAAAAMCMAACYT 795

Qy         94 TATCTCTCTGCCCCATCTCTA-GTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATA 152
Db        796 TTTTCCTTTTACAWHYCTTTAHCTTTTWTTCYTTTCTTCCCTAWAAYTCCCCCTTCATTW 855

Qy        153 TTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTT 212
Db        856 TTTWTTMMTTCHAWCCCCTHWTTCYCYYCCMYHHHCCHTYYCTWAABTTTAATTTSCTTT 915

Qy        213 GTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTT 272
Db        916 TTTTTYSSTAMYSYTMMCATWTTHCMMHYYCYMAWWHHHHHHHHHHHYYMHYMMWAWTTT 975
```

```
Qy        273 TTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCA 332
Db        976 TTTTATTTTTWCAACSYTTAWAWWWWTWTTTTAAWTTWTATWAWAWTTTWTTTWTTTTCT 1035

Qy        333 GTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTA 392
Db       1036 TATTTTTAAAAAWWTAAAAAAAAAAAAAATAAAAAAAWAWHWAAAAAAAAAWTTAATAAAA 1095

Qy        393 AAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCA 452
Db       1096 AAAAAAAAAATTWAATAAAWWAAAAWAAAAAATWAAAAAAAAAAWAAAAAAAAAWAAAA 1155

Qy        453 CAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTT 492
Db       1156 AWAAWTWTATWTTWTWTWATTWTWAWHWWAAWWAATWAWT 1195
```

RESULT 44
CF238805
LOCUS CF238805 1626 bp mRNA linear EST 05-AUG-2003
DEFINITION AGENCOURT_15099447 NICHD_XGC_Emb6 Xenopus tropicalis cDNA clone IMAGE:6995950 5', mRNA sequence.
ACCESSION CF238805
VERSION CF238805.1 GI:33442013
KEYWORDS EST.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae; Xenopus; Silurana.
REFERENCE 1 (bases 1 to 1626)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Robert M. Grainger
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM14680 row: g column: 21
High quality sequence start: 71
High quality sequence stop: 316.

FEATURES Location/Qualifiers
source 1. .1626
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="IMAGE:6995950"
/tissue_type="neurula"
/dev_stage="embryo, stages 14-19"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NICHD_XGC_Emb6"
/note="Vector: pCMV-SPORT6.1; Site_1: NotI; Site_2: EcoRV; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2.1 kb. Constructed by Invitrogen. Note: This is a Xenopus Gene Collection (XGC) library."

ORIGIN

Query Match 5.1%; Score 55.2; DB 7; Length 1626;
Best Local Similarity 42.3%; Pred. No. 0.17;
Matches 159; Conservative 0; Mismatches 217; Indels 0; Gaps 0;

```
Qy        135 GAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCA 194
Db        388 GAGATATATATAAAAAAATTTTTTTTNTTTNTTTNTTTTTTNTTTTNTTAAAAAAAAAAA 447

Qy        195 TCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACA 254
Db        448 AAANATAAATTTTTTTTTTNTTTTTTTTTTTTTTTNTNTNNNNAAAAAAAAAAAAAAAAA 507

Qy        255 TACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGA 314
Db        508 AAAAATTTTAATTTTTTTTTTTNNNNNNNNNNANAAAAAAAAAAAAAAAAAAAATTTATAAA 567

Qy        315 AAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTA 374
Db        568 TATTTTTTTNNNNTNNNNNAAAAAAAAAAAAAAAAAAAAAATAATATATTATNTNTNNTAA 627

Qy        375 ATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCAC 434
Db        628 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAAAAAAAAAAAAANAANAAATNTNA 687

Qy        435 TGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTT 494
Db        688 AAAAAAAAAAAAATAATAATTAAAAGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAANAAAAT 747

Qy        495 ATGTACTATTAAAACA 510
Db        748 TTTATNNAATTAAAAA 763
```

RESULT 45
CNS00FUH
LOCUS CNS00FUH 996 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC: BACR31O21 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL071063
VERSION AL071063.1 GI:4951105
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 996)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES Location/Qualifiers
source 1. .996
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR31O21"
/clone_lib="RPCI-98"
/note="end : TET3"

ORIGIN

Query Match 5.0%; Score 55; DB 9; Length 996;
Best Local Similarity 36.4%; Pred. No. 0.18;

```
  Matches  119;  Conservative   51;  Mismatches  157;  Indels    0;  Gaps    0;

Qy        209 TTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATT 268
              :|:  | :|||| :      |: :||| :|||   :: |   : | : |   ||| |||
Db        576 WTWAHTMMATATTTWAAAAAAAWAWTTTCWAAAATTMWAAWTWMTAAWTAWAAATTWATT 635

Qy        269 TGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTA 328
              :  ::          |:|||| || |||: :||| ||:| | | ||| |: | | | |:
Db        636 ATWAAYWAYAAAAAAAAWAAAAWGTTTAAWTMAATACAWTAAWATAAATAWWAAAAAATW 695

Qy        329 GCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAA 388
                |    |||| ||: | |||| | :    :: | |    : ||          :|
Db        696 AWWAAAAAAAAAAAAAAWAATAAAAAAAWWAATAWWAAAAAATAWAAAATAAAWTAATTWA 755

Qy        389 GGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATT 448
              |||||   ||   ||:  |  : ||  |   ::  :       |  |:::| :|| |
Db        756 TMAAAAAAAAAAAAAAAAAWATTTRWTTATATWAAWWAAWTAAWWTWTTKAWWWATWAAAT 815

Qy        449 ACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAA 508
              :  | ||| :  |:: |  ||| ||    | |||| :||::| :|:|  |   |:| |||
Db        816 WAAAAAAAAWAAAHWATTAAAAAAAAAAAAAAAAAWWTAWWTAWTWAAATYWAAYTTAAA 875

Qy        509 CATTTACTTTCCAAATATTCTGTCATT 535
              |: :       ::::|| | |  ||
Db        876 WTTWAWAAAAAATWWWWTTTTTTTTT 902


RESULT 46
CNS00BAO
LOCUS       CNS00BAO               1101 bp    DNA     linear   GSS 04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR22H09 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL056736
VERSION     AL056736.1  GI:4937404
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR22H09"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN

  Query Match             5.0%;  Score 55;  DB 9;  Length 1101;
  Best Local Similarity   38.1%;  Pred. No. 0.18;
  Matches  102;  Conservative   44;  Mismatches  122;  Indels    0;  Gaps    0;

Qy        162 ACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATA 221
              :| ||| |   |::| |: |  : || || :    ||:|| :  :| |||| |: ::| |
Db        826 MCBAATAAAAAADWATTYATTAWWAAAATTWTTATTTWAATWTSKTATTTTTTDKDKANA 885

Qy        222 CTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTG 281
                  ||  |   ||:|  |   : |  |   ||  | ::|||  ||| ||:|
Db        886 WAAATTTATTAMATTKATTAATTMATNAAATTATTTAASYATTTTTTTTTTWTAATAAMA 945

Qy        282 TTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAAT 341
                |  |:|||   :| |  ||    :|    | ||| :  | :| | |     ||   : |
Db        946 ATATAWAAAATAMTYATTAATAAWWTATBAAAAAAMWAAAAWTTTAAATTCATTTTYWTT 1005

Qy        342 TAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAA 401
              :|   :  :|:| : : :|||| |  |: | |:|: ::  |   |||   |:||| : |:
Db       1006 YATTWWTTWAWTTWTWYMATATCAATTWNCATWABAWWAAAWAAAAAAAAAWAAAAWAAW 1065

Qy        402 CCTAAAGGTATACTACCTGCTTTCTTAT 429
                :|||  |: | ||||:::::|| || |
Db       1066 AAHAAAATTWAAWTACCYKYWTTTTTTT 1093


RESULT 47
CNS025K2/c
LOCUS       CNS025K2               609 bp    DNA     linear   GSS 01-SEP-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            238G01 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL182171
VERSION     AL182171.1  GI:7820265
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE   1
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
  MEDLINE   20296633
   PUBMED   10835645
REFERENCE   2
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
            Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
            Saurin,W., Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
  MEDLINE   20359837
   PUBMED   10899143
REFERENCE   3  (bases 1 to 609)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .609
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="238G01"
```

```
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG238AD01LP1~end : T7"
ORIGIN

  Query Match             5.0%;  Score 54.8;  DB 9;  Length 609;
  Best Local Similarity   46.0%;  Pred. No. 0.19;
  Matches  165;  Conservative   15;  Mismatches  173;  Indels    6;  Gaps    1;

Qy        174 ATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTT 233
              ||| ||||  | || ||   || |:|||   ||| ||   ||| |||| |    |  |:|
Db        485 ATAATTTTTAAAAATATAATATATWAAAATTATTTTTAAATTATTATATTTKAATATTWT 426

Qy        234 TTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGT 293
              |||||  |    ||| | | | |    | || | ||   |     |:| || |   ||  |
Db        425 TTTTATKATKAATTATAAAAAAATTTNTTTTTAATTAAATAATTTWTTTTTATTTAAATT 366

Qy        294 ATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAAT 353
              |  | | ||    | | |        | ||:|:       ||||||  |  | | |||
Db        365 TTTTAAAGATTTTATAAAATTTTTKAATTTTAWTWAATTTAATAAAATTTATKAAATAAT 306

Qy        354 GAGTGCATATTACATTACTTAATCTTGC------AGTCAAAGGTAAAAAGTCAACCTAAA 407
              | |   ||||   ||  ||  | ||         | | |||  |||| | | ||  | |:
Db        305 WATTTTTTATTTTTTTTTTTTTTTTTTTTAATAWAWTAAAAATTAAATAATTAAAATTAW 246

Qy        408 GGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGG 467
                | || |   :  ||| || |      |    : |   | :|  |   ||||  :| |
Db        245 AATTTAATKAAWAWTTTTTTTTATTTTTTTTTWAAATTAAKAATAATTATTTAKWTATTT 186

Qy        468 AAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATAT 526
              |     | |    :||  |||||| |:|: || :|||||    |||  ||:  | ||||
Db        185 ATTATTTTTTTTTWAATTTAATTTATWAWTTATKATTAATTTTTTTTTTTKAAATATAT 127


RESULT 48
CL078758/c
LOCUS       CL078758               1168 bp    DNA     linear   GSS 31-DEC-2003
DEFINITION  CH216-153C16_RM4.1 CH216 Xenopus tropicalis genomic clone
            CH216-153C16, genomic survey sequence.
ACCESSION   CL078758
VERSION     CL078758.1  GI:40534671
KEYWORDS    GSS.
SOURCE      Xenopus tropicalis (western clawed frog)
  ORGANISM  Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 1168)
  AUTHORS   Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
            Mardis,E. and Wilson,R.
  TITLE     A physical map of the xenopus tropicalis genome
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 175000   Std Error: 0.00
            Seq primer: RM4 ctcaagggcatcggtcgagc
            Class: BAC ends
            High quality sequence start: 151
            High quality sequence stop: 204.
FEATURES             Location/Qualifiers
     source          1. .1168
                     /organism="Xenopus tropicalis"
                     /mol_type="genomic DNA"
                     /strain="Nigerian frog"
                     /db_xref="taxon:8364"
                     /clone="CH216-153C16"
                     /sex="male"
                     /cell_line="Stock 248 F7A2, inbred N7"
                     /clone_lib="CH216"
                     /note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis
                     BAC library"
ORIGIN

  Query Match             5.0%;  Score 54.8;  DB 9;  Length 1168;
  Best Local Similarity   46.2%;  Pred. No. 0.21;
  Matches  176;  Conservative    0;  Mismatches  203;  Indels    2;  Gaps    1;

Qy        133 ATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCT 192
              ||   |||      | |  |     ||  ||  || | | | | ||  ||| || || ||
Db       1057 ATATTTTTAAAATAAATTTTNNAATTTTTTTAATTTAATAATTAAATTTTTAAAATATNA 998

Qy        193 CATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTA--CA 250
              | ||       |  | |||||||  |  ||  |||||||  |||    | |   |
Db        997 AAAATTTTTTTATTAAAAATATTAATATTATTATTTTTTTTTTTTTAAATTTATAAATNA 938

Qy        251 GACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCA 310
              ||  |   |  ||||  ||| |  |   ||   |||      | || || |  | | |
Db        937 TTTATTAAAAAAANTATTTTTTTATATTAAATTATTAAATAAATTTAAAAATTTTAATAA 878

Qy        311 GAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTA 370
              || |   |||  |       || ||| | ||  | |  | | | |      |  ||||
Db        877 ATTTAAAAATTTATTATTAATTATTTAAATTAAANAAAATTAAAAATTANNNATTAATTA 818

Qy        371 CTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATC 430
              |||     |  |||  |  || | ||  |||  |||| ||     ||| |||
Db        817 TATAAATNATAAAATAAAAATTNTAAATNAAATAAAAATTATAATAAANTTTTTTTTAAA 758

Qy        431 GCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATT 490
              |   ||||  ||  |  | || |||| ||||  || | ||   | | ||  | |
Db        757 TNAAATAAAATTTTAAAAATTAAAATTTTTTTTTTAAAATTTATTNAATANAAATTTAAA 698

Qy        491 TTTTATGTACTATTAAAACAT 511
              || ||    |   | |  ||
Db        697 ATTAATTANNTNANANATAAT 677


RESULT 49
CNS0161D/c
LOCUS       CNS0161D               1225 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN15C18 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL106171
VERSION     AL106171.1  GI:5620504
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1225)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1225
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN15C18"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : SP6"
ORIGIN
```

```
  Query Match             5.0%;  Score 54.8;  DB 9;  Length 1225;
  Best Local Similarity   35.1%;  Pred. No. 0.21;
  Matches  101;  Conservative   55;  Mismatches  132;  Indels    0;  Gaps    0;

Qy        278 TTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGA 337
              ||: ||  |||||  ||||:| : ||   ::  | | |:   | ||||: |    : | |
Db       1178 TTWTTTATAAAAATATATAWATWTATTATWWWAAAWAWTTTTTATATWATTWTAWATAAA 1119

Qy        338 AAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAG 397
              |:|  ||    |::: :    | : :| | :: |:: : ::    | |||:  :| |
Db       1118 AWAAAAATATTAWWWAWWWAAWAAWAWAWAAWWATWWTWTWKAATATAAAARTAWATATA 1059

Qy        398 TCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAAT 457
              | ||  |:|:  ::|: :|     | | || |   :  |  |:| | ||| :|  :  ||
Db       1058 TAAAAATWAWAWWWTWAWAAAATATATTTTTTTTTWTTTTAWAAAAAAAAWAAAWATAAA 999

Qy        458 TTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTT 517
              ||  :: |  |::: |    | || | :| ||||: :|: |  |   ||:: :|||  ||
Db        998 TTATWWATAAAWWWTAAAAAAAAATAAWTTATTTWAWAWTTTTTTRAAAWWTWTTTTTTT 939

Qy        518 TCCAAATATTCTGTCATTCAGGAGTATGGAAGTATCGATGGCTTCTTT 565
              |  :  || :   | :||    : | : : :: |    :   || |||
Db        938 TTWWTTTAAWAAWTWWTTATWTRTTTWTRTRRRARAATWWWTTTTTTT 891


RESULT 50
CNS00EO7/c
LOCUS       CNS00EO7                1101 bp    DNA     linear   GSS 04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC:
            BACR29P01 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL069440
VERSION     AL069440.1  GI:4949583
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR29P01"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN

  Query Match             5.0%;  Score 54.6;  DB 9;  Length 1101;
  Best Local Similarity   36.0%;  Pred. No. 0.23;
  Matches  183;  Conservative   64;  Mismatches  260;  Indels    1;  Gaps    1;

Qy         31 ACTCTCTATGTTCTTATAAGCAAATGTTACAATGAGATATGAGTTTTAAAGCCAGATCTT 90
              | | : | | |   || ||   :|   || ||    | :: | || :||| |   :| ||
Db        954 AATWWTTTTTTATATACAATTTWAAAATAAAAAWTAACWWAATTTAWAAAWCATTWTTTT 895

Qy         91 CCTTATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAA 150
              ::|   | :    ||: : || :|    || | |   |:   |        :   |:|
Db        894 TAAWWTAATTTWATTACAWTWTTAWWTAAAAAAATWTTAAAWTAAAWAAAAARWATTAWA 835

Qy        151 TATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTT 210
              ||    : || : |  : |  |:| | || :: |   | | :  ||       :|
Db        834 ATTTAYATWATATWTAAAWWTATAWATTATTMWWAATWTTATAWNATTTTTTTTWTAWAA 775

Qy        211 TTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTG 270
              :| ||:||: :      || |:|::|| || | |     |   | |  |  |:||  |:
Db        774 WTWTTWATWAWTAATTTTAAWTWWTTAAATAAAAAAAATTTATTTTTATTTWTTATTWAA 715

Qy        271 TTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGC 330
              :|||  ||| :    ::      |||:  |||:   |  | | : | :| | |||
Db        714 AWTTTTWTTTTWAATTWWYTTTAATAWTTAAAWTWTTAAAWAWWTAAWTTAAATAAATWT 655

Qy        331 CAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCA-GTCAAAG 389
              |  || :|:||: | ||| | |  :  ||| :  ||  ||:  ||  |  |  ||
Db        654 ATTTAAAATWTWAAWTATAAATTTAWADWTTATAWWTTTTTTTWWCWTTWAATATATAAA 595

Qy        390 GTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTA 449
              |:|||| | ||  |:    |||| ::  |     : |::   | |   ||:: |   |
Db        594 ATWAAAAATTAAATTWTTTTTATATWWWATAAAAAMATWWWTTATTTATAAWWAATTATT 535

Qy        450 CCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAAC 509
              : :    || || |  |  |: | |   ||:  || | | |:|| || || |:  |
Db        534 TAWAWTTGATTTTTATTTATTTWTTTTTTAAAWTTTTATTATATWATTTATTAATWTTAW 475

Qy        510 ATTTACTTTCCAAATATTCTGTCATTCA 537
              | |  |||    ||||  |  :| : |
Db        474 TTATYTTTTTTTTATATCTTTCMAAWTA 447


RESULT 51
CNS0167M
LOCUS       CNS0167M                1201 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN15M24 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL106396
VERSION     AL106396.1  GI:5621701
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1201
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN15M24"
                     /clone_lib="DrosBAC"
```

```
                     /plasmid="pBeloBAC11"
                     /note="end : T7"
ORIGIN

  Query Match             5.0%;  Score 54.6;  DB 9;  Length 1201;
  Best Local Similarity   37.7%;  Pred. No. 0.23;
  Matches  180;  Conservative   53;  Mismatches  237;  Indels    8;  Gaps    1;

Qy         62 ATGAGATATGAGTTTTAAAGCCAGATCTTCCTTATCTCTCTGCCCCATCTCTAGTTCTTG 121
Db        671 RKKTKKTKKTTGTDAAAAAAATNTTTTTTTTTTTTTTTTTTTTTTTTTTTATDWAATTW 730

Qy        122 AAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTT 181
Db        731 TTTTTTGTTTTATGNATTTAAWTTATATTTAWWATTTTTTWWWTWTTTWTTTTTTATTTT 790

Qy        182 GTAGAAGA--------TCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTT 233
Db        791 NAAAAARAATTAAATTTTTTTTTTTTWAAAAATWTTATAATWAAAWTTTTTWWTTTTTT 850

Qy        234 TTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGT 293
Db        851 TTTTTTTTTWTTTTATTTTATAATTTTTAWATTTTTAAWTTTTTTTTTTAAWRAAAAAA 910

Qy        294 ATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAAT 353
Db        911 WTWTWTTTATAWATAAAAAAAAWTTTWATATTWTTTTTTTTTTWAAAAAWWATWAAAAAT 970

Qy        354 GAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATA 413
Db        971 TTTTAAAAATTTTWATAATTAWTTTATTTTTAAAATTTTTTTTTWTTTWTWTATAAAAAA 1030

Qy        414 CTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAA 473
Db       1031 AAAAAAAAAAWATTWTATANWAAAWAAATATTNTADAAAWWWTTTTTTTTTTTTTTTWN 1090

Qy        474 TCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGT 531
Db       1091 TTATATTWTATTTAATATWWWWTTTWTTWTTTATATWWTATTTTWTTTTTTATANTTT 1148


RESULT 52
BM605197
LOCUS       BM605197                576 bp    mRNA    linear   EST 25-FEB-2002
DEFINITION  17000687076007 A.Gam.ad.cDNA.blood1 Anopheles gambiae cDNA clone
            19600449700714 5', mRNA sequence.
ACCESSION   BM605197
VERSION     BM605197.1  GI:18903301
KEYWORDS    EST.
SOURCE      Anopheles gambiae (African malaria mosquito)
  ORGANISM  Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
            Anopheles.
REFERENCE   1  (bases 1 to 576)
  AUTHORS   Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L.,
            Charlab,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
  TITLE     Celera Anopheles gambiae EST project
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Holt R.A.
            Celera Genomics
            45 w. Gude Dr., Rockville, MD 20850, USA
            Tel: 2404533151
            Fax: 2404534580
            Email: HoltRA@celera.com
            Plate: NU01004AAY  row: G  column: 04
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .576
                     /organism="Anopheles gambiae"
                     /mol_type="mRNA"
                     /strain="RSP-ST (Reduced susc. to Permethrin - std.
                     chromosome)"
                     /db_xref="taxon:7165"
                     /clone="19600449700714"
                     /dev_stage="Adult"
                     /lab_host="DH10b"
                     /clone_lib="A.Gam.ad.cDNA.blood1"
                     /note="Vector: pSport1; Site_1: Sal1; Site_2: Not1; Whole
                     adult mosquitoes (mixed sex) frozen on liquid nitrogen 24
                     hours  after human blood feeding. cDNA inserts >500 bp
                     cloned directionally  into pSport 1. Not 1 site is 3'.
                     Clones available through the Malaria  Research and
                     Reference Reagent Resource Center (www.malaria.mr4.org)"
ORIGIN

  Query Match             5.0%;  Score 54.4;  DB 4;  Length 576;
  Best Local Similarity   46.0%;  Pred. No. 0.23;
  Matches  184;  Conservative    0;  Mismatches  216;  Indels    0;  Gaps    0;

Qy        141 GGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAA 200
Db        173 GTTTATTAAATAATAAATTATTATAATAATTTTTCTCGAATTTTAAAGATTTAATTATAT 232

Qy        201 AGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGC 260
Db        233 ATAAAAGTTATTGTGGAATAACTATTTTAAATATGTAATTAGAAATGAAATGTTAATCGT 292

Qy        261 TATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATC 320
Db        293 TTTAAAATATATCTAGTTTTTTAAGAAATGAATTTAATTTAGCTTATTTATTTTATTAAG 352

Qy        321 ATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTG 380
Db        353 TTAATTTTTTAATTTAATAATTAAATAAAGTAATATTTTAAGGGATGAGCTTTAAAATAA 412

Qy        381 CAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAA 440
Db        413 AATTTTATATTTTTTATAATTTTTAAATAAATATAAGCTTAAAAATAGCTATTATTAATA 472

Qy        441 TAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTAC 500
Db        473 AATTTGTTATAATTTATTTTTTATAAAAAATTATTTAATTTAAATTAAATTATTTATTAA 532

Qy        501 TATTAAAACATTTACTTTCCAAATATTCTGTCATTCAGGA 540
Db        533 AATTTAAATTTTTAATTAATAAAAATTTAGTAATTATGGA 572

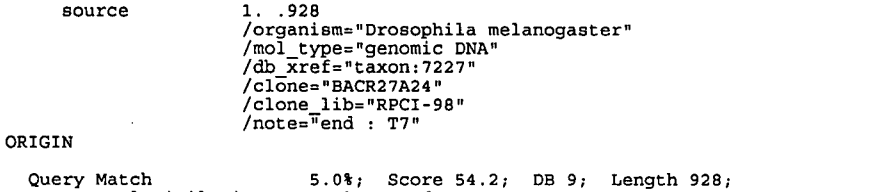

RESULT 53
CNS00DKY/c
LOCUS       CNS00DKY                928 bp    DNA     linear   GSS 04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR27A24 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL071865
VERSION     AL071865.1  GI:4948170
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 928)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
```

Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES Location/Qualifiers
source 1. .928
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR27A24"
/clone_lib="RPCI-98"
/note="end : T7"

ORIGIN

Query Match 5.0%; Score 54.2; DB 9; Length 928;
Best Local Similarity 31.6%; Pred. No. 0.27;
Matches 126; Conservative 88; Mismatches 180; Indels 5; Gaps 1;

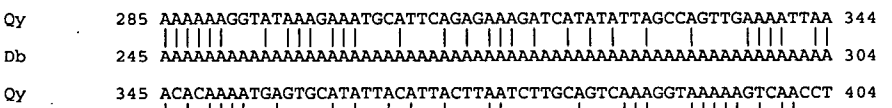


RESULT 54
CN834865
LOCUS CN834865 693 bp mRNA linear EST 02-JUN-2004
DEFINITION AGENCOURT_24959824 NIH_MGC_169 Mus musculus cDNA clone IMAGE:30916765 5', mRNA sequence.
ACCESSION CN834865
VERSION CN834865.1 GI:47940162
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 693)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Daniela S. Gerhard, Ph.D.
Office of Cancer Genomics
National Cancer Institute / NIH
Bldg. 31 Rm10A07 Bethesda, MD 20892
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Dr. Jonathan Kuo, NIMH
cDNA Library Preparation: Michael Brownstein Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: NDCM381 row: o column: 14
High quality sequence stop: 374.

FEATURES Location/Qualifiers
source 1. .693
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30916765"
/lab_host="DH10B (T1-phage-resistant)"
/clone_lib="NIH_MGC_169"
/note="Organ: Testicles; Vector: pDNR-LIB; Site_1: SfiI (ggccattatggcc); Site_2: SfiI (ggccgcctcggcc); cDNA made by oligo-dT priming and directionally cloned. 5' and 3' adaptors were used in cloning as follows: 5'-AAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGG-3' and 5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)NN-3'. Full-length enriched library was constructed using the Clontech Creator SMART kit and size-selected to contain the 0.5 kb size fraction. Library created in thelaboratory of M. Brownstein (NIMH, NIH). Note: this is a NIH_MGC Library."

ORIGIN

Query Match 5.0%; Score 54; DB 7; Length 693;
Best Local Similarity 50.2%; Pred. No. 0.29;
Matches 126; Conservative 0; Mismatches 125; Indels 0; Gaps 0;



RESULT 55
CNS00KHX
LOCUS CNS00KHX 905 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC: BACR17N06 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL077798
VERSION AL077798.1 GI:4957174
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 905)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

```
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES              Location/Qualifiers
     source           1. .905
                      /organism="Drosophila melanogaster"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:7227"
                      /clone="BACR17N06"
                      /clone_lib="RPCI-98"
                      /note="end : T7"
ORIGIN

  Query Match             5.0%;  Score 54;  DB 9;  Length 905;
  Best Local Similarity   36.0%;  Pred. No. 0.3;
  Matches  116;  Conservative   61;  Mismatches  144;  Indels    1;  Gaps    1;

Qy        206 TTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTC 265
              || :|| :|:  :|||      |:| ::::  : :       |:| |    ||: |: :|
Db        511 TTAWTTGKTKTTKATAAATGTTWGKKKKKRRWTWKTTKDKTKAAAATTATWAAKAWAAAK 570

Qy        266 ATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATA 325
              ||::::|:||: :::::| : |||| :| :|  || :  | | :   ||| |  ||| |
Db        571 ATKWKKTWTTDTKWWKWTAWKAAAAAAHAAWATAAATWATAATAWATAAAAAAAAATAAA 630

Qy        326 TTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTC 385
              |:|   |    ||||| ||: | ||||    |  :|  :| | |   |    ||      |
Db        631 TWAAAAAAAAAAAAAATAAAWWATAAAAAATATWTWTTAWAAAAAAAAAAAAAWAAAWAWAA 690

Qy        386 AAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGA- 444
              |||   |||||   ||  . ||  :||| :   :  |:: | ::   |    |:| | |
Db        691 AAAAAAAAAAAAAAAAAAAATAAAAWATAWWTTAWWTTWWWTAWWAWTAWAWAAWAAAAAT 750

Qy        445 AATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATT 504
              ||| :  |  ||:  :  | :  | ||:|      |:|: |||:| :: ||: ||  ::
Db        751 AATAWWAAWTAAWARWGATAWATATATWAAAAATTAWAWTATAWTAHWATAWTTATAWWA 810

Qy        505 AAAACATTTACTTTCCAAATAT 526
              || | |:  | : |  |||:::
Db        811 AATAWAWAAAAWATAWAAAWWW 832


RESULT 56
CNS017KX
LOCUS       CNS017KX                1101 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN37F10 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL108171
VERSION     AL108171.1  GI:5628475
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES              Location/Qualifiers
     source           1. .1101
                      /organism="Drosophila melanogaster"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:7227"
                      /clone="BACN37F10"
                      /clone_lib="DrosBAC"
                      /plasmid="pBeloBAC11"
                      /note="end : SP6"
ORIGIN

  Query Match             5.0%;  Score 54;  DB 9;  Length 1101;
  Best Local Similarity   23.0%;  Pred. No. 0.31;
  Matches  100;  Conservative  147;  Mismatches  185;  Indels    2;  Gaps    1;

Qy         91 CCTTATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAA 150
              ||   :::  :| : :: | | |   |::  :: | | |::| |   :|| |:   :::
Db        619 CCCCCYYHTMMTMMTYYTTATTTTYMTYYCCMMMTHTTTHWTTTMTWWTTCTTYTTTWWH 678

Qy        151 TATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTT 210
              :|:   : :|||      :         ||| :|      |  ::: || :  | |:| :|
Db        679 WAWAYYWTWTTATTTTCTHAAMCCATTTTTCCYAATTTTTTCYHYATTCMYAAAAWTHWT 738

Qy        211 TTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATT-- 268
              |: :: : : : :|: ||  |::|||||:|: |:  ::    ::  :: : : ::| :
Db        739 TYTYYCHAMAMAWCMTTTTYTYYTTTTAMAWAAYAMWMACMCHHACYMMYCMCYMACYHM 798

Qy        269 TGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTA 328
              | ||||:  ||| |::|:: |  :|  |  ::| ||  ::   | | : ::| | ||
Db        799 TWTTTTYYCTTTTTYMAMWTATTHWAAYACTTWWTYCAMAYMMYMATAMMAHWTTTTTTT 858

Qy        329 GCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAA 388
              | | |   :  :        :::: :   |:: ||      :: :|:  :| :|   ::|
Db        859 TCWAYTMMCMCCHMCTTTTTYMHWYWHHAYCWHTTTTTCCCMYMYAMAWWTCMAACWWWA 918

Qy        389 GGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATT 448
              ||::   : :  ::|    : : :: ::: :: |: |  : :| : :|: :: ::|::
Db        919 AWTAWWTTYYWWYWMYAYYYYYYYCYHCYYHHYYMTYCTYCHCYATWYMAWYHHMWMAHW 978

Qy        449 ACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAA 508
              :: ::|: ::::: ||:  :::::: : :| || : |::  |  |||| : ::|||  ::
Db        979 MHWWHAHYWWHYWMTTYYTWWMWMWAYAMACAAYMWAHHTATAATTATWYTYYATTCYWH 1038

Qy        509 CATTTACTTTCCAA 522
              :: :|:::  :|::
Db       1039 YMAHTMYHCCMCMM 1052


RESULT 57
BM645074
LOCUS       BM645074                560 bp    mRNA    linear   EST 26-FEB-2002
DEFINITION  17000687315640 A.Gam.ad.cDNA1 Anopheles gambiae cDNA clone
            19600449628059 5', mRNA sequence.
ACCESSION   BM645074
VERSION     BM645074.1  GI:18944585
KEYWORDS    EST.
SOURCE      Anopheles gambiae (African malaria mosquito)
  ORGANISM  Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
            Anopheles.
REFERENCE   1  (bases 1 to 560)
  AUTHORS   Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L.,
            Charlab,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
```

```
  TITLE     Celera Anopheles gambiae EST project
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Holt R.A.
            Celera Genomics
            45 w. Gude Dr., Rockville, MD 20850, USA
            Tel: 2404533151
            Fax: 2404534580
            Email: HoltRA@celera.com
            Plate: NU01003CY6  row: D  column: 09
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .560
                     /organism="Anopheles gambiae"
                     /mol_type="mRNA"
                     /strain="RSP-ST (Reduced susc. to Permethrin - std.
                     chromosome)"
                     /db_xref="taxon:7165"
                     /clone="19600449628059"
                     /dev_stage="Adult"
                     /lab_host="DH10b"
                     /clone_lib="A.Gam.ad.cDNA1"
                     /note="Vector: pSport1; Site_1: Sal1; Site_2: Not1; Whole
                     adult mosquitoes (mixed sex) frozen on liquid nitrogen.
                     cDNA  inserts >500 bp cloned directionally into pSport 1.
                     Not 1 site is  3'. Clones available through the Malaria
                     Research and Reference  Reagent Resource Center
                     (www.malaria.mr4.org)."
ORIGIN

  Query Match             4.9%;  Score 53.8;  DB 4;  Length 560;
  Best Local Similarity   46.4%;  Pred. No. 0.31;
  Matches  175;  Conservative    0;  Mismatches  202;  Indels    0;  Gaps    0;

Qy        148 AAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATT 207
              ||  ||  || |||| ||     ||  | |  || ||||||| | ||  || | | |
Db        171 AATAATAAATTATTATAATAATTTTTCTCGAATTTTAGAAGATTTAATTATATATAAAAG 230

Qy        208 GTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCAT 267
              | ||||  |   |||   || | | | | | |||  | | |     |  | | | ||
Db        231 TTATTGTGGAATAACTATTTTAAATATGTAATTAGAAATGAAATGTTAATCGTTTTACAA 290

Qy        268 TTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATT 327
              |  | | | ||| |  |||   | |||   |     ||| |   |  |   ||  ||
Db        291 TATATCTAGTTTTTTAAGAAATGAATTTAATTTAGCTTATTTATTTTATTAAGTTAATTT 350

Qy        328 AGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAA 387
              | || | ||||||| | |  | | |  |     |     |||| |   | | |
Db        351 TTTAATTTAATAATTAAATAAAGTAATATTTTAAGGGATGAGCTTTAAAATAAAATTTTA 410

Qy        388 AGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAAT 447
              |   |       ||||   |||  | ||       |          | | | |    |
Db        411 TATTTTTTATAATTTTTAAATAAATATAAGCTTAAAAATAGCTATTATTAATAAATTTGT 470

Qy        448 TACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAA 507
              ||  |   ||||| | |   |||| || | |   ||  ||| || ||   ||  ||| ||
Db        471 TATAATTTATTTTTTATAAAAAATTATTTAATTTAAATTAAATTATTTATTAAAATTTAA 530

Qy        508 ACATTTACTTTCCAAAT 524
              |  || |   |  ||||
Db        531 ATTTTAATAATAAAAT 547


RESULT 58
CNS00HTB
LOCUS       CNS00HTB               1101 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC:
            BACR35G21 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL074017
VERSION     AL074017.1  GI:4953669
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR35G21"
                     /clone_lib="RPCI-98"
                     /note="end : T7"
ORIGIN

  Query Match             4.9%;  Score 53.8;  DB 9;  Length 1101;
  Best Local Similarity   37.2%;  Pred. No. 0.34;
  Matches  159;  Conservative   58;  Mismatches  207;  Indels    3;  Gaps    1;

Qy         23 TAACTTTCACTCTCTATGTTCTTATAAGCAAATGTTACAATGAGATATGAGTTTTAAAGC 82
              ||| |:  |:|   ||| |:          | | |||     |   |  |  ::||    |
Db        647 TAATTKAGASTGAATATTTKGGGGAKGTAATAAGTTTTGTTTTAAACTCYSYTTAKGTGT 706

Qy         83 CAGATCTTCCTTATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGG 142
              : ||||| |  |    | |:  : |:    : || ::    :   |  | : |::  :
Db        707 TTSCTCTTCTTCSTTAATWTKATBAAKTSGGVATTTWWATKTKTATTTTTTWAAKKGASY 766

Qy        143 TTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAG 202
              ||     || |  |    :||   | : | ||     |  || ||  : |  ||||
Db        767 TTATTGKATGTATAAGGTATWTRATATTTGWTTTGTAARTTGAAATATKGSTTGGTAAAA 826

Qy        203 ACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATA---CAG 259
              | :||:|| | ||  |:|| |   || :|||||||    :      |: | |  |:
Db        827 AAWTTVTTATTTTTTTWTATTTWTTTTWTTTTTTTTTTTDTTTTRTWAAAATTTWTTTTTT 886

Qy        260 CTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGAT 319
              |:||  ||| ||||| :||| || :||::|  :|||:| |||||  |: ||: ||||
Db        887 ATWTTATTTTTTTTTTTKTTTTTTKVAAWWATAWATAWAAAAATGAGTKAAGRAAAAGGR 946

Qy        320 CATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTT 379
              |: | :::   | :::  |   ||    ||||  |  |  :  || |  |  |||   |
Db        947 AADKTTKDRAAAAAKWKKGATGAAAGAGAAAAAAAAACGGGKGRTAAARAAKGTAARTRT 1006

Qy        380 GCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAA 439
              :  ::   |  :||||| | ||   |||   :||     :: :::    | :      ::
Db       1007 RKGKKKKTAAAWAAAAAWTWAAAAAAAAAAAAWTATGTKTKKDKKWAAMTTTKKTAAAWRW 1066

Qy        440 ATAGAAA 446
              |:| |||
Db       1067 AWAAAAA 1073


RESULT 59
BZ196437/c
```

```
LOCUS       BZ196437                855 bp    DNA     linear   GSS 11-OCT-2002
DEFINITION  CH230-516N19.TVB CHORI-230 Segment 2 Rattus norvegicus genomic
            clone CH230-516N19, genomic survey sequence.
ACCESSION   BZ196437
VERSION     BZ196437.1  GI:23854489
KEYWORDS    GSS.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 855)
  AUTHORS   Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K.,
            Shvartsbeyn,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D.,
            Riggs,F., de Jong,P. and Fraser,C.M.
  TITLE     Rat BAC End Sequences from Library CHORI-230 MboI segment
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: CH230-516N19.TJB
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the rat BAC library CHORI-230
            (http://www.chori.org/bacpac/rat230.htm). For BAC library
            availability, please contact Pieter de Jong (pdejong@mail.cho.org).
            Clones may be purchased from BACPAC Resources
            (http://www.chori.org/bacpac/or ering_information.htm). BAC end
            page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
            Plate: 516  row: N  column: 19
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .855
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /strain="BN/SsNHsd/MCW"
                     /db_xref="taxon:10116"
                     /clone="CH230-516N19"
                     /sex="Female"
                     /cell_type="Brain"
                     /clone_lib="CHORI-230 Segment 2"
                     /note="Vector: pTARBAC1.3; Site_1: MboI; Site_2: MboI;
                     CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
                     Pieter de Jong"
ORIGIN

  Query Match             4.9%;  Score 53.6;  DB 8;  Length 855;
  Best Local Similarity   64.0%;  Pred. No. 0.36;
  Matches  114;  Conservative    0;  Mismatches   59;  Indels    5;  Gaps    2;

Qy        887 TTACCACCAGTGCCTGAGGCTCCAGAGGAGCCATTGCCTTCTCAAGGTCACATAGGTGGT 946
              | | |  ||| ||| | | |||||||||  ||  || || ||||||||||||| ||||||
Db        252 TCATCTTCAGGGCCAGGGTCTCCAGAGGGTCC--TGTCTGCTCAAGGTCACATGGGTGGT 195

Qy        947 ---GGGTGAGTTAGGACCAAATCTAGAATTCCTGACTCCAGTAACTTCTGAAGTCATTTT 1003
                 |  ||  ||||||  || |  | |||| ||||||  |||||||||| | |||
Db        194 AACAGCAGACAAAGGACCTTGTCCAAGCTACCTGTCTCCAGAGACTTCTGAAGTAACTTT 135

Qy       1004 GTTTTTTATTTTTATGGTTTTATTATAAGAATACTTGCTAAGCACACTTACCCCCTGC 1061
              | |||||||| || || || | || |   | |  | |   ||  | | |  |  ||
Db        134 GGTTTTTATTCTTCTGCTTCTGTTGTGTTACAAGATATTTGCCAAGCATGCTTCTAGC 77


RESULT 60
BH133234
LOCUS       BH133234                889 bp    DNA     linear   GSS 07-AUG-2001
DEFINITION  ENTNB03TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
            genomic, genomic survey sequence.
ACCESSION   BH133234
VERSION     BH133234.1  GI:15092295
KEYWORDS    GSS.
SOURCE      Entamoeba histolytica
  ORGANISM  Entamoeba histolytica
            Eukaryota; Entamoebidae; Entamoeba.
REFERENCE   1  (bases 1 to 889)
  AUTHORS   Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
  TITLE     Determination of clone end sequences from Entamoeba histolytica
            HM1:IMSS sheared DNA library (2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: bjloftus@tigr.org
            Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
            DNA library
            Seq primer: M13-Reverse
            Class: shotgun
            High quality sequence start: 33
            High quality sequence stop: 633.
FEATURES             Location/Qualifiers
     source          1. .889
                     /organism="Entamoeba histolytica"
                     /mol_type="genomic DNA"
                     /strain="HM1:IMSS"
                     /db_xref="taxon:5759"
                     /clone_lib="Entamoeba histolytica Sheared DNA"
                     /note="Vector: pHOS1; Site_1: Bst I; Constructed at The
                     Institute for Genomic Research (TIGR), Rockville, MD.
                     Genomic DNA isolated from broth cultures of E. histolytica
                     using a  method described by Clark and Diamond (Clark,
                     C.G., and Diamond, L.S.  (1993) Entamoeba histolytica: a
                     method for isolate identification.  Exp. Parasitol.
                     77:450.). The DNA was mechanically sheared to give a
                     tight size distribution (~2 kb). The v + i method used for
                     the library  construction is described in detail in Smith,
                     H.O. and Venter, J.C.  (Making small insert libraries for
                     whole genome shotgun sequencing  projects. In Genome
                     Sequencing: A Practical Approach, eds. M. Vaudin  and B.
                     Barell, Oxford University Press, 1999)."
ORIGIN

  Query Match             4.9%;  Score 53.6;  DB 8;  Length 889;
  Best Local Similarity   45.6%;  Pred. No. 0.37;
  Matches  188;  Conservative    0;  Mismatches  224;  Indels    0;  Gaps    0;

Qy        147 GAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACAT 206
              | | ||||| | |  | ||| |  | | || || || ||     |  |  | ||   ||
Db        463 GTATTATTGTTGAAAAAAAAATAAATGAAAGATTCGTTGATATATTTTTGGTGAATCAAT 522

Qy        207 TGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCA 266
                 || || |  ||    ||  ||  || |   | |   |    |||||  |    ||
Db        523 GAAATTATTCAATTATATTCTAAATGATTATTGTATAATATTTCAACATATATGATTTGT 582

Qy        267 TTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATAT 326
                | |  |  |||  |  | |||        ||  | |   || ||| | | | | |
Db        583 GATATAGTAAGTTCATCTAGAAATTCATACAAATATTTTTTTTAGACAGAAAACCTTATG 642

Qy        327 TAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCA 386
              |     ||   |      | | ||| || |    ||| |   | | |  | ||  | |
Db        643 TGTTTTGTGTTATTAAGTATAGAAATTGGGATGGTATAATCATTCATTCTGTTAAACTAC 702

Qy        387 AAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAA 446
                |||||  ||           |  ||| | |   ||| | ||  || |  ||| ||||
Db        703 TTGGTAATCAGATTTATGCTTGTCATAATTCTGACTTATTCATTACAATAAAAACTGAAA 762

Qy        447 TTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAA 506
              | |  | | |||| |   |  || |||    | ||     |   | |||| |  |||||
Db        763 TAAAGATATATTTCACAATCAAAGTAAAGAGAAAATGATTTTCATATTATATTTTATTAT 822

Qy        507 AACATTTACTTTCCAAATATTCTGTCATTCAGGAGTATGGAAGTATCGATGG 558
```

```
Db        823 AAAAATACTAATGGATTGGCGATATTATTCCATAATATTATAATATATATGG 874


RESULT 61
CNS005ZC
LOCUS       CNS005ZC                966 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR12A21 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL061991
VERSION     AL061991.1  GI:4943787
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 966)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .966
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR12A21"
                     /clone_lib="RPCI-98"
                     /note="end : T7"
ORIGIN

  Query Match             4.9%;  Score 53.6;  DB 9;  Length 966;
  Best Local Similarity   44.5%;  Pred. No. 0.37;
  Matches  199;  Conservative   19;  Mismatches  226;  Indels    3;  Gaps    2;

Qy        113 TAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTT 172
Db         24 TANAATTANAATTATTTTAATATATTTTTTATTAAAAATAAAAATAAAAAAATAATATTT 83

Qy        173 AATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATT 232
Db         84 WTTTTTTTTAAAATAAATTAAATATAAATTTTATAAAAATTGAATTAAAAAATTATAAAT 143

Qy        233 TTTTTAAA-AGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAG 291
Db        144 ATAAWAAAWATWAAATTAAAAAAAAAAAAAATTTAATTTTAAAAATAAAAATAAAAWAAW 203

Qy        292 GTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAA 351
Db        204 TWATTAAAAAAGAAAAAWAWAAAAAAAAATAAAWTAAAAWAAAWAGRAAAWTTATWAAAA 263

Qy        352 ATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTA 411
Db        264 AWTAATTAAWAAAWWAATTTTTAAATATTWAAAAATTTATAAAAAWAWAAWATAATAAAA 323

Qy        412 TACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATT--TTGGAA 469
Db        324 RTKAAGTGNTTTTTTTTTTGTTTTATAAATAWWAAAAAWTATAATTTAWATGAATTGAAA 383

Qy        470 ATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCT 529
Db        384 TTAATTTTAAATAAAWAAAWATGTAATATTTTTTAATTGTATTATTTATAGAAAGATTGG 443

Qy        530 GTCATTCAGGAGTATGGAAGTATCGAT 556
Db        444 GKAATTAARTGTATSTGAGGTAAAGAT 470


RESULT 62
CG750135/c
LOCUS       CG750135                1592 bp    DNA     linear   GSS 24-OCT-2003
DEFINITION  P044-3-D09.za Ppa EcoRI BAC Library Pristionchus pacificus genomic,
            genomic survey sequence.
ACCESSION   CG750135
VERSION     CG750135.1  GI:37971278
KEYWORDS    GSS.
SOURCE      Pristionchus pacificus
  ORGANISM  Pristionchus pacificus
            Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
            Neodiplogasteridae; Pristionchus.
REFERENCE   1  (bases 1 to 1592)
  AUTHORS   Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K.,
            Buntjer,J., van der Meulen,M. and Sommer,R.J.
  TITLE     An integrated physical and genetic map of the nematode Pristionchus
            pacificus
  JOURNAL   Mol. Genet. Genomics 269 (5), 715-722 (2003)
  MEDLINE   22835951
   PUBMED   12884007
COMMENT     Contact: Sommer RJ
            Evolutionary Biology
            Max-Planck-Institute for Developmental Biology
            Spemannstr. 37-39, Tuebingen D-72076, Germany
            Tel: 00497071601371
            Fax: 00497071601498
            Email: ralf.sommer@tuebingen.mpg.de
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .1592
                     /organism="Pristionchus pacificus"
                     /mol_type="genomic DNA"
                     /strain="California"
                     /db_xref="taxon:54126"
                     /clone_lib="Ppa EcoRI BAC Library"
                     /note="The library was generated by a partial digest of
                     the genomic DNA with EcoRI and cloning into the BAC
                     vector."
ORIGIN

  Query Match             4.9%;  Score 53.6;  DB 9;  Length 1592;
  Best Local Similarity   45.4%;  Pred. No. 0.39;
  Matches  202;  Conservative    0;  Mismatches  242;  Indels    1;  Gaps    1;

Qy        108 ATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAAT 167
Db       1353 ATATTAATTTAATTTAATTTATATTTTAATANTATATAAAATATAATAATAAATAATATA 1294

Qy        168 CAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCT 227
Db       1293 AAATAAAATTTAATTATAATAATATTAAATATAAAAAATTTTATNTTTAAAATATATTTA 1234

Qy        228 TGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAA 287
Db       1233 TAATATTAATTAANANAATTAAAAATATAAAAATATATATATTAATAAATNTAAAATAAA 1174

Qy        288 AAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACA 347
Db       1173 TATAATANTAATAAAAAAATATTAAAATAAATTTTATATTAAATATAAAATAANAATATA 1114
```

```
Qy        348 CAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAA 407
              || || |    ||| |   ||  | |  ||  | | |      || |  | |   | |
Db       1113 TAATATTAAAATATAAAATTATAAATTAA-TTAAAATAATNATAAATATTTAATTTTTAT 1055

Qy        408 GGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGG 467
              |   |    | |   |    | | ||||| ||  ||    | ||  ||||
Db       1054 ATATTTAAATNATATATTAAAAATAAATATAAATATAATATAAATAATATAATATTTAAA 995

Qy        468 AAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATT 527
              ||| |||   |  ||  | |      ||| || |   | | |  || |||   | ||||
Db        994 AAAAAATAATAATAATAAAATNNAAATATATAATTAAATATAAAATAATTTTNTATTATT 935

Qy        528 CTGTCATTCAGGAGTATGGAAGTAT 552
              | |   | |    | |  || |||
Db        934 ATATTTATTATATTTNTAAAATTAT 910


RESULT 63
CNS010MP/c
LOCUS       CNS010MP                734 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN04L20 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL099163
VERSION     AL099163.1  GI:5610774
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 734)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .734
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN04L20"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : T7"
ORIGIN

  Query Match             4.9%;  Score 53.2;  DB 9;  Length 734;
  Best Local Similarity   35.5%;  Pred. No. 0.44;
  Matches  139;  Conservative   56;  Mismatches  197;  Indels    0;  Gaps    0;

Qy        144 TGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGA 203
              | | ||||||  |  |:||  ||: | |  |    |:||| ||     |  | :||  |
Db        442 TTAAAAATATAAAGAAWTAATAAWAATATTTGTGWTTTTTATGTTGTTATAAWWTATKWA 383

Qy        204 CATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTAT 263
              :||||  | :|: | ||      |:  ||:| || |:  :  :||   :  |      | :
Db        382 TWTTGTAATAKTWTTTTATGTKTTWTTTTWTATTTAWTWWTTWTATTTWTTTTTTWWTTK 323

Qy        264 TCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATA 323
              |:|| | | |  ||| :|  ::|||   ||| :|:  |  |||     |    :|  |:
Db        322 ATAWTTTTGTATTWTTTTDTGKWWAAATTAATATWGWTTTTKATTTTTTGTTGWWTTTTW 263

Qy        324 TATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAG 383
              |  |      : :::    |:::| |    :|  :|  | :| | || | :|: | ||
Db        262 TKTATAWWWWWWKTTTTTKWWATATTKTWTTTKTTAAAWTAATATAATWTWTTWTTWTTT 203

Qy        384 TCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAG 443
              | |:   |:| :|   ||  : :|:::::  ||       || |||    |    ||
Db        202 TWAWKWTTWATWAWAAAAATWTRAKKWWWTTTAAAAAAATTTTTAAATKAAAAAAATATW 143

Qy        444 AAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTAT 503
                : |    ||| |: || ||   ::||:: | :       : |  ||| |  :  | |
Db        142 TTKDAAAAKAAAAATWAATATTWTKDWTAWWTTTWWTTTTTTKKAAATTTGAAWKTKTTT 83

Qy        504 TAAAACATTTACTTTCCAAATATTCTGTCATT 535
              |: |  |||:  | |    ||||  | |  |:
Db         82 TWTATWATTWTATATAAKTATATAATATKTTW 51


RESULT 64
BZ397987/c
LOCUS       BZ397987                744 bp    DNA     linear   GSS 30-APR-2003
DEFINITION  EINCO14TF EI_10_12_KB Entamoeba invadens genomic clone EINCO14,
            genomic survey sequence.
ACCESSION   BZ397987
VERSION     BZ397987.1  GI:30244534
KEYWORDS    GSS.
SOURCE      Entamoeba invadens
  ORGANISM  Entamoeba invadens
            Eukaryota; Entamoebidae; Entamoeba.
REFERENCE   1  (bases 1 to 744)
  AUTHORS   Wang,Z., Samuelson,J., Clark,C.G., Eichinger,D., Paul,J., van
            Dellen,K., Hall,N., Anderson,I. and Loftus,B.
  TITLE     Gene discovery in the Entamoeba invadens genome
  JOURNAL   Mol. Biochem. Parasitol. 129 (1), 23-31 (2003)
  MEDLINE   22684048
   PUBMED   12798503
COMMENT     Other_GSSs: EINCO14TR
            Contact: Brendan Loftus
            Department of Eukaryotic Genomics
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-3543
            Fax: 301-838-0208
            Email: enta@tigr.org
            DNA was provided by Daniel Eichinger
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .744
                     /organism="Entamoeba invadens"
                     /mol_type="genomic DNA"
                     /strain="IP-1"
                     /db_xref="taxon:33085"
                     /clone="EINCO14"
                     /clone_lib="EI_10_12_KB"
                     /note="Vector: pHOS2; Site_1: BstXI; Total genomic DNA was
                     isolated from early log phase trophozoites of E. invadens
                     IP-1 using a Qiagen plant DNA extraction kit. A shotgun
                     medium-size plasmid library (average insert size of 10 -
                     12 kb) was generated by random mechanical shearing of E.
                     invadens genomic DNA, repairing the ends of DNA fragments
                     with T4 Polymerase, adding BstXI adaptors and ligating
                     into the BstXI site of a pUC-derived vector pHOS2."
ORIGIN

  Query Match             4.9%;  Score 53.2;  DB 8;  Length 744;
  Best Local Similarity   48.4%;  Pred. No. 0.44;
  Matches  179;  Conservative    0;  Mismatches  188;  Indels    3;  Gaps    1;

Qy        158 CATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAA 217
              |||||  |||| |||  |  | ||         ||| || | |||    ||  ||| | |
Db        376 CATTAAGAATCTGTTGGTCTTATTAGTTTTATTCTAATTTAAAAATAGTTTGTTTATAAT 317

Qy        218 TATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGG 277
              |||       ||||     |  || |  |   |   |      | ||  | |  ||||
```

```
Db        316 TATTGAATAGTGATGACAATTCAAAATTTATAATTTACTTTTTTTTTATTATAGTTTTTT 257

Qy        278 TTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGA 337

Db        256 ATCTATCAAAAAATGTTCAAAGGTATTTATCAAAAATAAATTATTTAATAATTTAAGTAT 197

Qy        338 AAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAG 397

Db        196 AACTCTTAAACTATTCAAGTCTCTGTTATGTTTTTTAAGTATGTTTGTTTTTTTGGTATT 137

Qy        398 TCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGC---AAATAGAAATTACCACA 454

Db        136 TAAACTAATAGTAACGGGAAATATTTTATTAAAAAGATGATTTAACTGTCAATTTGTACA 77

Qy        455 AATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTA 514

Db         76 ATTATAATTTTGTAAATTTGTTTAGTTAATCAAATTAAACATATACTATTACTAAATGTA 17

Qy        515 CTTTCCAAAT 524

Db         16 TATTACTTAT 7
```

```
RESULT 65
CNS04D0K/c
LOCUS       CNS04D0K                945 bp    DNA     linear   GSS 01-SEP-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            101H21 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL285149
VERSION     AL285149.1  GI:8023560
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE   1
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
  MEDLINE   20296633
   PUBMED   10835645
REFERENCE   2
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
            Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
            Saurin,W., Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
  MEDLINE   20359837
   PUBMED   10899143
REFERENCE   3  (bases 1 to 945)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .945
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="101H21"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0BG101CD11LP1~end : T7"
ORIGIN
```

```
  Query Match             4.9%;  Score 53.2;  DB 9;  Length 945;
  Best Local Similarity   38.5%;  Pred. No. 0.45;
  Matches  193;  Conservative   51;  Mismatches  253;  Indels    4;  Gaps    2;

Qy         27 TTTCACTCTCTATGTTCTTATAAGCAAATGTTACAATGAGATATGAGTTTTAAAGCCAGA 86

Db        930 TTTTAAWATTTTTTTTTTTTTTTTATATTATTTTTATAATTTTATTTTTTATAATTTTW 871

Qy         87 TCTTCCTTATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGA 146

Db        870 TTTTTTTTTTTTAWATTTTATTTATTANWTANWTTATAANTTTTATWNTAATAWATAATW 811

Qy        147 GAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACAT 206

Db        810 TTTAATWTTTTTATTATTAWAAATTTTTTTATTTATWWATWTATTTTATATTTAATTTAT 751

Qy        207 TGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCA 266

Db        750 ATTTTWTATTATATATTTTTTTTTNTWTTATWAWWTTTWWTTTWTTTTTATTTTATTTTT 691

Qy        267 TTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATAT 326

Db        690 ATTWTWTTTTATTTTTTATATTWTAWWWTTWWTAWWWW-WTWTATWTWTTWTATAAATAT 632

Qy        327 TAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCA 386

Db        631 ATTTATTTTTATWTTTATTWTTAAAAATATTTWTTTTTAAAATTTTTTATATTATTWTTA 572

Qy        387 AAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAA 446

Db        571 AATAAAAAATATWAAAATWW---TWTAWAWAATWTTATTATWTTTTTTTAAAAAWWWTWT 515

Qy        447 TTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAA 506

Db        514 TATTTTTTATTTWTWTTTTWAWAATWNTTTAAWTTTTATTAATTATAAAAAAAAWAAWWW 455

Qy        507 AACATTTACTTTCCAAATATT 527

Db        454 TAAWRNTANTAAAAAAAAANW 434
```

```
RESULT 66
CNS0145P
LOCUS       CNS0145P                1043 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN11G11 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL103735
VERSION     AL103735.1  GI:5615346
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1043)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
```

```
     source          1. .1043
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN11G11"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : T7"
ORIGIN

  Query Match             4.9%;  Score 53.2;  DB 9;  Length 1043;
  Best Local Similarity   37.0%;  Pred. No. 0.46;
  Matches  123;  Conservative   47;  Mismatches  162;  Indels    0;  Gaps    0;

Qy         87 TCTTCCTTATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGA 146
              | ::  :|:| | | |      || | |: || :|  : : | | :|:|    ||   :|
Db        628 TTWWWTWTWTTTATWTWWWAAATATATWTTTTWTWTWTWTTTTTWTWTATATTATWWTWT 687

Qy        147 GAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACAT 206
                :|||    ||  ||    :   | |||     | |: |  ||  | |    ||:|      |
Db        688 WTWATAAATATTTTTTATWTATATTTATATWTATWTTTWAATTTATATATTTWANATANT 747

Qy        207 TGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCA 266
              | |||| ||  | |  |   ||  |||||||||||| |    | : : ::: |  :::: :
Db        748 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAAAAAAWAAAAWWWWWWWWWAAWWWWWWW 807

Qy        267 TTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATAT 326
              : :  :  |  ||| ||::       ||:| | | ::  ||  | | ||| |  | | | |
Db        808 WAWAAWAATTTTTTTTTTHHTTYCCYYTAWATATATWWAWATAWAWAWAAAAAAAAAAAAA 867

Qy        327 TAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCA 386
              |   |    ||||  ||| | ||||  : :  | :  | : :| |:|:     :  |
Db        868 AAAAAAAAAAAAAAAAAAAAAAAAAAAMAHAYATACMAAAAMCWAATHAWAAMAYAHAAAA 927

Qy        387 AAGGTAAAAAGTCAACCTAAAGGTATACTACC 418
              :|  ||||||   ||   |||   | :: :::
Db        928 WATATAAAAAAAAAAAAAAAAAAAAAMMMMMM 959


RESULT 67
CL508886
LOCUS       CL508886               1258 bp    DNA     linear   GSS 01-APR-2004
DEFINITION  SAIL_804_E01.v1 SAIL Collection Arabidopsis thaliana genomic clone
            SAIL_804_E01.v1, genomic survey sequence.
ACCESSION   CL508886
VERSION     CL508886.1  GI:46006206
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1258)
  AUTHORS   Sessions,A., Burke,E., Presting,G., Aux,G., McElver,J., Patton,D.,
            Dietrich,B., Ho,P., Bacwaden,J., Ko,C., Clarke,J.D., Cotton,D.,
            Bullis,D., Snell,J., Miguel,T., Hutchison,D., Kimmerly,B.,
            Mitzel,T., Katagiri,F., Glazebrook,J., Law,M. and Goff,S.A.
  TITLE     A high-throughput Arabidopsis reverse genetics system
  JOURNAL   Plant Cell 14 (12), 2985-2994 (2002)
  MEDLINE   22356987
   PUBMED   12468722
COMMENT     Contact: Sessions A
            Applied Trait Genetics
            Syngenta Biotechnology Inc.
            3054 Cornwallis Rd., Research Triangle Park, NC 27709, USA
            Email: allen.sessions@syngenta.com
            ABRC Stock Number CS835931; T-DNA left border flanking sequences of
            Syngenta Arabidopsis Insertion Library (SAIL) lines are available
            through the Arabidopsis Biological Resource Center (ABRC).
            Sequences represent a pool of amplified genomic regions and not
            single contiguous sequences.
            Class: TDNA tagged.
FEATURES             Location/Qualifiers
```

```
     source          1. .1258
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /ecotype="Columbia"
                     /db_xref="taxon:3702"
                     /clone="SAIL_804_E01.v1"
                     /clone_lib="SAIL Collection"
                     /note="T-DNA left border sequences were isolated using a
                     modified TAIL-PCR strategy"
ORIGIN

  Query Match             4.9%;  Score 53.2;  DB 9;  Length 1258;
  Best Local Similarity   40.2%;  Pred. No. 0.47;
  Matches  282;  Conservative    0;  Mismatches  419;  Indels    1;  Gaps    1;

Qy        138 TTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCT 197
              |||  || |    | ||  | |||      |   |||  | ||||    || |  | ||
Db        526 TTTTTTTTATTTTTTTTTTTTATTTTTTTTTTTTTATAATTTTTTATTAATAAATTATAN 585

Qy        198 TAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATAC 257
              |       | || | |||  ||  |          || || || |     || |   |||
Db        586 TTTTTNTTTAATTATTTTATAATTTTTATAAATATTATTAAATAATATATATAATAATAT 645

Qy        258 AGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAG 317
              |  ||||  ||  |       | || ||  |       ||  |||    |  | | | | |
Db        646 ATTTATTTTTTATATAAATTNTATTTNAANTATATNTAAATAAAAATTTTATATANATAT 705

Qy        318 ATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATC 377
              ||  | ||| |   | || ||  || ||  |||||  |     |  | | | ||  || |
Db        706 ATTTTTTATAATATAATTNAATTTTTAAATAAAAATAAAAATAATTNATAATATATATTA 765

Qy        378 TTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGC 437
              |  | | | |  ||  || | |   ||||   |||     |  | |        |
Db        766 ATTAATTTATAT-TATTAAATAATTATAAATTAATAAATATTAATATAAATANNAANNAT 824

Qy        438 AAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATG 497
              | || |||| |     ||    ||  ||  ||||||     ||       | ||  |
Db        825 TATTATAAATAAATTTAATANNTANNTTNATAATAATTAATAAATTANATAATTAATTAA 884

Qy        498 TACTATTAAAACATTTACTTTCCAAATATTCTGTCATTCAGGAGTATGGAAGTATCGATG 557
              || ||| | |||||  ||   || ||    | |   |  | ||       |
Db        885 NTNTAATAATAAATTTAAATTATTAAAATANAATAAAATANAANTAAAATTAAAATATAA 944

Qy        558 GCTTCTTTAAAATGAAGCAGGAGGGTCTGGCAGAGAGTATCTATGAAATAAGTTCCTCTG 617
              |  | | ||| ||  |  |      |   | | |      |  | | | |  |
Db        945 ATATAATAATAATAAAATAAAATTAAATAATATAAATANATAAANATAAATATANTANAT 1004

Qy        618 ACCTTCACGCTTAATTTTCTGAATGGAGTGGAGCAAATTACTTCAAGCTTCACTTAACTT 677
              |   |     | ||| | |  |    |  |    |    | ||     | ||  |
Db       1005 AATATTNATATATATTNTATATTTAATATATAAATTAAATAATAAAATAATATATATATA 1064

Qy        678 GCATATGAAATGAACCGTACAAAAATACAAGAGTGTCAGGAGAAAGTTATGCTCTGGTAA 737
              | |  ||      |  | ||| || |  | |  | ||  || |  | |  |
Db       1065 TNAAAATAANNNNTTANAANTTANATAAAATANAATAATAATAATNTTNTNNTATAATTN 1124

Qy        738 ATATTTTGCAAAACAGATAAAAGATAATACTAGAGCTCTGTCCTCAAAGAGTTAAGCAGC 797
              ||| ||   || |   |  ||| |||| | |  |         | |||   | ||
Db       1125 ATAATTNTTAATANNAAATAAAAATAAAATTNAAATAAAANATTAAAAATATNAANATAA 1184

Qy        798 TAATCTAAGGAGGTAAACTCTATGTCAGCAGGATGAACTGCT 839
              | |   |  |  |  |   ||    |  |  || || |  |
Db       1185 AATTAANAATAAATNTAAAATANANAATTAAAATAAATTANT 1226


RESULT 68
BW541022
LOCUS       BW541022               668 bp    mRNA    linear   EST 31-AUG-2004
DEFINITION  BW541022 Yutaka Satou unpublished cDNA library (cstb) Ciona
            savignyi cDNA clone cstb014117 5', mRNA sequence.
ACCESSION   BW541022
VERSION     BW541022.1  GI:51720875
KEYWORDS    EST.
SOURCE      Ciona savignyi
```

```
  ORGANISM  Ciona savignyi
            Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
            Phlebobranchia; Cionidae; Ciona.
REFERENCE   1  (bases 1 to 668)
  AUTHORS   Satou,Y. and Satoh,N.
  TITLE     Expressed genes in Ciona savignyi
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Yutaka Satou
            Department of Zoology
            Kyoto University
            Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
            Tel: 81-75-753-4095
            Fax: 81-75-705-1113
            Email: yutaka@ascidian.zool.kyoto-u.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .668
                     /organism="Ciona savignyi"
                     /mol_type="mRNA"
                     /db_xref="taxon:51511"
                     /clone="cstb014117"
                     /dev_stage="tailbud stage"
                     /clone_lib="Yutaka Satou unpublished cDNA library (cstb)"
ORIGIN

  Query Match             4.9%;  Score 53;  DB 5;  Length 668;
  Best Local Similarity   50.4%;  Pred. No. 0.48;
  Matches  209;  Conservative    0;  Mismatches  200;  Indels    6;  Gaps    3;

Qy        109 TCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAATC 168
              |||  |  ||||  |   ||| | ||  |||   |       ||   ||||   |||| |
Db         59 TCTGAAAATCTTCTAAGCTCTAAAATATGTTCTCTCTTAGGATCAAGATCTAGACAACTA 118

Qy        169 AGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTT 228
                |||   | | || || |||||     ||| |   | || | |  || || |   |
Db        119 GAAAAATTTGTGTATAAAATATCTCCGAAAAAATAAGATTTTGTATATATTTAATAATTA 178

Qy        229 GATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAA 288
              || |   || |||     |   | |||||  || | || ||    |  ||| || | |
Db        179 TATATCAATATAAGTATAAAATTATATACAAGTAGTGATATGCCAATCATTTATTTATAG 238

Qy        289 AAGGTATAAAGAAATGCATTCAGAGAAAGA-TCATATATTAGCCAGTTGAAAATTAAACA 347
              | | || | |||   ||  | | ||||  |||| | |||   | |  | || | |
Db        239 CTGATTTATATAAAATTATAAAAATAAAGGGTCATCTTTTATTTAATATAGAAATTTAAT 298

Qy        348 CAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAA 407
              | |  | |  || ||  |||| |||||  |  | | ||  ||||||| |
Db        299 ATAGAAAAATATATTTTTTATTA-TTAATAGTATATTTAAGTGTAAAAACTTGTTAATTT 357

Qy        408 GGTATACTACCTGCTTTCTTATC----GCACTGCAAATAGAAATTACCACAAATTTTATT 463
              | || ||. |  ||| |||     | |    || || ||  |   |   |.|   ||
Db        358 TATTTAATAATTATTTTTATATGTAAAGTATATAAATTAAAATGTTAAAATGAGTAATTT 417

Qy        464 TTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTT 518
              |  ||| | | |   || | ||||| | |  | |  | |||| | || ||||||
Db        418 ATAAAAAAAGTTTTTTAACAAATAATATCTATTTTTATTTTAATAAATATACTTT 472


RESULT 69
CNS04E5M/c
LOCUS       CNS04E5M                976 bp    DNA     linear   GSS 01-SEP-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            103P02 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL286627
VERSION     AL286627.1  GI:8025084
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE   1
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
  MEDLINE   20296633
   PUBMED   10835645
REFERENCE   2
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
            Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
            Saurin,W., Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
  MEDLINE   20359837
   PUBMED   10899143
REFERENCE   3  (bases 1 to 976)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .976
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="103P02"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0BG103DH01LP1~end : T7"
ORIGIN

  Query Match             4.9%;  Score 53;  DB 9;  Length 976;
  Best Local Similarity   35.2%;  Pred. No. 0.5;
  Matches  132;  Conservative   53;  Mismatches  190;  Indels    0;  Gaps    0;

Qy        148 AAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATT 207
              |:||  | |   | | || |   | | |  |  | | || |  | || ||| |   |||
Db        975 AWATTATAAATTTAAAAATTTTTTAATTTTTAATAAAAAAAAAATAATATTATATTTATT 916

Qy        208 GTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCAT 267
              |||   |||  | |       |   |: | | |  ||  | | |:|  :   | | |
Db        915 AATTTAAAAATTAAATTTWAANAAAATWATTATATAWTTTNTAAANAWAWWATAAATAAA 856

Qy        268 TTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATT 327
                :|: :  ::|  | | |||| ||||||| |||| |||          ||  |  ||
Db        855 AATWTWAWAWWWTWATAATAAAATTTATAAAAAAATTWATTTNTTATTTAATWTTTAATA 796

Qy        328 AGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAA 387
              |  |    |::|:| |:      | : :    : ||:  |  :   ||| |:  : | :
Db        795 ATAAAAAATAWWAWTTAWWTATNATWAWAAWAWAATWTWAAAWAAAAATWTWAAWATWTW 736

Qy        388 AGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAAT 447
              |  ||| |    :|  || |  :| :  |  |   ||  : :   | |  :|||| | |
Db        735 AATTAATATTATWAATTATATWWANWTAAATTTTATTAAWTWTNNATTAAWAATAAATAA 676

Qy        448 TACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAA 507
                :  : :|:    | : :  ||| ||: : | ||||  :  |||:: ::::  : ::  :
Db        675 AWWAHWWAWANAAAAWAWAWAAAAAAWWWAAAAAAAKKKTTTTTHWNWKKKKKKKKKKR 616

Qy        508 ACATTTACTTTCCAA 522
              :   :: ::|: : |
Db        615 VGGGBKBBYTBTVBA 601


RESULT 70
CNS014DY/c
LOCUS       CNS014DY                1021 bp    DNA     linear   GSS 26-JUL-1999
```

```
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN11O04 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL104032
VERSION     AL104032.1  GI:5615643
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1021)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1021
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN11O04"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : T7"
ORIGIN

  Query Match             4.9%;  Score 53;  DB 9;  Length 1021;
  Best Local Similarity   47.1%;  Pred. No. 0.5;
  Matches  179;  Conservative    4;  Mismatches  195;  Indels    2;  Gaps    1;

Qy        148 AAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATT 207
              | |||||  | |||  |||| : || | |  ||  || |   |   |  | || |  ||
Db        437 ATATATTATTAATTTAAAATTWATTTAAAAATTATTAAATTTTTAAAATATTAATATTTT 378

Qy        208 GTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCAT 267
              ||||   || ||       || |||| |||   | |  ||      ||  |  | ||  |
Db        377 TTTTTTAAAAAATTAAAAATTTATTTATTTTTTAAAATTTTTTTCAATTAAATTTTTTTT 318

Qy        268 TTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATT 327
              |   ||||   ||  || | |:|  |  ||  ||      |      | |   || [ |
Db        317 TAAATTTTTTATTNATCTATAWATATTAAATAAATAAATATTTTTTTTATAAATTAAAAT 258

Qy        328 AGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAA 387
              |    ||  |||  |||   |||||   |  ||| |   | | ||  |  |  |   ||
Db        257 AAAWTTTTTTAAAAAAAATTAAAAATTTTTATATAATTTTTAAATTTTTAATTAATAAAA 198

Qy        388 AGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAAT 447
              |   ||||  | ||         ||           ||| |||    | |  ||  | |  |
Db        197 AAAAAAAATATAAAATATTTTTTAACAATATATTTTTATTAATAAAATATAACAAAATTT 138

Qy        448 TACCACAAATT--TTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTA 505
              ||  | | | |  | | | |  |||| || | | :| | |||| |  : || |   ||||
Db        137 TAAAAAATAATAATAAATATWAAAATTATTTAAAWACAAATAAATAAWAATTTTAAATTA 78

Qy        506 AAACATTTACTTTCCAAATA 525
              ||  |||||  ||  ||| |
Db         77 AATAATTTATATTAAAAAAA 58


RESULT 71
CNS02T50
LOCUS       CNS02T50                1027 bp    DNA     linear   GSS 01-SEP-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            163M16 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL212733
VERSION     AL212733.1  GI:7871552
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE   1
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
  MEDLINE   20296633
   PUBMED   10835645
REFERENCE   2
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
            Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
            Saurin,W., Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
  MEDLINE   20359837
   PUBMED   10899143
REFERENCE   3  (bases 1 to 1027)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1027
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="163M16"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG163BG08LP1~end : T7"
ORIGIN

  Query Match             4.9%;  Score 53;  DB 9;  Length 1027;
  Best Local Similarity   39.4%;  Pred. No. 0.5;
  Matches  167;  Conservative   37;  Mismatches  218;  Indels    2;  Gaps    1;

Qy        125 TGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTA 184
              | : | || | | ||| ||      :||||  |: |  | :||  | ::|::|       |
Db        416 TTWATAATTTTATTTTTGTAATWTWATATAAAWWAAAARWAAAAAAWWAWWAAAAAAAAA 475

Qy        185 GAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGAC 244
              || |    || ||||| |  || |||| |   | |       ||   ||||||  ||
Db        476 AAAAAAAAAATATTAAAAAWTTTWTTTTATATTTTTTTATTTTTTTATTTTTTTAATNT 535

Qy        245 CTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATG 304
              || |   ||      |:|| | |  | |    ||| |  | :|      ||| || |
Db        536 TNTATATTTATTTTTTTWTTTAATAATATAATTTTTNTAAATTWATAAATAAAAWAATTA 595

Qy        305 CATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATT 364
              ||  | | :||  |  |||||      | || | | |  |      ||  |   |  | ||
Db        596 TATATATAAWAATTTTTTATATATTTTATTTTATATT--ATNNATAATTATTTTTTTTTT 653

Qy        365 ACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTT 424
              : :::: :::|      |   |||  :|||||:| :|       | :  ::  :  |:|
Db        654 WWWWWWWWWWAAAAAAAAAAAAAAAAWAAAAARTRRATTGTTTTWTTWTWWWWWWWATWT 713
```

```
Qy        425 CTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAAC 484
              :||:     |  ||| | |||  |  | ||   |||  |||  ||   |   |  | |
Db        714 WWTAWGTATTTNTAAAAATAAAAAAATAAAATAATTAAATTGTTAAATNTAATANTAGAT 773

Qy        485 ATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGTCATTCAGGAGTAT 544
              :|:  | |  |     : |:::  || |    |  || |:|    | ||| :|    ||
Db        774 WTWTAATATATTTATTAWATWRWGWATGTNTNATAAAATTWTAWAATNATTTWGATAGAT 833

Qy        545 GGAA 548
                ||
Db        834 TAAA 837


RESULT 72
CNS01ZGM
LOCUS       CNS01ZGM                1029 bp    DNA     linear   GSS 01-SEP-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            220I12 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL174271
VERSION     AL174271.1  GI:7812328
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE   1
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
  MEDLINE   20296633
   PUBMED   10835645
REFERENCE   2
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
            Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
            Saurin,W., Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
  MEDLINE   20359837
   PUBMED   10899143
REFERENCE   3  (bases 1 to 1029)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1029
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="220I12"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG220BE06SP1~end :
                     PUC-Ori"
ORIGIN

  Query Match             4.9%;  Score 53;  DB 9;  Length 1029;
  Best Local Similarity   43.7%;  Pred. No. 0.5;
  Matches  187;  Conservative   17;  Mismatches  223;  Indels    1;  Gaps    1;

Qy        109 TCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAATC 168
              ||| ||  | || |  | | |   ||   |||  ||      |:||  |  ||  |  |

Db        415 TCTTTATATATTTATTTTTTTTTAATTTTTTTATTTTTTTTTTWTTTTTTTTTTTATTTT 474

Qy        169 AGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTT 228
              ||  |  |||| |      | |  | ||        || |||| ||  | | |   ||
Db        475 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTNTTTTTTTTTTTTTTTTTTTTTT 534

Qy        229 GATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAA 288
              |||||||          ||    |  |     |||  || | | | |  ||    |  |||
Db        535 TTTTTTTTTTTTTTTTTTTTTTTATTTTTTTATATATATATTTATATTTTTATATATAAA 594

Qy        289 AAGGTATAAAGAAATGCATT-CAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACA 347
              |   ||||| |||   |||  | |  ||  | :| | ||       |   |||||:   :
Db        595 TAAAAATAAAAAAAATAATTATATATTAATTTTWTTTTTTTTTTTTTTATTAAATTWTTTW 654

Qy        348 CAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAA 407
              | ||| | |  | |     ||    ||  ||  |||   |  ||||| : ||   |||
Db        655 TATAATAATTTTAAAAATTTTTTTAAAAATTTTAAGTATTAAAAAAAAAAWAAAAWAAAA 714

Qy        408 GGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGG 467
              | |      |  ||| ||:      :  | ||  :  ||:    ||| | |:||||
Db        715 TATTTTTATAATTTTTTWTTWAWTWTWWATATATTAWTTTTWTTTTAAAATATWTTTTTA 774

Qy        468 AAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATT 527
              ||  || | | ::|| ||   ||||:|| |: ||:       |:|:  |   ||| ||
Db        775 TAAATATATAAAWWAATATTTATTTTWATATWATAWATTTTTTTWTWAATAATAAAAATA 834

Qy        528 CTGTCATT 535
                |  |||
Db        835 ATAAWATT 842


RESULT 73
CNS00Z38/c
LOCUS       CNS00Z38                1101 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN01A24 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL097166
VERSION     AL097166.1  GI:5608777
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN01A24"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : SP6"
ORIGIN

  Query Match             4.9%;  Score 53;  DB 9;  Length 1101;
  Best Local Similarity   37.1%;  Pred. No. 0.51;
  Matches   99;  Conservative   47;  Mismatches  121;  Indels    0;  Gaps    0;
```

```
Qy        151 TATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTT 210
              |||  ||  ||| |:::   :|||||  |:: |  :| :| |  | :  :  | |:| :|
Db       1093 TATATATTTTTAWAWWWTTTWTAATATATWWTTTAWATWTTTWTTAWATWTAAAAWTTWT 1034

Qy        211 TTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTG 270
              || || :|: | |    |  ||  |:: :|: |  ::| |    |       |:  : ||
Db       1033 TTATTTWTWAATTTTTATTTTTAATWWTWAWTAATWWATATTTTTTATATAAWAAWATTT 974

Qy        271 TTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGC 330
              |||||  :|| ||      |   : |||| ||||    :   |  |:  | | ||:::
Db        973 TTTTTTAWTTTTTWTTTTATTTWTTAAAAAAATATTAWTTWATTAWTTTAAAATWWWTTA 914

Qy        331 CAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGG 390
              :|    |||| ||| | ||||  |      ||:|  ||| ||:|: ||    : : :
Db        913 MAAAAAAAAAAAAAAAAAAAAAAAAAAATMAATAATWAATTTAATTWAKWTTTTTTWAWTWWT 854

Qy        391 TAAAAAGTCAACCTAAAGGTATACTAC 417
              |: :   |:::: ::: ||:| | |
Db        853 TWTWTTTTMMMMHMHHHAYTAWAWTTC 827


RESULT 74
BU497983
LOCUS       BU497983                351 bp    mRNA    linear   EST 11-SEP-2002
DEFINITION  PfESToab91a12.y1 Plasmodium falciparum 3D7 asexual cDNA Plasmodium
            falciparum 3D7 cDNA 5', mRNA sequence.
ACCESSION   BU497983
VERSION     BU497983.1  GI:22794177
KEYWORDS    EST.
SOURCE      Plasmodium falciparum 3D7
  ORGANISM  Plasmodium falciparum 3D7
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 351)
  AUTHORS   Tang,K., Cole,R., Chakrabarti,D., Haywood,R., Clifton,S., Pape,D.,
            Marra,M., Hillier,L., Martin,J., Wylie,T., Dante,M., Theising,B.,
            Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Jentes,E., Ronko,I.,
            Tsagareishvili,R., Belaygorod,L., Franklin,C., Carr,L., Grow,A.,
            Maguire,L., Richey,J., Wadkins,J., Kennedy,S., Levinso,D.,
            Waterston,R., Wilson,R. and Sibley,D.
  TITLE     WashU Plasmodium EST Project
  JOURNAL   Unpublished (2001)
COMMENT     Contact: L. David Sibley
            WashU Plasmodium EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Library was constructed by Debopam Chakrabarti DNA sequencing by:
            Washington University Genome Sequencing Center For information on
            obtaining a clone please contact: L. David Sibley
            (sibley@borcim.wustl.edu), Washington University
            Seq primer: -40UP from Gibco.
FEATURES             Location/Qualifiers
     source          1. .351
                     /organism="Plasmodium falciparum 3D7"
                     /mol_type="mRNA"
                     /db_xref="taxon:36329"
                     /lab_host="DH10B (GeneHog, Invitrogen, Inc.)"
                     /clone_lib="Plasmodium falciparum 3D7 asexual cDNA"
                     /note="Vector: pBluescript SK plus; Site_1: EcoRI; Site_2:
                     XhoI; Library was constructed by Debopam Chakrabarti.
                     Total RNA samples were isolated from mixed stage
                     saponin(0.1%)-lysed P. falciparum 3D7 infected
                     erythrocytes by the acidic guanidinium-phenol chloroform
                     method. The poly A+ RNA was isolated by the polyAT-Tract
                     mRNA isolation system (Promega, WI) using streptavidin
                     Magnesphere particles. Directional cDNA libraries were
                     constructed by oligo d(T) priming of poly(A)+ RNA (5mg)
                     into EcoRI and XhoI sites of l ZapII vector using the Zap
                     cDNA synthesis kit (Stragene, CA). The average size of the
                     cDNA inserts in the library was between 1.0 and 1.5kb.
                     Clones were mass excised using the ExAssist helper phage
                     (Stratagene), the phagemids were precitptated with PEG
                     8000 and extracted with phenol/chloroform. Phagemid DNA
                     was electroporated into DH10B cells."
ORIGIN

  Query Match             4.8%;  Score 52.8;  DB 5;  Length 351;
  Best Local Similarity   49.1%;  Pred. No. 0.5;
  Matches  167;  Conservative    0;  Mismatches  172;  Indels    1;  Gaps    1;

Qy        149 AATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTG 208
              |||  |    |||| |  ||  |  |||  |||||| ||  |   |  ||||     ||
Db         13 AATTATATATATTATATGTCTATATATATATTTGTAAAACTTTAGAATTTAATATTTTTA 72

Qy        209 TTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATT 268
              ||  |||| |  | |   || ||||||  | ||    | ||| | ||| |    || |||
Db         73 TTAGGTTATTTAATTTTATTTATTTTTAAAGAACTTTTAACATATATATATAATTTTATT 132

Qy        269 TGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTA 328
              || ||   | ||    |  ||| ||  |||||||| | |     ||    | | |
Db        133 ATACGTTTGTACATACACTTTACATATGAAT-AATGCATTTATAATTTAATACATTTTAA 191

Qy        329 GCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAA 388
              ||  |  |   |    |   | | | ||||  ||||  |  | ||  | |
Db        192 TTTTATTATATTTCTTATTTTATTATATTTCTTATTTTATTATATTTTATTTTATTTTCT 251

Qy        389 GGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATT 448
              ||||| || ||  ||  || ||| | | | |    |||  || |  ||| |   |
Db        252 CTTAAAATGTTAATATATTGGCATATTCCTTCCGGATTTAATGCTCATTAAAGATTCAAC 311

Qy        449 ACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAA 488
              ||| | ||| ||    | ||    | |||| | ||
Db        312 TAGTTAAAATATATATTTACCAAAAAAAAAAAAAAAAAAA 351


RESULT 75
BM628628
LOCUS       BM628628                577 bp    mRNA    linear   EST 26-FEB-2002
DEFINITION  17000687498822 A.Gam.ad.cDNA1 Anopheles gambiae cDNA clone
            19600449641805 5', mRNA sequence.
ACCESSION   BM628628
VERSION     BM628628.1  GI:18928139
KEYWORDS    EST.
SOURCE      Anopheles gambiae (African malaria mosquito)
  ORGANISM  Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
            Anopheles.
REFERENCE   1  (bases 1 to 577)
  AUTHORS   Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L.,
            Charlab,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
  TITLE     Celera Anopheles gambiae EST project
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Holt R.A.
            Celera Genomics
            45 w. Gude Dr., Rockville, MD 20850, USA
            Tel: 2404533151
            Fax: 2404534580
            Email: HoltRA@celera.com
            Plate: NU01003CY2  row: A  column: 03
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .577
                     /organism="Anopheles gambiae"
                     /mol_type="mRNA"
                     /strain="RSP-ST (Reduced susc. to Permethrin - std.
                     chromosome)"
                     /db_xref="taxon:7165"
                     /clone="19600449641805"
                     /dev_stage="Adult"
                     /lab_host="DH10b"
                     /clone_lib="A.Gam.ad.cDNA1"
                     /note="Vector: pSport1; Site_1: Sal1; Site_2: Not1; Whole
```

adult mosquitoes (mixed sex) frozen on liquid nitrogen.
cDNA inserts >500 bp cloned directionally into pSport 1.
Not 1 site is 3'. Clones available through the Malaria
Research and Reference Reagent Resource Center
(www.malaria.mr4.org)."

ORIGIN

Query Match 4.8%; Score 52.8; DB 4; Length 577;
Best Local Similarity 45.8%; Pred. No. 0.52;
Matches 183; Conservative 0; Mismatches 217; Indels 0; Gaps 0;

```
Qy        141 GGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAA 200
              | ||   ||||| | |    |   ||| | ||       | |  ||||| | ||  ||
Db        173 GTTTATTAAATAATAAATTATTATAATAATTTTTCTCGAATTTTAAAGATTTAATTATAT 232

Qy        201 AGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGC 260
              | | |    | ||||  |   |||   || | | | | |  |||  | | |      | |
Db        233 ATAAAAGTTATTGTGGAATAACTATTTTAAATATGTAATTAGAAATGAAATGTTAATCGT 292

Qy        261 TATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATC 320
              | |  | |   | | | ||| |   |||    | |||   |    ||| |    | |
Db        293 TTTAAAATATATCTAGTTTTTTAAGAAATGAATTTAATTTAGCTTATTTATTTTATTAAG 352

Qy        321 ATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTG 380
              ||  ||    | || | |||||| | |  |  | |  |    |    ||||  |
Db        353 TTAATTTTTTAATTTAATAATTAAATAAAGTAATATTTTAAGGGATGAGCTTTAAAATAA 412

Qy        381 CAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAA 440
              | |  |   |    |       ||||  |||  | ||      |        | | |
Db        413 AATTTTATATTTTTTATAATTTTTAAATAAATATAAGCTTAAAAATAGCTATTATTAATA 472

Qy        441 TAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTAC 500
              |    |||  |   ||||| | |   |||| || | |   ||  ||| || ||   ||
Db        473 AATTTGTTATAATTTATTTTTTATAAAAAATTATTTAATTTAAATTAAATTATTTATTAA 532

Qy        501 TATTAAAACATTTACTTTCCAAATATTCTGTCATTCAGGA 540
              |||  |||  |||| |    ||| |||  || |||  |||
Db        533 AATTTAAATTTTTAATAATAAAAAATTTAGTAATTATGGA 572
```

RESULT 76
AG516492/c
LOCUS AG516492 843 bp DNA linear GSS 04-JUN-2004
DEFINITION Mus musculus molossinus DNA, clone:MSMg01-421P22.TJ, genomic survey sequence.
ACCESSION AG516492
VERSION AG516492.1 GI:48223905
KEYWORDS GSS.
SOURCE Mus musculus molossinus
ORGANISM Mus musculus molossinus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
TITLE BAC end Sequences of Library MSMg01
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 843)
AUTHORS Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the mouse BAC library MSMg01. For BAC library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
Tsukuba Institude, Bio Resource Center,
The Institute of Physical and Chemical Research (RIKEN) 3-1-1
Koyadai, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, fax: 81-298-36-9199
e-mail: abe@rtc.riken.jp
PRIMERS
Sequencing : TJ
LIBRARY
Vector : pBACe3.6
R.Site 1 : EcoRI
R.Site 2 : EcoRI.
FEATURES Location/Qualifiers
source 1. .843
/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/sub_species="molossinus"
/db_xref="taxon:57486"
/clone="MSMg01-421P22.TJ"
/sex="male"
/tissue_type="mixture of kidney and spleen"
/clone_lib="MSMg01 Mouse Male BAC Library"

ORIGIN

Query Match 4.8%; Score 52.8; DB 9; Length 843;
Best Local Similarity 46.0%; Pred. No. 0.55;
Matches 162; Conservative 0; Mismatches 190; Indels 0; Gaps 0;

```
Qy        142 GTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAA 201
              | | | ||||||  |  |  | ||||  |||          | | |   ||  ||||
Db        485 GGTAAAAAATATATAAAAAAATAAATTAAATAAANAAAAAAAATAATGTTTAAATAATAAA 426

Qy        202 GACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCT 261
              | ||| |       || ||| |   |  || |   |||||  |    | | | | | | |
Db        425 TATATTATAAAAAAAAAATAATAATTATATATAAATAAAAAAAAAAAAAAAAAAAATAATT 366

Qy        262 ATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCA 321
              ||  |||  ||| |  ||| |  ||||||   | | | | ||  ||  | | ||| |
Db        365 ATANATTNTTTTNTNNTTTTTNTAAAAAAAAAAAATAAATATTAATAAAAAAAAAAAAAT 306

Qy        322 TATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGC 381
              | |  |   |   ||||  ||| | ||||  |   | |  | ||     ||| |
Db        305 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAAAAAAAATATTNAAATATACA 246

Qy        382 AGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAAT 441
              |  ||||  ||| |    |   | |  | || | |          | ||| |
Db        245 AAACAAAACAAAACAAAACAAAAACAAACAAACAAACAAAACACCGCTCCTTTGGA 186

Qy        442 AGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTT 493
              |  |  | |  ||| || |    |  |||  |    | | ||   | | |||
Db        185 AAGAGATCCTCCAAGTTGTTCGGTTCAAAGCAGGCAACATAAGGGAGTGTTT 134
```

RESULT 77
BH177346
LOCUS BH177346 860 bp DNA linear GSS 19-OCT-2001
DEFINITION 009_G_05-rev SmBAC1 Schistosoma mansoni genomic clone 009G05 5', genomic survey sequence.
ACCESSION BH177346
VERSION BH177346.1 GI:16275973
KEYWORDS GSS.
SOURCE Schistosoma mansoni
ORGANISM Schistosoma mansoni
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.
REFERENCE 1 (bases 1 to 860)
AUTHORS Le Paslier,M.C., Pierce,R.J., Merlin,F., Hirai,H., Wu,W., Williams,D.L., Johnston,D., LoVerde,P.T. and Le Paslier,D.
TITLE Construction and characterization of a Schistosoma mansoni bacterial artificial chromosome library
JOURNAL Genomics 65 (2), 87-94 (2000)
MEDLINE 20247247
PUBMED 10783255
COMMENT Other_GSSs: 009_G_05-21
Contact: Pierce RJ
INSERM U 167
Institut Pasteur de Lille
1 rue du Professeur A. Calmette, 59019-Lille, France
Tel: (33) (0)3 20877783
Fax: (33) (0)3 20877888
Email: Raymond.Pierce@pasteur-lille.fr

```
            CNS sequencing ID=DG0AA009AD03BP1
            Plate: 009  row: G  column: 05
            Seq primer: M13 reverse primer
            Class: BAC ends
            High quality sequence stop: 860.
FEATURES              Location/Qualifiers
     source           1. .860
                      /organism="Schistosoma mansoni"
                      /mol_type="genomic DNA"
                      /strain="Puerto-Rican"
                      /db_xref="taxon:6183"
                      /clone="009G05"
                      /sex="mixed"
                      /dev_stage="cercariae"
                      /lab_host="Biomphalaria glabrata"
                      /clone_lib="SmBAC1"
                      /note="Vector: pBeloBAC 11; Site_1: Hind III; Partially
                      Hind III digested and size-selected S. mansoni  cercarial
                      DNA was ligated into Hind III digested pBeloBAC 11 vector
                      and used to transform E. coli DH10B. The complete library
                      contains 23808 clones from 4 independent
                      sizing-ligation-transformations.  Average insert size
                      ranges from 70-127 kb and genome coverage is 7.9-fold."
ORIGIN

  Query Match             4.8%;  Score 52.8;  DB 8;  Length 860;
  Best Local Similarity   43.5%;  Pred. No. 0.55;
  Matches  128;  Conservative   25;  Mismatches  138;  Indels    3;  Gaps    1;

Qy        243 ACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAA 302
              | ||:|      |: |  ||||  |||  |  |  ||      | ||||  || ||| : |
Db        506 AACTWATTTTATTMTAATTATTTTTTTTMATAATAATTAAAAAATAAAAAATAMAAAAMTA 565

Qy        303 TGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATA 362
              |||  |  | :| |  | : |  |    |  | || ||||   | ||  : |    ||
Db        566 NAAATTAAATGKWAAAAAAAAWAAAAAAAAWATNTTAATATTAYNAAAAAYYWAAAATAATY 625

Qy        363 TTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCT 422
              |||||  || ||  |:::     | :    ||||      || ::    | |     :
Db        626 TTACANTATTTTTTBYYMNCCCCYMCMCAAAAANNAACCCCCMMCMMCAAAAATTAWTWA 685

Qy        423 TTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAA 482
              || | ||   : :  |||||   |: :  | |||||| || |  ::| |||   : :|
Db        686 TTTTAATTWAWWWWTAAATAATTAWAWAAATAAATTTTTTTTATTTWWAAAATTNTWAWAT 745

Qy        483 ---ACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGTCA 533
                | |:|||||  |:|  || |  ||||  ||||  |||  |::| || | | |
Db        746 TTTAAAWAATTTATWANTTAATTATAAATAATTTTTTTTTTAAWWTTTTWTWTNA 799


RESULT 78
CNS07JWU
LOCUS       CNS07JWU                 860 bp    DNA     linear   GSS 30-NOV-2001
DEFINITION  T3 end of clone 009AD03 of library SmBAC1 from strain Puerto-Rican
            of Schistosoma mansoni, genomic survey sequence.
ACCESSION   AL614304
VERSION     AL614304.1  GI:16027528
KEYWORDS    GSS.
SOURCE      Schistosoma mansoni
  ORGANISM  Schistosoma mansoni
            Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
            Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.
REFERENCE   1  (bases 1 to 860)
  AUTHORS   Le Paslier,M.C., Pierce,R.J., Merlin,F., Hirai,H., Wu,W.,
            Williams,D.L., Johnston,D., LoVerde,P.T. and Le Paslier,D.
  TITLE     Construction and characterization of a Schistosoma mansoni
            bacterial artificial chromosome library
  JOURNAL   Genomics 65 (2), 87-94 (2000)
  MEDLINE   20247247
   PUBMED   10783255
REFERENCE   2  (bases 1 to 860)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-OCT-2001) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Partially Hind III digested and size-selected S. mansoni cercarial
            DNA was ligated into Hind III digested pBeloBAC 11 vector and used
            to transform E. coli DH10B. The complete library contains 23808
            clones from 4 independent sizing-ligation-transformations. Average
            insert size ranges from 70-127 kb and genome coverage is 7.9-fold.
FEATURES              Location/Qualifiers
     source           1. .860
                      /organism="Schistosoma mansoni"
                      /mol_type="genomic DNA"
                      /strain="Puerto-Rican"
                      /db_xref="taxon:6183"
                      /clone="009AD03"
                      /clone_lib="SmBAC1"
                      /note="end : T3"
ORIGIN

  Query Match             4.8%;  Score 52.8;  DB 9;  Length 860;
  Best Local Similarity   43.5%;  Pred. No. 0.55;
  Matches  128;  Conservative   25;  Mismatches  138;  Indels    3;  Gaps    1;

Qy        243 ACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAA 302
              | ||:|      |: |  ||||  |||  |  |  ||      | ||||  || ||| : |
Db        506 AACTWATTTTATTMTAATTATTTTTTTTMATAATAATTAAAAAATAAAAAATAMAAAAMTA 565

Qy        303 TGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATA 362
              |||  |  | :| |  | : |  |    |  | || ||||   | ||  : |    ||
Db        566 NAAATTAAATGKWAAAAAAAAWAAAAAAAAWATNTTAATATTAYNAAAAAYYWAAAATAATY 625

Qy        363 TTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCT 422
              |||||  || ||  |:::     | :    ||||      || ::    | |     :
Db        626 TTACANTATTTTTTBYYMNCCCCYMCMCAAAAANNAACCCCCMMCMMCAAAAATTAWTWA 685

Qy        423 TTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAA 482
              || | ||   : :  |||||   |: :  | |||||| || |  ::| |||   : :|
Db        686 TTTTAATTWAWWWWTAAATAATTAWAWAAATAAATTTTTTTTATTTWWAAAATTNTWAWAT 745

Qy        483 ---ACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGTCA 533
                | |:|||||  |:|  || |  ||||  ||||  |||  |::| || | | |
Db        746 TTTAAAWAATTTATWANTTAATTATAAATAATTTTTTTTTTAAWWTTTTWTWTNA 799


RESULT 79
B10881/c
LOCUS       B10881                   935 bp    DNA     linear   GSS 14-MAY-1997
DEFINITION  F24H6-Sp6.1 IGF Arabidopsis thaliana genomic clone F24H6, genomic
            survey sequence.
ACCESSION   B10881
VERSION     B10881.1  GI:2092003
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 935)
  AUTHORS   Feng,J., Dewar,K., Buehler,E., Kim,C., Li,Y., Shinn,P., Sun,H. and
            Ecker,J.
  TITLE     BAC End Sequences at ATGC
  JOURNAL   Unpublished (1997)
COMMENT     Other_GSSs: F24H6-Sp6, F24H6-T7.2, F24H6-T7, F24H6-T7.1
            Contact: Ecker J.
            Arabidopsis Thaliana Genome Center
            University of Pennsylvania
            Dept. of Biology, University of Pennsylvania, Philadelphia, PA
            19104
            Tel: 215-898-9384
            Fax: 215-898-8780
            Email: jecker@atgenome.bio.upenn.edu
            Seq primer: Sp6
            Class: BAC ends
```

```
          High quality sequence start: 98
          High quality sequence stop: 122.
FEATURES             Location/Qualifiers
     source          1. .935
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /ecotype="Columbia"
                     /db_xref="taxon:3702"
                     /clone="F24H6"
                     /sex="hermaphrodite"
                     /clone_lib="IGF"
                     /note="Vector: BeloBACII; Site_1: EcoRI; Site_2: EcoRI;
                     Produced by Thomas Altmann"
ORIGIN

  Query Match             4.8%;  Score 52.8;  DB 8;  Length 935;
  Best Local Similarity   42.1%;  Pred. No. 0.55;
  Matches  234;  Conservative    0;  Mismatches  317;  Indels    5;  Gaps    1;

Qy        216 AATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTT 275
              || |||     |  |   |||  ||| |   || | | |||       | |||| ||||
Db        900 AAAATATAAAATAAAAANTTTATAAAAATAATAAAAAAATAATAAATATTATTTATTTTA 841

Qy        276 GGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTT 335
              |   | || |||  || ||| | |   ||| | | | | |  | | |||    | ||
Db        840 TAATAANTAAANAAAAATAAAAAAATAATAATTAAAATATAAAAAAAAAATTTATAAATT 781

Qy        336 GAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAA 395
              ||||||    | |||   | |  ||||   || |  | |      | |  |    |  |
Db        780 TAAAATTNNNAAAAAAAAAATTTAATATATTATAATATTAAAAAAAAATATANAANATTA 721

Qy        396 AGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCAC-- 453
              |    |  ||  ||||  |  |  ||     |      |  || | | |  |  |
Db        720 ATAATAAAAAATATTATAAAAAATAATTANAAAAAATTTTTATAAAAAATAAAAATAATT 661

Qy        454 ---AAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACA 510
              ||||  |||     |||  |  | | | |    || |||   |  | |||  |||| |
Db        660 AAAAAATAATATAAAANAAAATAAATTAAATATATAAAATTTANNTTAAATATAAAAATA 601

Qy        511 TTTACTTTCCAAATATTCTGTCATTCAGGAGTATGGAAGTATCGATGGCTTCTTTAAAAT 570
              ||   | |  |   |     | ||  |     |   ||  ||  |    || ||   | |
Db        600 TTATATATTAATTAAAAAAATTATATATTTANAATNAAAAATAAAAAAATTATTATTATT 541

Qy        571 GAAGCAGGAGGGTCTGGCAGAGAGTATCTATGAAATAAGTTCCTCTGACCTTCACGCTTA 630
              |  | |      |  | |   ||  |||  || |  ||          ||
Db        540 TTAAAAAAAAAAAATAAAATAATATAAATATATAAAATTTTATNNNTTNNTTTTNNNNT 481

Qy        631 ATTTTCTGAATGGAGTGGAGCAAATTACTTCAAGCTTCACTTAACTTGCATATGAAATGA 690
              | |   |  |  |  | |||  |   ||  ||   |  |     |   |||
Db        480 NNTNTNNNNTTAAAAAAAAAAAAAAAAAAAAAAAAAATTTTNTNNATNNTTNAAATNAATTT 421

Qy        691 ACCGTACAAAAATACAAGAGTGTCAGGAGAAAGTTATGCTCTGGTAAATATTTTGCAAAA 750
              |   || | || |  |  | | |    |  |   |||    |  |||||||     || |
Db        420 AAAATAAATAANTNAATAAATATATTAAATATAATATATATTAATAAATATAAATTAATA 361

Qy        751 CAGATAAAAGATAATA 766
              | ||||   | ||||
Db        360 AAAATAATTTACAATA 345


RESULT 80
CD385477/c
LOCUS       CD385477               1164 bp    mRNA    linear   EST 30-MAY-2003
DEFINITION  AGENCOURT_14304575 NIH_MGC_173 Homo sapiens cDNA 5', mRNA sequence.
ACCESSION   CD385477
VERSION     CD385477.1  GI:31219639
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1164)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
            National Cancer Institute / NIH
            Bldg. 31 Rm10A07 Bethesda, MD 20892
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Jamie Thompson, University of WI
             cDNA Library Preparation: Gina Zastrow-Hayes
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: NDKM67  row: p  column: 16
            High quality sequence start: 2
            High quality sequence stop: 323.
FEATURES             Location/Qualifiers
     source          1. .1164
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /tissue_type="embryonic trophoblasts, made from WA01 stem
                     cells"
                     /lab_host="DH10B TonA"
                     /clone_lib="NIH_MGC_173"
                     /note="Vector: pDONR201; Site_1: attP2; Site_2: attP1;
                     LIBR_PRIMING - oligo dT; METHOD - full-length enriched;
                     LIBR PROVIDER - Bradfield"
ORIGIN

  Query Match             4.8%;  Score 52.8;  DB 6;  Length 1164;
  Best Local Similarity   45.6%;  Pred. No. 0.57;
  Matches  141;  Conservative    0;  Mismatches  168;  Indels    0;  Gaps    0;

Qy        210 TTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTT 269
              |||  ||      |  ||  || |   |||| |  | | | | ||     |||  |  |
Db        797 TTTTATATCTCTNTTTCTCTATCTAACCAAAATATATAATATATATNANNATATATAAAT 738

Qy        270 GTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAG 329
              |||||  |||  |     |  ||||| |||   |       |   || |  ||||
Db        737 TTTTTTTTTTTTATTTTTTATATAATAAAAAAAAAAAANNNNNTNATTTATAAATTATTTT 678

Qy        330 CCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAG 389
              |    ||||  ||| | ||||  |    | |  | |       ||       |    |
Db        677 TTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAANAAAANAAANNNAAAAAAAAAAAAATATT 618

Qy        390 GTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTA 449
              | |  | | |    |  |    |  |  |  ||| ||  |    |    ||| | |||  |
Db        617 ATTATTATTTATTTTTTANANTAAAAAAATTTTTTTTTTTTTATNAAAAAAAAAAAANA 558

Qy        450 CCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAAC 509
              | ||| |||  ||||  | ||| | | |  ||  ||  |||||    |  |  |||||
Db        557 TAATAAAATTTTTTTTTAATATATTTTAATTAATAATTTTTTTTATNTTTTTTATAAAAN 498

Qy        510 ATTTACTTT 518
              |||  |||
Db        497 TTTTTTTTT 489


RESULT 81
AG392554/c
LOCUS       AG392554               1268 bp    DNA     linear   GSS 03-JUN-2004
DEFINITION  Mus musculus molossinus DNA, clone:MSMg01-210G22.TJ, genomic survey
            sequence.
ACCESSION   AG392554
VERSION     AG392554.1  GI:48016805
KEYWORDS    GSS.
SOURCE      Mus musculus molossinus
  ORGANISM  Mus musculus molossinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
```

```
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     BAC end Sequences of Library MSMg01
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1268)
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the mouse BAC library MSMg01. For BAC
            library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
            Tsukuba Institude, Bio Resource Center,
            The Institute of Physical and Chemical Research (RIKEN) 3-1-1
            Koyadai, Tsukuba, 305-0074 Japan
            phone: 81-298-36-9189, fax: 81-298-36-9199
            e-mail: abe@rtc.riken.jp
            PRIMERS
            Sequencing : TJ
            LIBRARY
            Vector     : pBACe3.6
            R.Site 1   : EcoRI
            R.Site 2   : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .1268
                     /organism="Mus musculus molossinus"
                     /mol_type="genomic DNA"
                     /sub_species="molossinus"
                     /db_xref="taxon:57486"
                     /clone="MSMg01-210G22.TJ"
                     /sex="male"
                     /tissue_type="mixture of kidney and spleen"
                     /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

  Query Match             4.8%;  Score 52.8;  DB 9;  Length 1268;
  Best Local Similarity   45.8%;  Pred. No. 0.57;
  Matches  220;  Conservative    0;  Mismatches  251;  Indels    9;  Gaps    1;

Qy         35 TCTATGTTCTTATAAGCAAATGTTACAATGAGATATGAGTTTTAAAGCCAGATCTTCCTT 94
              | | | ||  | ||  ||  | | | |       |  |  |  |||   | ||    |||
Db        596 TTTTTATTTCTCTATCCATCTTTCATATATCATCAAAAAATAAAAAAAAAAATTCATCTT 537

Qy         95 ATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATATT 154
              || ||    ||   | | |  ||||   | | |||  ||| | |||   | | | |   |
Db        536 ATTTCAAGNCCTTCTATATTNTTCTATTATTCTCTTCTATCATTTTATATTATATACTCT 477

Qy        155 GATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGT 214
                | | ||||||||   ||   || | || ||   |  || | | |   |||| |
Db        476 ATTACTCACAAATCATACTATTTCCTTTTTAAACATAAAAAAATATAAATAAATTTTTAT 417

Qy        215 TAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTT 274
              |||| |      ||  ||| |   ||| |  |||       ||   ||||| || |   ||
Db        416 TAATTTTTAATATTATTTTATACTAAAAAATTTATCTTTTTAGTCCTATTTATATCCATT 357

Qy        275 ---------TGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATA 325
                      |   ||  | |   |    |||| |||   |   | | |   | || ||
Db        356 AAAATATAATATATTTATTATTTATTTAATAAAAAAAAATAACTACAAATCATTAATTTA 297

Qy        326 TTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTC 385
              |||   | ||  |||| ||| | |||  | |       | |  |   ||     |
Db        296 TTATAAAATTATAAATACAACCCTAAAATATTNNTAGAAAAAAAAAAAAAAATCATAATAT 237

Qy        386 AAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAA 445
              |||  || | ||| |    |||  |||| |       ||  |||   | |    |||  |
Db        236 AAAATTATATAGTTATGAGAAAAATATATAATAATAATTTATATATAATTTATTATATTA 177

Qy        446 ATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTA 505
                |||   | | | ||  |   ||| |     |       ||  || | |  || |||||
Db        176 TATACATAATAATATAAATAAATAATTAAAAGAAGTGTTATTTTTATAAAAATAATATTA 117
```

```
RESULT 82
CNS04P5D
LOCUS       CNS04P5D                855 bp    DNA     linear   GSS 01-SEP-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            125H20 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL300874
VERSION     AL300874.1  GI:8177894
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE   1
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
  MEDLINE   20296633
   PUBMED   10835645
REFERENCE   2
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
            Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
            Saurin,W., Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
  MEDLINE   20359837
   PUBMED   10899143
REFERENCE   3  (bases 1 to 855)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .855
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="125H20"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0BG125DD10LP1~end : T7"
ORIGIN

  Query Match             4.8%;  Score 52.6;  DB 9;  Length 855;
  Best Local Similarity   44.2%;  Pred. No. 0.61;
  Matches  145;  Conservative   11;  Mismatches  172;  Indels    0;  Gaps    0;

Qy        238 AAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAA 297
              |||| |    | | | | | |  || |  | :  ||    ||   | ||| :   || ||
Db        500 AAAAAAAAAAAAAAAAAAAAAAWATAATATTAWAATTAATNTTAAATTAAATWTTTTAAAA 559

Qy        298 AGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGT 357
              | ||:   |   | | ||| |  | | |  |   | || :||| :|:| | ||||  | |
Db        560 AAAAWAWTAAAWAAAAAAAAAAAAAAAAAAAAAATAAATTAWAAAAWAWAAAAAAAAAWATT 619

Qy        358 GCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTAC 417
                |||||   | | ||||      |   | |  |||||   ||  :| :  ||:| ||
Db        620 TTATATTTWTTAATTTAAAANATNANANATAAAAAAAAAAAAAAAAWATWAATAWAWTAT 679

Qy        418 CTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTC 477
                 || | |      |  | ||| |||  |  | |||     |      ||||||
Db        680 TATTATTTTAAAATATTTATATATAAAAAAAATAATAAAAAAAAAAAAAWAAATAAAWAA 739
```

```
Qy        478 AGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGTCATTCA 537
Db        740 AAATNAAATAWAAAAATAAAATAATTTAAAAAAAWAATATTTTTTTNATTTTATAATAAA 799

Qy        538 GGAGTATGGAAGTATCGATGGCTTCTTT 565
Db        800 AAATAAATAAATAATAAAAAATTTTTTT 827
```

RESULT 83
CNS016LI
LOCUS CNS016LI 1101 bp DNA linear GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC BACN16D22 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL106896
VERSION AL106896.1 GI:5624374
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBeloBAC11.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACN16D22"
/clone_lib="DrosBAC"
/plasmid="pBeloBAC11"
/note="end : T7"
ORIGIN

Query Match 4.8%; Score 52.6; DB 9; Length 1101;
Best Local Similarity 31.7%; Pred. No. 0.62;
Matches 113; Conservative 84; Mismatches 159; Indels 1; Gaps 1;

```
Qy         50 GCAAATGTTACAATGAGATATGAGTTTTAAAGCCAGATCTTCCTTATCTCTCTGCCCCAT 109
Db        729 GWRAGGGWWAWAAAAAAATTKATDTTTTTTTTTTWTTTWTTTTTTBTTTTATWTTTTAKDT 788

Qy        110 CTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCA 169
Db        789 TTTTTTTWATATTTWTTWTTDATWTTTTTTTTTWAWTATATATTTWAWTATAATTAWAAT 848

Qy        170 GTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTG 229
Db        849 WWTATTAWTATAWWATAWTWWWWWTRTAWWWTDWTWWWWATTAKTTWATWATATTAWTWA 908

Qy        230 ATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAA 289
Db        909 WAWWTAWAWWAWAATTATWTWTWATTATDWRT-TTASTATWTTTDTWTWATAWRAKAWRW 967

Qy        290 AGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACA 349
Db        968 ATWDWAWAWRATAWDTATWAATADATTWATATAAAWWTWWDTATTGWRWAWTADATTAWD 1027

Qy        350 AAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAA 406
```

```
Db       1028 AAATRTAAATWTWTTKWTTKWKTTTKWTTTKTAKTKAKAGRAWWWWRTKAARRWWRW 1084
```

RESULT 84
CL147481
LOCUS CL147481 1118 bp DNA linear GSS 05-JAN-2004
DEFINITION ISB1-234H4_T7.1 ISB1 Xenopus tropicalis genomic clone ISB1-234H4, genomic survey sequence.
ACCESSION CL147481
VERSION CL147481.1 GI:40641116
KEYWORDS GSS.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae; Xenopus; Silurana.
REFERENCE 1 (bases 1 to 1118)
AUTHORS Kremitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T., Mardis,E. and Wilson,R.
TITLE A physical map of the xenopus tropicalis genome
JOURNAL Unpublished (2003)
COMMENT Contact: Richard K Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Insert Length: 75000 Std Error: 0.00
Seq primer: T7 TAATACGACTCACTATAGGG
Class: BAC ends
High quality sequence start: 79
High quality sequence stop: 315.
FEATURES Location/Qualifiers
source 1. .1118
/organism="Xenopus tropicalis"
/mol_type="genomic DNA"
/db_xref="taxon:8364"
/clone="ISB1-234H4"
/clone_lib="ISB1"
/note="Vector: pBeloBAC11; ISB-1 Xenopus tropicalis BAC Library Segment 1"
ORIGIN

Query Match 4.8%; Score 52.6; DB 9; Length 1118;
Best Local Similarity 43.8%; Pred. No. 0.62;
Matches 213; Conservative 0; Mismatches 266; Indels 7; Gaps 1;

```
Qy         33 TCTCTATGTTCTTATAAGCAAATGTTACAATGAGATATGAGTTTTAAAGCCAGATCTTCC 92
Db        453 TTTNTTTTTTTTTTTTTNNNNNTNTNNNNTTNNTTTTTTTTTTTTTTAATTTNTTTTTN 512

Qy         93 TTATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATA 152
Db        513 TTTNTTTTTTTTTATTTNTNTAATTNTTAATATTTATTATAAAATTTTAANNTTATTATT 572

Qy        153 TTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTT 212
Db        573 AAAAATTTATTAAAAATTAATAATTATTTTTAAATAATTTATAATTAAATTTTTTNTATA 632

Qy        213 GTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTT 272
Db        633 TTAAATAATTATTTTTATTATTTTTAATTTATTATATAAANTTAATAAAAAAAATAAATT 692

Qy        273 TTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCA 332
Db        693 TTTTTAATATTTATAAATATTATAAAATATTTAAAATATTTAAAAATTTTTTTT------ 746

Qy        333 GTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTA 392
Db        747 -TTTATATTAATANTTATATTTTTATATTAATTATTTTTTTTATTTAATAATAANAATTT 805

Qy        393 AAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCA 452
Db        806 ATAATTTTTTTTATTTTTATAAAAAAAAAATTTTTTTANATTTTAATATAAAAAATTTA 865
```

```
Qy        453 CAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATT 512
              |  |||| || |    || |    | || | ||  | |  ||| |  ||| |||| |||
Db        866 TATTTTTTTTTATATTTATTAAAATATAATATATTTTATNATATTTTATATAAAAATATT 925

Qy        513 TACTTT 518
              | |||
Db        926 AAATTT 931


RESULT 85
CC264939/c
LOCUS       CC264939               1272 bp    DNA     linear   GSS 13-MAY-2003
DEFINITION  CH261-19L17_Sp6.1 CH261 Gallus gallus genomic clone CH261-19L17,
            genomic survey sequence.
ACCESSION   CC264939
VERSION     CC264939.1  GI:30612358
KEYWORDS    GSS.
SOURCE      Gallus gallus (chicken)
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 1272)
  AUTHORS   Kremitzki,C., Higginbotham,J., Wylie,K., Carter,J., McPherson,J.,
            Warren,W., Graves,T., Mardis,E. and Wilson,R.
  TITLE     Gallus gallus BAC End Reads
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K. Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 182000   Std Error: 0.00
            Seq primer: Sp6 ATTTAGGTGACACTATAG
            Class: BAC ends
            High quality sequence start: 34
            High quality sequence stop: 261.
FEATURES             Location/Qualifiers
     source          1. .1272
                     /organism="Gallus gallus"
                     /mol_type="genomic DNA"
                     /strain="Red Jungle Fowl"
                     /db_xref="taxon:9031"
                     /clone="CH261-19L17"
                     /sex="female"
                     /cell_line="UCD001, inbred 256"
                     /clone_lib="CH261"
                     /note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;
                     CH261 Female Chicken library - for library and clone
                     ordering information: http://www.chori.org/bacpac"
ORIGIN

  Query Match             4.8%;  Score 52.6;  DB 8;  Length 1272;
  Best Local Similarity   47.4%;  Pred. No. 0.63;
  Matches  190;  Conservative    0;  Mismatches  206;  Indels    5;  Gaps    1;

Qy        173 AATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATT 232
              || | |  |  | | || | ||  |||    |||    | |  ||||| |     ||
Db        858 AAAATTAATAAAATATATATAATTATAATATTATTAAAATTTATATATAATTAAAAGAAA 799

Qy        233 TTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGG 292
              |  ||||| |   || | | ||| |  ||     |  | |  | | || || |||
Db        798 TAAATAAAAAATAATAAATAAATATATTTAAATTATATAATATTATATATTAAATAAAAT 739

Qy        293 TATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAA 352
              |||  ||||  | |   | ||| |  | || |     | |  |||  |||  | || |
Db        738 ATTAATAAAATTAAATATTAAAAATAATAAATTT-----AATAAAAATATAATAATAATA 684

Qy        353 TGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTAT 412
              |    || || || | | |    |          ||  || | | | ||  ||||   ||
Db        683 AAAAATTTTAATAAATAATTAATAATATTTTAATAATATATATATTTAAATTAAATAAAT 624

Qy        413 ACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATA 472
```

```
              | |      | | ||||   |     |||| ||||||   ||||  ||      | ||
Db        623 AATTAAAAATATTTTATAATATATATAATAAAAATTAAATAAAATAAAATANAATTATTA 564

Qy        473 ATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGTC 532
              | | | |||| || || ||  |  || |||||           ||  || || | | |
Db        563 TTATAATAAAATATTATATTAAAATTANTATTATTTATAAATAATTAAAATTAATATATT 504

Qy        533 ATTCAGGAGTATGGAAGTATCGATGGCTTCTTTAAAATGAA 573
              ||     |  |   |  |||   |   || |  |||| ||
Db        503 ATAATAAAAAAAAAATTTATAATTTAATTATAAAAAATAAA 463


RESULT 86
CN474833/c
LOCUS       CN474833               714 bp    mRNA    linear   EST 22-APR-2004
DEFINITION  USDA-FP_123923 Diaprepes abbreviatus, Diaprepes root weevil
            Diaprepes abbreviatus cDNA clone SLtDaF057_B10 5', mRNA sequence.
ACCESSION   CN474833
VERSION     CN474833.1  GI:46495277
KEYWORDS    EST.
SOURCE      Diaprepes abbreviatus (Diaprepes root weevil)
  ORGANISM  Diaprepes abbreviatus
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Coleoptera; Polyphaga; Cucujiformia;
            Curculionidae; Entiminae; Entimini; Diaprepes.
REFERENCE   1  (bases 1 to 714)
  AUTHORS   Lapointe,S.L., Hunter,W.B. and Dang,P.M.
  TITLE     Gene expression in Diaprepes abbreviatus (Coleoptera:
            Curculionidae)
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Stephen Lapointe, Wayne Hunter, Phat Dang
            US. Horticultural Research Lab, Ft. Pierce, FL.
            USDA-ARS
            2001 South Rock Rd., Ft. Pierce, FL 34945, USA
            Tel: (772) 462-5914, (772) 462-5940
            Fax: (772) 462-5986
            Email: SLapointe@ushrl.ars.usda.gov
            Seq primer: T3 Primer.
FEATURES             Location/Qualifiers
     source          1. .714
                     /organism="Diaprepes abbreviatus"
                     /mol_type="mRNA"
                     /db_xref="taxon:13040"
                     /clone="SLtDaF057_B10"
                     /sex="Adult Female"
                     /tissue_type="Whole body"
                     /dev_stage="Teneral"
                     /lab_host="XL1-Blue"
                     /clone_lib="Diaprepes abbreviatus, Diaprepes root weevil"
                     /note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2:
                     XhoI; A high quality EST with at least 200 contiguous
                     bases at Trace Tuner score of 20 or better. Construction
                     by PM DANG and ANNA SARA HILL, USDA, ARS, U.S.
                     Horticultural Res. Lab, Ft. Pierce, FL, USA."
ORIGIN

  Query Match             4.8%;  Score 52.4;  DB 7;  Length 714;
  Best Local Similarity   47.1%;  Pred. No. 0.66;
  Matches  161;  Conservative    0;  Mismatches  181;  Indels    0;  Gaps    0;

Qy        150 ATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGT 209
              | |  ||| ||     |  | |||  |  || |||| |       ||      | ||
Db        545 ACAAAGATAATGTGCGGTGTATGAATTATAATGAAGAATAGAATTAATTTTTCGTTTAGT 486

Qy        210 TTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTT 269
              | | ||||||| |   || |||| || |||  |  |     .|||||||    |  |
Db        485 GTAGATAATATATTTTTTTTATTTATTAAAAGAAATTGTATAGCATACAGAGTGTGCATA 426

Qy        270 GTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAG 329
              |||    | |   || |   |||||   |||  | |     ||| | |||  | |
Db        425 TAGTTTAAAATTTGACAACACTTTATAATACAATTTAATATTCTAAAAAGCATGAAATGT 366

Qy        330 CCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAG 389
```

```
Db        365 TTTTTTAAACAAATTTATTAATTTGATATGTTTTTCAATATTTAAAGTTAACAGTATCTT 306

Qy        390 GTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTA 449

Db        305 GTAATTATTCGTACTACCATTTTTATTACACCCTGTTTAATGTTTTATCGTTTTATATAT 246

Qy        450 CCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTT 491

Db        245 TTACAAATTATATTCTGGAAATATTAATCGCAAAATTAAATT 204
```

RESULT 87
CN474867/c

```
LOCUS       CN474867                715 bp    mRNA    linear   EST 22-APR-2004
DEFINITION  USDA-FP_123957 Diaprepes abbreviatus, Diaprepes root weevil
            Diaprepes abbreviatus cDNA clone SLtDaF057_H09 5', mRNA sequence.
ACCESSION   CN474867
VERSION     CN474867.1  GI:46495311
KEYWORDS    EST.
SOURCE      Diaprepes abbreviatus (Diaprepes root weevil)
  ORGANISM  Diaprepes abbreviatus
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Coleoptera; Polyphaga; Cucujiformia;
            Curculionidae; Entiminae; Entimini; Diaprepes.
REFERENCE   1  (bases 1 to 715)
  AUTHORS   Lapointe,S.L., Hunter,W.B. and Dang,P.M.
  TITLE     Gene expression in Diaprepes abbreviatus (Coleoptera:
            Curculionidae)
  JOURNAL   Unpublished (2004)
COMMENT     Contact: Stephen Lapointe, Wayne Hunter, Phat Dang
            US. Horticultural Research Lab, Ft. Pierce, FL.
            USDA-ARS
            2001 South Rock Rd., Ft. Pierce, FL 34945, USA
            Tel: (772) 462-5914, (772) 462-5940
            Fax: (772) 462-5986
            Email: SLapointe@ushrl.ars.usda.gov
            Seq primer: T3 Primer.
FEATURES             Location/Qualifiers
     source          1. .715
                     /organism="Diaprepes abbreviatus"
                     /mol_type="mRNA"
                     /db_xref="taxon:13040"
                     /clone="SLtDaF057_H09"
                     /sex="Adult Female"
                     /tissue_type="Whole body"
                     /dev_stage="Teneral"
                     /lab_host="XL1-Blue"
                     /clone_lib="Diaprepes abbreviatus, Diaprepes root weevil"
                     /note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2:
                     XhoI; A high quality EST with at least 200 contiguous
                     bases at Trace Tuner score of 20 or better. Construction
                     by PM DANG and ANNA SARA HILL, USDA, ARS, U.S.
                     Horticultural Res. Lab, Ft. Pierce, FL, USA."
ORIGIN
```

```
  Query Match             4.8%;  Score 52.4;  DB 7;  Length 715;
  Best Local Similarity   47.1%;  Pred. No. 0.66;
  Matches  161;  Conservative    0;  Mismatches  181;  Indels    0;  Gaps    0;

Qy        150 ATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGT 209

Db        545 ACAAAGATAATGTGCGGTGTATGAATTATAATGAAGAATAGAATTAATTTTTCGTTTAGT 486

Qy        210 TTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTT 269

Db        485 GTAGATAATATATTTTTTTTATTTATTAAAAGAAATTGTATAGCATACAGAGTGTGCATA 426

Qy        270 GTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAG 329

Db        425 TAGTTTAAAATTTGACAACACTTTATAATACAATTTAATATTCTAAAAAGCATGAAATGT 366

Qy        330 CCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAG 389
```

```
Db        365 TTTTTTAAACAAATTTATTAATTTGATATGTTTTTCAATATTTAAAGTTAACAGTATCTT 306

Qy        390 GTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTA 449

Db        305 GTAATTATTCGTACTACCATTTTTATTACACCCTGTTTAATGTTTTATCGTTTTATATAT 246

Qy        450 CCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTT 491

Db        245 TTACAAATTATATTCTGGAAATATTAATCGCAAAATTAAATT 204
```

RESULT 88
BQ390789

```
LOCUS       BQ390789                477 bp    mRNA    linear   EST 22-MAY-2002
DEFINITION  NISC_mq15b03.x1 NICHD_XGC_Emb5 Xenopus tropicalis cDNA clone
            IMAGE:5308372 3', mRNA sequence.
ACCESSION   BQ390789
VERSION     BQ390789.1  GI:21078476
KEYWORDS    EST.
SOURCE      Xenopus tropicalis (western clawed frog)
  ORGANISM  Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE   1  (bases 1 to 477)
  AUTHORS   NIH-XCG http://image.llnl.gov/image/html/xenopuslib_info.shtml.
  TITLE     National Institute of Child Health and Human Development, National
            Cancer Insitute, Xenopus Gene Collection
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            cDNA Library Preparation:
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL
             DNA Sequencing by: National Institutes of Health Intramural
            Sequencing Center (NISC)
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            info@image.llnl.gov
            Plate: LLAM11781  row: D  column: 5
            Seq primer: -21M13 forward primer (ABI).
FEATURES             Location/Qualifiers
     source          1. .477
                     /organism="Xenopus tropicalis"
                     /mol_type="mRNA"
                     /db_xref="taxon:8364"
                     /clone="IMAGE:5308372"
                     /tissue_type="gastrula"
                     /dev_stage="embryo, stages 10-13"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NICHD_XGC_Emb5"
                     /note="Vector: pCMV-SPORT6.1; Site_1: NotI; Site_2: EcoRV;
                     Cloned unidirectionally. Primer: Oligo dT. Average insert
                     size 2.0 kb. Constructed by Invitrogen. Note: This is a
                     Xenopus Gene Collection (XGC) library."
ORIGIN
```

```
  Query Match             4.8%;  Score 52.2;  DB 5;  Length 477;
  Best Local Similarity   47.9%;  Pred. No. 0.69;
  Matches  150;  Conservative    0;  Mismatches  163;  Indels    0;  Gaps    0;

Qy        206 TTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTC 265

Db         88 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 147

Qy        266 ATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATA 325

Db        148 TTTTTTTTTTTTTTTTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 207

Qy        326 TTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTC 385

Db        208 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATTAAAAAATAAAAAAAAAAAAA 267
```

```
Qy        386 AAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAA 445
              |||   |||||   ||   |||   | |  |          || |    |    ||| | ||
Db        268 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATTTTTTAAAAAAAAAAAAA 327

Qy        446 ATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTA 505
              |  |  | |||    |      || |||| | | |||| | |    ||| | ||  |  |
Db        328 AAAAAAAAAAAAAAAAAACCAAAATTAATTTTAAAAAAAAAAGGGGTTTTTTTAAAAAAA 387

Qy        506 AAACATTTACTTT 518
              |||  |   | |||
Db        388 AAAAAAAAAATTT 400


RESULT 89
BM582546
LOCUS       BM582546                617 bp    mRNA    linear   EST 22-FEB-2002
DEFINITION  17000687275996 A.Gam.ad.cDNA.blood1 Anopheles gambiae cDNA clone
            19600449705634 5', mRNA sequence.
ACCESSION   BM582546
VERSION     BM582546.1  GI:18871013
KEYWORDS    EST.
SOURCE      Anopheles gambiae (African malaria mosquito)
  ORGANISM  Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
            Anopheles.
REFERENCE   1  (bases 1 to 617)
  AUTHORS   Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L.,
            Charlab,R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
  TITLE     Celera Anopheles gambiae EST project
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Holt R.A.
            Celera Genomics
            45 w. Gude Dr., Rockville, MD 20850, USA
            Tel: 2404533151
            Fax: 2404534580
            Email: HoltRA@celera.com
            Plate: NU01004AA3  row: D  column: 04
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .617
                     /organism="Anopheles gambiae"
                     /mol_type="mRNA"
                     /strain="RSP-ST (Reduced susc. to Permethrin - std.
                     chromosome)"
                     /db_xref="taxon:7165"
                     /clone="19600449705634"
                     /dev_stage="Adult"
                     /lab_host="DH10b"
                     /clone_lib="A.Gam.ad.cDNA.blood1"
                     /note="Vector: pSport1; Site_1: Sal1; Site_2: Not1; Whole
                     adult mosquitoes (mixed sex) frozen on liquid nitrogen 24
                     hours  after human blood feeding. cDNA inserts >500 bp
                     cloned directionally  into pSport 1. Not 1 site is 3'.
                     Clones available through the Malaria  Research and
                     Reference Reagent Resource Center (www.malaria.mr4.org)"
ORIGIN

  Query Match             4.8%;  Score 52;  DB 4;  Length 617;
  Best Local Similarity   45.4%;  Pred. No. 0.79;
  Matches  187;  Conservative    0;  Mismatches  225;  Indels    0;  Gaps    0;

Qy        141 GGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAA 200
              | ||   ||||| | |    |   ||| | ||        | |  ||||| | ||  ||
Db        205 GTTTATTAAATAATAAATTATTATAATAATTTTTCTCGAATTTTAAAGATTTAATTATAT 264

Qy        201 AGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGC 260
              | | |  | ||||  |   |||   || | | | | |  ||| | | |     | |
Db        265 ATAAAAGTTATTGTGGAATAACTATTTTAAATATGTAATTAGAAATGAAATGTTAATCGT 324

Qy        261 TATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATC 320
              | |  | |   | | | ||| |   |||    | |||    |    ||| |    | |
Db        325 TTTAAAATATATCTAGTTTTTTAAGAAATGAATTTAATTTAGCTTATTTATTTTATTAAG 384

Qy        321 ATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTG 380
              ||  ||    | || | |||||||| | |  |  | | |  |     ||||  |
Db        385 TTAATTTTTTAATTTAATAATTAAATAAAGTAATATTTTAAGGGATGAGCTTTAAAATAA 444

Qy        381 CAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAA 440
              | | |  |    |        |||| |||  | ||       |         |  | | |
Db        445 AATTTTATATTTTTTATAATTTTTAAATAAATATAAGCTTAAAAATAGCTATTATTAATA 504

Qy        441 TAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTAC 500
              |    ||| |   ||||| | |   |||| || | |   ||  ||| || ||    ||
Db        505 AATTTGTTATAATTTATTTTTTATAAAAAATTATTTAATTTAAATTAAATTATTTATTAA 564

Qy        501 TATTAAAACATTTACTTTCCAAATATTCTGTCATTCAGGAGTATGGAAGTAT 552
              ||| |||  |||| |    ||| |||  |  | | | ||  |    |||||
Db        565 AATTTAAAATTTTAATAATAAAAAATTTAGGTAATTATGATAAAATTAGTAT 616


RESULT 90
BF278572/c
LOCUS       BF278572                839 bp    mRNA    linear   EST 07-MAR-2001
DEFINITION  GA__Eb0033C04f Gossypium arboreum 7-10 dpa fiber library Gossypium
            arboreum cDNA clone GA__Eb0033C04f, mRNA sequence.
ACCESSION   BF278572
VERSION     BF278572.1  GI:11209642
KEYWORDS    EST.
SOURCE      Gossypium arboreum
  ORGANISM  Gossypium arboreum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
REFERENCE   1  (bases 1 to 839)
  AUTHORS   Wing,R.A., Frisch,D., Yu,Y., Main,D., Rambo,T., Simmons,J.,
            Henry,D., Wood,T.C., Leslie,A. and Wilkins,T.A.
  TITLE     An integrated analysis of the genetics, development, and evolution
            of the cotton fiber
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: TAATACGACTCACTATAGGG
            High quality sequence start: 5
            High quality sequence stop: 300.
FEATURES             Location/Qualifiers
     source          1. .839
                     /organism="Gossypium arboreum"
                     /mol_type="mRNA"
                     /strain="AKA"
                     /cultivar="8400"
                     /db_xref="taxon:29729"
                     /clone="GA__Eb0033C04f"
                     /tissue_type="Fibers isolated from bolls harvested 7-10
                     dpa"
                     /lab_host="E. coli"
                     /clone_lib="Gossypium arboreum 7-10 dpa fiber library"
                     /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"
ORIGIN

  Query Match             4.8%;  Score 52;  DB 2;  Length 839;
  Best Local Similarity   49.1%;  Pred. No. 0.82;
  Matches  166;  Conservative    0;  Mismatches  170;  Indels    2;  Gaps    1;

Qy        179 TTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTA 238
              | |  | || |||  ||  ||||| | |   |   ||| | || ||   |  | ||
Db        781 TATAAATAACATCAAATTCTAAAGATAATACTAAATTACTCTAATCTGATACTCTTACAT 722

Qy        239 AAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAA 298
              | || | |   |   ||  || || | |   | |  || |||  | | |  || ||
Db        721 ATAGTCATGGGACGTATTAAGTTAATTTGTAAATCTATTTTAGTTTGATATATATAAAAT 662
```

```
Qy        299 GAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTG 358
              | ||          |  |     || |   || | |||   |||| || |  |||| | ||
Db        661 AATATTATAAATAAATACATACACACCATAACTAGTCATAAATCAATCTTAAAAGGTGTT 602

Qy        359 CATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAA--AGGTATACTA 416
              ||| ||  || |||  | ||     || |    ||||||    |||   |  |  | ||
Db        601 TATAATAATTTTCTTTGTATTTATCACATAATACAAAAGTATTTCTATTTAACTTAATTA 542

Qy        417 CCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCT 476
              |  | ||  |   || |  ||  |  |  ||  ||||||| ||         | ||
Db        541 TTTTATCTCAGAAATCAATATAATGATTATATATAACAAATTATAACCAAACATTATAAA 482

Qy        477 CAGAAAACATAATTTTTTATGTACTATTAAAACATTTA 514
              || |||  ||||  | ||  || | ||||     | ||
Db        481 CACAAATTATAATAATATAATTAATTTTAATTTTTATA 444


RESULT 91
CNS02T50/c
LOCUS       CNS02T50                1027 bp    DNA     linear   GSS 01-SEP-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            163M16 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL212733
VERSION     AL212733.1  GI:7871552
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE   1
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
  MEDLINE   20296633
   PUBMED   10835645
REFERENCE   2
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
            Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
            Saurin,W., Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
  MEDLINE   20359837
   PUBMED   10899143
REFERENCE   3  (bases 1 to 1027)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .1027
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="163M16"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0AG163BG08LP1~end : T7"
ORIGIN

  Query Match             4.8%;  Score 52;  DB 9;  Length 1027;
  Best Local Similarity   38.2%;  Pred. No. 0.84;
  Matches  149;  Conservative   35;  Mismatches  206;  Indels    0;  Gaps    0;

Qy        146 AGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACA 205
              | |||||||       | | || | |||  |    |  |:|| |||  ||     | | |
Db        954 ANAAATATTWTATTAAATATATAATTTATAAAAAATAAASAAAATCAAATAAATCAWAAA 895

Qy        206 TTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTC 265
              ||     ||| | |   |             ||:| ||   | : | |:|   :| ||
Db        894 AAATTAAAAAAATTTTCAAACACACACACCAAMAAACAAAAAHAAAAWAAYAMTTTTT 835

Qy        266 ATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATA 325
              | |   | :   |   |: |:||       | | | :  :: |   | ||| ||  ||:|
Db        834 AATCTATCWAAATNATTWTAWAATTTTATNANACATWCWYWATWTAATAAATATATTAWA 775

Qy        326 TTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTC 385
              :       | | | ||||  |   ||| | |  || ||    |  |||    | |:
Db        774 WATCTANTATTANATTTAACAATTTAATTATTTTATTTTTTTATTTTTANAAATACWTAW 715

Qy        386 AAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAA 445
              :|:  ::::::   ||   ||   | :|:|   |  ||| || |     |  ::::: ::
Db        714 WAWATWWWWWWWAWAAWAAAACAATYYAYTTTTTWTTTTTTTTTTTTTTTTWWWWWWWWW 655

Qy        446 ATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTA 505
              :  |  | ||||  |  |    ||   |    | || | ||||    ||:|    |||
Db        654 WAAAAAAAAAATAATTATNNATAATATAAAATAAAATATATAAAAAATTWTTATATATAT 595

Qy        506 AAACATTTACTTTCCAAATATTCTGTCATT 535
              ||  :|||  |||  :||| |     |||
Db        594 AATTWTTTTATTTATWAATTTANAAAAATT 565


RESULT 92
CK242493
LOCUS       CK242493                691 bp    mRNA    linear   EST 11-DEC-2003
DEFINITION  rx10h04.y1 Meloidogyne paranaensis egg SMART pGEM Meloidogyne
            paranaensis cDNA 5', mRNA sequence.
ACCESSION   CK242493
VERSION     CK242493.1  GI:39748083
KEYWORDS    EST.
SOURCE      Meloidogyne paranaensis
  ORGANISM  Meloidogyne paranaensis
            Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
            Tylenchoidea; Heteroderidae; Meloidogyninae; Meloidogyne.
REFERENCE   1  (bases 1 to 691)
  AUTHORS   McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J.,
            Wylie,T., Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B.,
            Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Franklin,C.,
            Tsagareishvili,R., Ronko,I., Kennedy,S., Maguire,L., Beck,C.,
            Underwood,K., Steptoe,M., Allen,M., Person,B., Swaller,T.,
            Harvey,N., Schurk,R., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
            McCann,R., Waterston,R. and Wilson,R.
  TITLE     The Washington Univ. Nematode EST Project, 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: McCarter JP
            The Washington Univ. Nematode EST Project, 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Cloned unidirectionally. Poly(A)+ RNA was concentrated and purified
            using Dynabeads (Dynal) and mRNA eluted for first strand synthesis.
            First strand cDNA was created using MMLV RT (Powerscript, Clontech)
            and primed with oligo(dT) with XhoI site (primer CDSIII/3-XhoI) and
            5'SMART 'anchor' added using chimeric DNA-RNA oligo
            (SMART-NotI-r-GGG). 12 PCR cycles were done using first strand and
            primers specific to SMART oligo(5'PCR-primer) and 3'
            end(XhoI-No-dT). Double stranded cDNA was digested using XhoI/NotI,
            fractioned on Chroma-spin 400 columns (Clontech) and ligated to
            digested pGEM-11zf(+) plasmid. Chemically competent DH10B cells
            were used as host cells. Library constructed by Joanna Carlson of
            North Carolina State University.
            Seq primer: Sp6.
```

```
FEATURES             Location/Qualifiers
     source          1. .691
                     /organism="Meloidogyne paranaensis"
                     /mol_type="mRNA"
                     /db_xref="taxon:189293"
                     /tissue_type="whole organism"
                     /dev_stage="egg"
                     /lab_host="DH10B"
                     /clone_lib="Meloidogyne paranaensis egg SMART pGEM"
                     /note="Vector: pGEM-11zf(+) (Promega); Site_1: XhoI;
                     Site_2: NotI; Cloned unidirectionally. Poly(A)+ RNA was
                     concentrated and purified using Dynabeads (Dynal) and mRNA
                     eluted for first strand synthesis. First strand cDNA was
                     created using MMLV RT (Powerscript, Clontech) and primed
                     with oligo(dT) with XhoI site (primer CDSIII/3-XhoI) and
                     5'SMART 'anchor' added using chimeric DNA-RNA oligo
                     (SMART-NotI-r-GGG). 12 PCR cycles were done using first
                     strand and primers specific to SMART oligo(5'PCR-primer)
                     and 3' end(XhoI-No-dT). Double stranded cDNA was digested
                     using XhoI/NotI, fractioned on Chroma-spin 400 columns
                     (Clontech) and ligated to digested pGEM-11zf(+) plasmid.
                     Chemically competent DH10B cells were used as host cells.
                     Library constructed by Joanna Carlson."
ORIGIN

  Query Match             4.8%;  Score 51.8;  DB 7;  Length 691;
  Best Local Similarity   46.0%;  Pred. No. 0.89;
  Matches  176;  Conservative    0;  Mismatches  207;  Indels    0;  Gaps    0;

Qy        144 TGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGA 203
              || | ||  ||  |       | | | || |       | || || | |  |||| |
Db        278 TGGGGAAGTTTATTTGGCTAATTTAATTGAAGAAAAACAAAAAATATATGAATTTAATGT 337

Qy        204 CATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTAT 263
                | | |||  | | ||     ||    | || ||||   | |  ||   |    ||
Db        338 TTTGGCTTTAATTAAATTGGAAAATGGAAATGTTGAAAGTTGTGAAGGATTAATTCAAAT 397

Qy        264 TCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATA 323
              | || |  |||| |   |   |||| | | | ||  | |    || | ||||| |  | |
Db        398 TAATGTTATTTTTGGGGGGAGAAAAGAAGGAGAAGAAGAAAAGTTAAAAGAAAAAGAAGA 457

Qy        324 TATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAG 383
                |        ||| | | | |  | || | | |  | ||| || |    |||      |
Db        458 AAAACCTTTTTTGGATAATCAGAATCAATTTAATTTAAATTCCAATCGAGAATTAAATAC 517

Qy        384 TCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAG 443
              || ||   |  |     |  ||  | | |  |  | ||  | |   || | ||||
Db        518 TCCAATACCACCAACATTCAAAACAGAAGAAGAAAAGATTATTGAAACAACAGAAAATTT 577

Qy        444 AAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTAT 503
                |  | || ||||  ||    |||  | | ||||  | |  |  |     | ||||
Db        578 TGTGGAAAATAATTTTTTGTTAAATAATTTTGTGAGAAGGCCTCCTCCTCCTCCTTCTAT 637

Qy        504 TAAAACATTTACTTTCCAAATAT 526
              ||||  |  || ||   |||| |
Db        638 TAAACAAAATATTTCAAAAATTT 660


RESULT 93
BM162809
LOCUS       BM162809                 787 bp    mRNA    linear   EST 04-DEC-2001
DEFINITION  EST565332 PyBS Plasmodium yoelii yoelii cDNA clone PYCKX22 5' end,
            mRNA sequence.
ACCESSION   BM162809
VERSION     BM162809.1  GI:17308490
KEYWORDS    EST.
SOURCE      Plasmodium yoelii yoelii
  ORGANISM  Plasmodium yoelii yoelii
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 787)
  AUTHORS   Carlton,J.M., Daly,T.M., Long,C.A., Bergman,L.W., Vaidya,A.B.,
            Fraser,C.M. and Carucci,D.J.
  TITLE     Plasmodium yoelii EST project at TIGR
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Jane Carlton
            Parasite Genomics Group
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-530-9319
            Fax: 301-838-0208
            Email: carlton@tigr.org
            For clone info, please contact the Malaria Research and Reference
            Reagent Resource Center, ATCC
            http://www.malaria.mr4.org/mr4pages/index.html
            Seq primer: ADF.
FEATURES             Location/Qualifiers
     source          1. .787
                     /organism="Plasmodium yoelii yoelii"
                     /mol_type="mRNA"
                     /strain="17XL"
                     /sub_species="yoelii"
                     /db_xref="taxon:73239"
                     /clone="PYCKX22"
                     /dev_stage="Asexual blood stages"
                     /lab_host="E. coli XL-1 Blue"
                     /clone_lib="PyBS"
                     /note="Vector: pAD-GAL4;. At 20-25% parasitemia, blood was
                     collected from BALB/cByJ mice infected with Py17XL
                     parasites, and leukocytes removed by passage over
                     microcrystalline cellulose collumns. Total RNA was
                     isolated using the guanidinium isothiocyanate method, and
                     mRNA isolated using oligo(dT)-cellulose chromatography.
                     First strand cDNA synthesis was completed using a 50-base
                     primer and reverse transcriptase in the presence of
                     5-methyl dCTP. After second strand synthesis, uneven
                     termini were treated with Pfu DNA polymerase and EcoRI
                     adaptors ligated to the blunt ends. The sample was cleaved
                     with XhoI and separated on a Sephacryl S-500 column.
                     Size-fractionated cDNA was precipitated and ligated to
                     HybriZAP arms directionally using EcoRI-XhoI cleaved arms.
                     After packaging, the phagemid vector (pAD-GAL4) was
                     excised from the HybriZAP vector and plasmid DNA
                     isolated."
ORIGIN

  Query Match             4.8%;  Score 51.8;  DB 4;  Length 787;
  Best Local Similarity   46.9%;  Pred. No. 0.9;
  Matches  161;  Conservative    0;  Mismatches  182;  Indels    0;  Gaps    0;

Qy        215 TAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTT 274
              | ||||| |     |  | | |  |||  |  | | |||  ||    ||     |  |
Db        220 TCATATATTAATATATTATATACAAACAAAATAAAAGATTTAATTATAAATTATGCTACA 279

Qy        275 TGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGT 334
              |  ||   |  |||||    |||      ||  | ||||   | ||| |||||   ||
Db        280 TCATTATTAGAAAAATATAATACTTGCATTGAATCAGGACATGATAATATAGATGCAGAA 339

Qy        335 TGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAA 394
              |  |  |||   | ||| | || ||   | |||| | ||| ||      ||    |
Db        340 ATACGAAGAAAAGAACAATTAATGAATTCAAAATTATTAAATTTTTATAATAATTTTTTT 399

Qy        395 AAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACA 454
              |    ||   |      ||| | |      | ||||||  |     ||     | ||
Db        400 ATTATAAAAAATCAAAATATTCCTAAAAGTGTTATCGAATATAAAGATAATATATATGAT 459

Qy        455 AATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTA 514
              |||  ||  |    | |||||  ||  |||      | ||||  || |||  || | |||||
Db        460 AATAATAATAATAATATAATTTCCAAAATTTATAATTTTACTGGACTAACAATAAATTTA 519

Qy        515 CTTTCCAAATATTCTGTCATTCAGGAGTATGGAAGTATCGATG 557
                |  ||| | ||| | |      |   |  || ||   |||
Db        520 AGATTTAAACAATCTCTTAGAAGTAAAGTTCAAAATAATAATG 562


RESULT 94
CL477665/c
```

```
LOCUS       CL477665               983 bp    DNA     linear   GSS 01-APR-2004
DEFINITION  SAIL_278_D12.v1 SAIL Collection Arabidopsis thaliana genomic clone
            SAIL_278_D12.v1, genomic survey sequence.
ACCESSION   CL477665
VERSION     CL477665.1  GI:45944762
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 983)
  AUTHORS   Sessions,A., Burke,E., Presting,G., Aux,G., McElver,J., Patton,D.,
            Dietrich,B., Ho,P., Bacwaden,J., Ko,C., Clarke,J.D., Cotton,D.,
            Bullis,D., Snell,J., Miguel,T., Hutchison,D., Kimmerly,B.,
            Mitzel,T., Katagiri,F., Glazebrook,J., Law,M. and Goff,S.A.
  TITLE     A high-throughput Arabidopsis reverse genetics system
  JOURNAL   Plant Cell 14 (12), 2985-2994 (2002)
  MEDLINE   22356987
   PUBMED   12468722
COMMENT     Contact: Sessions A
            Applied Trait Genetics
            Syngenta Biotechnology Inc.
            3054 Cornwallis Rd., Research Triangle Park, NC 27709, USA
            Email: allen.sessions@syngenta.com
            ABRC Stock Number CS812871; T-DNA left border flanking sequences of
            Syngenta Arabidopsis Insertion Library (SAIL) lines are available
            through the Arabidopsis Biological Resource Center (ABRC).
            Sequences represent a pool of amplified genomic regions and not
            single contiguous sequences.
            Class: TDNA tagged.
FEATURES             Location/Qualifiers
     source          1. .983
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /ecotype="Columbia"
                     /db_xref="taxon:3702"
                     /clone="SAIL_278_D12.v1"
                     /clone_lib="SAIL Collection"
                     /note="T-DNA left border sequences were isolated using a
                     modified TAIL-PCR strategy"
ORIGIN

  Query Match             4.8%;  Score 51.8;  DB 9;  Length 983;
  Best Local Similarity   36.7%;  Pred. No. 0.92;
  Matches  149;  Conservative    0;  Mismatches  257;  Indels    0;  Gaps    0;

Qy        132 TATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATC 191
              ||  |  ||  |  |  |  || |  |  |  |||     ||||    | ||    |
Db        928 TAANAAATTAATAAATNNAANTTAANNANAAANAATNTNNAAATACCCCTNTANTNTNTN 869

Qy        192 TCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAG 251
              |      |  |  || |||| | ||   |           | || ||        | | |
Db        868 NTANNNANANNAATTTTTTTTTTAAAANAAAAAAAAAAANATNTTATATNNTTAATAATAN 809

Qy        252 ACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAG 311
              | ||| |   ||   |||  ||||              |||  |||| | |||    |
Db        808 AAATAAAAAAATATTTTTNNTTTTTTNAAAANNNNTAAAATTATANAAAAANTATTNTTA 749

Qy        312 AGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTAC 371
              | |||  |  |    |    |       ||  ||       |       |   |
Db        748 AAAAATTTTTTNNNNTTNNNANNNNNNATNTAAANNTNNNTNTNNTNNNNTNTAAAANN 689

Qy        372 TTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCG 431
              | |||      |   ||    ||||  |  |      |   ||          ||    |
Db        688 TAAATNNNAAAANAAATTTAAAAANATATATTNNNANAAATNNNNNNATTTTAAAAAAAA 629

Qy        432 CACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTT 491
              |       | ||  |   | |||  |  ||     || |  | | | ||   | | |||
Db        628 AAAANANNNAATAANATTATAAAAAATAAATAAAANAAGTTTTTTNAAAATTTAAANTTT 569

Qy        492 TTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGTCATTCA 537
              |  |  ||  ||  |      |   |  ||||| | |   | | |
```

```
Db        568 TAAAATTAAAATATATNTTNATTTGTAAAAAATAGTATAAAAATNA 523


RESULT 95
CD386564/c
LOCUS       CD386564               1067 bp    mRNA    linear   EST 30-MAY-2003
DEFINITION  AGENCOURT_14286445 NIH_MGC_173 Homo sapiens cDNA 5', mRNA sequence.
ACCESSION   CD386564
VERSION     CD386564.1  GI:31222277
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1067)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
            National Cancer Institute / NIH
            Bldg. 31 Rm10A07 Bethesda, MD 20892
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Jamie Thompson, University of WI
            cDNA Library Preparation: Gina Zastrow-Hayes
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: NDKM74  row: e  column: 02
            High quality sequence start: 8
            High quality sequence stop: 462.
FEATURES             Location/Qualifiers
     source          1. .1067
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /tissue_type="embryonic trophoblasts, made from WA01 stem
                     cells"
                     /lab_host="DH10B TonA"
                     /clone_lib="NIH_MGC_173"
                     /note="Vector: pDONR201; Site_1: attP2; Site_2: attP1;
                     LIBR_PRIMING - oligo dT; METHOD - full-length enriched;
                     LIBR_PROVIDER - Bradfield"
ORIGIN

  Query Match             4.8%;  Score 51.8;  DB 6;  Length 1067;
  Best Local Similarity   43.0%;  Pred. No. 0.93;
  Matches  176;  Conservative    0;  Mismatches  233;  Indels    0;  Gaps    0;

Qy        127 TCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGA 186
              | | |||| |  ||   |      |||||  | |||   | |      |||   |  || |
Db        803 TATTATATAATATTTTATANTTNATATTTTTAATTTATATTTTTAATATATAATAATANA 744

Qy        187 AGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCT 246
              |    ||  ||||    || || | ||  || |        | ||||||     |
Db        743 NAAAAAAATAATAAAATTTTTATTATATTTTTAAAAATAAAAAAATTTTTTTTTTAATAA 684

Qy        247 TACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCA 306
              |  | | ||| |  | ||  ||| |||     |    | | ||  |    ||   | | |
Db        683 TTAAAAAATATATATTTTTTTTTTTTTTANATTAATATAATAATATAAATTNTTTATTTTA 624

Qy        307 TTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTAC 366
              ||    |      || |  ||||      ||  |||  ||| | |||   | |   ||| |
Db        623 TTTTNANNTTTATAAATTATTTTTTTATTATAAAAAAAAAAAAAAATATAATTTTTATAAA 564

Qy        367 ATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCT 426
              || |   ||  ||      |||||   ||  ||||   ||  |   |  ||| |
Db        563 ATAAAAAAAATTTATTTTTTTNAAAAAAAAANAAAAATAAAAAAATTNTTTTTTTTTTTT 504

Qy        427 TATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACAT 486
                ||    |      | |||            || ||||     |  | |        |
```

```
Db        503 NNTCCTTTTNTTTTAAAAAAANNNNNNNNNNNTTTTTTTTTTTTTTTTTTTTTTTTTTT 444

Qy        487 AATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGTCATT 535
              |||||| | |  | ||      |||  |||     | || | |  ||
Db        443 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 395


RESULT 96
B11102
LOCUS       B11102                 1187 bp    DNA     linear   GSS 14-MAY-1997
DEFINITION  F19C22-T7 IGF Arabidopsis thaliana genomic clone F19C22, genomic
            survey sequence.
ACCESSION   B11102
VERSION     B11102.1  GI:2092386
KEYWORDS    GSS.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1187)
  AUTHORS   Feng,J., Dewar,K., Buehler,E., Kim,C., Li,Y., Shinn,P., Sun,H. and
            Ecker,J.
  TITLE     BAC End Sequences at ATGC
  JOURNAL   Unpublished (1997)
COMMENT     Other GSSs: F19C22-Sp6
            Contact: Ecker J.
            Arabidopsis Thaliana Genome Center
            University of Pennsylvania
            Dept. of Biology, University of Pennsylvania, Philadelphia, PA
            19104
            Tel: 215-898-9384
            Fax: 215-898-8780
            Email: jecker@atgenome.bio.upenn.edu
            Seq primer: T7
            Class: BAC ends
            High quality sequence start: 72
            High quality sequence stop: 353.
FEATURES             Location/Qualifiers
     source          1. .1187
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /ecotype="Columbia"
                     /db_xref="taxon:3702"
                     /clone="F19C22"
                     /sex="hermaphrodite"
                     /clone_lib="IGF"
                     /note="Vector: BeloBACII; Site_1: EcoRI; Site_2: EcoRI;
                     Produced by Thomas Altmann"
ORIGIN

  Query Match             4.8%;  Score 51.8;  DB 8;  Length 1187;
  Best Local Similarity   46.1%;  Pred. No. 0.94;
  Matches  184;  Conservative    0;  Mismatches  214;  Indels    1;  Gaps    1;

Qy        138 TTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCT 197
              ||    | | ||||||| ||  | |  |||   ||||    |||  |  |  |     | |
Db        762 TTNTTNTNATAAATATTAATTTTAAATAATNTTTTAAATTATTTTAAATATTTAATTTTT 821

Qy        198 TAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTA-CAGACATA 256
              ||||      | || | ||  |||  |    |  |||||||||  |   |    ||  | | | |
Db        822 TAAAATATATTTTATTTTNTTATTTTNTTTATATTTTTTTATATTATTATATTAAAAAAA 881

Qy        257 CAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAA 316
                | || |||| | |||    | || | ||||  |||      |   ||   | |||
Db        882 TTTTTTTTAATTTATATTTTAAATATTTATAAAAAATATTTNNTTNTTTTTTTTAATAAA 941

Qy        317 GATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAAT 376
              || |  | ||    | |  |  |||||  | || ||      ||||||   |  | |||
Db        942 AATAAATTTTTTTTTAATATANTATTAATAAAAATATTTTATAATATTAATATTTTAAAT 1001

Qy        377 CTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTG 436
                |    | ||   | |    |       |   |    |   | | | |||    |
```

```
Db       1002 TATAATTTTAATTTTTATTTATATTTTATATTATTATATTAATTATATTTTANTTTAAAT 1061

Qy        437 CAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTAT 496
              | |      |||    |  | || ||||  |  | | |    | | |    ||  || || |
Db       1062 TATAATTTTATTTTTAATATTTATATTATATATAAATAATTATTATTATTATATTATTNT 1121

Qy        497 GTACTATTAAAACATTTACTTTCCAAATATTCTGTCATT 535
                | |   |  |||| | |  | ||||  | | |||
Db       1122 AATAAAATTTTATTTTTATTATATATATATAATTTAATT 1160


RESULT 97
BE420740/c
LOCUS       BE420740               1299 bp    mRNA    linear   EST 24-JUL-2000
DEFINITION  HWM002.A09 ITEC HWM Barley Leaf Library Hordeum vulgare subsp.
            vulgare cDNA clone HWM002.A09, mRNA sequence.
ACCESSION   BE420740
VERSION     BE420740.1  GI:9418583
KEYWORDS    EST.
SOURCE      Hordeum vulgare subsp. vulgare
  ORGANISM  Hordeum vulgare subsp. vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Pooideae; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 1299)
  AUTHORS   Anderson,O.A., Appels,R., Bailey,P., Blake,T., Close,T.,
            Cloutier,S., Dubcovsky,J., Feuillet,C., Gale,M., Graner,A.,
            Gustafson,P., Herrmann,R.G., Holton,T., Jacquemin,J.M., Jia,J.,
            Joudrier,P., Langridge,P., Lazo,G.R., Lin,J.J., McGuire,P.,
            Ogihara,Y., Pecchioni,N., Qualset,C., Schuch,W., Selvaraj,G.,
            Shariflou,M., Sorrells,M., Warburton,M. and Wenzel,G.
  TITLE     International Triticeae EST Cooperative (ITEC): Production of
            Expressed Sequence Tags for Species of the Triticeae
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Herrmann RG
            Botanisches Institut der LMU
            Menzinger Str. 67, D-80638 Munchen GERMANY
            Fax: 49 30 171683
            Email: herrmann@botanik.biologie.uni-muenchen.de
            International Triticeae EST Cooperative (ITEC)
            http://wheat.pw.usda.gov/genome.
FEATURES             Location/Qualifiers
     source          1. .1299
                     /organism="Hordeum vulgare subsp. vulgare"
                     /mol_type="mRNA"
                     /cultivar="barke"
                     /sub_species="vulgare"
                     /db_xref="taxon:112509"
                     /clone="HWM002.A09"
                     /tissue_type="leaf"
                     /dev_stage="14 day old"
                     /clone_lib="ITEC HWM Barley Leaf Library"
                     /note="Vector: pBluescriptSK(-); 850 bp average insert
                     size."
ORIGIN

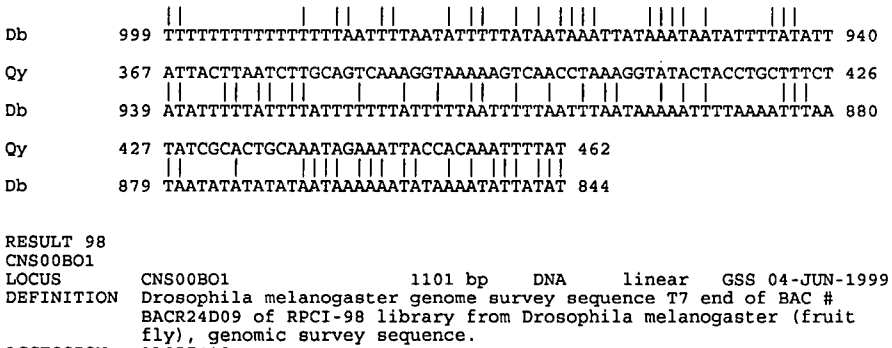
  Query Match             4.8%;  Score 51.8;  DB 2;  Length 1299;
  Best Local Similarity   44.3%;  Pred. No. 0.95;
  Matches  149;  Conservative    0;  Mismatches  187;  Indels    0;  Gaps    0;

Qy        127 TCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGA 186
              | | ||||   ||   | | ||||||  || |||| ||   | ||| |    |  || |
Db       1179 TATAATATATTTTAAAATTAAAAATATAAATAATTAAAATAAAATTATTTTAATAATATA 1120

Qy        187 AGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCT 246
              |  | | ||  ||   | ||  | |   ||  |||         | |  ||||       |
Db       1119 AATTATTATAATATTTAAATAATATAAATATAATATATAAAATAATAATTTANNTTNNTT 1060

Qy        247 TACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCA 306
                                  |||   |||  ||| || ||||||||   ||   |||
Db       1059 NNNNNNNNNNNNNNNNNNTTTTTNNNTTTTTTTTTTTTAAAAAAAGGAGGAATATAATTTT 1000

Qy        307 TTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTAC 366
```



```
Db        999 TTTTTTTTTTTTTTTTAATTTTAATATTTTTATAATAAATTATAAATAATATTTTATATT 940

Qy        367 ATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCT 426

Db        939 ATATTTTTATTTTATTTTTTTATTTTTAATTTTAATTTAATAAAAATTTTAAAATTTAA 880

Qy        427 TATCGCACTGCAAATAGAAATTACCACAAATTTTAT 462

Db        879 TAATATATATATAATAAAAAATATAAAATATTATAT 844


RESULT 98
CNS00BO1
LOCUS       CNS00BO1                1101 bp    DNA     linear   GSS 04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR24D09 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL057419
VERSION     AL057419.1  GI:4937885
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR24D09"
                     /clone_lib="RPCI-98"
                     /note="end : T7"
ORIGIN

  Query Match             4.7%;  Score 51.6;  DB 9;  Length 1101;
  Best Local Similarity   40.2%;  Pred. No. 1;
  Matches  185;  Conservative   48;  Mismatches  217;  Indels   10;  Gaps    2;

Qy        159 ATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTG------TTTT 212

Db        630 AYASCASAAAATTWTMTWTWTTTSTAAAAAATATAAWBSSVSTGBSATTKTASAWTTATT 689

Qy        213 GTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTT 272

Db        690 TWTWABATTGSSSTKBMWAKBSTTTATAAAAAAAAAAAAATTTWTTTTTATWTTTTTTTT 749

Qy        273 TTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCA 332

Db        750 TTTTTTWTWWTWTATTAATTTATATAAATTAAWTWAAAAAAAAWAAAAATWATTWAAAAA 809

Qy        333 GTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTA 392

Db        810 AWAATAAAAAAAAAAAAAAAATAAAAWATAATWAWTWTATAAATTTAATWAAATAATAAAW 869

Qy        393 AAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCA 452

Db        870 AWAAATAAAAWAAAATTTATTTAAAWAATATTAAAATAATTWTTTATATTTTTTTTATWA 929

Qy        453 C----AAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAA 508

Db        930 ATTTWAATTAAAAATATAAATTAAATTAAAWAAAATWWAAAAATTWAAAAAWATTTAAAA 989

Qy        509 CATTTACTTTCCAAATATTCTGTCATTCAGGAGTATGGAAGTATCGATGGCTTCTTTAAA 568

Db        990 ATTTWAATTAAAAWAAWTTTWATTTTTTTTAAAAATWATATTAATWATWAAWATAWWAAA 1049

Qy        569 ATGAAGCAGGAGGGTCTGGCAGAGAGTATCTATGAAATAA 608

Db       1050 AAAAATTWAAATAATAWAAWATTAATTATTTTWATATTWA 1089


RESULT 99
CC218891
LOCUS       CC218891                1164 bp    DNA     linear   GSS 12-MAY-2003
DEFINITION  CH261-14M11_Sp6.1 CH261 Gallus gallus genomic clone CH261-14M11,
            genomic survey sequence.
ACCESSION   CC218891
VERSION     CC218891.1  GI:30540572
KEYWORDS    GSS.
SOURCE      Gallus gallus (chicken)
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1  (bases 1 to 1164)
  AUTHORS   Kremitzki,C., Higginbotham,J., Wylie,K., Carter,J., McPherson,J.,
            Warren,W., Graves,T., Mardis,E. and Wilson,R.
  TITLE     Gallus gallus BAC End Reads
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K. Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 182000   Std Error: 0.00
            Seq primer: Sp6 ATTTAGGTGACACTATAG
            Class: BAC ends
            High quality sequence start: 122
            High quality sequence stop: 310.
FEATURES             Location/Qualifiers
     source          1. .1164
                     /organism="Gallus gallus"
                     /mol_type="genomic DNA"
                     /strain="Red Jungle Fowl"
                     /db_xref="taxon:9031"
                     /clone="CH261-14M11"
                     /sex="female"
                     /cell_line="UCD001, inbred 256"
                     /clone_lib="CH261"
                     /note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;
                     CH261 Female Chicken library - for library and clone
                     ordering information: http://www.chori.org/bacpac"
ORIGIN

  Query Match             4.7%;  Score 51.6;  DB 8;  Length 1164;
  Best Local Similarity   38.9%;  Pred. No. 1;
  Matches  193;  Conservative    0;  Mismatches  297;  Indels    6;  Gaps    1;

Qy         18 TGCAGTAACTTTCACTCTCTATGTTCTTATAAGCAAATGTTACAATGAGATATGAGTTTT 77

Db        244 TCCATTATCTTTTTTTTTTTTTTTTTTTTAAGAAACAGCTTGATTTGTCTGTGTGGGGG 303

Qy         78 AAAGCCAGATCTTCCTTATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAG 137
```

```
Db        304 GGGAGGAANNNNNNNTNNNNNNNNNANNANNNNNNNNNNNANTTANANNNNANANAATNNA 363

Qy        138 TTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCT 197

Db        364 TTATTTTAAAANAAATAAATATAATAAATATTAANAAAATTTTTTTTTANNNNNTTNNTT 423

Qy        198 TAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATAC 257

Db        424 TNAATTAATATTTTTTTTATTATATTTATNTTNNTTTATTTA------NTATNTATNTNN 477

Qy        258 AGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAG 317

Db        478 TNNTATTTATTTTTTATTNANTTTTATAATTATATAATNATATANTATTTATATATATNT 537

Qy        318 ATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATC 377

Db        538 ATTAAATTNNTTTNAAATATANTATTTATATATTATATAATTATTNTATATTTTATTATA 597

Qy        378 TTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGC 437

Db        598 AAATTTTTATTATNATNATTTTTTTNTATNTTNNTATTTNNNAAATTTNAATTTTTTTAA 657

Qy        438 AAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATG 497

Db        658 ANNTATTANTTTATATTTATTAATTATTTNAAATTANNNAATATATATNATTTATNTATT 717

Qy        498 TACTATTAAAACATTT 513

Db        718 TTTTANTAATATTTAT 733
```

```
RESULT 100
BZ048202
LOCUS       BZ048202                754 bp    DNA     linear   GSS 09-OCT-2002
DEFINITION  lkg33f10.b1 B.oleracea002 Brassica oleracea genomic, genomic survey
            sequence.
ACCESSION   BZ048202
VERSION     BZ048202.1  GI:23645459
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 754)
  AUTHORS   Delehaunty,K., Fewell,G., Fulton,L., McCombie,W.R., Miner,T.,
            Nash,W., Rabinowicz,P.D. and Wilson,R.K.
  TITLE     Whole genome shotgun reads from Brassica oleracea
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Richard K. Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Plate: lkg33  row: f  column: 10
            Seq primer: -21UPpOT forward
            Class: shotgun
            High quality sequence start: 16
            High quality sequence stop: 551.
FEATURES             Location/Qualifiers
     source          1. .754
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3712"
                     /clone_lib="B.oleracea002"
                     /note="Vector: pOTw13; Whole genome shotgun library from
                     flowering buds. DNA was purified  from a crude nuclear
                     prep using Brassica oleracea TO1000DH3 buds  provided by
                     Thosmas Osborn at the University of Wisconsin. Genomic
                     DNA was provided by Pablo Rabinowicz (CSHL) and the
                     shotgun library  prepared at Washington University Genome
                     Sequencing Center."
ORIGIN
```

```
  Query Match             4.7%;  Score 51.4;  DB 8;  Length 754;
  Best Local Similarity   48.5%;  Pred. No. 1.1;
  Matches  172;  Conservative    0;  Mismatches  181;  Indels    2;  Gaps    1;

Qy        183 TAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAG 242

Db        269 TTGAATGTTTCGTAACAATATATTTATTATAATTTTTTAATTAAAATATATTTTTACCAT 328

Qy        243 ACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAA 302

Db        329 TATGTGTTTTCTGTTAAATATTTTTATAATAATGATTGTTTTCTAAAAAGGATAAATATG 388

Qy        303 TGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATA 362

Db        389 ACTATTTTAAAAAAAAT--TTAAATTTTTTATTGATAATTGTACTAAAATTAGTTGTATA 446

Qy        363 TTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCT 422

Db        447 TAATATTCATGATTCATTTTTTAATAAAATTACTTAAATTTTGAAGGTATATAATACTGA 506

Qy        423 TTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAA 482

Db        507 AACTTTTGAAATAGTTTATATAGTTTAGAAATATTTGATTCCTAACAATTATTTTATGAA 566

Qy        483 ACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGTCATTCA 537

Db        567 TCATTATATATATTATATAAATTACATAGTTATTTACCATAAATTATTTTGTGCA 621
```

```
Search completed: April 22, 2005, 04:39:43
Job time : 2203.35 secs
```

This Page Blank (uspto)

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 20, 2005, 22:49:00 ; Search time 355.089 Seconds
(without alignments)
18171.562 Million cell updates/sec

Title: US-09-939-209A-3_COPY_11056_12145
Perfect score: 1090
Sequence: 1 tgctttccgtgtggtggtgc.....taactctaggatctcaggtg 1090

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_16Dec04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		1090	100.0	20300	6	ABK47337	Abk47337 Genomic n
2	c	54.8	5.0	6118	6	ABL33031	Abl33031 Human imm
3		54.4	5.0	461	5	ADL43834	Adl43834 Human ova
4	c	52.8	4.8	6759	6	ABL32738	Abl32738 Human imm
5	c	52.8	4.8	198522	11	ACN44010	Acn44010 Human gen
6		52.4	4.8	6191	6	ABL33216	Abl33216 Human imm
7		52.4	4.8	6191	6	ABK31306	Abk31306 Signal tr
8		52.4	4.8	6191	6	ABL70281	Abl70281 Chemicall
9		52.4	4.8	6191	6	ABN80160	Abn80160 Human che
10	c	52.2	4.8	499	13	ACN55408	Acn55408 Cotton an
11	c	51.8	4.8	8056	8	ABZ10100	Abz10100 Haematopo
12		51.4	4.7	15373	6	ABL32467	Abl32467 Human imm
13	c	51.2	4.7	12393	6	ABL33263	Abl33263 Human imm
14		51	4.7	8056	8	ABZ10246	Abz10246 Haematopo
15		50.8	4.7	8056	8	ABZ10100	Abz10100 Haematopo
16	c	50.6	4.6	8222	8	ACF62816	Acf62816 Colon can
17	c	50.6	4.6	8222	8	ACF62794	Acf62794 Colon can
18	c	50.6	4.6	11222	10	ADB54190	Adb54190 Pretreate
19	c	50.6	4.6	11222	10	ADB54318	Adb54318 Pretreate
20	c	50.4	4.6	436	5	ADI72306	Adi72306 Human ova
21	c	50.4	4.6	436	5	ADL37450	Adl37450 Human ova
22	c	50.4	4.6	34688	6	ABQ67060	Abq67060 Human ang
23	c	50.2	4.6	8056	8	ABZ10246	Abz10246 Haematopo
24		50.2	4.6	50000	6	ABL55644	Abl55644 AmEPV gen
25	c	50	4.6	18988	4	AAS46342	Aas46342 Tumour su
26	c	50	4.6	18988	6	ABL32701	Abl32701 Human imm
27	c	50	4.6	18988	6	ABL34509	Abl34509 Human met
28	c	50	4.6	18988	6	ABL70204	Abl70204 Chemicall
29	c	50	4.6	18988	7	ADS99770	Ads99770 Complemen
30	c	49.6	4.6	5195	6	ABL33597	Abl33597 Human imm
31	c	49.6	4.6	6887	6	ABN80014	Abn80014 Human che
32		49.6	4.6	8169	4	AAS46288	Aas46288 Tumour su
33		49.6	4.6	8169	13	ADS89254	Ads89254 Oligonucl
34		49.6	4.6	8169	13	ADS89528	Ads89528 Oligonucl
35	c	49.6	4.6	20486	6	ABL34611	Abl34611 Human met
36	c	49.6	4.6	20486	7	ADS99872	Ads99872 Complemen
37	c	49.6	4.6	110000	6	ABX08336_14	Continuation (15 o
38	c	49.6	4.6	110000	12	ADJ25985_14	Continuation (15 o
39	c	49.6	4.6	110000	12	ADN97989_14	Continuation (15 o
40	c	49.6	4.6	110000	12	ADO50281_14	Continuation (15 o
41	c	49.4	4.5	2000	6	ABZ16669	Abz16669 Arabidops
42		49.4	4.5	134499	3	AAF22286	Aaf22286 BAC conta
43		49.2	4.5	6102	6	ABL33499	Abl33499 Human imm
44	c	49.2	4.5	6244	6	ABL32484	Abl32484 Human imm
45	c	49.2	4.5	34688	6	ABQ67059	Abq67059 Human ang

ALIGNMENTS

RESULT 1
ABK47337
ID ABK47337 standard; DNA; 20300 BP.
XX
AC ABK47337;
XX
DT 18-JUN-2002 (first entry)
XX
DE Genomic nucleotide sequence encoding human RGS-4 protein.
XX
KW RGS-4; schizophrenia; human; regulator of G protein signalling 4;
KW neuroleptic; polymorphism; schizotypy; schizoid; chromosome 1q21-22;
KW gene; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT variation replace(4121,T)
FT /*tag= a
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace(4123,A)
FT /*tag= b
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace(4368,C)
FT /*tag= c
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace(4621,C)
FT /*tag= d
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace(4790,T)
FT /*tag= e
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace(4816,T)
FT /*tag= f
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace(4970,T)
FT /*tag= g
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace(5055,G)
FT /*tag= h
FT /standard_name= "Single-nucleotide polymorphism"
FT variation replace(5295,A)
FT /*tag= i

```
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(5695,A)
FT                   /*tag= j
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(7375,T)
FT                   /*tag= k
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(7759,A)
FT                   /*tag= l
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(8596,A)
FT                   /*tag= m
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(9602. .9610,CA)
FT                   /*tag= n
FT                   /note= "Deletion polymorphism"
FT   variation       replace(9892,A)
FT                   /*tag= o
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(9963,A)
FT                   /*tag= p
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(10132,A)
FT                   /*tag= q
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(11056,C)
FT                   /*tag= r
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(11091,T)
FT                   /*tag= s
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(11106,A)
FT                   /*tag= t
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(11774,T)
FT                   /*tag= u
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(12143,A)
FT                   /*tag= v
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(12145,T)
FT                   /*tag= w
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(14367,G)
FT                   /*tag= x
FT                   /standard_name= "Single-nucleotide polymorphism"
FT   variation       replace(17027. .17029,GA)
FT                   /*tag= y
FT                   /note= "Deletion polymorphism"
FT   variation       replace(17630,T)
FT                   /*tag= Z
FT                   /standard_name= "Single-nucleotide polymorphism"
XX
PN   WO200216653-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US026622.
XX
PR   24-AUG-2000; 2000US-0228021P.
XX
PA   (UYPI-) UNIV PITTSBURGH.
XX
PI   Levitt PR,  Mirnics K,  Kodavali VC,  Nimgaonkar VL;
XX
DR   WPI; 2002-292070/33.
XX
PT   Diagnosing, assessing susceptibility and treating schizophrenia, involves
PT   observing regulator of G-protein signaling 4, RGS4 levels in a subject.
XX
PS   Claim 1; Page 20-33; 112pp; English.
XX
CC   This invention relates to a novel method for diagnosing schizophrenia or
CC   determining susceptibility to schizophrenia in a human. The method
CC   comprises obtaining from a patient a DNA sample and detecting variations
CC   in the regulator of G-protein signalling 4 (RGS4) gene. Alternatively,
CC   the method involves measuring RGS4 mRNA or protein levels in a tissue
CC   sample from the patient and determining if there is a reduced level. The
CC   method of the invention is useful for diagnosing and determining
CC   susceptibility to schizophrenia. The invention also comprises a method
CC   that is useful for treating schizophrenia which includes a prophylactic
CC   treatment. The method of genotyping polymorphic variants in the RGS-4
CC   gene is applied to diagnosing pathologies of the schizophrenic spectrum,
CC   such as in particular schizotypy, schizoid individuals, etc. This method
CC   offers the possibility of diagnosing schizophrenia by a biological test
CC   and no longer exclusively by clinical evaluations. The present sequence
CC   represents the genomic DNA encoding the human regulator of G-protein
CC   signalling 4 (RGS4) protein used in the method of the invention. The gene
CC   for the RGS4 protein is located on human chromosome 1q21-22
XX
SQ   Sequence 20300 BP; 6157 A; 4102 C; 3775 G; 6266 T; 0 U; 0 Other;

  Query Match           100.0%;  Score 1090;  DB 6;  Length 20300;
  Best Local Similarity  100.0%;  Pred. No. 9e-236;
  Matches 1090;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TGCTTTCCGTGTGGTGGTGCAGTAACTTTCACTCTCTATGTTCTTATAAGCAAATGTTAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11056 TGCTTTCCGTGTGGTGGTGCAGTAACTTTCACTCTCTATGTTCTTATAAGCAAATGTTAC 11115

Qy         61 AATGAGATATGAGTTTTAAAGCCAGATCTTCCTTATCTCTCTGCCCCATCTCTAGTTCTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11116 AATGAGATATGAGTTTTAAAGCCAGATCTTCCTTATCTCTCTGCCCCATCTCTAGTTCTT 11175

Qy        121 GAAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11176 GAAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTT 11235

Qy        181 TGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11236 TGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAA 11295

Qy        241 AGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11296 AGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGA 11355

Qy        301 AATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11356 AATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCA 11415

Qy        361 TATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11416 TATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTG 11475

Qy        421 CTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11476 CTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGA 11535

Qy        481 AAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGTCATTCAGGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11536 AAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGTCATTCAGGA 11595

Qy        541 GTATGGAAGTATCGATGGCTTCTTTAAAATGAAGCAGGAGGGTCTGGCAGAGAGTATCTA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11596 GTATGGAAGTATCGATGGCTTCTTTAAAATGAAGCAGGAGGGTCTGGCAGAGAGTATCTA 11655

Qy        601 TGAAATAAGTTCCTCTGACCTTCACGCTTAATTTTCTGAATGGAGTGGAGCAAATTACTT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11656 TGAAATAAGTTCCTCTGACCTTCACGCTTAATTTTCTGAATGGAGTGGAGCAAATTACTT 11715

Qy        661 CAAGCTTCACTTAACTTGCATATGAAATGAACCGTACAAAAATACAAGAGTGTCAGGAGA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11716 CAAGCTTCACTTAACTTGCATATGAAATGAACCGTACAAAAATACAAGAGTGTCAGGAGA 11775

Qy        721 AAGTTATGCTCTGGTAAATATTTTGCAAAACAGATAAAAGATAATACTAGAGCTCTGTCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      11776 AAGTTATGCTCTGGTAAATATTTTGCAAAACAGATAAAAGATAATACTAGAGCTCTGTCC 11835

Qy        781 TCAAAGAGTTAAGCAGCTAATCTAAGGAGGTAAACTCTATGTCAGCAGGATGAACTGCTC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11836 TCAAAGAGTTAAGCAGCTAATCTAAGGAGGTAAACTCTATGTCAGCAGGATGAACTGCTC 11895

Qy        841 TTCCCTTTCCTCCTCAATAAATTGCAAATCATCTAGTCCAACATCTTTACCACCAGTGCC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11896 TTCCCTTTCCTCCTCAATAAATTGCAAATCATCTAGTCCAACATCTTTACCACCAGTGCC 11955

Qy        901 TGAGGCTCCAGAGGAGCCATTGCCTTCTCAAGGTCACATAGGTGGTGGGTGAGTTAGGAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11956 TGAGGCTCCAGAGGAGCCATTGCCTTCTCAAGGTCACATAGGTGGTGGGTGAGTTAGGAC 12015

Qy        961 CAAATCTAGAATTCCTGACTCCAGTAACTTCTGAAGTCATTTTGTTTTTTATTTTTATGG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12016 CAAATCTAGAATTCCTGACTCCAGTAACTTCTGAAGTCATTTTGTTTTTTATTTTTATGG 12075

Qy       1021 TTTTATTATAAGAATACTTGCTAAGCACACTTACCCCCTGCATTGATTAATAACTCTAGG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12076 TTTTATTATAAGAATACTTGCTAAGCACACTTACCCCCTGCATTGATTAATAACTCTAGG 12135

Qy       1081 ATCTCAGGTG 1090
              ||||||||||
Db      12136 ATCTCAGGTG 12145
```

RESULT 2
ABL33031/c
ID ABL33031 standard; DNA; 6118 BP.
XX
AC ABL33031;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 1004.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 1004; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 6118 BP; 1751 A; 141 C; 1229 G; 2997 T; 0 U; 0 Other;

```
  Query Match             5.0%;  Score 54.8;  DB 6;  Length 6118;
  Best Local Similarity   48.4%;  Pred. No. 0.027;
  Matches  210;  Conservative    0;  Mismatches  222;  Indels    2;  Gaps    2;

Qy          94 TATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATAT 153
               ||   || |  ||    || | || || || |  |   |||    ||  |  |  || |
Db        2830 TAAACCTATAACCTTAATCAAATTATTTAAATTCCCTTTATCTCCTTTATAAACTAAAAA 2771

Qy         154 TGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTG 213
               | || || || |  |    |||  | ||  | |  | || |  | | | |||     |
Db        2770 TAATAATAAC-ACCCTCCAAATTATATTAAAAATTACCTAAATTAATATACACAAAATCC 2712

Qy         214 TTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTT 273
               |||  ||    ||||  | ||||  ||| ||  || | | |     ||| |  |   |
Db        2711 TTATAATTACACCTTTTTATTTTATAAACACTATATAAATACTTCAATATACTATATATA 2652

Qy         274 TTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAG 333
               |        | | |      ||| |   |  | | |   | | |||| | | ||
Db        2651 TATAAAAATCTATACCCAAAATATATTTAAAAAATAAATAATAAATCAAAAAATATTTTA 2592

Qy         334 TTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAA 393
               ||   | || |   || |  | |  | |||   |||  | || ||   | ||    |
Db        2591 TTATTACTTCATACAAATAACATTTTAAATTCTTTTAAATTATTTTCTCTTTAATTCAAT 2532

Qy         394 AAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCAC 453
               ||    |  ||||  ||||||  | |    ||||  |    ||||| |||||   |
Db        2531 AACAATATTATAAACAAATACTAAATCCC-AAATATCCTAAAATAAATAAAAATTTTAAA 2473

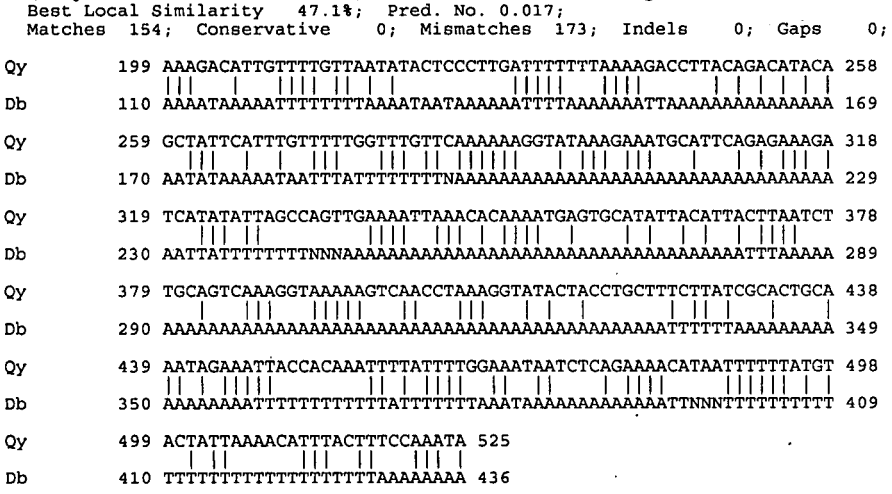
Qy         454 AAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTT 513
               ||||  |||  |||||||| ||| |||| | ||| |  ||   | |  || |||
Db        2472 CAATTAATTTTCATAAATAATATCAAAAAATAAAATATCATAAACATTTCTATAACTCCA 2413

Qy         514 ACTTTCCAAATATT 527
               | |   | ||||||
Db        2412 ATTACACTAATATT 2399
```

RESULT 3
ADL43834
ID ADL43834 standard; DNA; 461 BP.
XX
AC ADL43834;
XX
DT 20-MAY-2004 (first entry)
XX
DE Human ovarian cancer DNA marker #17724.
XX
KW Human; ovarian cancer; ds; tumour; cytostatic; DNA marker.
XX
OS Homo sapiens.
XX
PN WO200170979-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US009126.
XX
PR 21-MAR-2000; 2000US-0191031P.
PR 25-MAY-2000; 2000US-0207124P.
PR 15-JUN-2000; 2000US-0211940P.
PR 07-JUL-2000; 2000US-0216820P.
PR 25-JUL-2000; 2000US-0220661P.
PR 21-DEC-2000; 2000US-0257672P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX

PI Lee J, Lillie J;
XX
DR WPI; 2001-611502/70.
XX
PT Novel isolated nucleic acid molecules (markers) overexpressed in ovarian
PT cancer cells as compared to their normal non-cancerous ovarian cells are
PT used to characterize stage, grade, histological type of ovarian cancer.
XX
PS Disclosure; SEQ ID NO 17724; 106pp; English.
XX
CC The invention relates to nucleic acid markers which are overexpressed in
CC ovarian cancer cells as compared to their expression in normal (i.e. non-
CC cancerous) ovarian cells. The invention also relates to polypeptides
CC encoded by the markers, antibodies that selectively bind to the
CC polypeptides, a method of inhibiting ovarian cancer in a patient at risk
CC of developing ovarian cancer involving inhibiting expression of a gene
CC corresponding to a marker of the invention and a method of treating a
CC patient afflicted with ovarian cancer comprising providing to cells of
CC the patient an antisense oligonucleotide complementary to a marker of the
CC invention. The markers are useful for assessing if a patient is afflicted
CC with ovarian cancer, which involves comparing the level of expression of
CC a marker in a patient sample and a normal level of expression of the
CC marker in a control non-ovarian cancer sample. A difference between the
CC expression levels indicates ovarian cancer. The level of expression of a
CC marker corresponds to a secreted protein or to a transcribed
CC polynucleotide or its portion. The level of expression of the marker is
CC assessed by detecting the presence in the sample, a protein or protein
CC fragment corresponding to the marker. The presence of protein or protein
CC fragment is detected using an antibody that specifically binds with the
CC protein or protein fragment. Alternatively, the level of expression of
CC the marker is assessed by detecting the presence of a transcribed
CC polynucleotide which anneals with the marker or anneals with a portion of
CC the polynucleotide comprising the marker, under stringent conditions. The
CC marker is also used for monitoring the progression of ovarian cancer in a
CC patient which involves detecting expression of the marker in a patient
CC sample at a first point in time, repeating the method at a subsequent
CC time and comparing the level of expression. The method is carried out
CC using an ovarian tissue sample. A composition comprising a marker,
CC polypeptide or antibody of the invention is used to treat ovarian cancer.
CC This sequence represents a human ovarian cancer DNA marker of the
CC invention.
XX
SQ Sequence 461 BP; 263 A; 28 C; 25 G; 138 T; 0 U; 7 Other;

Query Match 5.0%; Score 54.4; DB 5; Length 461;
Best Local Similarity 47.1%; Pred. No. 0.017;
Matches 154; Conservative 0; Mismatches 173; Indels 0; Gaps 0;



RESULT 4
ABL32738/c
ID ABL32738 standard; DNA; 6759 BP.
XX
AC ABL32738;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 711.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 711; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 6759 BP; 2159 A; 60 C; 1367 G; 3173 T; 0 U; 0 Other;

Query Match 4.8%; Score 52.8; DB 6; Length 6759;
Best Local Similarity 46.7%; Pred. No. 0.077;
Matches 168; Conservative 0; Mismatches 192; Indels 0; Gaps 0;

```
Qy        397 GTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAA 456
              | | |   | | ||| |   | | |||   |  || |  || |||| |
Db       2626 AAACAACATACAAAAAAATACTTATAATATAATCCTTATTTCAAAAATAAATACTAATTT 2567

Qy        457 TTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACT 516
              ||  |  |   |  ||| | |  | | | |   |  | ||||  || ||  | || |
Db       2566 TTAAAAATCTATATAAATATAAACACAAACACACATAAACATACTCATATAATCTATAAT 2507


RESULT 5
ACN44010/c
ID   ACN44010 standard; DNA; 198522 BP.
XX
AC   ACN44010;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Human genomic sequence hCG1643869.
XX
KW   Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO2003073826-A2.
XX
PD   12-SEP-2003.
XX
PF   28-FEB-2003; 2003WO-US006235.
XX
PR   01-MAR-2002; 2002US-00087192.
XX
PA   (SAGR-) SAGRES DISCOVERY.
XX
PI   Morris DW;
XX
DR   WPI; 2003-328604/31.
XX
PT   Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
PT   comprises a nucleotide sequence.
XX
PS   Claim 1; SEQ ID NO 244; 0pp; English.
XX
CC   The present invention relates to novel DNA and protein sequences which
CC   are associated with carcinomas. The sequences are useful for: (i) for
CC   screening drug candidates; (ii) for screening of bioactive agent capable
CC   of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
CC   a bioactive agent capable of modulating the activity of CAP; (iv) for
CC   evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
CC   carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
CC   carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
CC   (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
CC   determining Carcinoma Associated (CA) gene copy number. In addition, the
CC   CA genes are useful as DNA vaccines and the CAP are useful as markers of
CC   carcinoma including lymphoma. The present sequence is one such CA coding
CC   sequence. Note: This patent is an equivalent to basic patent
CC   US2002182586A1, for which no sequence data was published
XX
SQ   Sequence 198522 BP; 58663 A; 35063 C; 36453 G; 68323 T; 0 U; 20 Other;

  Query Match             4.8%;  Score 52.8;  DB 11;  Length 198522;
  Best Local Similarity   46.9%;  Pred. No. 0.18;
  Matches  238;  Conservative    0;  Mismatches  262;  Indels    8;  Gaps    2;

Qy         33 TCTCTATGTTCTTATAAGCAAATGTTACAATGAGATATGAGTTTTAAAGCCAGATCTTCC 92
              | | |||  | | |  |  |||   || | | | |||| |  | |  |   | |   |
Db      84632 TATATATAATGTAAATATAAAAATATATATTTATATATAATATATTTATATATAATATAT 84573

Qy         93 TTATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATA 152
              |||| |          || | ||  | |  |    | |   |  | |||  |  |   |||
Db      84572 TTATATTATATTTATATTTATAAATTTATATAAATTTATAAATAATTTTATAAAATTATA 84513

Qy        153 TTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTT 212
                |  | ||| |||| | || |||    | || |  |  | |   |||| | | |  | |

Db      84512 ATTTTATTTATAAATAAATTTATATAAATTTATATAAATTTA---TAAATATAATTATAT 84456

Qy        213 GTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTT 272
              | |||||| | | |   | | | | ||  |   ||   | ||| |   , | || |  |
Db      84455 ATAAATATATTGCATGTTTATGTATTAATAAATATATTAATATATAATATATTATATACT 84396

Qy        273 TTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCA 332
              |       | |||||  ||| |   | |  ||  |     |  | ||||||| |
Db      84395 ATATAAAATATATAAAAATATATTATATACTATATAAAATATATAAAAATATATTATATA 84336

Qy        333 GTTGAAAA-----TTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAA 387
              |  | ||      |||| | | | |   | |    || | ||  |||   |  | | |
Db      84335 CTATATAAAATATATAAAAATATATTATATACTATATAAAATATATAAAAATATATTATA 84276

Qy        388 AGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAAT 447
              || | |     |  |||| |||| ||  | || | | |       |  ||||| |
Db      84275 TACTATATAAAATATATAAAAATATATTATATACTATATAAAATATATAAAAATATATTA 84216

Qy        448 TACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAA 507
              ||    | ||   || |   |||  || | | | |  |||  ||  ||| |  ||| |
Db      84215 TATACTATATAAAATATATAAAAATATATAAAACATTATATATTAATATATTATATATAG 84156

Qy        508 ACATTTACTTTCCAAATATTCTGTCATT 535
              || |||| |  | |   | | |  ||
Db      84155 TTATATACTATATATAGTATATATATTT 84128


RESULT 6
ABL33216
ID   ABL33216 standard; DNA; 6191 BP.
XX
AC   ABL33216;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 1189.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP007537.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful for
PT   diagnosis and treatment of diseases associated with abnormal cytosine
PT   methylation.
XX
PS   Claim 1; SEQ ID NO 1189; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
```

```
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention
XX
SQ   Sequence 6191 BP; 2084 A; 36 C; 1162 G; 2909 T; 0 U; 0 Other;

  Query Match             4.8%;  Score 52.4;  DB 6;  Length 6191;
  Best Local Similarity   47.6%;  Pred. No. 0.093;
  Matches  219;  Conservative    0;  Mismatches  236;  Indels    5;  Gaps    2;

Qy        108 ATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAAT 167
              || | || ||  | |  | | |   || | ||| |||       |  | | ||   ||
Db       2650 ATTTGTAATTTGTAATTTTTTTTTAATTAATTTTGTTTTTTTTATGTTTTTAATAGTTAT 2709

Qy        168 CAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCT 227
              ||  |   | | | | |  | |     |    ||  || || | || | | | |   |
Db       2710 TTTTTTTTTGTGTATTGTAATTTTTTGTTGGTTGATTTTTTTGTTTTTAAAAATTTATT 2769

Qy        228 TGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAA 287
              |  ||||| |     |   |  | | |      ||     || |||||| ||| |||
Db       2770 TTTTTTTTATTTTTTATTGTTAATATTTTGTTATAAAGGATTTTTTTTGTTTTATTCGTT 2829

Qy        288 AAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACA 347
              || |  | ||    |  |  |||  || |  ||| ||| |  | |||| ||| ||
Db       2830 TTAGTTTAATAGTGTAGTTTGTAGATGAATAATATAGATTGG--ATTTGAGAATAAATAG 2887

Qy        348 CAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAA 407
              ||||    | ||| || ||   | ||      |   ||  ||  || || ||    ||
Db       2888 GTAAATATAGGTATAATAGATAGGTAAA---AAAAAAAAATTGTTGAATGTAAAAAGGAA 2944

Qy        408 GGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGG 467
              | |||     |   || |||  | | || |||||    ||     ||||||| |||
Db       2945 TGAATAAATGGTAAGTTGTTAAGGAAGTGTAAATATTTTTTTAGGAAAATTTTTTTTGAC 3004

Qy        468 AAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATT 527
              |||  | | |  | ||  | |||| |    ||||   |  |||  ||   |  | ||
Db       3005 GTTTAAAGTAAAATTAGATGGTGTTTTTTTGTGTATTTGTAATTTTTTTTGTAATTTTTT 3064

Qy        528 CTGTCATTCAGGAGTATGGAAGTATCGATGGCTTCTTTAA 567
              |     | | |||||   | |   | | | || ||| |
Db       3065 TTAATTGTTATGTGTATTTTAATTATGGTAGTTTTTTTTA 3104


RESULT 7
ABK31306
ID   ABK31306 standard; DNA; 6191 BP.
XX
AC   ABK31306;
XX
DT   23-APR-2002  (first entry)
XX
DE   Signal transduction associated gene modified DNA #75.
XX
KW   Human; signal transduction associated gene; cytosine methylation state;
KW   CpG island; signal transduction associated disease; solid tumour; cancer;
KW   antitumour; cytostatic; mutant; ds.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO200200926-A2.
XX
PD   03-JAN-2002.
XX
PF   29-JUN-2001; 2001WO-EP007472.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-147896/19.
XX
PT   Oligonucleotide for diagnosis and therapy of diseases associated with
PT   signal transduction e.g. cancer, comprises chemically modified genomic
PT   sequences of genes associated with signal transduction.
XX
PS   Claim 1; SEQ ID NO 149; 24pp; English.
XX
CC   The present invention relates to chemically modified DNA sequences of
CC   signal transduction associated genes. The DNA sequences are chemically
CC   modified using a solution of bisulphite, hydrogen sulphite or disulphite.
CC   Also disclosed are oligonucleotides and/or PNA oligomers for detecting
CC   the cytosine methylation state (CpG islands) of these genes, and a method
CC   for the diagnosis and/or therapy of genetic and epigenetic parameters of
CC   genes associated with signal transduction. The genomic DNA can be
CC   obtained from cells or cellular components which contain DNA, e.g. cell
CC   lines, biopsies, blood, sputum, stool, urine, cerebral-spinal fluid,
CC   tissue embedded in paraffin such as tissue from eyes, intestine, kidney,
CC   brain, heart, prostate, lung, breast or liver, histologic object slides,
CC   and all their possible combinations. The sequences of the invention are
CC   useful for the diagnosis and therapy of diseases associated with signal
CC   transduction e.g. solid tumours and cancer. ABK31158-ABK31545 represent
CC   chemically pretreated genomic DNA sequences of different genes associated
CC   with signal transduction, or their complementary sequences. Note: The
CC   sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from the
CC   European Patent Office
XX
SQ   Sequence 6191 BP; 2084 A; 36 C; 1162 G; 2909 T; 0 U; 0 Other;

  Query Match             4.8%;  Score 52.4;  DB 6;  Length 6191;
  Best Local Similarity   47.6%;  Pred. No. 0.093;
  Matches  219;  Conservative    0;  Mismatches  236;  Indels    5;  Gaps    2;

Qy        108 ATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAAT 167
              || | || ||  | |  | | |   || | ||| |||       |  | | ||   ||
Db       2650 ATTTGTAATTTGTAATTTTTTTTTAATTAATTTTGTTTTTTTTATGTTTTTAATAGTTAT 2709

Qy        168 CAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCT 227
              ||  |   | | | | |  | |     |    ||  || || | || | | | |   |
Db       2710 TTTTTTTTTGTGTATTGTAATTTTTTGTTGGTTGATTTTTTTGTTTTTAAAAATTTATT 2769

Qy        228 TGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAA 287
              |  ||||| |     |   |  | | |      ||     || |||||| ||| |||
Db       2770 TTTTTTTTATTTTTTATTGTTAATATTTTGTTATAAAGGATTTTTTTTGTTTTATTCGTT 2829

Qy        288 AAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACA 347
              || |  | ||    |  |  |||  || |  ||| ||| |  | |||| ||| ||
Db       2830 TTAGTTTAATAGTGTAGTTTGTAGATGAATAATATAGATTGG--ATTTGAGAATAAATAG 2887

Qy        348 CAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAA 407
              ||||    | ||| || ||   | ||      |   ||  ||  || || ||    ||
Db       2888 GTAAATATAGGTATAATAGATAGGTAAA---AAAAAAAAATTGTTGAATGTAAAAAGGAA 2944

Qy        408 GGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGG 467
              | |||     |   || |||  | | || |||||    ||     ||||||| |||
Db       2945 TGAATAAATGGTAAGTTGTTAAGGAAGTGTAAATATTTTTTTAGGAAAATTTTTTTTGAC 3004

Qy        468 AAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATT 527
              |||  | | |  | ||  | |||| |    ||||   |  |||  ||   |  | ||
Db       3005 GTTTAAAGTAAAATTAGATGGTGTTTTTTTGTGTATTTGTAATTTTTTTTGTAATTTTTT 3064

Qy        528 CTGTCATTCAGGAGTATGGAAGTATCGATGGCTTCTTTAA 567
              |     | | |||||   | |   | | | || ||| |
Db       3065 TTAATTGTTATGTGTATTTTAATTATGGTAGTTTTTTTTA 3104


RESULT 8
ABL70281
ID   ABL70281 standard; DNA; 6191 BP.
XX
AC   ABL70281;
```

XX
DT 01-JUL-2002 (first entry)
XX
DE Chemically treated cell signalling DNA sequence#86.
XX
KW Cell signalling; cytosine methylation; cell signalling disease; cancer;
KW tumour; cytostatic; ds.
XX
OS Unidentified.
XX
PN WO200202807-A2.
XX
PD 10-JAN-2002.
XX
PF 29-JUN-2001; 2001WO-EP007471.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-154758/20.
XX
PT Nucleic acid, useful for diagnosis and therapy of diseases associated
PT with cell signaling e.g. cancer, comprises chemically modified genomic
PT sequences of genes associated with cell signaling.
XX
PS Claim 1; SEQ ID NO 171; 24pp + Sequence Listing; English.
XX
CC The invention relates to a nucleic acid comprising a sequence of at least
CC 18 bases of a segment of chemically pretreated DNA of genes associated
CC with cell signalling. The activity of the modified sequences of the
CC invention may be described as cytostatic. The object of the invention is
CC to provide the chemically modified DNA of genes associated with cell
CC signalling, as well as oligonucleotides and/or PNA-oligomers for
CC detecting cytosine methylations, as well as a method which is
CC particularly suitable for the diagnosis and/or therapy of genetic and
CC epigenetic parameters of genes associated with cell signalling. The
CC chemically modified DNA provided by the invention is useful for diagnosis
CC and therapy of diseases such as solid tumours and cancer. The sequences
CC given in records ABL70111-ABL70626 represent chemically pre-treated
CC genomic DNA's of genes associated with cell signalling. Note: The
CC sequence data for this patent is not represented in the printed
CC specification, but is based on sequence information supplied by the
CC European Patent Office
XX
SQ Sequence 6191 BP; 2084 A; 36 C; 1162 G; 2909 T; 0 U; 0 Other;

Query Match 4.8%; Score 52.4; DB 6; Length 6191;
Best Local Similarity 47.6%; Pred. No. 0.093;
Matches 219; Conservative 0; Mismatches 236; Indels 5; Gaps 2;

```
Qy       108 ATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAAT 167
Db      2650 ATTTGTAATTTGTAATTTTTTTTTAATTAATTTTGTTTTTTTTATGTTTTTAATAGTTAT 2709

Qy       168 CAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCT 227
Db      2710 TTTTTTTTTGTGTATTGTAATTTTTTGTTGGTTGATTTTTTTGTTTTTAAAAATTTATT 2769

Qy       228 TGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAA 287
Db      2770 TTTTTTTTATTTTTTATTGTTAATATTTTGTTATAAAGGATTTTTTTTGTTTTATTCGTT 2829

Qy       288 AAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACA 347
Db      2830 TTAGTTTAATAGTGTAGTTTGTAGATGAATAATATAGATTGG--ATTTGAGAATAAATAG 2887

Qy       348 CAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAA 407
Db      2888 GTAAATATAGGTATAATAGATAGGTAAA---AAAAAAAAATTGTTGAATGTAAAAAGGAA 2944

Qy       408 GGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGG 467
Db      2945 TGAATAAATGGTAAGTTGTTAAGGAAGTGTAAATATTTTTTTAGGAAAATTTTTTTTGAC 3004

Qy       468 AAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATT 527
Db      3005 GTTTAAAGTAAAATTAGATGGTGTTTTTTTGTGTATTTGTAATTTTTTTTGTAATTTTTT 3064

Qy       528 CTGTCATTCAGGAGTATGGAAGTATCGATGGCTTCTTTAA 567
Db      3065 TTAATTGTTATGTGTATTTTAATTATGGTAGTTTTTTTTA 3104
```

RESULT 9
ABN80160
ID ABN80160 standard; DNA; 6191 BP.
XX
AC ABN80160;
XX
DT 15-JUL-2002 (first entry)
XX
DE Human chemically modified disease associated gene SEQ ID NO 177.
XX
KW Human; development; homeobox gene; HOX; diabetes; cancer; apoptosis;
KW heart disease; epilepsy; histone deacetylation; muscular dystrophy;
KW dwarfism; single nucleotide polymorphism; SNP; cytosine methylation;
KW antidiabetic; cytostatic; anticonvulsant; ds.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200200927-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007536.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130908/17.
XX
PT Novel nucleic acid useful for diagnosis and therapy of diseases
PT associated with development genes such as diabetes, comprises a sequence
PT of a segment of chemically pretreated DNA of genes associated with
PT development.
XX
PS Claim 1; SEQ ID NO 177; 27pp; English.
XX
CC The invention relates to a nucleic acid (I) comprising a sequence at
CC least 18 bases in length of a segment of chemically pretreated DNA (II)
CC of genes associated with development selected from 87 genes listed in the
CC specification such as ACCPN, ADFN, or AFD1 and comprising one of 350
CC sequences (ABN79984-ABN80333) or their complements. The invention is
CC useful for the diagnosis or therapy of diseases associated with
CC development genes, in particular disease related to homeobox containing
CC genes (HOX), like diabetes, cancer, apoptosis related diseases, syndromes
CC associated with congenital heart disease, epilepsy, diseases related to
CC histone deacetylation, Currarino syndrome, diseases related with the
CC development of the brain and limb girdle muscular dystrophy and dwarfism.
CC Oligomers specific to each of the genes are useful for detecting the
CC methylation state of all CpG dinucleotides within the 350 sequences or
CC (II) and their complementary sequences, as primer oligonucleotides for
CC the amplification of the 350 sequences, (II) and/or their complements and
CC as oligomer probes for detecting the cytosine methylation state and/or
CC single nucleotide polymorphisms (SNPs). Note: The sequence data for this
CC patent did not form part of the printed specification but is based on
CC sequence information supplied to Derwent by the European Patent Office

```
XX
SQ   Sequence 6191 BP; 2084 A; 36 C; 1162 G; 2909 T; 0 U; 0 Other;

  Query Match             4.8%;  Score 52.4;  DB 6;  Length 6191;
  Best Local Similarity   47.6%;  Pred. No. 0.093;
  Matches  219;  Conservative    0;  Mismatches  236;  Indels    5;  Gaps    2;

Qy        108 ATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAAT 167
              || | || ||  | |  | | |   || | ||| |||          |  | | ||   ||
Db       2650 ATTTGTAATTTGTAATTTTTTTTTAATTAATTTTGTTTTTTTTATGTTTTAATAGTTAT 2709

Qy        168 CAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCT 227
                ||  |   | | | | | | |    |    ||  || || | || | | | |   |
Db       2710 TTTTTTTTTTGTGTATTGTAATTTTTTGTTGGTTGATTTTTTTGTTTTTAAAAATTTATT 2769

Qy        228 TGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAA 287
              |  ||||| |      |    | | |      ||     || |||||| ||| |||
Db       2770 TTTTTTTTATTTTTTATTGTTAATATTTTGTTATAAAGGATTTTTTTTGTTTTATTCGTT 2829

Qy        288 AAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACA 347
              || |  | ||    |  |  |||  || |  ||| ||| |  | |||| ||| ||
Db       2830 TTAGTTTAATAGTGTAGTTTGTAGATGAATAATATAGATTGG--ATTTGAGAATAAATAG 2887

Qy        348 CAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAA 407
              ||||    | ||| || ||   | ||      |   ||  ||  || || ||    ||
Db       2888 GTAAATATAGGTATAATAGATAGGTAAA---AAAAAAAAATTGTTGAATGTAAAAAGGAA 2944

Qy        408 GGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGG 467
              | |||     |   || |||  | | || |||||    ||     ||||||| |||
Db       2945 TGAATAAATGGTAAGTTGTTAAGGAAGTGTAAATATTTTTTTAGGAAAATTTTTTTTGAC 3004

Qy        468 AAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATT 527
              |||  | | |  | ||  | |||| |    ||||   |  |||  ||   |  | ||
Db       3005 GTTTAAAGTAAAATTAGATGGTGTTTTTTTGTGTATTTGTAATTTTTTTTGTAATTTTTT 3064

Qy        528 CTGTCATTCAGGAGTATGGAAGTATCGATGGCTTCTTTAA 567
              |    | | | ||||   | |   | | | || ||| |
Db       3065 TTAATTGTTATGTGTATTTTAATTATGGTAGTTTTTTTTA 3104


RESULT 10
ACN55408/c
ID   ACN55408 standard; cDNA; 499 BP.
XX
AC   ACN55408;
XX
DT   02-DEC-2004  (first entry)
XX
DE   Cotton androecium tissue EST Clone ID: LIB3828-023-Q6-N6-C10, SEQ:10189.
XX
KW   Cotton; plant; EST; expressed sequence tag; transgenic plant; androecium;
KW   variety Nucotton33B; library LIB3828; molecular tag; molecular marker;
KW   genetic mapping; molecular mapping; seed germination; plant growth;
KW   plant quality; plant yield; plant breeding; tissue printing; ss.
XX
OS   Gossypium hirsutum.
XX
PN   US2004123340-A1.
XX
PD   24-JUN-2004.
XX
PF   12-DEC-2001; 2001US-00021323.
XX
PR   14-DEC-2000; 2000US-0255619P.
XX
PA   (DEIK/) DEIKMAN J.
PA   (FENG/) FENG P C C.
PA   (FINC/) FINCHER K L.
PA   (ZIEG/) ZIEGLER T E.
XX
PI   Deikman J,  Feng PCC,  Fincher KL,  Ziegler TE;
XX
DR   WPI; 2004-479808/45.
XX
PT   New isolated nucleic acid molecule that encodes a plant protein or its
PT   fragment, useful for isolating a variety of agronomically significant
PT   genes associated with plant growth, quality or yield, and as molecular
PT   tags to map genes.
XX
PS   Claim 1; SEQ ID NO 10189; 34pp; English.
XX
CC   The invention relates to 17880 cotton expressed sequence tags (ESTs;
CC   ACN45220-ACN63099). The ESTs were isolated from cDNA libraries generated
CC   from primed or non-primed seeds from variety DP50B, mature seeds from
CC   variety Coker 312 Boswell 96 Field, and androecium tissue, gynoecium
CC   tissue, developing fibres, carpel walls and septa from variety
CC   Nucotton33B. The invention also relates to substantially purified
CC   proteins or their fragments encoded by nucleic acid molecules of the
CC   invention, and to transformed plants having a nucleic acid construct
CC   comprising a nucleic acid of the invention. The cotton ESTs are useful as
CC   molecular tags to isolate genetic regions, to isolate genes, to map
CC   genes, to determine gene function and to determining whether genes are
CC   members of a particular gene family. The nucleic acid molecules may be
CC   used for isolating a variety of agronomically significant genes
CC   associated with plant growth, quality, yield, and could also serve as
CC   links in metabolic and catabolic pathways. The nucleic acid molecules are
CC   also useful for identifying genes important in initiating and maintaining
CC   seed germination or that may be used to mitigate stresses encountered
CC   during seed germination. The ESTs additionally enable the acquisition of
CC   promoters and cis-regulatory elements which will be useful to express
CC   agronomically significant genes in these tissues and/or other tissues,
CC   and also permits the acquisition of molecular markers useful in breeding
CC   schemes, genetic and molecular mapping, and in cloning of agronomically
CC   significant genes. The nucleic acid molecules are further useful for
CC   detecting the expression level or pattern of a protein or mRNA and for
CC   detecting the presence or quantity of a protein by tissue printing. The
CC   present sequence represents a specifically claimed EST isolated from a
CC   cotton variety Nucotton33B androecium tissue cDNA library (LIB3828). The
CC   sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from the US
CC   patent office at seqdata.uspto.gov/sequence.html?DocID=US20040123340
XX
SQ   Sequence 499 BP; 245 A; 2 C; 49 G; 203 T; 0 U; 0 Other;

  Query Match             4.8%;  Score 52.2;  DB 13;  Length 499;
  Best Local Similarity   47.7%;  Pred. No. 0.055;
  Matches  153;  Conservative    0;  Mismatches  168;  Indels    0;  Gaps    0;

Qy         87 TCTTCCTTATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGA 146
              | ||  || | | | | |||  | | |     || |  | || | | |   |||  ||
Db        359 TTTTTTTTTTTTTTTTCCCCTTTTTTTTTAAATTTATTTTTCCCTTTTTTTTTTTTTTT 300

Qy        147 GAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACAT 206
                | ||  || ||      |   ||  |  |||| |       | |  | ||     | |
Db        299 TTTTTTTTTTCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCCCCCCTT 240

Qy        207 TGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCA 266
              | |||| ||  | |  |   ||  |||||||       ||| | | | | |  | ||
Db        239 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAAAAAAAAAAATTTTTTTT 180

Qy        267 TTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATAT 326
              ||| |||||  |||     ||||||   | ||| |||   |   | | | | |  ||| ||
Db        179 TTTTTTTTTTTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAT 120

Qy        327 TAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCA 386
                |   |    ||||  ||| | ||||  |    | |  | | |   ||      |  |
Db        119 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 60

Qy        387 AAGGTAAAAAGTCAACCTAAA 407
              ||   |||||  ||  |||
Db         59 AAAAAAAAAAAAAAAAAAAAA 39


RESULT 11
ABZ10100/c
ID   ABZ10100 standard; DNA; 8056 BP.
```

XX
AC ABZ10100;
XX
DT 16-JAN-2003 (first entry)
XX
DE Haematopoietic cell proliferation disorder related DNA sequence #240.
XX
KW Human; haematopoietic cell proliferation disorder; cytostatic;
KW gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;
KW cytosine methylation state; gene; ds.
XX
OS Homo sapiens.
XX
PN WO200277272-A2.
XX
PD 03-OCT-2002.
XX
PF 26-MAR-2002; 2002WO-EP003401.
XX
PR 26-MAR-2001; 2001US-0278333P.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J;
PI Olek A, Piepenbrock C, Adorjan P, Grabs G, Lesche R, Leu E;
PI Lewin A, Lipscher E, Maier S, Model F, Mueller V, Otto T, Pelet C;
PI Schwope I, Ziebarth H;
XX
DR WPI; 2003-018942/01.
XX
PT Detecting and differentiating between hematopoietic cell proliferative
PT disorders, comprises contacting a target nucleic acid with a reagent that
PT distinguishes between methylated and non-methylated CpG dinucleotides.
XX
PS Claim 28; SEQ ID NO 240; 117pp; English.
XX
CC The present invention describes a method for detecting and
CC differentiating between haematopoietic cell proliferative disorders
CC associated with at least 1 gene and/or their regulatory regions in a
CC subject. The method comprises contacting a target nucleic acid in a
CC biological sample obtained from the subject with at least 1 reagent,
CC which distinguishes between methylated and non-methylated CpG
CC dinucleotides within the target nucleic acid. ABZ09861 to ABZ11118
CC represent specifically claimed nucleotide sequences from the present
CC invention. Oligonucleotides from the present invention can be used: for
CC differentiating between healthy haematopoietic cells and proliferative
CC disorder haematopoietic cells; for differentiating between acute
CC lymphocytic leukaemia and acute myelogenous leukaemia; as probes for
CC determining the cytosine methylation state and/or single nucleotide
CC polymorphisms (SNPs) of haematopoietic cell proliferation disorder
CC related sequences and their complements; and as primers for the
CC amplification of haematopoietic cell proliferation disorder related DNA
CC sequences. The nucleotide sequences from the present invention can also
CC be used for detecting a predisposition to, differentiation between
CC subclasses, diagnosis, prognosis, treatment and/or monitoring of
CC haematopoietic cell proliferative disorders. The present method enables a
CC highly specific classification of haematopoietic cell proliferative
CC disorders allowing for improved and informed treatment of patients
XX
SQ Sequence 8056 BP; 3711 A; 371 C; 371 G; 3603 T; 0 U; 0 Other;

Query Match 4.8%; Score 51.8; DB 8; Length 8056;
Best Local Similarity 44.5%; Pred. No. 0.14;
Matches 206; Conservative 0; Mismatches 257; Indels 0; Gaps 0;

Qy 111 TCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAG 170
Db 473 TTTATTTTTTAAAATTAATTTTTTTTTATTATTTTTAAAAAAAATAAAATAATTAATAAA 414

Qy 171 TTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGA 230
Db 413 TATATTATTAAATTAAATTTTAAATTATAAATATTTTATTTCGTTTTTTTTTTTTTTAA 354

Qy 231 TTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAA 290
Db 353 TTATTATTATTTTTATTTATTTTTATTAATTATTTTTTTAAAAAATAATTTTAAAAAAAA 294

Qy 291 GGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAA 350
Db 293 TAAAAATTTTTTTTTATTAAAATAAATTTAAAAATTTTAAAATTTTTAAATTTTTAAAAA 234

Qy 351 AATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGT 410
Db 233 TTTTAAAATATTTTTTAAAAATTATAATAATAATTAAAATTAATAATTAAAAAAAATAAT 174

Qy 411 ATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAA 470
Db 173 AAAAAATAAATAAAAAAATAAAAAATAAATAAAAAAAATATAAAAAATTAAAAATAATAA 114

Qy 471 TAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTG 530
Db 113 TTAAATATATAAAAAAATATAAAAATTAAAAAATTAAAAAAAAAATAAAAAAAAAATTTT 54

Qy 531 TCATTCAGGAGTATGGAAGTATCGATGGCTTCTTTAAAATGAA 573
Db 53 TATTATTTTAATTAATTAATATTAATTTTTTATTTATTTTAAA 11

RESULT 12
ABL32467
ID ABL32467 standard; DNA; 15373 BP.
XX
AC ABL32467;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 440.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 440; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,

```
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention
XX
SQ   Sequence 15373 BP; 5097 A; 89 C; 2465 G; 7722 T; 0 U; 0 Other;

  Query Match             4.7%;  Score 51.4;  DB 6;  Length 15373;
  Best Local Similarity   45.4%;  Pred. No. 0.2;
  Matches  223;  Conservative    0;  Mismatches  266;  Indels    2;  Gaps    1;

Qy         28 TTCACTCTCTATGTTCTTATAAGCAAATGTTACAATGAGATATGAGTTTTAAAGCCAGAT 87
              || | | |   | || ||| |  ||| |||   |    ||      |  | |      |
Db       4150 TTTATTTTTGTTATTTTTAGGAAAAAAAGTTGATAAATTATTATTTGTAGATATTTTTTT 4209

Qy         88 CTTCCTTATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAG 147
               ||  ||| |    |    || | |  ||  |     |      || |   ||  |||||
Db       4210 GTTTTATATTTTATAGATATATGTTTTTTTAATTTTTTTAGAAATGTTTTTTAAGTTGAT 4269

Qy        148 AAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATT 207
               ||| ||   | |  || | | ||  |||  ||||  || |  | |  ||       |||
Db       4270 TAATTTTTTAGAATTTAATTAATTTTTTAGAATTGTTTAAAAGATTAATTTTTTAGAATT 4329

Qy        208 GTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCAT 267
              ||||     ||  | |   | || || |||||||||  |||      || |  | || |
Db       4330 GTTTAAAAGATTAATTTTTTAGAATTGTTTAAAAGA--TTAATTTTTTAGAATTGTTTAA 4387

Qy        268 TTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATT 327
                | ||    |||    |   ||  | |  |||||||  ||   ||     | |||| ||
Db       4388 AAGATTAATTTTTTAGAATTTAATTTTTTTAGAAATGTTTTTTAAGTTGATTAATATTTT 4447

Qy        328 AGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAA 387
              |    ||| ||||| |||   |  |      | |  |  |||  || | |     | |
Db       4448 ATTTTGTTAAAAATAAAAATAATTAGATAATTAGAAAATTTTAAGTATTATATATTTAAT 4507

Qy        388 AGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAAT 447
              ||  || |  |  || |  | | |   |   | | |              | || |||
Db       4508 TATTAGTAATTGTATATATAATAAAATTTTTTTAGTGTTAAATTATAAAAATATTGAAGA 4567

Qy        448 TACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAA 507
              |  | |  || | | ||  |    || |   |||  ||  |   || | || | ||| |
Db       4568 GAATATATTTTATTTGTTTTATTATATTTGGTAAATAATTTTAAATTGTATAGTTTTATA 4627

Qy        508 ACATTTACTTT 518
              |     |||
Db       4628 AATAAAATTTT 4638


RESULT 13
ABL33263/c
ID   ABL33263 standard; DNA; 12393 BP.
XX
AC   ABL33263;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 1236.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP007537.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful for
PT   diagnosis and treatment of diseases associated with abnormal cytosine
PT   methylation.
XX
PS   Claim 1; SEQ ID NO 1236; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention
XX
SQ   Sequence 12393 BP; 3484 A; 219 C; 2406 G; 6282 T; 0 U; 2 Other;

  Query Match             4.7%;  Score 51.2;  DB 6;  Length 12393;
  Best Local Similarity   49.4%;  Pred. No. 0.21;
  Matches  160;  Conservative    0;  Mismatches  163;  Indels    1;  Gaps    1;

Qy        209 TTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATT 268
              |||  ||||||   |   ||     | |||| | |  ||  | |   | |    ||| |
Db       1820 TTTATTTAATAACATTATTTTTCAATATTAATACATATTCTATAACAAAAAAACTTCGTA 1761

Qy        269 TGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTA 328
              | ||||   ||| || ||| ||   | ||| |||   | | | |  || | || | | |
Db       1760 TTTTTTATTTTTTTTTAAATAAAAAAAAAAATAAACAAAATAAAAATAATAACAAAAAAAA 1701

Qy        329 GCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAA 388
                |    ||||    | | | ||    |  |||| | || ||  || |      |   |
Db       1700 AAAAAAATAAAAAACCAAAAATAAAACTTATATATAAAATAACAAAAACAACATCTTTTA 1641

Qy        389 GGTAAAAAGTCAACCTAAAGG-TATACTACCTGCTTTCTTATCGCACTGCAAATAGAAAT 447
                | |||| |     ||||   || | ||  | |   |||          |  ||| |  |
Db       1640 AATCAAAACTTTTTTTAAACCATAAATTATTTACAAACTTTAAAATATTTAAAAAATACA 1581

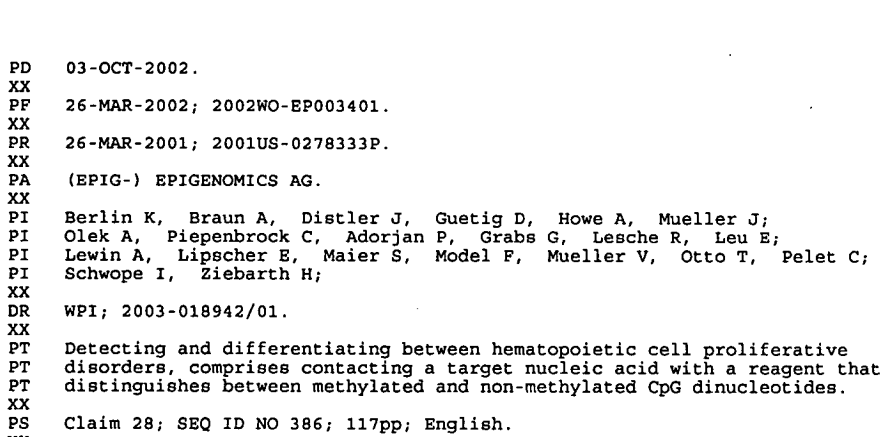
Qy        448 TACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAA 507
              ||  |||||||||         | |||   | || | | || |     | ||||| ||
Db       1580 TAATACAAATTTTAAAAATTTTAAAATACTAAAATATAAAAATAAAATTTTACTATAAAT 1521

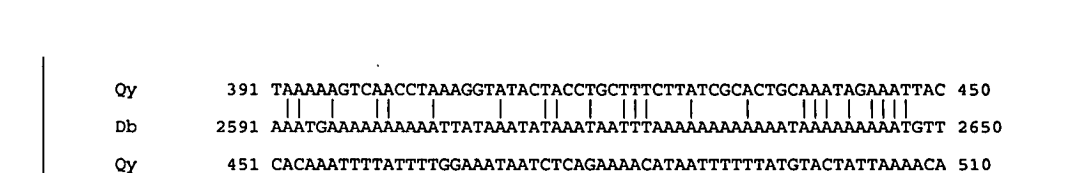
Qy        508 ACATTTACTTTCCAAATATTCTGT 531
              | ||| | || |   | | | | |
Db       1520 AAATTAATTTCCTTTAAAATATAT 1497


RESULT 14
ABZ10246
ID   ABZ10246 standard; DNA; 8056 BP.
XX
AC   ABZ10246;
XX
DT   16-JAN-2003  (first entry)
XX
DE   Haematopoietic cell proliferation disorder related DNA sequence #386.
XX
KW   Human; haematopoietic cell proliferation disorder; cytostatic;
KW   gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;
KW   cytosine methylation state; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200277272-A2.
XX
```

PD 03-OCT-2002.
XX
PF 26-MAR-2002; 2002WO-EP003401.
XX
PR 26-MAR-2001; 2001US-0278333P.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J;
PI Olek A, Piepenbrock C, Adorjan P, Grabs G, Lesche R, Leu E;
PI Lewin A, Lipscher E, Maier S, Model F, Mueller V, Otto T, Pelet C;
PI Schwope I, Ziebarth H;
XX
DR WPI; 2003-018942/01.
XX
PT Detecting and differentiating between hematopoietic cell proliferative
PT disorders, comprises contacting a target nucleic acid with a reagent that
PT distinguishes between methylated and non-methylated CpG dinucleotides.
XX
PS Claim 28; SEQ ID NO 386; 117pp; English.
XX
CC The present invention describes a method for detecting and
CC differentiating between haematopoietic cell proliferative disorders
CC associated with at least 1 gene and/or their regulatory regions in a
CC subject. The method comprises contacting a target nucleic acid in a
CC biological sample obtained from the subject with at least 1 reagent,
CC which distinguishes between methylated and non-methylated CpG
CC dinucleotides within the target nucleic acid. ABZ09861 to ABZ11118
CC represent specifically claimed nucleotide sequences from the present
CC invention. Oligonucleotides from the present invention can be used: for
CC differentiating between healthy haematopoietic cells and proliferative
CC disorder haematopoietic cells; for differentiating between acute
CC lymphocytic leukaemia and acute myelogenous leukaemia; as probes for
CC determining the cytosine methylation state and/or single nucleotide
CC polymorphisms (SNPs) of haematopoietic cell proliferation disorder
CC related sequences and their complements; and as primers for the
CC amplification of haematopoietic cell proliferation disorder related DNA
CC sequences. The nucleotide sequences from the present invention can also
CC be used for detecting a predisposition to, differentiation between
CC subclasses, diagnosis, prognosis, treatment and/or monitoring of
CC haematopoietic cell proliferative disorders. The present method enables a
CC highly specific classification of haematopoietic cell proliferative
CC disorders allowing for improved and informed treatment of patients
XX
SQ Sequence 8056 BP; 3711 A; 0 C; 371 G; 3974 T; 0 U; 0 Other;

Query Match 4.7%; Score 51; DB 8; Length 8056;
Best Local Similarity 45.1%; Pred. No. 0.2;
Matches 242; Conservative 0; Mismatches 285; Indels 10; Gaps 1;

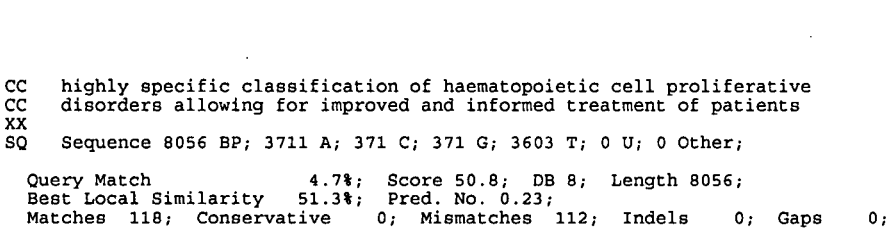




RESULT 15
ABZ10100
ID ABZ10100 standard; DNA; 8056 BP.
XX
AC ABZ10100;
XX
DT 16-JAN-2003 (first entry)
XX
DE Haematopoietic cell proliferation disorder related DNA sequence #240.
XX
KW Human; haematopoietic cell proliferation disorder; cytostatic;
KW gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;
KW cytosine methylation state; gene; ds.
XX
OS Homo sapiens.
XX
PN WO200277272-A2.
XX
PD 03-OCT-2002.
XX
PF 26-MAR-2002; 2002WO-EP003401.
XX
PR 26-MAR-2001; 2001US-0278333P.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J;
PI Olek A, Piepenbrock C, Adorjan P, Grabs G, Lesche R, Leu E;
PI Lewin A, Lipscher E, Maier S, Model F, Mueller V, Otto T, Pelet C;
PI Schwope I, Ziebarth H;
XX
DR WPI; 2003-018942/01.
XX
PT Detecting and differentiating between hematopoietic cell proliferative
PT disorders, comprises contacting a target nucleic acid with a reagent that
PT distinguishes between methylated and non-methylated CpG dinucleotides.
XX
PS Claim 28; SEQ ID NO 240; 117pp; English.
XX
CC The present invention describes a method for detecting and
CC differentiating between haematopoietic cell proliferative disorders
CC associated with at least 1 gene and/or their regulatory regions in a
CC subject. The method comprises contacting a target nucleic acid in a
CC biological sample obtained from the subject with at least 1 reagent,
CC which distinguishes between methylated and non-methylated CpG
CC dinucleotides within the target nucleic acid. ABZ09861 to ABZ11118
CC represent specifically claimed nucleotide sequences from the present
CC invention. Oligonucleotides from the present invention can be used: for
CC differentiating between healthy haematopoietic cells and proliferative
CC disorder haematopoietic cells; for differentiating between acute
CC lymphocytic leukaemia and acute myelogenous leukaemia; as probes for
CC determining the cytosine methylation state and/or single nucleotide
CC polymorphisms (SNPs) of haematopoietic cell proliferation disorder
CC related sequences and their complements; and as primers for the
CC amplification of haematopoietic cell proliferation disorder related DNA
CC sequences. The nucleotide sequences from the present invention can also
CC be used for detecting a predisposition to, differentiation between
CC subclasses, diagnosis, prognosis, treatment and/or monitoring of
CC haematopoietic cell proliferative disorders. The present method enables a

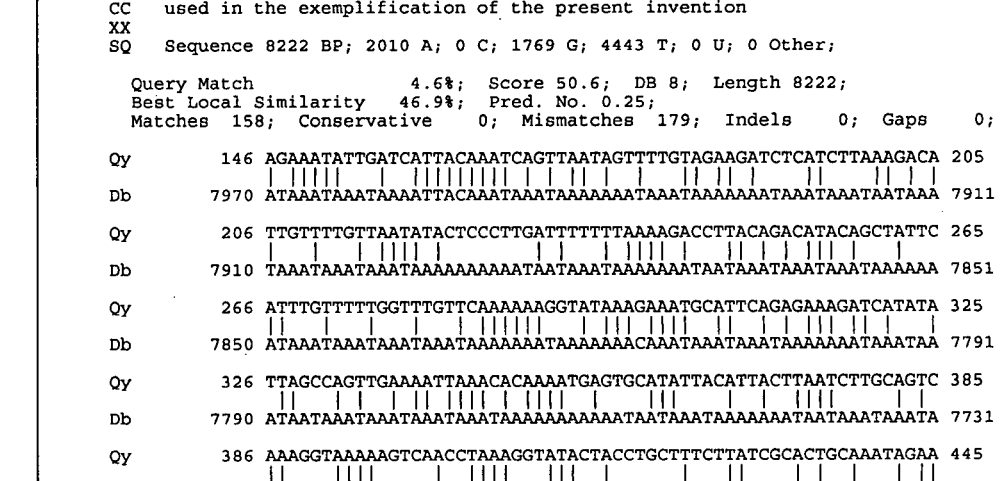
CC highly specific classification of haematopoietic cell proliferative
CC disorders allowing for improved and informed treatment of patients
XX
SQ Sequence 8056 BP; 3711 A; 371 C; 371 G; 3603 T; 0 U; 0 Other;

Query Match 4.7%; Score 50.8; DB 8; Length 8056;
Best Local Similarity 51.3%; Pred. No. 0.23;
Matches 118; Conservative 0; Mismatches 112; Indels 0; Gaps 0;



RESULT 16
ACF62816/c
ID ACF62816 standard; DNA; 8222 BP.
XX
AC ACF62816;
XX
DT 09-OCT-2003 (first entry)
XX
DE Colon cancer analysis related genomic DNA SEQ ID NO:65.
XX
KW Human; colon cancer; oestrogen receptor; myoglobin; p21; p27; p16; p53;
KW progesterone receptor; pcna; CEA; cdc2; c-erbB2; methylation; CpG;
KW characterisation; classification; diagnosis; differentiation;
KW colon cell proliferative disorder; gene; ds.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO2003014388-A2.
XX
PD 20-FEB-2003.
XX
PF 09-AUG-2002; 2002WO-EP008939.
XX
PR 09-AUG-2001; 2001DE-01039283.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Distler J, Model F, Taubert H;
XX
DR WPI; 2003-256600/25.
XX
PT Determining methylation status of CpG dinucleotides using modified
PT genomic sequences, oligonucleotides and/or PNA-oligomers, useful in the
PT characterization, grading, staging and/or diagnosis of colon cancer.
XX
PS Claim 22; Page 115-117; 219pp; English.
XX
CC The present invention describes a method for determining the methylation
CC status of CpG dinucleotides within the genes for oestrogen receptor, p21,
CC p27, p16, progesterone receptor, myoglobin, pcna, cdc2, c-erB2, p53
CC and/or CEA, which comprises contacting the target nucleic acid with a
CC reagent that distinguishes between methylated and non-methylated CpG
CC dinucleotides, and determining from the methylation status of the CpG
CC positions the presence of a colon cancer. A set of oligomers or peptide
CC nucleic acid (PNA)-oligomers can be used as probes for determining the
CC cytosine methylation state and/or single nucleotide polymorphisms (SNP)
CC of a corresponding genomic DNA by analysis of a chemically pretreated
CC genomic DNA. The pretreated genomic DNA is useful for the determination
CC of the methylation status of a corresponding genomic DNA and/or detection
CC of SNPs. The methods and pretreated genomic DNA are also useful for the
CC characterisation, classification, diagnosis and differentiation of colon
CC cell proliferative disorders. ACF62752 to ACF63278 represent sequences
CC used in the exemplification of the present invention
XX
SQ Sequence 8222 BP; 2010 A; 0 C; 1769 G; 4443 T; 0 U; 0 Other;

Query Match 4.6%; Score 50.6; DB 8; Length 8222;
Best Local Similarity 46.9%; Pred. No. 0.25;
Matches 158; Conservative 0; Mismatches 179; Indels 0; Gaps 0;



RESULT 17
ACF62794/c
ID ACF62794 standard; DNA; 8222 BP.
XX
AC ACF62794;
XX
DT 09-OCT-2003 (first entry)
XX
DE Colon cancer analysis related genomic DNA SEQ ID NO:43.
XX
KW Human; colon cancer; oestrogen receptor; myoglobin; p21; p27; p16; p53;
KW progesterone receptor; pcna; CEA; cdc2; c-erbB2; methylation; CpG;
KW characterisation; classification; diagnosis; differentiation;
KW colon cell proliferative disorder; gene; ds.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO2003014388-A2.
XX
PD 20-FEB-2003.
XX
PF 09-AUG-2002; 2002WO-EP008939.
XX
PR 09-AUG-2001; 2001DE-01039283.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Distler J, Model F, Taubert H;
XX
DR WPI; 2003-256600/25.
XX
PT Determining methylation status of CpG dinucleotides using modified
PT genomic sequences, oligonucleotides and/or PNA-oligomers, useful in the

PT characterization, grading, staging and/or diagnosis of colon cancer.
XX
PS Claim 22; Page 81-83; 219pp; English.
XX
CC The present invention describes a method for determining the methylation
CC status of CpG dinucleotides within the genes for oestrogen receptor, p21,
CC p27, p16, progesterone receptor, myoglobin, pcna, cdc2, c-erB2, p53
CC and/or CEA, which comprises contacting the target nucleic acid with a
CC reagent that distinguishes between methylated and non-methylated CpG
CC dinucleotides, and determining from the methylation status of the CpG
CC positions the presence of a colon cancer. A set of oligomers or peptide
CC nucleic acid (PNA)-oligomers can be used as probes for determining the
CC cytosine methylation state and/or single nucleotide polymorphisms (SNP)
CC of a corresponding genomic DNA by analysis of a chemically pretreated
CC genomic DNA. The pretreated genomic DNA is useful for the determination
CC of the methylation status of a corresponding genomic DNA and/or detection
CC of SNPs. The methods and pretreated genomic DNA are also useful for the
CC characterisation, classification, diagnosis and differentiation of colon
CC cell proliferative disorders. ACF62752 to ACF63278 represent sequences
CC used in the exemplification of the present invention
XX
SQ Sequence 8222 BP; 2010 A; 116 C; 1769 G; 4327 T; 0 U; 0 Other;

Query Match 4.6%; Score 50.6; DB 8; Length 8222;
Best Local Similarity 46.9%; Pred. No. 0.25;
Matches 158; Conservative 0; Mismatches 179; Indels 0; Gaps 0;

```
Qy       146 AGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACA 205
Db      7970 ATAAATAAATAAAATTACAAATAAATAAAAAAATAAATAAAAAAATAAATAAATAATAAA 7911

Qy       206 TTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTC 265
Db      7910 TAAATAAATAAATAAAAAAAAAAATAATAAATAAAAAAATAATAAATAAATAAATAAAAAA 7851

Qy       266 ATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATA 325
Db      7850 ATAAATAAATAAATAAATAAAAAAATAAAAAAACAAATAAATAAATAAAAAAATAAATAA 7791

Qy       326 TTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTC 385
Db      7790 ATAATAAATAAATAAATAAATAAAAAAAAAAAATAATAAATAAAAAAATAATAAATAAATA 7731

Qy       386 AAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAA 445
Db      7730 AATAAAAAAATAAATAAATAAATAAATAAAAAAATAATAAATAAATAAATAAATTAAAAA 7671

Qy       446 ATTACCACAAATTTTATTTTGGAAATAATCTCAGAAA 482
Db      7670 ATAAATAAATAAACACATAAATAAATAAATAAATAAA 7634
```

RESULT 18
ADB54190/c
ID ADB54190 standard; DNA; 11222 BP.
XX
AC ADB54190;
XX
DT 04-DEC-2003 (first entry)
XX
DE Pretreated genomic DNA region 114.
XX
KW colon cell proliferative disorder; non methylated CpG dinucleotide;
KW cytostatic; cancer; adenoma; carcinoma; cytosine methylation state; ds.
XX
OS Unidentified.
XX
PN WO2003072821-A2.
XX
PD 04-SEP-2003.
XX
PF 27-FEB-2003; 2003WO-EP002035.
XX
PR 27-FEB-2002; 2002EP-00004551.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Adorjan P, Burger M, Maier S, Nimmrich I, Becker E, Lesche R;
PI Rujan T, Schmitt A;
XX
DR WPI; 2003-731620/69.
XX
PT Detecting and differentiating between colon cell proliferative disorders
PT associated with a gene or its regulatory regions comprises contacting a
PT target nucleic acid in a biological sample obtained from the subject with
PT a reagent.
XX
PS Claim 32; SEQ ID NO 246; 74pp; English.
XX
CC The invention relates to a novel method for detecting and differentiating
CC between colon cell proliferative disorders associated with at least one
CC gene or its regulatory regions. The method comprises contacting a target
CC nucleic acid in a biological sample obtained from the subject with at
CC least one reagent or a series of reagents, where the reagent or series of
CC reagents, distinguishes between methylated and non methylated CpG
CC dinucleotides within the target nucleic acid. The molecules of the
CC invention demonstrate cytostatic activity whilst the method may useful
CC for detecting and differentiating between colon cell proliferative
CC disorders, including cancers such as colon adenoma and colon carcinoma.
CC The PNA (peptide nucleic acid)-oligomers are useful as probes for
CC determining cytosine methylation state or single nucleotide
CC polymorphisms. The current sequence is that of the pretreated genomic DNA
CC region of the invention. This sequence is not shown within the
CC specification but is taken from Wipoweb.
XX
SQ Sequence 11222 BP; 2767 A; 164 C; 2498 G; 5793 T; 0 U; 0 Other;

Query Match 4.6%; Score 50.6; DB 10; Length 11222;
Best Local Similarity 46.9%; Pred. No. 0.27;
Matches 158; Conservative 0; Mismatches 179; Indels 0; Gaps 0;

```
Qy       146 AGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACA 205
Db      7970 ATAAATAAATAAAATTACAAATAAATAAAAAAATAAATAAAAAAATAAATAAATAATAAA 7911

Qy       206 TTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTC 265
Db      7910 TAAATAAATAAATAAAAAAAAAAATAATAAATAAAAAAATAATAAATAAATAAATAAAAAA 7851

Qy       266 ATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATA 325
Db      7850 ATAAATAAATAAATAAATAAAAAAATAAAAAAACAAATAAATAAATAAAAAAATAAATAA 7791

Qy       326 TTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTC 385
Db      7790 ATAATAAATAAATAAATAAATAAAAAAAAAAAATAATAAATAAAAAAATAATAAATAAATA 7731

Qy       386 AAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAA 445
Db      7730 AATAAAAAAATAAATAAATAAATAAATAAAAAAATAATAAATAAATAAATAAATTAAAAA 7671

Qy       446 ATTACCACAAATTTTATTTTGGAAATAATCTCAGAAA 482
Db      7670 ATAAATAAATAAACACATAAATAAATAAATAAATAAA 7634
```

RESULT 19
ADB54318/c
ID ADB54318 standard; DNA; 11222 BP.
XX
AC ADB54318;
XX
DT 04-DEC-2003 (first entry)
XX
DE Pretreated genomic DNA region 242.
XX

KW colon cell proliferative disorder; non methylated CpG dinucleotide;
KW cytostatic; cancer; adenoma; carcinoma; cytosine methylation state; ds.
XX
OS Unidentified.
XX
PN WO2003072821-A2.
XX
PD 04-SEP-2003.
XX
PF 27-FEB-2003; 2003WO-EP002035.
XX
PR 27-FEB-2002; 2002EP-00004551.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Adorjan P, Burger M, Maier S, Nimmrich I, Becker E, Lesche R;
PI Rujan T, Schmitt A;
XX
DR WPI; 2003-731620/69.
XX
PT Detecting and differentiating between colon cell proliferative disorders
PT associated with a gene or its regulatory regions comprises contacting a
PT target nucleic acid in a biological sample obtained from the subject with
PT a reagent.
XX
PS Claim 32; SEQ ID NO 374; 74pp; English.
XX
CC The invention relates to a novel method for detecting and differentiating
CC between colon cell proliferative disorders associated with at least one
CC gene or its regulatory regions. The method comprises contacting a target
CC nucleic acid in a biological sample obtained from the subject with at
CC least one reagent or a series of reagents, where the reagent or series of
CC reagents, distinguishes between methylated and non methylated CpG
CC dinucleotides within the target nucleic acid. The molecules of the
CC invention demonstrate cytostatic activity whilst the method may useful
CC for detecting and differentiating between colon cell proliferative
CC disorders, including cancers such as colon adenoma and colon carcinoma.
CC The PNA (peptide nucleic acid)-oligomers are useful as probes for
CC determining cytosine methylation state or single nucleotide
CC polymorphisms. The current sequence is that of the pretreated genomic DNA
CC region of the invention. This sequence is not shown within the
CC specification but is taken from Wipoweb.
XX
SQ Sequence 11222 BP; 2767 A; 0 C; 2498 G; 5957 T; 0 U; 0 Other;

Query Match 4.6%; Score 50.6; DB 10; Length 11222;
Best Local Similarity 46.9%; Pred. No. 0.27;
Matches 158; Conservative 0; Mismatches 179; Indels 0; Gaps 0;

```
Qy       146 AGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACA 205
            | |||||   |  |||||||||| | | || |  |   || || |    ||    || | |
Db      7970 ATAAATAAATAAAATTACAAATAAATAAAAAAATAAATAAAAAAATAAATAAATAATAAA 7911

Qy       206 TTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTC 265
            |   |   | |||| |        | |   | |||| |   || | | ||| |   |
Db      7910 TAAATAAATAAATAAAAAAAAAAATAATAAATAAAAAAATAATAAATAAATAAATAAAAAA 7851

Qy       266 ATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATA 325
            ||   |   |   |   | ||||||   | ||| ||||  ||  | | ||| || |   |
Db      7850 ATAAATAAATAAATAAATAAAAAAATAAAAAAACAAATAAATAAATAAAAAAATAAATAA 7791

Qy       326 TTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTC 385
              ||   | |  | || |||| | ||||  |    |||     |  |  ||||     | |
Db      7790 ATAATAAATAAATAAATAAATAAAAAAAAAAAATAATAAATAAAAAAATAATAAATAAATA 7731

Qy       386 AAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAA 445
            ||    ||||     |  ||||   |||  |       |   ||    | |  |   | ||
Db      7730 AATAAAAAAAATAAATAAATAAATAAATAAAAAAATAATAAATAAATAAATAAATTAAAAA 7671

Qy       446 ATTACCACAAATTTTATTTTGGAAATAATCTCAGAAA 482
            || |  | | |      |    ||||||    | |||
Db      7670 ATAAATAAATAAACACATAAATAAATAAATAAATAAA 7634
```

RESULT 20
ADI72306/c
ID ADI72306 standard; DNA; 436 BP.
XX
AC ADI72306;
XX
DT 20-MAY-2004 (first entry)
XX
DE Human ovarian cancer DNA marker #5048.
XX
KW Human; ovarian cancer; ds; tumour; cytostatic; DNA marker.
XX
OS Homo sapiens.
XX
PN WO200170979-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US009126.
XX
PR 21-MAR-2000; 2000US-0191031P.
PR 25-MAY-2000; 2000US-0207124P.
PR 15-JUN-2000; 2000US-0211940P.
PR 07-JUL-2000; 2000US-0216820P.
PR 25-JUL-2000; 2000US-0220661P.
PR 21-DEC-2000; 2000US-0257672P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Lee J, Lillie J;
XX
DR WPI; 2001-611502/70.
XX
PT Novel isolated nucleic acid molecules (markers) overexpressed in ovarian
PT cancer cells as compared to their normal non-cancerous ovarian cells are
PT used to characterize stage, grade, histological type of ovarian cancer.
XX
PS Disclosure; SEQ ID NO 5048; 106pp; English.
XX
CC The invention relates to nucleic acid markers which are overexpressed in
CC ovarian cancer cells as compared to their expression in normal (i.e. non-
CC cancerous) ovarian cells. The invention also relates to polypeptides
CC encoded by the markers, antibodies that selectively bind to the
CC polypeptides, a method of inhibiting ovarian cancer in a patient at risk
CC of developing ovarian cancer involving inhibiting expression of a gene
CC corresponding to a marker of the invention and a method of treating a
CC patient afflicted with ovarian cancer comprising providing to cells of
CC the patient an antisense oligonucleotide complementary to a marker of the
CC invention. The markers are useful for assessing if a patient is afflicted
CC with ovarian cancer, which involves comparing the level of expression of
CC a marker in a patient sample and a normal level of expression of the
CC marker in a control non-ovarian cancer sample. A difference between the
CC expression levels indicates ovarian cancer. The level of expression of a
CC marker corresponds to a secreted protein or to a transcribed
CC polynucleotide or its portion. The level of expression of the marker is
CC assessed by detecting the presence in the sample, a protein or protein
CC fragment corresponding to the marker. The presence of protein or protein
CC fragment is detected using an antibody that specifically binds with the
CC protein or protein fragment. Alternatively, the level of expression of
CC the marker is assessed by detecting the presence of a transcribed
CC polynucleotide which anneals with the marker or anneals with a portion of
CC the polynucleotide comprising the marker, under stringent conditions. The
CC marker is also used for monitoring the progression of ovarian cancer in a
CC patient which involves detecting expression of the marker in a patient
CC sample at a first point in time, repeating the method at a subsequent
CC time and comparing the level of expression. The method is carried out
CC using an ovarian tissue sample. A composition comprising a marker,
CC polypeptide or antibody of the invention is used to treat ovarian cancer.
CC This sequence represents a human ovarian cancer DNA marker of the
CC invention.
XX
SQ Sequence 436 BP; 62 A; 24 C; 27 G; 197 T; 0 U; 126 Other;

```
  Query Match             4.6%;  Score 50.4;  DB 5;  Length 436;
  Best Local Similarity   34.4%;  Pred. No. 0.14;
  Matches  114;  Conservative    0;  Mismatches  217;  Indels    0;  Gaps    0;

Qy         78 AAAGCCAGATCTTCCTTATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAG 137
              |||                                                      | |
Db        390 AAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNANANANNNNNNNNTNTTTTTT 331

Qy        138 TTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCT 197
              |||  |  | ||         |  | |||             || |  ||      |
Db        330 TTTTTTNAAAAANNNNNNNNNAAAAAAAAANNNNNGGGGGNNNTTTTTNAAANNNNNTTNN 271

Qy        198 TAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATAC 257
              | | | | | |      ||  |  |   |  |||||||         | | | | |
Db        270 NANAAAAAATNTNNNNNNAAATTGNTNTNTNNNTTTTTTTNNNNNNNNNAAAAAAAAAA 211

Qy        258 AGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAG 317
              |  | ||  ||| |||||    |   |          | ||| |||   |   | | |||
Db        210 ANNTTTTTTTTTTTTTTTTANAATNCNCCNNNNNNNNAAAAAAAAAAAAAAAAAAAAAAA 151

Qy        318 ATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATC 377
              |  | | |  |   |    ||||  ||| | ||||  |   | |  | |  |   ||
Db        150 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 91

Qy        378 TTGCAGTCAAAGGTAAAAAGTCAACCTAAAG 408
                |   |||   |||||   ||   ||||
Db         90 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAG 60


RESULT 21
ADL37450/c
ID   ADL37450 standard; DNA; 436 BP.
XX
AC   ADL37450;
XX
DT   20-MAY-2004  (first entry)
XX
DE   Human ovarian cancer DNA marker #11340.
XX
KW   Human; ovarian cancer; ds; tumour; cytostatic; DNA marker.
XX
OS   Homo sapiens.
XX
PN   WO200170979-A2.
XX
PD   27-SEP-2001.
XX
PF   21-MAR-2001; 2001WO-US009126.
XX
PR   21-MAR-2000; 2000US-0191031P.
PR   25-MAY-2000; 2000US-0207124P.
PR   15-JUN-2000; 2000US-0211940P.
PR   07-JUL-2000; 2000US-0216820P.
PR   25-JUL-2000; 2000US-0220661P.
PR   21-DEC-2000; 2000US-0257672P.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Lee J,  Lillie J;
XX
DR   WPI; 2001-611502/70.
XX
PT   Novel isolated nucleic acid molecules (markers) overexpressed in ovarian
PT   cancer cells as compared to their normal non-cancerous ovarian cells are
PT   used to characterize stage, grade, histological type of ovarian cancer.
XX
PS   Disclosure; SEQ ID NO 11340; 106pp; English.
XX
CC   The invention relates to nucleic acid markers which are overexpressed in
CC   ovarian cancer cells as compared to their expression in normal (i.e. non-
CC   cancerous) ovarian cells. The invention also relates to polypeptides
CC   encoded by the markers, antibodies that selectively bind to the
CC   polypeptides, a method of inhibiting ovarian cancer in a patient at risk
CC   of developing ovarian cancer involving inhibiting expression of a gene
CC   corresponding to a marker of the invention and a method of treating a
CC   patient afflicted with ovarian cancer comprising providing to cells of
CC   the patient an antisense oligonucleotide complementary to a marker of the
CC   invention. The markers are useful for assessing if a patient is afflicted
CC   with ovarian cancer, which involves comparing the level of expression of
CC   a marker in a patient sample and a normal level of expression of the
CC   marker in a control non-ovarian cancer sample. A difference between the
CC   expression levels indicates ovarian cancer. The level of expression of a
CC   marker corresponds to a secreted protein or to a transcribed
CC   polynucleotide or its portion. The level of expression of the marker is
CC   assessed by detecting the presence in the sample, a protein or protein
CC   fragment corresponding to the marker. The presence of protein or protein
CC   fragment is detected using an antibody that specifically binds with the
CC   protein or protein fragment. Alternatively, the level of expression of
CC   the marker is assessed by detecting the presence of a transcribed
CC   polynucleotide which anneals with the marker or anneals with a portion of
CC   the polynucleotide comprising the marker, under stringent conditions. The
CC   marker is also used for monitoring the progression of ovarian cancer in a
CC   patient which involves detecting expression of the marker in a patient
CC   sample at a first point in time, repeating the method at a subsequent
CC   time and comparing the level of expression. The method is carried out
CC   using an ovarian tissue sample. A composition comprising a marker,
CC   polypeptide or antibody of the invention is used to treat ovarian cancer.
CC   This sequence represents a human ovarian cancer DNA marker of the
CC   invention.
XX
SQ   Sequence 436 BP; 62 A; 24 C; 27 G; 197 T; 0 U; 126 Other;

  Query Match             4.6%;  Score 50.4;  DB 5;  Length 436;
  Best Local Similarity   34.4%;  Pred. No. 0.14;
  Matches  114;  Conservative    0;  Mismatches  217;  Indels    0;  Gaps    0;

Qy         78 AAAGCCAGATCTTCCTTATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAG 137
              |||                                                      | |
Db        390 AAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNANANANNNNNNNNTNTTTTTT 331

Qy        138 TTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCT 197
              |||  |  | ||         |  | |||             || |  ||      |
Db        330 TTTTTTNAAAAANNNNNNNNNAAAAAAAAANNNNNGGGGGNNNTTTTTNAAANNNNNTTNN 271

Qy        198 TAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATAC 257
              | | | | | |      ||  |  |   |  |||||||         | | | | |
Db        270 NANAAAAAATNTNNNNNNAAATTGNTNTNTNNNTTTTTTTNNNNNNNNNAAAAAAAAAA 211

Qy        258 AGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAG 317
              |  | ||  ||| |||||    |   |          | ||| |||   |   | | |||
Db        210 ANNTTTTTTTTTTTTTTTTANAATNCNCCNNNNNNNNAAAAAAAAAAAAAAAAAAAAAAA 151

Qy        318 ATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATC 377
              |  | | |  |   |    ||||  ||| | ||||  |   | |  | |  |   ||
Db        150 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 91

Qy        378 TTGCAGTCAAAGGTAAAAAGTCAACCTAAAG 408
                |   |||   |||||   ||   ||||
Db         90 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAG 60


RESULT 22
ABQ67060/c
ID   ABQ67060 standard; DNA; 34688 BP.
XX
AC   ABQ67060;
XX
DT   28-AUG-2002  (first entry)
XX
DE   Human angiogenesis associated polynucleotide SEQ ID NO 90.
XX
KW   Human; angiogenesis; methylation; eye disease; glaucoma; tumour;
KW   inflammation; rheumatoid arthritis; diabetic retinopathy; antiulcerds;
KW   macular degeneration; inflammatory bowel disease; Crohn's disease;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
```

```
KW   antiarteriosclerotic; ds.
XX
OS   Homo sapiens.
XX
PN   WO200246454-A2.
XX
PD   13-JUN-2002.
XX
PF   06-DEC-2001; 2001WO-EP014320.
XX
PR   06-DEC-2000; 2000DE-01061338.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Schacht O;
XX
DR   WPI; 2002-500450/53.
XX
PT   New nucleic acid fragments from chemically treated angiogenesis-
PT   associated genes, useful for determining methylation status, e.g. in
PT   diagnosis or treatment of cancer.
XX
PS   Claim 1; SEQ ID NO 90; 41pp + Sequence Listing; German.
XX
CC   The invention relates to a nucleic acid (I) comprising a segment of 18
CC   bases of chemically pretreated DNA of angiogenesis-associated genes (II)
CC   having sequences (ABQ66971-ABQ67178) or their complements. (I), also
CC   related oligomers, are used to evaluate the methylation status and/or
CC   single-nucleotide polymorphisms, in angiogenesis-related genes, for
CC   diagnosis and treatment of eye diseases, proliferative retinopathy,
CC   neovascular glaucoma, solid tumours, inflammation, rheumatoid arthritis,
CC   diabetic retinopathy, macular degeneration caused by neovascularisation,
CC   psoriasis, arteriosclerosis, inflammatory bowel diseases, ulcers and
CC   Crohn's disease. Note: The sequence data for this patent did not form
CC   part of the printed specification, but was obtained in electronic format
CC   directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 34688 BP; 11855 A; 444 C; 5972 G; 16417 T; 0 U; 0 Other;

  Query Match             4.6%;  Score 50.4;  DB 6;  Length 34688;
  Best Local Similarity   45.0%;  Pred. No. 0.4;
  Matches  189;  Conservative    0;  Mismatches  231;  Indels    0;  Gaps    0;

Qy        151 TATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTT 210
              | || || ||||||   |   ||||  |  | |   |      | |  |  |   ||| |
Db      26204 TTTTAATAATTACACTCCCAATAATTTTAATATTCCAACGTACTTAATTTACTTATTTTC 26145

Qy        211 TTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTG 270
              ||||||||  ||  || |||    ||||| | | |  | |    ||   |    |||
Db      26144 ACGTTAATATTTTCTATTTATTAAAATAAAAAAACATTAAAATTATCATTAAAAACTTTA 26085

Qy        271 TTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGC 330
              | |||    | ||  || |      | |  ||   |   | | |||  | |  |  |
Db      26084 TCTTTAACATATTATAATATTTATTTCAATAACTAACCTATAAAAAATTAAATTCCTTTA 26025

Qy        331 CAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGG 390
              |    ||    | | | |   | |    | | | |||||  |  |  ||
Db      26024 TAAAAATAATAATTAAAATATTTTTATCTAATAAAAAAAAATTAATAATAAAACCATTTC 25965

Qy        391 TAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTAC 450
              |   ||   ||   |||   ||| | |    ||        |   ||||  || |
Db      25964 TCCGAAAAAAATTAAAAACAATAATTCAATTTTAAAATATTATCGCAAAATTAAATTATT 25905

Qy        451 CACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACA 510
              |  ||| ||||||    |||| || |    ||   |||| |   || || || ||||
Db      25904 CTAAAAATTTATTAATAAAATTATATACTTAATATTAATCTAAAATATAATACTAAATTT 25845

Qy        511 TTTACTTTCCAAATATTCTGTCATTCAGGAGTATGGAAGTATCGATGGCTTCTTTAAAAT 570
                ||| |  ||| |||   |   |     | |||  ||  |       |||| | | ||
Db      25844 AAAACTATAAAAAAATTAAATATATTTAAAATATATAAAAACGTTAAACTTCTATTATAT 25785

RESULT 23
ABZ10246/c
ID   ABZ10246 standard; DNA; 8056 BP.
XX
AC   ABZ10246;
XX
DT   16-JAN-2003  (first entry)
XX
DE   Haematopoietic cell proliferation disorder related DNA sequence #386.
XX
KW   Human; haematopoietic cell proliferation disorder; cytostatic;
KW   gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;
KW   cytosine methylation state; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200277272-A2.
XX
PD   03-OCT-2002.
XX
PF   26-MAR-2002; 2002WO-EP003401.
XX
PR   26-MAR-2001; 2001US-0278333P.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Berlin K,  Braun A,  Distler J,  Guetig D,  Howe A,  Mueller J;
PI   Olek A,  Piepenbrock C,  Adorjan P,  Grabs G,  Lesche R,  Leu E;
PI   Lewin A,  Lipscher E,  Maier S,  Model F,  Mueller V,  Otto T,  Pelet C;
PI   Schwope I,  Ziebarth H;
XX
DR   WPI; 2003-018942/01.
XX
PT   Detecting and differentiating between hematopoietic cell proliferative
PT   disorders, comprises contacting a target nucleic acid with a reagent that
PT   distinguishes between methylated and non-methylated CpG dinucleotides.
XX
PS   Claim 28; SEQ ID NO 386; 117pp; English.
XX
CC   The present invention describes a method for detecting and
CC   differentiating between haematopoietic cell proliferative disorders
CC   associated with at least 1 gene and/or their regulatory regions in a
CC   subject. The method comprises contacting a target nucleic acid in a
CC   biological sample obtained from the subject with at least 1 reagent,
CC   which distinguishes between methylated and non-methylated CpG
CC   dinucleotides within the target nucleic acid. ABZ09861 to ABZ11118
CC   represent specifically claimed nucleotide sequences from the present
CC   invention. Oligonucleotides from the present invention can be used: for
CC   differentiating between healthy haematopoietic cells and proliferative
CC   disorder haematopoietic cells; for differentiating between acute
CC   lymphocytic leukaemia and acute myelogenous leukaemia; as probes for
CC   determining the cytosine methylation state and/or single nucleotide
CC   polymorphisms (SNPs) of haematopoietic cell proliferation disorder
CC   related sequences and their complements; and as primers for the
CC   amplification of haematopoietic cell proliferation disorder related DNA
CC   sequences. The nucleotide sequences from the present invention can also
CC   be used for detecting a predisposition to, differentiation between
CC   subclasses, diagnosis, prognosis, treatment and/or monitoring of
CC   haematopoietic cell proliferative disorders. The present method enables a
CC   highly specific classification of haematopoietic cell proliferative
CC   disorders allowing for improved and informed treatment of patients
XX
SQ   Sequence 8056 BP; 3711 A; 0 C; 371 G; 3974 T; 0 U; 0 Other;

  Query Match             4.6%;  Score 50.2;  DB 8;  Length 8056;
  Best Local Similarity   44.3%;  Pred. No. 0.31;
  Matches  205;  Conservative    0;  Mismatches  258;  Indels    0;  Gaps    0;

Qy        111 TCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAG 170
              | || || || || |   |  | |    ||  ||   ||| |   |  || |  ||| |
Db        473 TTTATTTTTTAAAATTAATTTTTTTTTATTATTTTTAAAAAAAATAAAATAATTAATAAA 414

Qy        171 TTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGA 230
              |  ||  ||   |  ||  |   ||  |||| |  || |||  ||  | |  |   || |
```

```
Db        413 TATATTATTAAATTAAATTTTAAATTATAAATATTTTATTTCATTTTTTTTTTTTTTTAA 354

Qy        231 TTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAA 290

Db        353 TTATTATTATTTTTATTTATTTTTATTAATTATTTTTTTAAAAAATAATTTTAAAAAAAA 294

Qy        291 GGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAA 350

Db        293 TAAAAATTTTTTTTTATTAAAATAAATTTAAAAATTTAAAATTTTTAAATTTTTAAAAA 234

Qy        351 AATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGT 410

Db        233 TTTTAAAATATTTTTTAAAAATTATAATAATAATTAAAATTAATAATTAAAAAAAATAAT 174

Qy        411 ATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAA 470

Db        173 AAAAAATAAATAAAAAAATAAAAAATAAATAAAAAAAATATAAAAAATTAAAAATAATAA 114

Qy        471 TAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTG 530

Db        113 TTAAATATATAAAAAAATATAAAAATTAAAAAATTAAAAAAAAAATAAAAAAAAAATTTT 54

Qy        531 TCATTCAGGAGTATGGAAGTATCGATGGCTTCTTTAAAATGAA 573

Db         53 TATTATTTTAATTAATTAATATTAATTTTTTATTTATTTTAAA 11


RESULT 24
ABL55644
ID   ABL55644 standard; DNA; 50000 BP.
XX
AC   ABL55644;
XX
DT   01-JUL-2002  (first entry)
XX
DE   AmEPV genome fragment#2.
XX
KW   AmEPV; gene therapy; viral vector; chromosome mapping; gene mapping;
KW   genetic deficiency disorder; ds.
XX
OS   Amsacta moorei entomopoxvirus.
XX
PN   WO200212526-A2.
XX
PD   14-FEB-2002.
XX
PF   10-AUG-2001; 2001WO-US025287.
XX
PR   10-AUG-2000; 2000US-0224479P.
PR   14-SEP-2000; 2000US-00662254.
XX
PA   (UYFL ) UNIV FLORIDA.
XX
PI   Moyer RW,  Li Y,  Bawden AL;
XX
DR   WPI; 2002-227161/28.
XX
PT   Novel recombinant entomopox virus vector useful for delivering
PT   polynucleotide encoding protein to vertebrate cell, comprises
PT   polynucleotide encoding protein operably linked with heterologous
PT   promoter sequence.
XX
PS   Disclosure; Page 150-175; 326pp; English.
XX
CC   The invention relates to a recombinant entomopox virus (EPV) vector,
CC   comprising a polynucleotide encoding a protein operably linked with a
CC   heterologous promoter sequence. The invention also concerns methods for
CC   providing gene therapy for genetic deficiency disorders. Vectors of the
CC   invention are useful for delivering a polynucleotide encoding a protein
CC   to a vertebrate cell preferably a mammalian cell, such as a human cell.
CC   The vector is introduced into the vertebrate cell by infection in a viral
CC   particle, or by transfection, transduction, or injection either in vitro
CC   or in vivo. The vector is useful for the delivery and expression of
CC   biologically useful proteins in gene therapy protocols, and for
CC   delivering large DNA segments for engineering of vertebrate cells.
CC   Polynucleotides of the invention have applications in techniques such as
CC   their use as insertion sites for foreign genes of interest, hybridisation
CC   probes, for chromosome and gene mapping, in PCR technologies, and in the
CC   production of sense or antisense nucleic acids. Vectors of the invention
CC   provide for stable integration and expression of heterologous DNA in host
CC   cells, and are adapted for accepting large heterologous polynucleotide
CC   inserts which can be delivered in an infected or transformed cell and
CC   expressed in a stable fraction. The current sequence represents a
CC   fragment of the genome of the genus B entomopoxvirus from amsacta moorei
CC   (AmEPV)
XX
SQ   Sequence 50000 BP; 20514 A; 4505 C; 4614 G; 20367 T; 0 U; 0 Other;

  Query Match             4.6%;  Score 50.2;  DB 6;  Length 50000;
  Best Local Similarity   47.9%;  Pred. No. 0.49;
  Matches  145;  Conservative    0;  Mismatches  158;  Indels    0;  Gaps    0;

Qy        228 TGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAA 287

Db      39496 TTAATTTTATAGTTTTTTTATTATTAAGCCAATTATTTATTTTTCTATAATTTTTTATAC 39555

Qy        288 AAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACA 347

Db      39556 ACACATACGAAGCATTGAAAAATCCAGTTTTTTTATCAATTACTAATTGATAACTATTTA 39615

Qy        348 CAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAA 407

Db      39616 TATGAGCAGAATAAAATGTATTTTTTATTTTTTCATCATATATTATAGAAGACATTTTAA 39675

Qy        408 GGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGG 467

Db      39676 CTATTATTATCTTATTTTATTTTTCAAAAAACATATTTTTTACATTTAATACTGTTTTTA 39735

Qy        468 AAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATT 527

Db      39736 TAATATGTTTTGATATTGAATTTTTATCAATATTATTTAAATATTTTAATAAATTATTGT 39795

Qy        528 CTG 530

Db      39796 CTG 39798


RESULT 25
AAS46342/c
ID   AAS46342 standard; DNA; 18988 BP.
XX
AC   AAS46342;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Tumour suppressor gene derived chemically modified sequence #64.
XX
KW   Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer;
KW   tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW   cytosine methylation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200168912-A2.
XX
PD   20-SEP-2001.
XX
PF   15-MAR-2001; 2001WO-EP002955.
XX
PR   15-MAR-2000; 2000DE-01013847.
PR   06-APR-2000; 2000DE-01019058.
PR   07-APR-2000; 2000DE-01019173.
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
```

XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2001-602752/68.
XX
PT Fragments of chemically modified genes associated with tumor suppressor
PT genes and oncogenes, useful in designing primers and probes for analyzing
PT diseases associated with cytosine methylation state e.g. cancer.
XX
PS Claim 1; SEQ ID NO 64; 27pp; English.
XX
CC The invention relates to a nucleic acid comprising a sequence of 18
CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC bisulphite, of genes associated with tumour suppression and oncogenes
CC having a sequence taken from 536 (actually 533 since numbers 408, 458 and
CC 500 are missing from the sequence listing) sequences (Ss) and sequences
CC complementary to (Ss). The nucleic acid may be a peptide nucleic acid-
CC oligomer (PNA) of at least 9 nucleotides and may form part of a set of
CC probes for detecting the cytosine methylation state and/or single
CC nucleotide polymorphisms and also to be used in an array for analysing
CC diseases associated with CpG dinucleotides e.g. cancers and tumours. The
CC probes can also be used in a method for ascertaining genetic and/or
CC epigenetic parameters for the diagnosis and/or therapy of existing
CC diseases or the predisposition to specific diseases, by analysing
CC cytosine methylations. The parameters may be compared to another set of
CC genetic and/or epigenetic parameters, the differences serving as basis
CC for diagnosis and/or prognosis events which are disadvantageous to
CC patients. The present sequence is one of the 533 genomic sequences
CC derived from tumour suppressor genes and oncogenes. Sequences with even
CC numbered Seq ID numbers are the complementary sequence of the
CC corresponding odd numbered sequence (e.g. ID 2 and ID1, ID 536 and ID
CC 535, except for those whose partner sequence is missing). Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 18988 BP; 5467 A; 567 C; 4014 G; 8940 T; 0 U; 0 Other;

Query Match 4.6%; Score 50; DB 4; Length 18988;
Best Local Similarity 47.4%; Pred. No. 0.43;
Matches 183; Conservative 0; Mismatches 200; Indels 3; Gaps 1;

```
Qy       146 AGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACA 205
Db      8192 AAAAAAAATAACCATTATTAATTATTAACTACTCTACCAAAATATCATTTAAATTACTTC 8133

Qy       206 TTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTC 265
Db      8132 TATTTATAAAAAAAAATACATTAAATAAAATAATCTCCATTCACAAAAAATCAACAATCC 8073

Qy       266 ATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATA 325
Db      8072 ATTAAACTTCTCCTTAAACAAAATTTATCCTAAAATATAAATAAAAAAAAAATAATTAAA 8013

Qy       326 TTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTC 385
Db      8012 AAATAAAAAATATATAAAAAAAAAAAAAAAAAAAAAATAACACAAAAATACCTCACCCTACC 7953

Qy       386 AAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAA 445
Db      7952 AAAAATAAACACACAACCGCTACCTACCTATCCTA---TCTTATCGTTCCCAAAAACACA 7896

Qy       446 ATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTA 505
Db      7895 AAACAACTAAAACCCGTCATCTTCGTAATACCAACAAACATATTTTATAAATTCATATAA 7836

Qy       506 AAACATTTACTTTCCAAATATTCTGT 531
Db      7835 ATAACTAAATTTTATATAAAAACTAT 7810
```

RESULT 26
ABL32701/c

ID ABL32701 standard; DNA; 18988 BP.
XX
AC ABL32701;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 674.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 674; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 18988 BP; 5467 A; 567 C; 4014 G; 8940 T; 0 U; 0 Other;

Query Match 4.6%; Score 50; DB 6; Length 18988;
Best Local Similarity 47.4%; Pred. No. 0.43;
Matches 183; Conservative 0; Mismatches 200; Indels 3; Gaps 1;

```
Qy       146 AGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACA 205
Db      8192 AAAAAAAATAACCATTATTAATTATTAACTACTCTACCAAAATATCATTTAAATTACTTC 8133

Qy       206 TTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTC 265
Db      8132 TATTTATAAAAAAAAAATACATTAAATAAAATAATCTCCATTCACAAAAAATCAACAATCC 8073

Qy       266 ATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATA 325
Db      8072 ATTAAACTTCTCCTTAAACAAAATTTATCCTAAAATATAAATAAAAAAAAAATAATTAAA 8013

Qy       326 TTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTC 385
Db      8012 AAATAAAAAATATATAAAAAAAAAAAAAAAAAAAAAATAACACAAAAATACCTCACCCTACC 7953

Qy       386 AAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAA 445
```

```
Db       7952 AAAAATAAACACACAACCGCTACCTACCTATCCTA---TCTTATCGTTCCCAAAAACACA 7896

Qy        446 ATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTA 505
              |        |||      |   |      ||||  ||  |||||||  |||  |  |   |  |||| |
Db       7895 AAACAACTAAAACCCGTCATCTTCGTAATACCAACAAACATATTTTATAAATTCATATAA 7836

Qy        506 AAACATTTACTTTCCAAATATTCTGT 531
              | |  |  | |||  | | |  || |
Db       7835 ATAACTAAATTTTATATAAAAACTAT 7810


RESULT 27
ABL34509/c
ID   ABL34509 standard; DNA; 18988 BP.
XX
AC   ABL34509;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human metastasis associated gene SEQ ID NO: 62.
XX
KW   Metastasis associated gene; cytostatic; gene therapy; cancer;
KW   cytosine methylation; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200177376-A2.
XX
PD   18-OCT-2001.
XX
PF   06-APR-2001; 2001WO-EP003970.
XX
PR   06-APR-2000; 2000DE-01019058.
PR   07-APR-2000; 2000DE-01019173.
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-010922/01.
XX
PT   New nucleic acid derived from chemically treated metastasis genes, useful
PT   for diagnosis of cancers by analysis of cytosine methylation, also for
PT   treatment.
XX
PS   Claim 1; SEQ ID NO 62; 23pp + Sequence Listing; English.
XX
CC   The present invention provides a number of human metastasis associated
CC   genes which are modified by cytosine methylation. The sequences can be
CC   used in the diagnosis and treatment of cancer. The present sequence is
CC   one of the genes of the invention
XX
SQ   Sequence 18988 BP; 5467 A; 567 C; 4014 G; 8940 T; 0 U; 0 Other;

  Query Match             4.6%;  Score 50;  DB 6;  Length 18988;
  Best Local Similarity   47.4%;  Pred. No. 0.43;
  Matches  183;  Conservative    0;  Mismatches  200;  Indels    3;  Gaps    1;

Qy        146 AGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACA 205
              | ||| | | | |||||  ||| | | | || | |   | || |||   |     |
Db       8192 AAAAAAAATAACCATTATTAATTATTAACTACTCTACCAAAATATCATTTAAATTACTTC 8133

Qy        206 TTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTC 265
              |  || |   || | |  |  |  ||    | |       | ||| | |  || | || |
Db       8132 TATTTATAAAAAAAAATACATTAAATAAAATAATCTCCATTCACAAAAAATCAACAATCC 8073

Qy        266 ATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATA 325
              |||     ||    ||  |||||    |   || | ||  ||  | | |||      || |
Db       8072 ATTAAACTTCTCCTTAAACAAAATTTATCCTAAAATATAAATAAAAAAAAAATAATTAAA 8013

Qy        326 TTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTC 385
              |   |    | |   ||| | ||||  |    |||  |||  | |   ||   |   |
Db       8012 AAATAAAAAATATATAAAAAAAAAAAAAAAAAAAAATAACACAAAAATACCTCACCCTACC 7953

Qy        386 AAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAA 445
              |||  |||| |  |||||    |  ||      |||     ||||||||  |   |||    |
Db       7952 AAAAATAAACACACAACCGCTACCTACCTATCCTA---TCTTATCGTTCCCAAAAACACA 7896

Qy        446 ATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTA 505
              |        |||      |  |     ||||  ||  |||||||  |||  |  |  |  |||| |
Db       7895 AAACAACTAAAACCCGTCATCTTCGTAATACCAACAAACATATTTTATAAATTCATATAA 7836

Qy        506 AAACATTTACTTTCCAAATATTCTGT 531
              | |  |  | |||  | | |  || |
Db       7835 ATAACTAAATTTTATATAAAAACTAT 7810


RESULT 28
ABL70204/c
ID   ABL70204 standard; DNA; 18988 BP.
XX
AC   ABL70204;
XX
DT   01-JUL-2002  (first entry)
XX
DE   Chemically treated cell signalling DNA sequence complementary to#47.
XX
KW   Cell signalling; cytosine methylation; cell signalling disease; cancer;
KW   tumour; cytostatic; ds.
XX
OS   Unidentified.
XX
PN   WO200202807-A2.
XX
PD   10-JAN-2002.
XX
PF   29-JUN-2001; 2001WO-EP007471.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-154758/20.
XX
PT   Nucleic acid, useful for diagnosis and therapy of diseases associated
PT   with cell signaling e.g. cancer, comprises chemically modified genomic
PT   sequences of genes associated with cell signaling.
XX
PS   Claim 1; SEQ ID NO 94; 24pp + Sequence Listing; English.
XX
CC   The invention relates to a nucleic acid comprising a sequence of at least
CC   18 bases of a segment of chemically pretreated DNA of genes associated
CC   with cell signalling. The activity of the modified sequences of the
CC   invention may be described as cytostatic. The object of the invention is
CC   to provide the chemically modified DNA of genes associated with cell
CC   signalling, as well as oligonucleotides and/or PNA-oligomers for
CC   detecting cytosine methylations, as well as a method which is
CC   particularly suitable for the diagnosis and/or therapy of genetic and
CC   epigenetic parameters of genes associated with cell signalling. The
CC   chemically modified DNA provided by the invention is useful for diagnosis
CC   and therapy of diseases such as solid tumours and cancer. The sequences
CC   given in records ABL70111-ABL70626 represent chemically pre-treated
CC   genomic DNA's of genes associated with cell signalling. Note: The
CC   sequence data for this patent is not represented in the printed
CC   specification, but is based on sequence information supplied by the
CC   European Patent Office
XX
SQ   Sequence 18988 BP; 5467 A; 567 C; 4014 G; 8940 T; 0 U; 0 Other;

  Query Match             4.6%;  Score 50;  DB 6;  Length 18988;
  Best Local Similarity   47.4%;  Pred. No. 0.43;
```

```
  Matches  183;  Conservative    0;  Mismatches  200;  Indels    3;  Gaps    1;

Qy        146 AGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACA 205
              | ||| | | | |||||  ||| | | | || | |   | || |||   |     |
Db       8192 AAAAAAAATAACCATTATTAATTATTAACTACTCTACCAAAATATCATTTAAATTACTTC 8133

Qy        206 TTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTC 265
              |  || |   || | |  | |  ||    | |       | ||| | |  || | || |
Db       8132 TATTTATAAAAAAAAATACATTAAATAAAATAATCTCCATTCACAAAAAATCAACAATCC 8073

Qy        266 ATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATA 325
              |||    ||    ||  |||||    |  || | ||  ||  | | |||     || |
Db       8072 ATTAAACTTCTCCTTAAACAAAATTTATCCTAAAATATAAATAAAAAAAAAATAATTAAA 8013

Qy        326 TTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTC 385
              |   |    | |   ||| | ||||  |    |||  |||  | |   ||   |   |
Db       8012 AAATAAAAAATATATAAAAAAAAAAAAAAAAAAAAAATAACACAAAAATACCTCACCCTACC 7953

Qy        386 AAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAA 445
              |||  |||| |  |||||   |  ||     |||    |||||||  |   |||    |
Db       7952 AAAAATAAACACACAACCGCTACCTACCTATCCTA---TCTTATCGTTCCCAAAAACACA 7896

Qy        446 ATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTA 505
              |       |||    |  |    ||||  ||  ||||||| ||| | |  |  ||| |
Db       7895 AAACAACTAAAACCCGTCATCTTCGTAATACCAACAAACATATTTTATAAATTCATATAA 7836

Qy        506 AAACATTTACTTTCCAAATATTCTGT 531
              | |  |  | |||  | | |  || |
Db       7835 ATAACTAAATTTTATATAAAAACTAT 7810


RESULT 29
ADS99770/c
ID   ADS99770 standard; DNA; 18988 BP.
XX
AC   ADS99770;
XX
DT   02-DEC-2004  (first entry)
XX
DE   Complement of bisulphite treated metastasis-associated human gene #31.
XX
KW   Human; ds; gene; Bisulphite; metastasis; cancer; cytostatic;
KW   DNA methylation; matrix-assisted laser desorption/ionisation; MALDI;
KW   electrospray; mass spectrometry; CpG dinucleotide; solid tumour.
XX
OS   Homo sapiens.
XX
PN   US2003148327-A1.
XX
PD   07-AUG-2003.
XX
PF   21-JAN-2003; 2003US-00240485.
XX
PR   06-APR-2000; 2000DE-01019058.
PR   07-APR-2000; 2000DE-01019173.
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
PR   06-APR-2001; 2001WO-EP003970.
XX
PA   (OLEK/) OLEK A.
PA   (PIEP/) PIEPENBROCK C.
PA   (BERL/) BERLIN K.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-010922/01.
XX
PT   New nucleic acid derived from chemically treated metastasis genes, useful
PT   for diagnosis of cancers by analysis of cytosine methylation, also for
PT   treatment.
XX
PS   Claim 1; SEQ ID NO 62; 9pp; English.
XX
CC   The invention relates to a nucleic acid comprising at least 18 bases from
CC   a segment of the chemically pretreated DNA of genes associated with
CC   metastasis, i.e. any of ADS99709-ADS99906 human genomic sequences or any
CC   of the 19 sequences appearing as ADS99911-ADS99929. SEQ ID 2,4,6 etc are
CC   the complements of SEQ ID 1,3,5, etc. Also included are an oligomer
CC   (particularly an oligonucleotide or peptide nucleic acid) comprising at
CC   least one base sequence of at least 9 bases which hybridises to (or is
CC   identical with) the sequences referred to above, producing an array of
CC   the oligomers on a carrier, obtaining genetic and/or epigenetic
CC   parameters for diagnosis and/or therapy of diseases (or predisposition to
CC   them) by analysis of cytosine methylation and a kit comprising a
CC   bisulphite (disulphite or hydrogen sulphite) and the oligomers. In the
CC   method of above 5-unmethylated cytosines in a genomic DNA sample are
CC   converted chemically to uracil, or another base with hybridisation
CC   properties different from those of cytosine, then fragments of the
CC   treated DNA amplified (particularly by polymerase chain reaction) using
CC   the oligomers and a polymerase (preferably heat stable) to produce
CC   labelled amplicons. These are tested for hybridisation to an array of
CC   oligomers and any hybridisation detected. The amplicons are labelled with
CC   fluorescent or radioactive markers, or with a detachable mass marker to
CC   allow their detection by mass spectrometry, specifically using the matrix
CC   -assisted laser desorption/ionisation (MALDI) or electrospray techniques.
CC   To improve detection in the mass spectrometer, fragments formed in the
CC   instrument have only a single net charge (positive or negative). The
CC   genomic DNA is from e.g. a cell line, biopsy sample, blood, or paraffin-
CC   embedded tissue sample. Oligonucleotides or peptide-nucleic acids that
CC   are complementary to (or identical with) parts of the nuclei acids listed
CC   above may be used as primers for amplification of the nucleic acids or
CC   their complements, and for determining cytosine methylation status and/or
CC   single nucleotide polymorphisms in metastasis-related genes. They can be
CC   used for analysis of diseases associated with methylation of CpG
CC   dinucleotides and to determine (epi)genetic parameters for diagnosis
CC   and/or therapy of disease (or predisposition). The genomic DNA sequences
CC   are useful for diagnosis and therapy of solid tumours and cancer. The
CC   present sequence is the complementary sequence to a bisulphite treated
CC   human gene associated with metastasis. Note: The sequence data for this
CC   patent did not form part of the printed specification, but was obtained
CC   in electronic format directly from USPTO at
CC   seqdata.uspto.gov/sequence.html?DocID=20030148327.
XX
SQ   Sequence 18988 BP; 5467 A; 567 C; 4014 G; 8940 T; 0 U; 0 Other;

  Query Match             4.6%;  Score 50;  DB 7;  Length 18988;
  Best Local Similarity   47.4%;  Pred. No. 0.43;
  Matches  183;  Conservative    0;  Mismatches  200;  Indels    3;  Gaps    1;

Qy        146 AGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACA 205
              | ||| | | | |||||  ||| | | | || | |   | || |||   |     |
Db       8192 AAAAAAAATAACCATTATTAATTATTAACTACTCTACCAAAATATCATTTAAATTACTTC 8133

Qy        206 TTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTC 265
              |  || |   || | |  | |  ||    | |       | ||| | |  || | || |
Db       8132 TATTTATAAAAAAAAATACATTAAATAAAATAATCTCCATTCACAAAAAATCAACAATCC 8073

Qy        266 ATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATA 325
              |||    ||    ||  |||||    |  || | ||  ||  | | |||     || |
Db       8072 ATTAAACTTCTCCTTAAACAAAATTTATCCTAAAATATAAATAAAAAAAAAATAATTAAA 8013

Qy        326 TTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTC 385
              |   |    | |   ||| | ||||  |    |||  |||  | |   ||   |   |
Db       8012 AAATAAAAAATATATAAAAAAAAAAAAAAAAAAAAAATAACACAAAAATACCTCACCCTACC 7953

Qy        386 AAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAA 445
              |||  |||| |  |||||   |  ||     |||    |||||||  |   |||    |
Db       7952 AAAAATAAACACACAACCGCTACCTACCTATCCTA---TCTTATCGTTCCCAAAAACACA 7896

Qy        446 ATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTA 505
              |       |||    |  |    ||||  ||  ||||||| ||| | |  |  ||| |
Db       7895 AAACAACTAAAACCCGTCATCTTCGTAATACCAACAAACATATTTTATAAATTCATATAA 7836

Qy        506 AAACATTTACTTTCCAAATATTCTGT 531
              | |  |  | |||  | | |  || |
Db       7835 ATAACTAAATTTTATATAAAAACTAT 7810
```

RESULT 30
ABL33597/c
ID ABL33597 standard; DNA; 5195 BP.
XX
AC ABL33597;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 1570.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 1570; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 5195 BP; 1437 A; 119 C; 1105 G; 2534 T; 0 U; 0 Other;

Query Match 4.6%; Score 49.6; DB 6; Length 5195;
Best Local Similarity 46.7%; Pred. No. 0.38;
Matches 157; Conservative 0; Mismatches 179; Indels 0; Gaps 0;

```
Qy       173 AATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATT 232
Db      2734 AATAAATATTTTATAAAATTTACATAAAATAACTAATTTCAATAAAAAAAATTATATCTA 2675

Qy       233 TTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGG 292
Db      2674 TTAATTAAACACCTACTATAAACAAATCTCTTAATATACCCAATAAAAAAATAAAAAAAA 2615

Qy       293 TATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAA 352
Db      2614 TAAAAACAAATACAAAAATAAATAAAAAAAACATAATCTATCTCTCCATAAAACATAACA 2555

Qy       353 TGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTAT 412
```

```
Db      2554 AAAAAATAAATAAAATTTTATATACATTATATACGATCTATAAAATATAATTTAAAACAT 2495

Qy       413 ACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATA 472
Db      2494 AATTTTACCTACAAAAAAAAAAACAAAAAAAATATTTATAAAAAAAATATTTACTTTATT 2435

Qy       473 ATCTCAGAAAACATAATTTTTTATGTACTATTAAAA 508
Db      2434 TTCTAATTCACGTTTTCATCATAAAAAATTTTAAAA 2399
```

RESULT 31
ABN80014/c
ID ABN80014 standard; DNA; 6887 BP.
XX
AC ABN80014;
XX
DT 15-JUL-2002 (first entry)
XX
DE Human chemically modified disease associated gene SEQ ID NO 31.
XX
KW Human; development; homeobox gene; HOX; diabetes; cancer; apoptosis;
KW heart disease; epilepsy; histone deacetylation; muscular dystrophy;
KW dwarfism; single nucleotide polymorphism; SNP; cytosine methylation;
KW antidiabetic; cytostatic; anticonvulsant; ds.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200200927-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007536.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130908/17.
XX
PT Novel nucleic acid useful for diagnosis and therapy of diseases
PT associated with development genes such as diabetes, comprises a sequence
PT of a segment of chemically pretreated DNA of genes associated with
PT development.
XX
PS Claim 1; SEQ ID NO 31; 27pp; English.
XX
CC The invention relates to a nucleic acid (I) comprising a sequence at
CC least 18 bases in length of a segment of chemically pretreated DNA (II)
CC of genes associated with development selected from 87 genes listed in the
CC specification such as ACCPN, ADFN, or AFD1 and comprising one of 350
CC sequences (ABN79984-ABN80333) or their complements. The invention is
CC useful for the diagnosis or therapy of diseases associated with
CC development genes, in particular disease related to homeobox containing
CC genes (HOX), like diabetes, cancer, apoptosis related diseases, syndromes
CC associated with congenital heart disease, epilepsy, diseases related to
CC histone deacetylation, Currarino syndrome, diseases related with the
CC development of the brain and limb girdle muscular dystrophy and dwarfism.
CC Oligomers specific to each of the genes are useful for detecting the
CC methylation state of all CpG dinucleotides within the 350 sequences or
CC (II) and their complementary sequences, as primer oligonucleotides for
CC the amplification of the 350 sequences, (II) and/or their complements and
CC as oligomer probes for detecting the cytosine methylation state and/or
CC single nucleotide polymorphisms (SNPs). Note: The sequence data for this
CC patent did not form part of the printed specification but is based on
CC sequence information supplied to Derwent by the European Patent Office
XX
SQ Sequence 6887 BP; 1594 A; 403 C; 1983 G; 2907 T; 0 U; 0 Other;

```
  Query Match             4.6%;  Score 49.6;  DB 6;  Length 6887;
  Best Local Similarity   48.0%;  Pred. No. 0.41;
  Matches  142;  Conservative    0;  Mismatches  154;  Indels    0;  Gaps    0;

Qy        234 TTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGT 293
              | | |||| |  || |   | |     | || | ||   || | ||| |  |||||  |
Db       3883 TCTAAAAATATTTTTCTTTCCTCGTCGTCTTTAATTTAATTCGATTTTTATAAAAAAATT 3824

Qy        294 ATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAAT 353
              ||||| ||||    | |   |||  |   |  |   |  | ||||   | |   |||
Db       3823 ATAAAAAAATAATATAATTAAAACTAAAATCAAAAAAAAAAATTAAAAAACATCTTTAAAA 3764

Qy        354 GAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATA 413
              | |   ||| |  ||  | |   || |  | | |    || | | ||  |||    |
Db       3763 AAATTTCTATAAATTTTATAATCGTTTCCCTAACAAACCAATATTAAATTTAATTAAAAT 3704

Qy        414 CTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAA 473
              ||     |||   |||  |    ||| | ||| |  ||     ||| ||||  | ||
Db       3703 TTATTAAATTTTACATCAAAAAAAAAAAAAAAAACTTTCAATTTATTTTTTTTAAATATCT 3644

Qy        474 TCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCT 529
              |    | | || |  |   ||     || || | ||||| || |  || | | |
Db       3643 CTTACTTATAAATTAAATAAAATTATAAATAAATATATTTATTTACTTAAAAATAT 3588


RESULT 32
AAS46288
ID   AAS46288 standard; DNA; 8169 BP.
XX
AC   AAS46288;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Tumour suppressor gene derived chemically modified sequence #10.
XX
KW   Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer;
KW   tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW   cytosine methylation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200168912-A2.
XX
PD   20-SEP-2001.
XX
PF   15-MAR-2001; 2001WO-EP002955.
XX
PR   15-MAR-2000; 2000DE-01013847.
PR   06-APR-2000; 2000DE-01019058.
PR   07-APR-2000; 2000DE-01019173.
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2001-602752/68.
XX
PT   Fragments of chemically modified genes associated with tumor suppressor
PT   genes and oncogenes, useful in designing primers and probes for analyzing
PT   diseases associated with cytosine methylation state e.g. cancer.
XX
PS   Claim 1; SEQ ID NO 10; 27pp; English.
XX
CC   The invention relates to a nucleic acid comprising a sequence of 18
CC   bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC   bisulphite, of genes associated with tumour suppression and oncogenes
CC   having a sequence taken from 536 (actually 533 since numbers 408, 458 and
CC   500 are missing from the sequence listing) sequences (Ss) and sequences
CC   complementary to (Ss). The nucleic acid may be a peptide nucleic acid-
CC   oligomer (PNA) of at least 9 nucleotides and may form part of a set of
CC   probes for detecting the cytosine methylation state and/or single
CC   nucleotide polymorphisms and also to be used in an array for analysing
CC   diseases associated with CpG dinucleotides e.g. cancers and tumours. The
CC   probes can also be used in a method for ascertaining genetic and/or
CC   epigenetic parameters for the diagnosis and/or therapy of existing
CC   diseases or the predisposition to specific diseases, by analysing
CC   cytosine methylations. The parameters may be compared to another set of
CC   genetic and/or epigenetic parameters, the differences serving as basis
CC   for diagnosis and/or prognosis events which are disadvantageous to
CC   patients. The present sequence is one of the 533 genomic sequences
CC   derived from tumour suppressor genes and oncogenes. Sequences with even
CC   numbered Seq ID numbers are the complementary sequence of the
CC   corresponding odd numbered sequence (e.g. ID 2 and ID1, ID 536 and ID
CC   535, except for those whose partner sequence is missing). Note: The
CC   sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 8169 BP; 2364 A; 116 C; 1712 G; 3969 T; 0 U; 8 Other;

  Query Match             4.6%;  Score 49.6;  DB 4;  Length 8169;
  Best Local Similarity   45.7%;  Pred. No. 0.43;
  Matches  172;  Conservative    0;  Mismatches  204;  Indels    0;  Gaps    0;

Qy         35 TCTATGTTCTTATAAGCAAATGTTACAATGAGATATGAGTTTTAAAGCCAGATCTTCCTT 94
              | | ||   ||| |    || |||| | |||| ||||| | |   |      | |   ||
Db       6987 TTTTTGAGTTTAGAGTTTAAGGTTAGAGTGAGTTATGATTATATTATTGTATTTTAGTTT 7046

Qy         95 ATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATATT 154
                 |    |    |  | | ||| || | |       |||  ||||   ||      | |
Db       7047 GGGTAATAGAGTGAGTTTTTGTTTTTAAGGAAAAAAATAAAAGTTGTATTTTTTGAGTTA 7106

Qy        155 GATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGT 214
                 |   | |  |  ||   || |||| | |    | |      |||| |  || | ||
Db       7107 TTTATAGATATTTATGTATTTAATTTTATGGTTTTTATGTGGGTAAATATTTTATGTTAA 7166

Qy        215 TAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTT 274
              ||   ||     ||  ||  ||||||| |  ||  |   ||      || ||||  | |
Db       7167 TAGGTTATATAATTATTTAATTTAAAAAAATTTGAAATAATGATTAGTTTTATTTAATAT 7226

Qy        275 TGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGT 334
              |    | ||   || | ||||   | |    || | | | || |  | | |   |
Db       7227 TTTAATTTTTGTAAGAAATATAGGTATAATTGTTAAAAAATAAATTAATTGTAAATGAAA 7286

Qy        335 TGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAA 394
              | ||  || |  |   |||   || |  ||| | |  | ||  ||   || |||    |
Db       7287 TAAATGTTTATAAATGAATTTATGTAATTTAAAGTTTTAAAATTTTATGTTAAATTATAT 7346

Qy        395 AAGTCAACCTAAAGGT 410
              |  | ||   || | |
Db       7347 ATATAAAATAAATGTT 7362


RESULT 33
ADS89254
ID   ADS89254 standard; DNA; 8169 BP.
XX
AC   ADS89254;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Oligonucleotide of the invention SEQ ID NO:270.
XX
KW   ss; cell proliferative disorder; breast; methylation; cytostatic;
KW   gene therapy; single nucleotide polymorphism; SNP.
XX
OS   Unidentified.
XX
PN   WO2004035803-A2.
XX
PD   29-APR-2004.
XX
PF   01-OCT-2003; 2003WO-EP010881.
XX
```

PR 01-OCT-2002; 2002DE-01045779.
PR 07-JAN-2003; 2003DE-01000096.
PR 17-APR-2003; 2003DE-01017955.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Foekens J, Harbeck N, Koenig T, Maier S, Martens J, Model F;
PI Nimmrich I, Rujan T, Schmitt A, Schmitt M, Look MP, Marx A;
XX
DR WPI; 2004-348468/32.
XX
PT Predicting responsiveness of a subject with breast cell proliferative
PT disorder, useful for treating or differentiating breast cell
PT proliferative disorders comprises analyzing methylation pattern of a
PT genomic DNA from the subject.
XX
PS Disclosure; SEQ ID NO 270; 104pp; English.
XX
CC The invention relates to a novel method for predicting the responsiveness
CC of a subject with a cell proliferative disorder of the breast tissues to
CC a therapy comprising analysing the methylation pattern of a target
CC nucleic acid by contacting at least one of the target nucleic acids in a
CC biological sample obtained from the subject prior to or during treatment.
CC The method of the invention has cytostatic activity, and may have a use
CC in gene therapy. The set of oligonucleotides comprising at least two of
CC the oligomers are useful for detecting the cytosine methylation state
CC and/or single nucleotide polymorphisms (SNPs) within the sequences. The
CC methods, nucleic acid, oligonucleotide, and kit are useful for the
CC treatment, characterisation, classification and/or differentiation, of
CC breast cell proliferative disorders. The method is also useful for
CC predicting the responsiveness of a subject with a cell proliferative
CC disorder of the breast tissues to a therapy. The present sequence is used
CC in the exemplification of the invention.
XX
SQ Sequence 8169 BP; 2364 A; 116 C; 1712 G; 3969 T; 0 U; 8 Other;

Query Match 4.6%; Score 49.6; DB 13; Length 8169;
Best Local Similarity 45.7%; Pred. No. 0.43;
Matches 172; Conservative 0; Mismatches 204; Indels 0; Gaps 0;

```
Qy        35 TCTATGTTCTTATAAGCAAATGTTACAATGAGATATGAGTTTTAAAGCCAGATCTTCCTT 94
             | | ||   ||| |    || |||| | |||| ||||| | |   |      | |   ||
Db      6987 TTTTTGAGTTTAGAGTTTAAGGTTAGAGTGAGTTATGATTATATTATTGTATTTTAGTTT 7046

Qy        95 ATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATATT 154
               |    |   | | ||| || | |      ||| |||   ||     | |
Db      7047 GGGTAATAGAGTGAGTTTTTGTTTTTAAGGAAAAAAATAAAAGTTGTATTTTTTGAGTTA 7106

Qy       155 GATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGT 214
               |  | |  |  ||   || |||| | |    | |     |||| |  || | ||
Db      7107 TTTATAGATATTTATGTATTTAATTTTATGGTTTTTATGTGGGTAAATATTTTATGTTAA 7166

Qy       215 TAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTT 274
             ||   ||     ||  ||  ||||||| |  ||  |   ||       || ||||  | |
Db      7167 TAGGTTATATAATTATTTAATTTAAAAAAATTTGAAATAATGATTAGTTTTATTTAATAT 7226

Qy       275 TGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGT 334
             |    | ||   || |  ||||   | |    || | | | | || |  | | |   |
Db      7227 TTTAATTTTTGTAAGAAATATAGGTATAATTGTTAAAAAATAAATTAATTGTAAATGAAA 7286

Qy       335 TGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAA 394
             | ||  || |  |   |||   || |  ||| | |  | ||  ||   || |||    |
Db      7287 TAAATGTTTATAAATGAATTTATGTAATTTAAAGTTTTAAAATTTTATGTTAAATTATAT 7346

Qy       395 AAGTCAACCTAAAGGT 410
             |  | ||   || | |
Db      7347 ATATAAAATAAATGTT 7362
```

RESULT 34
ADS89528
ID ADS89528 standard; DNA; 8169 BP.
XX
AC ADS89528;
XX
DT 18-NOV-2004 (first entry)
XX
DE Oligonucleotide of the invention SEQ ID NO:544.
XX
KW ss; cell proliferative disorder; breast; methylation; cytostatic;
KW gene therapy; single nucleotide polymorphism; SNP.
XX
OS Unidentified.
XX
PN WO2004035803-A2.
XX
PD 29-APR-2004.
XX
PF 01-OCT-2003; 2003WO-EP010881.
XX
PR 01-OCT-2002; 2002DE-01045779.
PR 07-JAN-2003; 2003DE-01000096.
PR 17-APR-2003; 2003DE-01017955.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Foekens J, Harbeck N, Koenig T, Maier S, Martens J, Model F;
PI Nimmrich I, Rujan T, Schmitt A, Schmitt M, Look MP, Marx A;
XX
DR WPI; 2004-348468/32.
XX
PT Predicting responsiveness of a subject with breast cell proliferative
PT disorder, useful for treating or differentiating breast cell
PT proliferative disorders comprises analyzing methylation pattern of a
PT genomic DNA from the subject.
XX
PS Disclosure; SEQ ID NO 544; 104pp; English.
XX
CC The invention relates to a novel method for predicting the responsiveness
CC of a subject with a cell proliferative disorder of the breast tissues to
CC a therapy comprising analysing the methylation pattern of a target
CC nucleic acid by contacting at least one of the target nucleic acids in a
CC biological sample obtained from the subject prior to or during treatment.
CC The method of the invention has cytostatic activity, and may have a use
CC in gene therapy. The set of oligonucleotides comprising at least two of
CC the oligomers are useful for detecting the cytosine methylation state
CC and/or single nucleotide polymorphisms (SNPs) within the sequences. The
CC methods, nucleic acid, oligonucleotide, and kit are useful for the
CC treatment, characterisation, classification and/or differentiation, of
CC breast cell proliferative disorders. The method is also useful for
CC predicting the responsiveness of a subject with a cell proliferative
CC disorder of the breast tissues to a therapy. The present sequence is used
CC in the exemplification of the invention.
XX
SQ Sequence 8169 BP; 2364 A; 0 C; 1712 G; 4085 T; 0 U; 8 Other;

Query Match 4.6%; Score 49.6; DB 13; Length 8169;
Best Local Similarity 45.7%; Pred. No. 0.43;
Matches 172; Conservative 0; Mismatches 204; Indels 0; Gaps 0;

```
Qy        35 TCTATGTTCTTATAAGCAAATGTTACAATGAGATATGAGTTTTAAAGCCAGATCTTCCTT 94
             | | ||   ||| |    || |||| | |||| ||||| | |   |      | |   ||
Db      6987 TTTTTGAGTTTAGAGTTTAAGGTTAGAGTGAGTTATGATTATATTATTGTATTTTAGTTT 7046

Qy        95 ATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATATT 154
               |    |   | | ||| || | |      ||| |||   ||     | |
Db      7047 GGGTAATAGAGTGAGTTTTTGTTTTTAAGGAAAAAAATAAAAGTTGTATTTTTTGAGTTA 7106

Qy       155 GATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGT 214
               |  | |  |  ||   || |||| | |    | |     |||| |  || | ||
Db      7107 TTTATAGATATTTATGTATTTAATTTTATGGTTTTTATGTGGGTAAATATTTTATGTTAA 7166

Qy       215 TAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTT 274
             ||   ||     ||  ||  ||||||| |  ||  |   ||       || ||||  | |
Db      7167 TAGGTTATATAATTATTTAATTTAAAAAAATTTGAAATAATGATTAGTTTTATTTAATAT 7226
```

```
Qy        275 TGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGT 334
              |    | ||   || |  ||||   | |    || | | | | | || |  | | |   |
Db       7227 TTTAATTTTTGTAAGAAATATAGGTATAATTGTTAAAAAATAAATTAATTGTAAATGAAA 7286

Qy        335 TGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAA 394
              | ||  || | |   |||   || | ||| | | | ||  ||   || |||     |
Db       7287 TAAATGTTTATAAATGAATTTATGTAATTTAAAGTTTTAAAATTTTATGTTAAATTATAT 7346

Qy        395 AAGTCAACCTAAAGGT 410
              | | ||   || | |
Db       7347 ATATAAAATAAATGTT 7362


RESULT 35
ABL34611/c
ID   ABL34611 standard; DNA; 20486 BP.
XX
AC   ABL34611;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human metastasis associated gene SEQ ID NO: 164.
XX
KW   Metastasis associated gene; cytostatic; gene therapy; cancer;
KW   cytosine methylation; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200177376-A2.
XX
PD   18-OCT-2001.
XX
PF   06-APR-2001; 2001WO-EP003970.
XX
PR   06-APR-2000; 2000DE-01019058.
PR   07-APR-2000; 2000DE-01019173.
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-010922/01.
XX
PT   New nucleic acid derived from chemically treated metastasis genes, useful
PT   for diagnosis of cancers by analysis of cytosine methylation, also for
PT   treatment.
XX
PS   Claim 1; SEQ ID NO 164; 23pp + Sequence Listing; English.
XX
CC   The present invention provides a number of human metastasis associated
CC   genes which are modified by cytosine methylation. The sequences can be
CC   used in the diagnosis and treatment of cancer. The present sequence is
CC   one of the genes of the invention
XX
SQ   Sequence 20486 BP; 6101 A; 360 C; 4446 G; 9577 T; 0 U; 2 Other;

  Query Match             4.6%;  Score 49.6;  DB 6;  Length 20486;
  Best Local Similarity   47.7%;  Pred. No. 0.53;
  Matches  145;  Conservative    0;  Mismatches  159;  Indels    0;  Gaps    0;

Qy        231 TTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAA 290
              ||||| |||   | |  | |   ||| | | ||| |  |   ||       |  || | |
Db      12378 TTTTTATAATTTAACAAATATTTATAAAACAATTTACATTAATTAACCCCTTAAAACACA 12319

Qy        291 GGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAA 350
              | |||       |     |  || |  ||     |  || |  || |  | | | |
Db      12318 AAAAAAAACCTTCAAAACATCATCAAAAATATTCTAAATACAATAAAATAACATATATAT 12259

Qy        351 AATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGT 410
              |||  |       ||  || |  | ||  |  | |  ||   ||   |  ||||   ||
Db      12258 AATAAAATATATTTTAATAAAATCATAATAAAATATAAAACAATTAACCAACAATAATCA 12199

Qy        411 ATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAA 470
              | |  ||||   ||||  ||     |  || |  ||||||  | || || ||||||  ||
Db      12198 AAATAACCTAACTTCTCTTCCACTTTTAATTTCAAATTAAAATAATTTATATTTTTTAAC 12139

Qy        471 TAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTG 530
              |     | |||  | |||||| || |  | |||| || | |        ||||  | |
Db      12138 ATAAAAAACAAATTAAAATTTTTAATCTTATTTTAATACTTATTACCGATAAATCATTTT 12079

Qy        531 TCAT 534
              ||
Db      12078 ATAT 12075


RESULT 36
ADS99872/c
ID   ADS99872 standard; DNA; 20486 BP.
XX
AC   ADS99872;
XX
DT   02-DEC-2004  (first entry)
XX
DE   Complement of bisulphite treated metastasis-associated human gene #82.
XX
KW   Human; ds; gene; Bisulphite; metastasis; cancer; cytostatic;
KW   DNA methylation; matrix-assisted laser desorption/ionisation; MALDI;
KW   electrospray; mass spectrometry; CpG dinucleotide; solid tumour.
XX
OS   Homo sapiens.
XX
PN   US2003148327-A1.
XX
PD   07-AUG-2003.
XX
PF   21-JAN-2003; 2003US-00240485.
XX
PR   06-APR-2000; 2000DE-01019058.
PR   07-APR-2000; 2000DE-01019173.
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
PR   06-APR-2001; 2001WO-EP003970.
XX
PA   (OLEK/) OLEK A.
PA   (PIEP/) PIEPENBROCK C.
PA   (BERL/) BERLIN K.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-010922/01.
XX
PT   New nucleic acid derived from chemically treated metastasis genes, useful
PT   for diagnosis of cancers by analysis of cytosine methylation, also for
PT   treatment.
XX
PS   Claim 1; SEQ ID NO 164; 9pp; English.
XX
CC   The invention relates to a nucleic acid comprising at least 18 bases from
CC   a segment of the chemically pretreated DNA of genes associated with
CC   metastasis, i.e. any of ADS99709-ADS99906 human genomic sequences or any
CC   of the 19 sequences appearing as ADS99911-ADS99929. SEQ ID 2,4,6 etc are
CC   the complements of SEQ ID 1,3,5, etc. Also included are an oligomer
CC   (particularly an oligonucleotide or peptide nucleic acid) comprising at
CC   least one base sequence of at least 9 bases which hybridises to (or is
CC   identical with) the sequences referred to above, producing an array of
CC   the oligomers on a carrier, obtaining genetic and/or epigenetic
CC   parameters for diagnosis and/or therapy of diseases (or predisposition to
CC   them) by analysis of cytosine methylation and a kit comprising a
CC   bisulphite (disulphite or hydrogen sulphite) and the oligomers. In the
CC   method of above 5-unmethylated cytosines in a genomic DNA sample are
CC   converted chemically to uracil, or another base with hybridisation
CC   properties different from those of cytosine, then fragments of the
CC   treated DNA amplified (particularly by polymerase chain reaction) using
CC   the oligomers and a polymerase (preferably heat stable) to produce
```

```
CC   labelled amplicons. These are tested for hybridisation to an array of
CC   oligomers and any hybridisation detected. The amplicons are labelled with
CC   fluorescent or radioactive markers, or with a detachable mass marker to
CC   allow their detection by mass spectrometry, specifically using the matrix
CC   -assisted laser desorption/ionisation (MALDI) or electrospray techniques.
CC   To improve detection in the mass spectrometer, fragments formed in the
CC   instrument have only a single net charge (positive or negative). The
CC   genomic DNA is from e.g. a cell line, biopsy sample, blood, or paraffin-
CC   embedded tissue sample. Oligonucleotides or peptide-nucleic acids that
CC   are complementary to (or identical with) parts of the nuclei acids listed
CC   above may be used as primers for amplification of the nucleic acids or
CC   their complements, and for determining cytosine methylation status and/or
CC   single nucleotide polymorphisms in metastasis-related genes. They can be
CC   used for analysis of diseases associated with methylation of CpG
CC   dinucleotides and to determine (epi)genetic parameters for diagnosis
CC   and/or therapy of disease (or predisposition). The genomic DNA sequences
CC   are useful for diagnosis and therapy of solid tumours and cancer. The
CC   present sequence is the complementary sequence to a bisulphite treated
CC   human gene associated with metastasis. Note: The sequence data for this
CC   patent did not form part of the printed specification, but was obtained
CC   in electronic format directly from USPTO at
CC   seqdata.uspto.gov/sequence.html?DocID=20030148327.
XX
SQ   Sequence 20486 BP; 6101 A; 360 C; 4446 G; 9577 T; 0 U; 2 Other;

  Query Match             4.6%;  Score 49.6;  DB 7;  Length 20486;
  Best Local Similarity   47.7%;  Pred. No. 0.53;
  Matches  145;  Conservative    0;  Mismatches  159;  Indels    0;  Gaps    0;

Qy        231 TTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAA 290
              ||||| |||   | |  | |   ||| | | ||| |  |   ||       |  || | |
Db      12378 TTTTTATAATTTAACAAATATTTATAAAACAATTTACATTAATTAACCCCTTAAAACACA 12319

Qy        291 GGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAA 350
              | |||      |   |  || |  ||     |  || |  || |  | | | |
Db      12318 AAAAAAAACCTTCAAAACATCATCAAAAATATTCTAAATACAATAAAATAACATATATAT 12259

Qy        351 AATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGT 410
              |||  |       ||  || |  | ||  |  | |  ||   ||  |  ||||   ||
Db      12258 AATAAAATATATTTTAATAAAATCATAATAAAATATAAAACAATTAACCAACAATAATCA 12199

Qy        411 ATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAA 470
              | |  ||||   ||||  ||     |  || |  ||||||  | || || ||||||  ||
Db      12198 AAATAACCTAACTTCTCTTCCACTTTTAATTTCAAATTAAAATAATTTATATTTTTTAAC 12139

Qy        471 TAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTG 530
              |    | |||  | |||||| || |  | ||||  || | |       ||||  | |
Db      12138 ATAAAAAACAAATTAAAATTTTTAATCTTATTTTAATACTTATTACCGATAAATCATTTT 12079

Qy        531 TCAT 534
              ||
Db      12078 ATAT 12075


RESULT 37
ABX08336_14/c
Continuation (15 of 17) of ABX08336 from base 1400001 (Human phosphodiesterase 4D (PDE4D
WP Sequence split into 17 fragments  LOCUS ABX08336 Accession Abx08336
WP     Fragment Name      Begin     End
WP     ABX08336_00            1    110000
WP     ABX08336_01       100001    210000
WP     ABX08336_02       200001    310000
WP     ABX08336_03       300001    410000
WP     ABX08336_04       400001    510000
WP     ABX08336_05       500001    610000
WP     ABX08336_06       600001    710000
WP     ABX08336_07       700001    810000
WP     ABX08336_08       800001    910000
WP     ABX08336_09       900001   1010000
WP     ABX08336_10      1000001   1110000
WP     ABX08336_11      1100001   1210000
WP     ABX08336_12      1200001   1310000
WP     ABX08336_13      1300001   1410000
WP     ABX08336_14      1400001   1510000
WP     ABX08336_15      1500001   1610000
WP     ABX08336_16      1600001   1691080

  Query Match             4.6%;  Score 49.6;  DB 6;  Length 110000;
  Best Local Similarity   46.8%;  Pred. No. 0.81;
  Matches  189;  Conservative    0;  Mismatches  214;  Indels    1;  Gaps    1;

Qy        130 CATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGA 189
              |||||   |||   ||||||||||   |  |    ||| ||   |||  |  | |  | |
Db      58405 CATATATATTTATATGAGAAATATATTTATTATATAATAAGCATATATATGAGAAATACA 58346

Qy        190 TCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTAC 249
              |   ||  |||| |  |   ||| || ||||     |  |||| | ||  | |  ||
Db      58345 TTATATAATAAATAGTTATATTTATTTATATGAGAATTATATTTATATATTATATATTTA 58286

Qy        250 AGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGG-TATAAAGAAATGCATT 308
              |||  |  |||    || | | |  ||   | | | | |  ||||   |     |
Db      58285 TATAATAAATATATAATATTATATATATTTATATAATATATGTATATAATATATAATACA 58226

Qy        309 CAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACAT 368
              | | ||  |  || |||      |  | | || || | |  ||||  ||  ||| || ||
Db      58225 TATATAATAAATATGTATATAATATATAATAAATATATAATAAATATGTATATAATATAT 58166

Qy        369 TACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTA 428
              |  |   |     |  | | |||  |      |                 | | | |
Db      58165 AATAAATATATATTTATACAATTTATAAGAAATATATATTTATATAAGAAATATATATAA 58106

Qy        429 TCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAA 488
              |   |    | ||| ||  ||  | | ||| ||| ||  | ||  | |    | | |||
Db      58105 TTTTATAAAATATATAATATATAATATATTATATATTATATATTTTGTATATATATATAT 58046

Qy        489 TTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGTC 532
              | |    |  ||| || | || || ||    |   ||| |||
Db      58045 AATATATATTTATATAAATATATATATTTAAAGAGGTTTCAGTC 58002


RESULT 38
ADJ25985_14/c
Continuation (15 of 17) of ADJ25985 from base 1400001 (Human phosphodiesterase 4D (PDE4D
WP Sequence split into 17 fragments  LOCUS ADJ25985 Accession Adj25985
WP     Fragment Name      Begin     End
WP     ADJ25985_00            1    110000
WP     ADJ25985_01       100001    210000
WP     ADJ25985_02       200001    310000
WP     ADJ25985_03       300001    410000
WP     ADJ25985_04       400001    510000
WP     ADJ25985_05       500001    610000
WP     ADJ25985_06       600001    710000
WP     ADJ25985_07       700001    810000
WP     ADJ25985_08       800001    910000
WP     ADJ25985_09       900001   1010000
WP     ADJ25985_10      1000001   1110000
WP     ADJ25985_11      1100001   1210000
WP     ADJ25985_12      1200001   1310000
WP     ADJ25985_13      1300001   1410000
WP     ADJ25985_14      1400001   1510000
WP     ADJ25985_15      1500001   1610000
WP     ADJ25985_16      1600001   1691139

  Query Match             4.6%;  Score 49.6;  DB 12;  Length 110000;
  Best Local Similarity   46.8%;  Pred. No. 0.81;
  Matches  189;  Conservative    0;  Mismatches  214;  Indels    1;  Gaps    1;

Qy        130 CATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGA 189
              |||||   |||   ||||||||||   |  |    ||| ||   |||  |  | |  | |
Db      58464 CATATATATTTATATGAGAAATATATTTATTATATAATAAGCATATATATGAGAAATACA 58405

Qy        190 TCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTAC 249
              |   ||  |||| |  |   ||| || ||||     |  |||| | ||  | |  ||
Db      58404 TTATATAATAAATAGTTATATTTATTTATATGAGAATTATATTTATATATTATATATTTA 58345

Qy        250 AGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGG-TATAAAGAAATGCATT 308
```

```
                  ||| |  |||    || | | |  ||   | | | | |  |||||   |    |
Db       58344 TATAATAAATATATAATATTATATATATTTATATAATATATGTATATAATATATAATACA 58285

Qy         309 CAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACAT 368
               | | ||  |  || |||     |  | | || || | |  ||||  ||  ||| || ||
Db       58284 TATATAATAAATATGTATATAATATATAATAAATATATAATAAATATGTATATAATATAT 58225

Qy         369 TACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTA 428
               |   |    |       | |  | | |||  |      |                 | | | |
Db       58224 AATAAATATATATTTATACAATTTATAAGAAATATATATTTATATAAGAAATATATATAA 58165

Qy         429 TCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAA 488
               |   |    | ||| ||  ||  | | ||| ||| ||  | ||  | |    | | |||
Db       58164 TTTTATAAAATATATAATATATAATATATTATATATTATATATTTTGTATATATATATAT 58105

Qy         489 TTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGTC 532
                | |    |  ||| || | || || ||    |   ||| |||
Db       58104 AATATATATTTATATAAATATATATATTTAAAGAGGTTTCAGTC 58061


RESULT 39
ADN97989_14/c
Continuation (15 of 17) of ADN97989 from base 1400001 (Human phosphodiesterase 4D genomi
WP Sequence split into 17 fragments  LOCUS ADN97989 Accession Adn97989
WP     Fragment Name      Begin     End
WP     ADN97989_00            1    110000
WP     ADN97989_01       100001    210000
WP     ADN97989_02       200001    310000
WP     ADN97989_03       300001    410000
WP     ADN97989_04       400001    510000
WP     ADN97989_05       500001    610000
WP     ADN97989_06       600001    710000
WP     ADN97989_07       700001    810000
WP     ADN97989_08       800001    910000
WP     ADN97989_09       900001   1010000
WP     ADN97989_10      1000001   1110000
WP     ADN97989_11      1100001   1210000
WP     ADN97989_12      1200001   1310000
WP     ADN97989_13      1300001   1410000
WP     ADN97989_14      1400001   1510000
WP     ADN97989_15      1500001   1610000
WP     ADN97989_16      1600001   1691138

  Query Match             4.6%;  Score 49.6;  DB 12;  Length 110000;
  Best Local Similarity   46.8%;  Pred. No. 0.81;
  Matches  189;  Conservative    0;  Mismatches  214;  Indels    1;  Gaps    1;

Qy         130 CATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGA 189
               |||||   |||   ||||||||||    |  |    ||| ||   |||  |  | |  | |
Db       58463 CATATATATTTATATGAGAAATATATTTATTATATAATAAGCATATATATGAGAAATACA 58404

Qy         190 TCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTAC 249
               |    ||  |||| |  |   ||| || ||||      |  |||| | ||  | |  ||
Db       58403 TTATATAATAAATAGTTATATTTATTTATATGAGAATTATATTTATATATTATATATTTA 58344

Qy         250 AGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGG-TATAAAGAAATGCATT 308
               |||  |  |||    || | | |  ||   | | | | |  |||||   |    |
Db       58343 TATAATAAATATATAATATTATATATATTTATATAATATATGTATATAATATATAATACA 58284

Qy         309 CAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACAT 368
               | | ||  |  || |||     |  | | || || | |  ||||  ||  ||| || ||
Db       58283 TATATAATAAATATGTATATAATATATAATAAATATATAATAAATATGTATATAATATAT 58224

Qy         369 TACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTA 428
               |   |    |       | |  | | |||  |      |                 | | | |
Db       58223 AATAAATATATATTTATACAATTTATAAGAAATATATATTTATATAAGAAATATATATAA 58164

Qy         429 TCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAA 488
               |   |    | ||| ||  ||  | | ||| ||| ||  | ||  | |    | | |||
Db       58163 TTTTATAAAATATATAATATATAATATATTATATATTATATATTTTGTATATATATATAT 58104

Qy         489 TTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGTC 532
                | |    |  ||| || | || || ||    |   ||| |||
Db       58103 AATATATATTTATATAAATATATATATTTAAAGAGGTTTCAGTC 58060


RESULT 40
ADO50281_14/c
Continuation (15 of 17) of ADO50281 from base 1400001 (Human phosphodiesterase 4D (PDE4D)
WP Sequence split into 17 fragments  LOCUS ADO50281 Accession Ado50281
WP     Fragment Name      Begin     End
WP     ADO50281_00            1    110000
WP     ADO50281_01       100001    210000
WP     ADO50281_02       200001    310000
WP     ADO50281_03       300001    410000
WP     ADO50281_04       400001    510000
WP     ADO50281_05       500001    610000
WP     ADO50281_06       600001    710000
WP     ADO50281_07       700001    810000
WP     ADO50281_08       800001    910000
WP     ADO50281_09       900001   1010000
WP     ADO50281_10      1000001   1110000
WP     ADO50281_11      1100001   1210000
WP     ADO50281_12      1200001   1310000
WP     ADO50281_13      1300001   1410000
WP     ADO50281_14      1400001   1510000
WP     ADO50281_15      1500001   1610000
WP     ADO50281_16      1600001   1691134

  Query Match             4.6%;  Score 49.6;  DB 12;  Length 110000;
  Best Local Similarity   46.8%;  Pred. No. 0.81;
  Matches  189;  Conservative    0;  Mismatches  214;  Indels    1;  Gaps    1;

Qy         130 CATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGA 189
               |||||   |||   ||||||||||    |  |    ||| ||   |||  |  | |  | |
Db       58462 CATATATATTTATATGAGAAATATATTTATTATATAATAAGCATATATATGAGAAATACA 58403

Qy         190 TCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTAC 249
               |    ||  |||| |  |   ||| || ||||      |  |||| | ||  | |  ||
Db       58402 TTATATAATAAATAGTTATATTTATTTATATGAGAATTATATTTATATATTATATATTTA 58343

Qy         250 AGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGG-TATAAAGAAATGCATT 308
               |||  |  |||    || | | |  ||   | | | | |  |||||   |    |
Db       58342 TATAATAAATATATAATATTATATATATTTATATAATATATGTATATAATATATAATACA 58283

Qy         309 CAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACAT 368
               | | ||  |  || |||     |  | | || || | |  ||||  ||  ||| || ||
Db       58282 TATATAATAAATATGTATATAATATATAATAAATATATAATAAATATGTATATAATATAT 58223

Qy         369 TACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTA 428
               |   |    |       | |  | | |||  |      |                 | | | |
Db       58222 AATAAATATATATTTATACAATTTATAAGAAATATATATTTATATAAGAAATATATATAA 58163

Qy         429 TCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAA 488
               |   |    | ||| ||  ||  | | ||| ||| ||  | ||  | |    | | |||
Db       58162 TTTTATAAAATATATAATATATAATATATTATATATTATATATTTTGTATATATATATAT 58103

Qy         489 TTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGTC 532
                | |    |  ||| || | || || ||    |   ||| |||
Db       58102 AATATATATTTATATAAATATATATATTTAAAGAGGTTTCAGTC 58059


RESULT 41
ABZ16669/c
ID   ABZ16669 standard; DNA; 2000 BP.
XX
AC   ABZ16669;
XX
DT   21-JAN-2003  (first entry)
XX
DE   Arabidopsis thaliana stress regulated gene SEQ ID NO 4474.
XX
KW   Arabidopsis thaliana; plant; gene; stress; transgenic; ds.
XX
OS   Arabidopsis thaliana.
XX
```

PN WO200216655-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US026685.
XX
PR 24-AUG-2000; 2000US-0227866P.
PR 26-JAN-2001; 2001US-0264647P.
PR 22-JUN-2001; 2001US-0300111P.
XX
PA (SCRI) SCRIPPS RES INST.
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Harper JF, Kreps J, Wang X, Zhu T;
XX
DR WPI; 2002-304127/34.
XX
PT Identifying a stress condition to which a plant cell has been exposed and
PT producing plants with increased tolerance to these abiotic stresses.
XX
PS Claim 144; SEQ ID NO 4474; 577pp + Sequence Listing; English.
XX
CC The invention relates to identifying a stress condition to which a plant
CC cell has been exposed, comprising: (a) contacting nucleic acid
CC representative of expressed polynucleotides in the plant cell with an
CC array or probes representative of the plant cell genome; and (b)
CC detecting a profile of expressed polynucleotides in the plant cell
CC characteristic of a stress response. The method is useful in the
CC production of transgenic plants, cells and seeds and in producing plants
CC with increased tolerance to abiotic stress. The present sequence is that
CC of an Arabidopsis thaliana stress regulated gene (ABZ12196-ABZ17574) used
CC in methods of the invention. Note: The sequence data for this patent is
CC not represented in the printed specification but is based on sequence
CC information supplied to Derwent by the European Patent Office
XX
SQ Sequence 2000 BP; 795 A; 259 C; 228 G; 718 T; 0 U; 0 Other;

Query Match 4.5%; Score 49.4; DB 6; Length 2000;
Best Local Similarity 45.4%; Pred. No. 0.33;
Matches 214; Conservative 0; Mismatches 256; Indels 1; Gaps 1;

Qy 120 TGAAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTT 179
Db 1375 TAAAATTAGTAATACAAAAATCGTGGTGATAATATTTTCATACTATTAATTTTAATAGTT 1316

Qy 180 TTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAA 239
Db 1315 TTTGTGATCGGTAACTATTAAAACCGTATATATATATATATA-TATATATATATATATAT 1257

Qy 240 AAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAG 299
Db 1256 ATATATATATATATATTTTGTAAACCGTATATTTTTTACCAGTAAATTAAAGAATTATCA 1197

Qy 300 AAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGC 359
Db 1196 ATGTGTTAACATAGGAAGAACGTTCCTTTTATTATTGAGATTTGATAATCCTTTTACCCC 1137

Qy 360 ATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCT 419
Db 1136 TGGTTCTAGTTTAGAAGTTTATCAATGTTCTTATATCCTTCAGTTAAACATAGGAACAAC 1077

Qy 420 GCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAG 479
Db 1076 GTTCCTTTTATTCTTGAGATTTGATAAATCAATTAAACATATTATCAGAAAAAATCTAAT 1017

Qy 480 AAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGTCATTCAGG 539
Db 1016 AATCATACAATATTTATTTTCTTATTAAACGTTATTTTCAGTAATATTTTCTTATTAAAC 957

Qy 540 AGTATGGAAGTATCGATGGCTTCTTTAAAATGAAGCAGGAGGGTCTGGCAG 590
Db 956 TGTTTTTTAGTAATATTTTTTTTTGATAAAATTAACAGATGTGATTTTCAG 906

RESULT 42
AAF22286
ID AAF22286 standard; DNA; 134499 BP.
XX
AC AAF22286;
XX
DT 20-MAR-2001 (first entry)
XX
DE BAC containing repeats from centromeres 1-4 #9.
XX
KW Centromere; michrosome; vector; ds.
XX
OS Arabidopsis thaliana.
XX
PN WO200055325-A2.
XX
PD 21-SEP-2000.
XX
PF 17-MAR-2000; 2000WO-US007392.
XX
PR 18-MAR-1999; 99US-0125219P.
PR 01-APR-1999; 99US-0127409P.
PR 18-MAY-1999; 99US-0134770P.
PR 13-SEP-1999; 99US-0153584P.
PR 17-SEP-1999; 99US-0154603P.
PR 16-DEC-1999; 99US-0172493P.
XX
PA (UYCH-) UNIV CHICAGO.
XX
PI Preuss D, Copenhaver G, Keith K;
XX
DR WPI; 2000-587529/55.
XX
PT Recombinant DNA construct comprising a plant centromere, useful for
PT producing stably inherited michrosomes which can serve as vectors for the
PT construction of transgenic plant and animal cells.
XX
PS Claim 102; Page 453-484; 1449pp; English.
XX
CC The present invention relates to a recombinant DNA construct of a plant
CC (Arabidopsis thaliana) centromere. The constructs are useful for
CC producing stably inherited michrosomes which can serve as vectors for the
CC construction of transgenic plant and animal cells expressing selected
CC proteins such as hormones, enzymes, interleukins, clotting factors,
CC cytokines, antibodies, and growth factors
XX
SQ Sequence 134499 BP; 41565 A; 25130 C; 25225 G; 42577 T; 0 U; 2 Other;

Query Match 4.5%; Score 49.4; DB 3; Length 134499;
Best Local Similarity 45.4%; Pred. No. 0.94;
Matches 214; Conservative 0; Mismatches 256; Indels 1; Gaps 1;

Qy 120 TGAAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTT 179
Db 59995 TAAAATTAGTAATACAAAAATCGTGGTGATAATATTTTCATACTATTAATTTTAATAGTT 60054

Qy 180 TTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAA 239
Db 60055 TTTGTGATCGGTAACTATTAAAACCGTATATATATATATATA-TATATATATATATATAT 60113

Qy 240 AAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAG 299
Db 60114 ATATATATATATATATTTTGTAAACCGTATATTTTTTACCAGTAAATTAAAGAATTATCA 60173

Qy 300 AAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGC 359
Db 60174 ATGTGTTAACATAGGAAGAACGTTCCTTTTATTATTGAGATTTGATAATCCTTTTACCCC 60233

Qy 360 ATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCT 419
Db 60234 TGGTTCTAGTTTAGAAGTTTATCAATGTTCTTATATCCTTCAGTTAAACATAGGAACAAC 60293

```
Qy        420 GCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAG 479
              | |   ||     |      |   |    || |      |||   |  | |  |||| ||||| |
Db      60294 GTTCCTTTTATTCTTGAGATTTGATAAATCAATTAAACATATTATCAGAAAAAATCTAAT 60353

Qy        480 AAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGTCATTCAGG 539
              ||       | | ||||| | ||  | |||| ||    ||     |||||| | | ||| |
Db      60354 AATCATACAATATTTATTTTCTTATTAAACGTTATTTTCAGTAATATTTTCTTATTAAAC 60413

Qy        540 AGTATGGAAGTATCGATGGCTTCTTTAAAATGAAGCAGGAGGGTCTGGCAG 590
              || |   ||||    |   || |   |||   | |||  | |  |  |||
Db      60414 TGTTTTTTAGTAATATTTTTTTTTGATAAAATTAACAGATGTGATTTTCAG 60464


RESULT 43
ABL33499
ID   ABL33499 standard; DNA; 6102 BP.
XX
AC   ABL33499;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 1472.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP007537.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful for
PT   diagnosis and treatment of diseases associated with abnormal cytosine
PT   methylation.
XX
PS   Claim 1; SEQ ID NO 1472; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention
XX
SQ   Sequence 6102 BP; 1825 A; 74 C; 1240 G; 2963 T; 0 U; 0 Other;

  Query Match             4.5%;  Score 49.2;  DB 6;  Length 6102;
  Best Local Similarity   47.4%;  Pred. No. 0.49;
  Matches  147;  Conservative    0;  Mismatches  163;  Indels    0;  Gaps    0;

Qy        261 TATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATC 320
              | ||  |  ||| |   | | |  ||||||  |||  |||| |   ||   | ||| |
Db       5758 TGTTTGTAAGTTGTATTTATATATAAAAAATTTATGGAGAATTTTTTTTTTAATAAAAAGA 5817

Qy        321 ATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTG 380
              |  |||||    ||| ||  |      |    || || | |||||   ||  |   |  ||
Db       5818 AAGTATTAAGTTGTTTAATTTATTTAAATTTATTAGAGAATATTTAGTTGGTGTTTAATG 5877

Qy        381 CAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAA 440
              |  |   |||||      |  | |   ||||       |   | || |   | |  | |
Db       5878 GGTTTTAGTTTAAAAGAGGGAATTGAGTTTATATATTGAGTAATTTTTTTTTTAATAGATA 5937

Qy        441 TAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTAC 500
              || |||||   |   | | ||||  |||   ||   ||      |||  | |||||  |
Db       5938 TATAAATTTATATGGAAATGATTTAAGAATATATAAAAGTGTGAATATATATTTATTTGT 5997

Qy        501 TATTAAAACATTTACTTTCCAAATATTCTGTCATTCAGGAGTATGGAAGTATCGATGGCT 560
              |   || | | |   |||  | |||| | | |   |   ||  | || |    | |
Db       5998 TTAAAAGAAAGTATTTTTATATATATATTTTTTTGTTGTGTTAGTGGAGAAGGTTTAGAA 6057

Qy        561 TCTTTAAAAT 570
              |  |||  ||
Db       6058 GCGATAATAT 6067


RESULT 44
ABL32484/c
ID   ABL32484 standard; DNA; 6244 BP.
XX
AC   ABL32484;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 457.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP007537.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful for
PT   diagnosis and treatment of diseases associated with abnormal cytosine
PT   methylation.
XX
PS   Claim 1; SEQ ID NO 457; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention
XX
```

```
SQ   Sequence 6244 BP; 1969 A; 15 C; 993 G; 3267 T; 0 U; 0 Other;

  Query Match             4.5%;  Score 49.2;  DB 6;  Length 6244;
  Best Local Similarity   47.4%;  Pred. No. 0.49;
  Matches  181;  Conservative    0;  Mismatches  198;  Indels    3;  Gaps    1;

Qy        146 AGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACA 205
              | | ||||  ||   |||| ||   |  |||  | | || |    |  |  |  |
Db       2605 ATATATATATATACATACATATATATATATATATATATATATACATACATATATATAT 2546

Qy        206 TTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTC 265
              | | |   ||      |   |  || | | || |    | ||||  | |||  |  |||
Db       2545 ATATATATATATATACATATATATATATATATATATATACATACATATATATATATAT 2486

Qy        266 ATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATA 325
              ||   |   |       |  | | |  |||| | | ||   ||   |   | ||| |  |
Db       2485 ATACATACATATACATATATATATATATATATATATATATTTTATAATCCATATCTTTAA 2426

Qy        326 TTAGCCAGTTGAAAATTAAACA---CAAAATGAGTGCATATTACATTACTTAATCTTGCA 382
              | | | |  | ||||  ||| |    |||||      | |||  ||||  || |||
Db       2425 TCATCAAAATAAAAAAAAAAAAAACAAAAATTTTAAAAAATTTAATTAAATATTCTAAAT 2366

Qy        383 GTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATA 442
              |||  |||||| || | | |     || | ||        |||   ||    | | |
Db       2365 TAAAAACTTAAAAAATCCATCAATTATAATTCCACAATAAAAATTAAAACAACCAATAAA 2306

Qy        443 GAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTA 502
              |||  ||  |   | |||  |    ||||| ||  | ||||  || | | ||   |
Db       2305 TAAAATAATAACTAATTTCATACTTAAATATTCAAACTAAACTAAAATCTATAACCAAAC 2246

Qy        503 TTAAAACATTTACTTTCCAAAT 524
              | |||| |  ||  || || ||
Db       2245 TAAAAAAAAATATATTACACAT 2224


RESULT 45
ABQ67059/c
ID   ABQ67059 standard; DNA; 34688 BP.
XX
AC   ABQ67059;
XX
DT   28-AUG-2002  (first entry)
XX
DE   Human angiogenesis associated polynucleotide SEQ ID NO 89.
XX
KW   Human; angiogenesis; methylation; eye disease; glaucoma; tumour;
KW   inflammation; rheumatoid arthritis; diabetic retinopathy; antiulcerds;
KW   macular degeneration; inflammatory bowel disease; Crohn's disease;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiarteriosclerotic; ds.
XX
OS   Homo sapiens.
XX
PN   WO200246454-A2.
XX
PD   13-JUN-2002.
XX
PF   06-DEC-2001; 2001WO-EP014320.
XX
PR   06-DEC-2000; 2000DE-01061338.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Schacht O;
XX
DR   WPI; 2002-500450/53.
XX
PT   New nucleic acid fragments from chemically treated angiogenesis-
PT   associated genes, useful for determining methylation status, e.g. in
PT   diagnosis or treatment of cancer.
XX
PS   Claim 1; SEQ ID NO 89; 41pp + Sequence Listing; German.
XX
CC   The invention relates to a nucleic acid (I) comprising a segment of 18
CC   bases of chemically pretreated DNA of angiogenesis-associated genes (II)
CC   having sequences (ABQ66971-ABQ67178) or their complements. (I), also
CC   related oligomers, are used to evaluate the methylation status and/or
CC   single-nucleotide polymorphisms, in angiogenesis-related genes, for
CC   diagnosis and treatment of eye diseases, proliferative retinopathy,
CC   neovascular glaucoma, solid tumours, inflammation, rheumatoid arthritis,
CC   diabetic retinopathy, macular degeneration caused by neovascularisation,
CC   psoriasis, arteriosclerosis, inflammatory bowel diseases, ulcers and
CC   Crohn's disease. Note: The sequence data for this patent did not form
CC   part of the printed specification, but was obtained in electronic format
CC   directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 34688 BP; 9920 A; 444 C; 6941 G; 17383 T; 0 U; 0 Other;

  Query Match             4.5%;  Score 49.2;  DB 6;  Length 34688;
  Best Local Similarity   43.6%;  Pred. No. 0.75;
  Matches  277;  Conservative    0;  Mismatches  353;  Indels    6;  Gaps    1;

Qy        131 ATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGAT 190
              |||  |  |  || |  | ||   |  | || || |    ||     || | | || |
Db      16782 ATAAAATATACTTTCAATATTAAATACAAATAAAATTACTATATATAATTCGAAAAAAAA 16723

Qy        191 CTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACA 250
              || ||  ||     |     ||   | | ||          ||| |  |   | | | | |
Db      16722 ATCTTCAAAACTCTAACACTTCAATTAAATTTATATAACATTATAATTTCATAAATCAAA 16663

Qy        251 GACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCA 310
              ||||  |  ||  |  | |||   || |  |||      |||| ||     |||
Db      16662 ACCATAAAAATCCTCTCTAATATTTAAATTATCTTAAATCCACTTAAAAAA------TCA 16609

Qy        311 GAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTA 370
              | ||  |  | |||  | | |  | || |   |    || || |    |||     ||
Db      16608 CAAAATAAAAAAATAAAAACAAACTTAATAAATATATAAATATAAAAAAATAAATATTTC 16549

Qy        371 CTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATC 430
              |  ||         |  |||||   | ||  ||    |||  |  |  ||| | | |
Db      16548 TAATAATTTCACCATTTAATTAAAACAACGACAAAATACAATAAAAATTATTTTTTAAAA 16489

Qy        431 GCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATT 490
              |   |      |||||  |   ||  | | |||  |||||    || | ||   ||
Db      16488 AAAACACCCTCTTAAATTCACCATAAACTAAATTTTAAAATATATACATATAAACAAAAC 16429

Qy        491 TTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGTCATTCAGGAGTATGGAAGT 550
              |   | || ||| |||||||| | |||  ||| | |    || | |  | | |  ||
Db      16428 TAAAAACTAATATAAAAACATTAAATTTTAAAAAAATAAAACAATAAATAATTTTTAAAA 16369

Qy        551 ATCGATGGCTTCTTTAAAATGAAGCAGGAGGGTCTGGCAGAGAGTATCTATGAAATAAGT 610
              ||  |    || |    | |   |  |  | | || | | | ||  |   |  | | |
Db      16368 TTCTCTTTAATCATATTTTTAATAAAAAAAATTTTCTTAAATACAATAAACACATAATTT 16309

Qy        611 TCCTCTGACCTTCACGCTTAATTTTCTGAATGGAGTGGAGCAAATTACTTCAAGCTTCAC 670
              || |      |  | ||  |  | ||     |  | | || |  |  |    |
Db      16308 TCATTCATATACTAAAATAAACATCATAAACCTTATTCAACTAAATTATAACAATAAATC 16249

Qy        671 TTAACTTGCATATGAAATGAACCGTACAAAAATACAAGAGTGTCAGGAGAAAGTTATGCT 730
              ||  |  | ||   ||| ||     |||||   |     | |||  | ||| | |   |
Db      16248 CTATTTATCTTAAATAATTAATATCTCAAAATATCCCATATTTCAAAAAAAATTAACAAT 16189

Qy        731 CTGGTAAATATTTTGCAAAACAGATAAAAGATAATA 766
              |   |    | |||   ||| |  ||||| | ||||
Db      16188 CAAATCTTAAATTTCTTAAAAATTTAAAAAAAAATA 16153


RESULT 46
ABL33416/c
ID   ABL33416 standard; DNA; 6265 BP.
XX
AC   ABL33416;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 1389.
```

XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 1389; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 6265 BP; 1614 A; 266 C; 1524 G; 2861 T; 0 U; 0 Other;

Query Match 4.5%; Score 48.8; DB 6; Length 6265;
Best Local Similarity 50.9%; Pred. No. 0.6;
Matches 116; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

Qy 262 ATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCA 321
Db 1700 AATCAATTCCTTATTTTTTAAACGAATACTCTAAAAATAAATAAAATTACCCAAAAAAAA 1641

Qy 322 TATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGC 381
Db 1640 ATCAAAAACACTTTTATAAACACTAACTTCATACAAATAAATAACACTAACTCCTATAAA 1581

Qy 382 AGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAAT 441
Db 1580 CCTTAAACTCACAAACTCATATTTAAAAAAATTTAAAATCTATATAATATAATTATACAC 1521

Qy 442 AGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAAT 489
Db 1520 TAAATTTATATCAAATTTTATTTCTATAATAATCCTTCAAAAAACAAT 1473

RESULT 47
ABL33758/c
ID ABL33758 standard; DNA; 7455 BP.
XX
AC ABL33758;
XX
DT 26-MAR-2002 (first entry)
XX

DE Human immune system associated gene SEQ ID NO: 1731.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 1731; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 7455 BP; 2348 A; 38 C; 1207 G; 3862 T; 0 U; 0 Other;

Query Match 4.5%; Score 48.8; DB 6; Length 7455;
Best Local Similarity 45.0%; Pred. No. 0.63;
Matches 182; Conservative 0; Mismatches 222; Indels 0; Gaps 0;

Qy 149 AATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTG 208
Db 5837 ATTTTATACCTATAAATTTTATTAAAAATATAAAATCAAAAAATACTCATAAAAAAATAA 5778

Qy 209 TTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATT 268
Db 5777 TTTTATTTTACTAACCTTAATAAACTAAAAAATAAATATATAAACATAAATATATAAATA 5718

Qy 269 TGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTA 328
Db 5717 TATATATCATAATTTTAAAAAAATATTTCCCCAATAAAAAATAATAAAAATATAAAATAA 5658

Qy 329 GCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAA 388
Db 5657 TAAAATCTTTATTTTCTACCATATCACAACTAAATAAAAACAACTTAAATAATATAAAAC 5598

Qy 389 GGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATT 448
Db 5597 AATAAAACAATAAAACATAACTATTATACAACATTTTTTATTATACAAAATAACATAACT 5538

Qy 449 ACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAA 508
Db 5537 ATTTAAAATAATAAAAAAATCAATAAAAAAATAATAAAAAATATTTACTTATAAAAAAAA 5478

```
Qy        509 CATTTACTTTCCAAATATTCTGTCATTCAGGAGTATGGAAGTAT 552
              ||  | |||  |||  |    | ||  |  | |||  |  | |
Db       5477 AATAAATTTTTAAAAATTAAACTTATAAAACAATATATACATTT 5434


RESULT 48
ABL32766
ID   ABL32766 standard; DNA; 19659 BP.
XX
AC   ABL32766;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 739.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP007537.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful for
PT   diagnosis and treatment of diseases associated with abnormal cytosine
PT   methylation.
XX
PS   Claim 1; SEQ ID NO 739; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention
XX
SQ   Sequence 19659 BP; 6080 A; 68 C; 3359 G; 10152 T; 0 U; 0 Other;

  Query Match             4.5%;  Score 48.8;  DB 6;  Length 19659;
  Best Local Similarity   48.5%;  Pred. No. 0.8;
  Matches  193;  Conservative    0;  Mismatches  202;  Indels    3;  Gaps    2;

Qy        111 TCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAG 170
              | || || || ||  |  |  | | |||||  || ||| | ||  |   ||| |||
Db      11664 TTTATTTGTTTAAAAGGTTTTTTTTAGTTTTATTAAGATAAATGAAGTGTTATAAAAGTA 11723

Qy        171 TTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGA 230
              || | |||  |||| | ||     | | | | | | |        |     |  |
Db      11724 TTTAAAGTAATGTATATTATAATGGTTAATATAAAATAGTATTGAGTATTTGATAGTGAA 11783

Qy        231 TTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAA 290
              |||||||    |  |         || |  | || | |  |   |  || |||    |
Db      11784 ATTTTTTATTTAAATTATAGTTTTTATAATTGTTTAATAATAAGTTTTTAGTTTTTTATT 11843

Qy        291 GGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAA 350
              || ||  |  ||   | | ||| ||  ||  | ||||    |    ||||||      ||
Db      11844 GGGATTTAATAAGAAAATTAGAAAATTATATTTTATTTATGAAAAAAAAATTTTTGTTAA 11903

Qy        351 AATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGT 410
              ||       ||| || || |  || | |  | |      |   | | | ||  | ||  |
Db      11904 TATATTATTATAGTAGATAAG-TATTATAAAATTTTTTTGGTTATATTAAAAATTAATAT 11962

Qy        411 ATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAA 470
              |||     |  ||| ||||      |  ||| | ||||||    || ||||    | ||||
Db      11963 ATAAATATTTATTTTTTATAATTTTAAAAAGATAAATTATT--AATTTTTGAAATAGAAA 12020

Qy        471 TAATCTCAGAAAACATAATTTTTTATGTACTATTAAAA 508
              |  | | | | || ||||| |  | |||   |||  ||
Db      12021 TTTTTTTATATAAGATAATGTAATTTGTTGGATTTTAA 12058


RESULT 49
ABD33504/c
ID   ABD33504 standard; DNA; 93544 BP.
XX
AC   ABD33504;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Human cancer-associated (CA) gene HD07-098.
XX
KW   Human; cancer-associated protein; CAP; cancer-associated gene; CA; gene;
KW   ds; cancer; cytostatic.
XX
OS   Homo sapiens.
XX
PN   WO2004058146-A2.
XX
PD   15-JUL-2004.
XX
PF   15-DEC-2003; 2003WO-US040081.
XX
PR   17-DEC-2002; 2002US-00322281.
XX
PA   (SAGR-) SAGRES DISCOVERY INC.
XX
PI   Morris DW,  Malandro MS;
XX
DR   WPI; 2004-499109/47.
XX
PT   Novel human cancer associated protein encoded within open reading frame
PT   of cancer associated gene, useful as targets for diagnosing cancer.
XX
PS   Claim 16; SEQ ID NO 676; 182pp; English.
XX
CC   The invention relates to cancer-associated proteins (CAP) and the cancer-
CC   associated (CA) nucleic acids encoding them. The invention also relates
CC   to a method for treating cancers involving administering to a patient an
CC   inhibitor of CAP, and a method of screening for anticancer activity in a
CC   potential drug involving providing a cell that expresses a CA gene,
CC   contacting a tissue sample derived from a cancer cell with an anticancer
CC   drug candidate and monitoring the effect of the anticancer drug candidate
CC   on expression of the CA gene. The CAP proteins are useful for detecting
CC   cancer associated with expression of a CAP protein in a test cell sample
CC   and for screening for a bioactive agent capable of modulating the
CC   activity of a CAP protein. The CA nucleic acids are useful for diagnosing
CC   cancer, involving determining the expression of a CA nucleic acid in a
CC   tissue. This sequence represents a human CA gene of the invention. Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 93544 BP; 26138 A; 20027 C; 21635 G; 25744 T; 0 U; 0 Other;

  Query Match             4.5%;  Score 48.6;  DB 13;  Length 93544;
  Best Local Similarity   47.4%;  Pred. No. 1.3;
```

```
  Matches  176;  Conservative    0;  Mismatches  194;  Indels    1;  Gaps    1;

Qy         33 TCTCTATGTTCTTATAAGCAAATGTTACAATGAGATATGAGTTTTAAAGCCAGATCTTCC 92
              ||  | | | | ||     |   | ||  | |   |-|   |  ||| |    || ||
Db      89520 TCAGTTTCTACATACTGCCTTGTATTTGATTTTTTTTTTTTTGGTAATGTACCATTTTTA 89461

Qy         93 TTATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATA 152
              || |||   | |     || ||  | || |  | | |    |   | |  ||   |||||
Db      89460 TTTTCTTCTTTCTTTGCCTGTATATTTTTAGTTATTTTCAGTTCCTATCATTTTAAAATA 89401

Qy        153 TTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTT 212
              || |||   | |  | |     | |  || ||  |  | | || | |  |   || ||||
Db      89400 TTTATCCCCAAATGTAATCCCTTCGAGTT-TATTATTTATAATGTCAGTGGTTTTATTTT 89342

Qy        213 GTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTT 272
              | ||||| ||   ||  |  || |     || |  |||| |    |   ||  |
Db      89341 TTAAATATTCTAGATTCTTACATTGATTCTTTCTGATTGACAAATTCATCCATATCAGAA 89282

Qy        273 TTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCA 332
              |||    |||| | | |||   | ||| || ||       ||||    ||||  ||
Db      89281 TTTTAAGTGTTAACACAAGCCCTCAAGCAAGGCTGCTGAAGAATTTGCATACTTTTCTAT 89222

Qy        333 GTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTA 392
              |||  ||| |  | |       |||   || | |||||||| |       ||  |||||||
Db      89221 GTTCCAAACTTGAAATTTCCAAAGTTTTTAATGCATTACTTTAGACCCAAGAGAAAGGTA 89162

Qy        393 AAAAGTCAACC 403
              ||  |   ||
Db      89161 CAACATTTTCC 89151


RESULT 50
ACN44850
ID   ACN44850 standard; DNA; 142318 BP.
XX
AC   ACN44850;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Human genomic sequence hCG33122.
XX
KW   Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO2003073826-A2.
XX
PD   12-SEP-2003.
XX
PF   28-FEB-2003; 2003WO-US006235.
XX
PR   01-MAR-2002; 2002US-00087192.
XX
PA   (SAGR-) SAGRES DISCOVERY.
XX
PI   Morris DW;
XX
DR   WPI; 2003-328604/31.
XX
PT   Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
PT   comprises a nucleotide sequence.
XX
PS   Claim 1; SEQ ID NO 1504; 0pp; English.
XX
CC   The present invention relates to novel DNA and protein sequences which
CC   are associated with carcinomas. The sequences are useful for: (i) for
CC   screening drug candidates; (ii) for screening of bioactive agent capable
CC   of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
CC   a bioactive agent capable of modulating the activity of CAP; (iv) for
CC   evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
CC   carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
CC   carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
CC   (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
CC   determining Carcinoma Associated (CA) gene copy number. In addition, the
CC   CA genes are useful as DNA vaccines and the CAP are useful as markers of
CC   carcinoma including lymphoma. The present sequence is one such CA coding
CC   sequence. Note: This patent is an equivalent to basic patent
CC   US2002182586A1, for which no sequence data was published
XX
SQ   Sequence 142318 BP; 38833 A; 25825 C; 26759 G; 47076 T; 0 U; 3825 Other;

  Query Match             4.5%;  Score 48.6;  DB 11;  Length 142318;
  Best Local Similarity   47.4%;  Pred. No. 1.4;
  Matches  176;  Conservative    0;  Mismatches  194;  Indels    1;  Gaps    1;

Qy         33 TCTCTATGTTCTTATAAGCAAATGTTACAATGAGATATGAGTTTTAAAGCCAGATCTTCC 92
              ||  | | | | ||     |   | ||  | |   |    |  ||| |    || ||
Db     137461 TCAGTTTCTACATACTGCCTTGTATTTGATTTTTTTTTTTTGGTAATGTACCATTTTTA 137520

Qy         93 TTATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATA 152
              || |||   | |     || ||  | || |  | | |    |   | |  ||   |||||
Db     137521 TTTTCTTCTTTCTTTGCCTGTATATTTTTAGTTATTTTCAGTTCCTATCATTTTAAAATA 137580

Qy        153 TTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTT 212
              || |||   | |  | |     | |  || ||  |  | | || | |  |   || ||||
Db     137581 TTTATCCCCAAATGTAATCCCTTCGAGTT-TATTATTTATAATGTCAGTGGTTTTATTTT 137639

Qy        213 GTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTT 272
              | ||||| ||   ||  |  || |     || |  |||| |    |   ||  |
Db     137640 TTAAATATTCTAGATTCTTACATTGATTCTTTCTGATTGACAAATTCATCCATATCAGAA 137699

Qy        273 TTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCA 332
              |||    |||| | | |||   | ||| || ||       ||||    ||||  ||
Db     137700 TTTTAAGTGTTAACACAAGCCCTCAAGCAAGGCTGCTGAAGAATTTGCATACTTTTCTAT 137759

Qy        333 GTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTA 392
              |||  ||| |  | |       |||   || | |||||||| |       ||  |||||||
Db     137760 GTTCCAAACTTGAAATTTCCAAAGTTTTTAATGCATTACTTTAGACCCAAGAGAAAGGTA 137819

Qy        393 AAAAGTCAACC 403
              ||  |   ||
Db     137820 CAACATTTTCC 137830


RESULT 51
ABL33396
ID   ABL33396 standard; DNA; 5641 BP.
XX
AC   ABL33396;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 1369.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP007537.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
```

PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 1369; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 5641 BP; 1442 A; 34 C; 930 G; 3235 T; 0 U; 0 Other;

Query Match 4.4%; Score 48.4; DB 6; Length 5641;
Best Local Similarity 47.4%; Pred. No. 0.72;
Matches 145; Conservative 0; Mismatches 161; Indels 0; Gaps 0;

```
Qy       171 TTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGA 230
Db      2111 TTATTATATTTTTTGTAGTTATTTTTTTAGTTTAATTTTGTTTTTATTTAGTTTTTTTAT 2170

Qy       231 TTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAA 290
Db      2171 GTTGTAATCGATAAATATGTTTATATATATATTTTTATATGTTATGGTTTTAACGATATAT 2230

Qy       291 GGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAA 350
Db      2231 TATATATATATTATTTTATATAAATGGTGTTTAAATAAGTTATAAGAAAAAATAAGAAAA 2290

Qy       351 AATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGT 410
Db      2291 TATGAATTTATATTGTTATATTTAATAGATATATGTAAAATATTATATAGATATATAGAT 2350

Qy       411 ATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAA 470
Db      2351 TATATTAATATATATTTATATATTTATATATATTTATAATATTATATTTTTATATTTTAT 2410

Qy       471 TAATCT 476
Db      2411 TATTGT 2416
```

RESULT 52
AAS45384/c
ID AAS45384 standard; DNA; 5689 BP.
XX
AC AAS45384;
XX
DT 18-DEC-2001 (first entry)
XX
DE Chemically pretreated genomic DNA associated with cell cycle #45.
XX
KW Cell cycle; human; CpG dinucleotide; cytosine methylation; HIV; aging;
KW human immunodeficiency virus; neurodegenerative disorder; solid tumour;
KW graft-versus-host disease; glomerular disease; Lewy body disease; cancer;
KW arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritic;
KW immunosuppressive; antitumour; cytostatic; antiarteriosclerotic; ds;
KW PCR primer.
XX
OS Homo sapiens.
XX
PN WO200168911-A2.
XX
PD 20-SEP-2001.
XX
PF 15-MAR-2001; 2001WO-EP002945.
XX
PR 15-MAR-2000; 2000DE-01013847.
PR 06-APR-2000; 2000DE-01019058.
PR 07-APR-2000; 2000DE-01019173.
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2001-602751/68.
XX
PT Designing primers and probes for analyzing diseases associated with
PT cytosine methylation state e.g. arthritis, cancer, aging,
PT arteriosclerosis comprising fragments of chemically modified genes
PT associated with cell cycle.
XX
PS Claim 1; SEQ ID NO 89; 28pp; English.
XX
CC Sequences AAS45296-AAS45520 represent chemically pretreated genomic DNA
CC molecules associated with the cell cycle and specific PCR primers of the
CC invention. The sequences are useful for detecting the methylation state
CC of all CpG dinucleotides in a sequence and therefore for analysing
CC associated diseases. By analysing cytosine methylations in the pretreated
CC DNA, genetic and/or epigenetic parameters for the diagnosis and therapy
CC of existing diseases or the predisposition to specific diseases can be
CC ascertained. The parameters may be compared to another set of genetic
CC and/or epigenetic parameters, the differences serving as basis for
CC diagnosis and/or prognosis events which are disadvantageous to patients.
CC The sequences of the invention are useful for the diagnosis and therapy
CC of HIV infection, neurodegenerative disorders, graft-versus-host disease,
CC aging, glomerular disease, Lewy body disease, arthritis,
CC arteriosclerosis, solid tumours and cancers
XX
SQ Sequence 5689 BP; 1490 A; 207 C; 1277 G; 2707 T; 0 U; 8 Other;

Query Match 4.4%; Score 48.4; DB 4; Length 5689;
Best Local Similarity 48.4%; Pred. No. 0.73;
Matches 121; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

```
Qy       238 AAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAA 297
Db      2030 AAAAAAAAATAAATTTATTTTTATATATTTTTTTTTTTTTATTATTTAAAAATAATTAAA 1971

Qy       298 AGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGT 357
Db      1970 AAAANATAAAAAAATTTAATAANNTTAAAAAAAAANNTANNATAAAAAAAAATAATAATA 1911

Qy       358 GCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTAC 417
Db      1910 ATATAAAAAATTTATTAATAAAAAAATATAAATAAAAAAAAATAAATATTAATTAAATAA 1851

Qy       418 CTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTC 477
Db      1850 TTTATTTTATATATATTAAAATAAAATAATTANTAAAAATTATAAAATAAAAATAAATAA 1791

Qy       478 AGAAAACATA 487
Db      1790 AAAAATAATA 1781
```

RESULT 53
AAS46426/c
ID AAS46426 standard; DNA; 5689 BP.
XX
AC AAS46426;
XX
DT 18-DEC-2001 (first entry)
XX
DE Tumour suppressor gene derived chemically modified sequence #148.
XX

```
KW   Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer;
KW   tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW   cytosine methylation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200168912-A2.
XX
PD   20-SEP-2001.
XX
PF   15-MAR-2001; 2001WO-EP002955.
XX
PR   15-MAR-2000; 2000DE-01013847.
PR   06-APR-2000; 2000DE-01019058.
PR   07-APR-2000; 2000DE-01019173.
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2001-602752/68.
XX
PT   Fragments of chemically modified genes associated with tumor suppressor
PT   genes and oncogenes, useful in designing primers and probes for analyzing
PT   diseases associated with cytosine methylation state e.g. cancer.
XX
PS   Claim 1; SEQ ID NO 148; 27pp; English.
XX
CC   The invention relates to a nucleic acid comprising a sequence of 18
CC   bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC   bisulphite, of genes associated with tumour suppression and oncogenes
CC   having a sequence taken from 536 (actually 533 since numbers 408, 458 and
CC   500 are missing from the sequence listing) sequences (Ss) and sequences
CC   complementary to (Ss). The nucleic acid may be a peptide nucleic acid-
CC   oligomer (PNA) of at least 9 nucleotides and may form part of a set of
CC   probes for detecting the cytosine methylation state and/or single
CC   nucleotide polymorphisms and also to be used in an array for analysing
CC   diseases associated with CpG dinucleotides e.g. cancers and tumours. The
CC   probes can also be used in a method for ascertaining genetic and/or
CC   epigenetic parameters for the diagnosis and/or therapy of existing
CC   diseases or the predisposition to specific diseases, by analysing
CC   cytosine methylations. The parameters may be compared to another set of
CC   genetic and/or epigenetic parameters, the differences serving as basis
CC   for diagnosis and/or prognosis events which are disadvantageous to
CC   patients. The present sequence is one of the 533 genomic sequences
CC   derived from tumour suppressor genes and oncogenes. Sequences with even
CC   numbered Seq ID numbers are the complementary sequence of the
CC   corresponding odd numbered sequence (e.g. ID 2 and ID1, ID 536 and ID
CC   535, except for those whose partner sequence is missing). Note: The
CC   sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 5689 BP; 1490 A; 207 C; 1277 G; 2707 T; 0 U; 8 Other;

  Query Match             4.4%;  Score 48.4;  DB 4;  Length 5689;
  Best Local Similarity   48.4%;  Pred. No. 0.73;
  Matches  121;  Conservative    0;  Mismatches  129;  Indels    0;  Gaps    0;

Qy        238 AAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAA 297
              |||| |   || |   ||     |||   ||| |||||   || || |||||   |  ||
Db       2030 AAAAAAAAATAAATTTATTTTTATATATTTTTTTTTTTTTATTATTTAAAAATAATTAAA 1971

Qy        298 AGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGT 357
              | ||    |   |    || |   |  |  |   |  |   |   ||| | | ||| |
Db       1970 AAAANATAAAAAAATTTAATAANNTTAAAAAAAAAANNTANNATAAAAAAAAATAATAATA 1911

Qy        358 GCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTAC 417
                |||  | |||  |||||     | |  ||   |||||    |  ||    |  | ||
Db       1910 ATATAAAAAATTTATTAATAAAAAAATATAAATAAAAAAAAAATAAATATTAATTAAATAA 1851

Qy        418 CTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTC 477
              |  |||  |||        | | |  |||||  | ||||| ||   |  ||||||
Db       1850 TTTATTTTATATATATTAAAATAAAATAATTANTAAAAATTATAAAATAAAAATAAATAA 1791

Qy        478 AGAAAACATA 487
              | |||  |||
Db       1790 AAAAATAATA 1781


RESULT 54
ABK28226/c
ID   ABK28226 standard; DNA; 5689 BP.
XX
AC   ABK28226;
XX
DT   23-APR-2002  (first entry)
XX
DE   DNA transcription associated complementary genomic DNA #50.
XX
KW   DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;
KW   PNA; cytosine methylation state; SNP; retroviral infection; gene; ds;
KW   single nucleotide polymorphism; adenosine deaminase deficiency; cancer;
KW   viral infection; Sezary syndrome; haematological disorder; tuberculosis;
KW   immunological disorder; Werner syndrome; developmental disorder;
KW   psoriasis; Rieger's syndrome; neurological disorder; erythropoiesis;
KW   neurodegenerative disorder; Waardenburg syndrome; Niemann-Pick disease;
KW   myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;
KW   angiogenesis; congenital heart disease; HDR syndrome; gene therapy;
KW   polyglutamine disorder; solid tumour.
XX
OS   Unidentified.
XX
PN   WO200192565-A2.
XX
PD   06-DEC-2001.
XX
PF   06-APR-2001; 2001WO-EP003973.
XX
PR   06-APR-2000; 2000DE-01019058.
PR   07-APR-2000; 2000DE-01019173.
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-090046/12.
XX
PT   New nucleic acids or oligomers, useful for diagnosing or treating
PT   diseases associated with DNA transcription, e.g. immunological disorders,
PT   Werner syndrome, psoriasis, myocardial infarction, solid tumors or
PT   cancer.
XX
PS   Claim 1; SEQ ID NO 100; 32pp; English.
XX
CC   The invention relates to a nucleic acid, which comprises a segment of the
CC   chemically pretreated DNA of genes associated with DNA transcription from
CC   one of 346 sequences, and an oligomer, in particular an oligonucleotide
CC   or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical
CC   to the chemically pretreated DNA of genes associated with DNA
CC   transcription. The set of oligomer probes are useful for detecting the
CC   cytosine methylation state and/or single nucleotide polymorphisms (SNPs)
CC   in a chemically pretreated genomic DNA. The nucleic acids are useful for
CC   diagnosing or treating diseases associated with DNA transcription
CC   (particularly with the methylation status), e.g. adenosine deaminase
CC   deficiency, viral infection, retroviral infection, Sezary syndrome,
CC   haematological disorders, immunological disorders, Werner syndrome,
CC   tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,
CC   neurological disorders, neurodegenerative disorders, Waardenburg
CC   syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial
CC   infarction, hypertension, angiogenesis, erythropoiesis, congenital heart
CC   disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours
```

```
CC   or cancer. Sequences ABK28127-ABK28472 represent DNA transcription
CC   associated genomic DNA molecules of the invention. Note: The sequence
CC   data for this patent did not form part of the printed specification but
CC   was obtained in electronic format directly from the European Patent
CC   Office
XX
SQ   Sequence 5689 BP; 1490 A; 207 C; 1277 G; 2707 T; 0 U; 8 Other;

  Query Match             4.4%;  Score 48.4;  DB 6;  Length 5689;
  Best Local Similarity   48.4%;  Pred. No. 0.73;
  Matches  121;  Conservative    0;  Mismatches  129;  Indels    0;  Gaps    0;

Qy        238 AAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAA 297
              |||| |   || |   ||      |||   ||| |||||   || || |||||   |  ||
Db       2030 AAAAAAAAATAAATTTATTTTTATATATTTTTTTTTTTTTATTATTTAAAAATAATTAAA 1971

Qy        298 AGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGT 357
              | ||   |  |   || |  | |  |  |  | |  ||| | | ||| |
Db       1970 AAAANATAAAAAAATTTAATAANNTTAAAAAAAAANNTANNATAAAAAAAAATAATAATA 1911

Qy        358 GCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTAC 417
               |||  | |||  |||||     | |  ||   |||||    |  ||     |  | ||
Db       1910 ATATAAAAAATTTATTAATAAAAAAATATAAATAAAAAAAAAATAAATATTAATTAAATAA 1851

Qy        418 CTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTC 477
              |  |||  |||       | | |  |||||  | ||||| ||   |  ||||||
Db       1850 TTTATTTTATATATATTAAAATAAAATAATTANTAAAAATTATAAAATAAAAATAAATAA 1791

Qy        478 AGAAAACATA 487
              | |||  |||
Db       1790 AAAAATAATA 1781


RESULT 55
AAS46429/c
ID   AAS46429 standard; DNA; 6106 BP.
XX
AC   AAS46429;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Tumour suppressor gene derived chemically modified sequence #151.
XX
KW   Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer;
KW   tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW   cytosine methylation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200168912-A2.
XX
PD   20-SEP-2001.
XX
PF   15-MAR-2001; 2001WO-EP002955.
XX
PR   15-MAR-2000; 2000DE-01013847.
PR   06-APR-2000; 2000DE-01019058.
PR   07-APR-2000; 2000DE-01019173.
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2001-602752/68.
XX
PT   Fragments of chemically modified genes associated with tumor suppressor
PT   genes and oncogenes, useful in designing primers and probes for analyzing
PT   diseases associated with cytosine methylation state e.g. cancer.
XX
PS   Claim 1; SEQ ID NO 151; 27pp; English.
XX
CC   The invention relates to a nucleic acid comprising a sequence of 18
CC   bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC   bisulphite, of genes associated with tumour suppression and oncogenes
CC   having a sequence taken from 536 (actually 533 since numbers 408, 458 and
CC   500 are missing from the sequence listing) sequences (Ss) and sequences
CC   complementary to (Ss). The nucleic acid may be a peptide nucleic acid-
CC   oligomer (PNA) of at least 9 nucleotides and may form part of a set of
CC   probes for detecting the cytosine methylation state and/or single
CC   nucleotide polymorphisms and also to be used in an array for analysing
CC   diseases associated with CpG dinucleotides e.g. cancers and tumours. The
CC   probes can also be used in a method for ascertaining genetic and/or
CC   epigenetic parameters for the diagnosis and/or therapy of existing
CC   diseases or the predisposition to specific diseases, by analysing
CC   cytosine methylations. The parameters may be compared to another set of
CC   genetic and/or epigenetic parameters, the differences serving as basis
CC   for diagnosis and/or prognosis events which are disadvantageous to
CC   patients. The present sequence is one of the 533 genomic sequences
CC   derived from tumour suppressor genes and oncogenes. Note: The sequence
CC   data for this patent did not form part of the printed specification, but
CC   was obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 6106 BP; 1938 A; 30 C; 849 G; 3289 T; 0 U; 0 Other;

  Query Match             4.4%;  Score 48.4;  DB 4;  Length 6106;
  Best Local Similarity   45.5%;  Pred. No. 0.74;
  Matches  172;  Conservative    0;  Mismatches  206;  Indels    0;  Gaps    0;

Qy        160 TTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATA 219
              | | ||| ||   || |   | | | ||   |  | |||||| |       | |  | | |
Db       4729 TAAAAAAACAAACAAAAACCTCGAAAAATTACATAACTTAAAAATTAAAATATAATTAAA 4670

Qy        220 TACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTT 279
              | |  |  | ||   ||  | |  | | | | | |  |   |  |  | |   |
Db       4669 AAATAACCAAAATTAAAAAATTTAACCAAAATAAACAAAATAAAAATTAAAATCTAAACT 4610

Qy        280 TGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAA 339
                  ||||||   | ||| |||    |||    ||  |      ||  |      || |||
Db       4609 AAAAAAAAAAAAAAAAAAAAAAAAAACTTCTTTTAATAAATTACTAAAAAAACCTTCAAA 4550

Qy        340 ATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTC 399
              |||   |||   |    | ||||  ||| | ||  ||  | |       | |
Db       4549 TAAAAAATAAAATAAAAATTAAATTAATTTATTCAAATTTTTAATTTCTCTTTATCCACT 4490

Qy        400 AACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTT 459
              |||  |||  ||    |    ||   | |    |   || |||  | ||   ||||| ||
Db       4489 AACTAAAAAATAATTAATAAATTTCAATTTAAAATACCATATAATATTTCAAACAAAATT 4430

Qy        460 TATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTC 519
              |   |  ||  ||  |   |||  |||| | |  |     ||  || | ||| |||||
Db       4429 AAAAATATAACAAAAATTTAAAATAATAAATATAAAAAATATAAAAATAAATTAACTTTA 4370

Qy        520 CAAATATTCTGTCATTCA 537
              ||| ||||  | ||| |
Db       4369 TAAAAATTCATTTATTTA 4352


RESULT 56
ABK40031/c
ID   ABK40031 standard; DNA; 6106 BP.
XX
AC   ABK40031;
XX
DT   21-MAY-2002  (first entry)
XX
DE   Human chemically pretreated gene sequence #57 strand 1.
XX
KW   Human; ds; bisulphite treatment; CpG; DNA methylation; cancer; tumour;
KW   cytostatic; ALDH6; CYP11A; CYP11B1; CYP3A3; DPYD; EPHX2; OCLN; TXNRD1;
KW   UGT8; MRP; pharmacogenomics; SNP; single nucleotide polymorphism.
XX
OS   Homo sapiens.
XX
```

PN WO200202806-A2.
XX
PD 10-JAN-2002.
XX
PF 29-JUN-2001; 2001WO-EP007470.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-154757/20.
XX
PT New nucleic acid, oligonucleotides and peptide nucleic acid-oligomers,
PT useful for detecting cytosine methylation state of genes associated with
PT pharmacogenomics and for therapy of diseases e.g. cancer.
XX
PS Claim 1; SEQ ID NO 113; 24pp; English.
XX
CC The invention relates to a nucleic acid comprising a sequence at least 18
CC bases in length of a segment of the chemically pretreated DNA of genes
CC associated with pharmacogenomics according to one of the sequences of the
CC genes ALDH6 (NM_000693), CYP11A (NM_000781), CYP11B1 (NM_000497), CYP3A3
CC (NM_000776 and NM_017460), DPYD (NM_000110), EPHX2 (NM_001979), OCLN
CC (NM_002538), TXNRD1 (NM_003330), UGT8 (NM_003360), MRP (NM_004996,
CC NM_019900, NM_019901, NM_019902, NM_019862, NM_019898, NM_019899) and
CC their complementary sequences, or a sequence (S1) chosen from 87
CC sequences and their complements. The chemical pretreatment is bisulphite
CC treatment to convert cytosines (but not methyl-cytosines) into uracils.
CC Also included are an oligomer (II) in particular an oligonucleotide or a
CC peptide nucleic acid (PNA)-oligomer, comprising in each case at least one
CC base sequence having a length of 9 nucleotides which hybridises to or is
CC identical to a chemically pretreated DNA of genes associated with
CC pharmacogenomics and their complements, arranged in an array for
CC analysing diseases associated with the methylation state (CpG) and/or
CC detecting SNPs (single nucleotide polymorphisms) of the 87 sequences. The
CC oligomers may also be used as PCR primers. The set of 87 nucleic acids
CC and their complements is useful for diagnosis and therapy of solid
CC tumours and cancer. The present sequence represents one the 87 DNA
CC sequences or its complement. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 6106 BP; 1938 A; 30 C; 849 G; 3289 T; 0 U; 0 Other;

Query Match 4.4%; Score 48.4; DB 6; Length 6106;
Best Local Similarity 45.5%; Pred. No. 0.74;
Matches 172; Conservative 0; Mismatches 206; Indels 0; Gaps 0;

Qy 160 TTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATA 219
Db 4729 TAAAAAAACAAACAAAAACCTCGAAAAATTACATAACTTAAAAATTAAAATATAATTAAA 4670

Qy 220 TACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTT 279
Db 4669 AAATAACCAAAATTAAAAAATTTAACCAAAATAAACAAAATAAAAATTAAAATCTAAACT 4610

Qy 280 TGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAA 339
Db 4609 AAAAAAAAAAAAAAAAAAAAAAAAAACTTCTTTTAATAAATTACTAAAAAAACCTTCAAA 4550

Qy 340 ATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTC 399
Db 4549 TAAAAAATAAAATAAAAATTAAATTAATTTATTCAAATTTTTAATTTCTCTTTATCCACT 4490

Qy 400 AACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTT 459
Db 4489 AACTAAAAAATAATTAATAAATTTCAATTTAAAATACCATATAATATTTCAAACAAAATT 4430

Qy 460 TATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTC 519
Db 4429 AAAAATATAACAAAAATTTAAAATAATAAATATAAAAAATATAAAAATAAATTAACTTTA 4370

Qy 520 CAAATATTCTGTCATTCA 537
Db 4369 TAAAAATTCATTTATTTA 4352

RESULT 57
ABL33472/c
ID ABL33472 standard; DNA; 6106 BP.
XX
AC ABL33472;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 1445.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 1445; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 6106 BP; 1938 A; 30 C; 849 G; 3289 T; 0 U; 0 Other;

Query Match 4.4%; Score 48.4; DB 6; Length 6106;
Best Local Similarity 45.5%; Pred. No. 0.74;
Matches 172; Conservative 0; Mismatches 206; Indels 0; Gaps 0;

Qy 160 TTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATA 219
Db 4729 TAAAAAAACAAACAAAAACCTCGAAAAATTACATAACTTAAAAATTAAAATATAATTAAA 4670

Qy 220 TACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTT 279
Db 4669 AAATAACCAAAATTAAAAAATTTAACCAAAATAAACAAAATAAAAATTAAAATCTAAACT 4610

```
Qy        280 TGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAA 339
              ||||||   | ||| |||    |||    ||  |     ||  |    || |||
Db       4609 AAAAAAAAAAAAAAAAAAAAAAAAAACTTCTTTTAATAAATTACTAAAAAAACCTTCAAA 4550

Qy        340 ATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTC 399
             |||   |||   |    | ||||  ||| | ||  ||  | |      | |
Db       4549 TAAAAAATAAAATAAAAATTAAATTAATTTATTCAAATTTTTAATTTCTCTTTATCCACT 4490

Qy        400 AACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTT 459
             |||  |||  ||    |      ||   | |    |   || |||  | ||   ||||| ||
Db       4489 AACTAAAAAATAATTAATAAATTTCAATTTAAAATACCATATAATATTTCAAACAAAATT 4430

Qy        460 TATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTC 519
              |   |  ||  ||  |   |||  |||| | |  |     ||  || | ||| |||||
Db       4429 AAAAATATAACAAAAATTTAAAATAATAAATATAAAAAATATAAAAATAAATTAACTTTA 4370

Qy        520 CAAATATTCTGTCATTCA 537
              ||| ||||  | ||| |
Db       4369 TAAAAATTCATTTATTTA 4352
```

RESULT 58
ABL32202/c
ID ABL32202 standard; DNA; 8246 BP.
XX
AC ABL32202;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 175.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 175; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 8246 BP; 2406 A; 196 C; 1801 G; 3843 T; 0 U; 0 Other;

```
  Query Match             4.4%;  Score 48.4;  DB 6;  Length 8246;
  Best Local Similarity   47.6%;  Pred. No. 0.8;
  Matches  175;  Conservative    0;  Mismatches  191;  Indels    2;  Gaps    1;

Qy        150 ATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGT 209
              ||||| ||| | | || |||   || | | ||  |  | |  |||  || || |  ||
Db       7840 ATATTTATCCTAATAACTCAAAAAAAAATATTACATTAAAAATCAAATTTAATAATTTTA 7781

Qy        210 TTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTT 269
              | ||  |||| |      |   |  |||| |      | | |           || | ||
Db       7780 AATATTCTTATAATATAAAAACCATACAAAAAAAATCCCTAAATAATATAAACTTTAATT 7721

Qy        270 GTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAG 329
              |||||  ||  || ||||     | ||  |  |  | | | |     | || ||  ||
Db       7720 ATTTTTCTTTAATTAAAAATTTAAAAAACCATCTATAATTATAATTCAAACAAATTATAT 7661

Qy        330 CCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAG 389
              |    || | ||||   ||||  | | |  |  | |||     ||| ||  |    ||
Db       7660 TAAAAAAAACAATAAAATTAAAA--AATTCCAACAAAATTCTAAAATATTTTATAATAAT 7603

Qy        390 GTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTA 449
              | |     || ||| |   ||| ||    |||||  |   |    ||||  ||| |
Db       7602 TAACACTAATAAACTACATAAATATTATAACCTTTCACAATAAAACTACAATACTAATAA 7543

Qy        450 CCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAAC 509
              |||        ||| |   |||  |||||  |  ||||   || |  | |||
Db       7542 TATACCTTTTCTAAAAATAAAAATAATCAATAAACACTAACTTTTCCTTAATTATTAAAT 7483

Qy        510 ATTTACTT 517
              ||  ||||
Db       7482 ATAAACTT 7475
```

RESULT 59
ABK31511/c
ID ABK31511 standard; DNA; 47108 BP.
XX
AC ABK31511;
XX
DT 23-APR-2002 (first entry)
XX
DE Signal transduction associated gene modified complementary DNA #177.
XX
KW Human; signal transduction associated gene; cytosine methylation state;
KW CpG island; signal transduction associated disease; solid tumour; cancer;
KW antitumour; cytostatic; mutant; ds.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200200926-A2.
XX
PD 03-JAN-2002.
XX
PF 29-JUN-2001; 2001WO-EP007472.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-147896/19.
XX
PT Oligonucleotide for diagnosis and therapy of diseases associated with
PT signal transduction e.g. cancer, comprises chemically modified genomic
PT sequences of genes associated with signal transduction.
XX
PS Claim 1; SEQ ID NO 354; 24pp; English.
XX
CC The present invention relates to chemically modified DNA sequences of

```
CC   signal transduction associated genes. The DNA sequences are chemically
CC   modified using a solution of bisulphite, hydrogen sulphite or disulphite.
CC   Also disclosed are oligonucleotides and/or PNA oligomers for detecting
CC   the cytosine methylation state (CpG islands) of these genes, and a method
CC   for the diagnosis and/or therapy of genetic and epigenetic parameters of
CC   genes associated with signal transduction. The genomic DNA can be
CC   obtained from cells or cellular components which contain DNA, e.g. cell
CC   lines, biopsies, blood, sputum, stool, urine, cerebral-spinal fluid,
CC   tissue embedded in paraffin such as tissue from eyes, intestine, kidney,
CC   brain, heart, prostate, lung, breast or liver, histologic object slides,
CC   and all their possible combinations. The sequences of the invention are
CC   useful for the diagnosis and therapy of diseases associated with signal
CC   transduction e.g. solid tumours and cancer. ABK31158-ABK31545 represent
CC   chemically pretreated genomic DNA sequences of different genes associated
CC   with signal transduction, or their complementary sequences. Note: The
CC   sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from the
CC   European Patent Office
XX
SQ   Sequence 47108 BP; 15349 A; 429 C; 8736 G; 22594 T; 0 U; 0 Other;

  Query Match             4.4%;  Score 48.4;  DB 6;  Length 47108;
  Best Local Similarity   46.9%;  Pred. No. 1.2;
  Matches  151;  Conservative    0;  Mismatches  171;  Indels    0;  Gaps    0;

Qy        249 CAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATT 308
              || | ||     | |   ||| ||| |   || || || | |     |  |       |||
Db      36805 CATAAATTTCTTTTTCTTTTTCTTTCTTTCTTTTTAAATATAAAATTTTAACCTACGATT 36746

Qy        309 CAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACAT 368
              |  |   |  ||| |   |   |    ||||| ||  | ||||  |    |||| |||
Db      36745 CTAAATCATTTCAAAACACAAAAACAAAAAAATAAACTATAAAAATAACCTATATAACAA 36686

Qy        369 TACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTA 428
              |  |      || |       |   || || | ||| | ||    || ||  |  |     |
Db      36685 TTATATTCTTTACTACTCTACTAAATAATTTAACATTAATTACTATTAAATTTTAAAAAA 36626

Qy        429 TCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAA 488
              |   | |  | ||| ||       |||||       | |   |||||  |     ||   | |
Db      36625 TACTATTTTACATATAATAATTAACAAAAAAATATATAATAATAAACAATTAATTTAAAC 36566

Qy        489 TTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGTCATTCAGGAGTATGGAA 548
              | |||||   |  ||| |   |||| |  || |  ||||   || ||  | |||  ||
Db      36565 ATATTTATTCTTTCATAACATTCTTACCTCTCATACCTTCTTAAATCCAAAAATATCAAA 36506

Qy        549 GTATCGATGGCTTCTTTAAAAT 570
              | || ||  |   |||||  |
Db      36505 ATTTCTATAACAAAATTAAACT 36484


RESULT 60
ACA64845
ID   ACA64845 standard; DNA; 115218 BP.
XX
AC   ACA64845;
XX
DT   27-JUN-2003  (first entry)
XX
DE   Human HNRNP GP43 DNA corresponding to AL034397.
XX
KW   Human; chronic inflammatory joint disease; infection; tumour;
KW   antiinflammatory; cytostatic; antiarthritic; antirheumatic;
KW   immunosuppressive; gene therapy; etiological pathogenicity; ds.
XX
OS   Homo sapiens.
XX
PN   DE10127572-A1.
XX
PD   05-DEC-2002.
XX
PF   30-MAY-2001; 2001DE-01027572.
XX
PR   30-MAY-2001; 2001DE-01027572.
XX
PA   (PATH-) PATHOARRAY GMBH.
XX
PI   Haeupl T,  Ungethuem U,  Blaess S;
XX
DR   WPI; 2003-240797/24.
XX
PT   Reagents for diagnosis, study and therapy of chronic inflammatory joint
PT   and other diseases, comprises any of many specified genes or derived
PT   proteins.
XX
PS   Claim 1; Page; 12pp; German.
XX
CC   This invention describes a novel reagent for diagnosis, molecular
CC   definition and therapy of chronic inflammatory joint diseases, and other
CC   inflammatory disorders, infective or tumour diseases in humans. The
CC   products of the invention have antiinflammatory, cytostatic,
CC   antiarthritic, antirheumatic and immunosuppressive activity and can be
CC   used for gene therapy. The reagent of the invention and any proteins and
CC   antibodies derived from it, are used (i) for analysing tissue and blood
CC   samples for medical diagnosis; (ii) for diagnosis and characterisation of
CC   chronic joint diseases, on the basis of molecular characterisation, and
CC   determining the etiological pathogenicity principle of as yet
CC   uncharacterised inflammatory diseases, also monitoring progression and/or
CC   treatment of disease, and optimisation of therapy and (iii) for
CC   developing treatments for inflammatory diseases, particularly of joints,
CC   infections and tumours. ACA64801-ACA64965 represent human polynucleotides
CC   used in the method of the invention
XX
SQ   Sequence 115218 BP; 35618 A; 23477 C; 22286 G; 33837 T; 0 U; 0 Other;

  Query Match             4.4%;  Score 48.4;  DB 8;  Length 115218;
  Best Local Similarity   45.5%;  Pred. No. 1.5;
  Matches  212;  Conservative    0;  Mismatches  251;  Indels    3;  Gaps    1;

Qy        138 TTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCT 197
              ||   ||   | |||||    |||| | || |  | ||| | || || || ||   || |
Db      17014 TTATATTATTATATATTATATATTATATATTATATTTATATTATTATATAATAT---ATAT 17070

Qy        198 TAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATAC 257
              || | |   | || | ||| ||||      ||   | || ||  | | |      |||
Db      17071 TATATATTATATTATATTATTATATAATATTTTATATTATATTATATTATTATATAATAT 17130

Qy        258 AGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAG 317
              |  |  |    | | ||   ||  |  | || |  ||| |     ||  |||     | |
Db      17131 ATATTATATGATATATTATATTATTATATAATATATATTATATGATATATTATATTACAT 17190

Qy        318 ATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATC 377
              |||| ||    |  | |||| |   | || |    |||| | ||    || |
Db      17191 TATATATTATATTATATGATATATTATATGATATATTATATTATATGATATATTATTATTA 17250

Qy        378 TTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGC 437
              |       | |   || |    |  | |    |||   | |   | | ||    |
Db      17251 TATATTATATAATATAATATAATATATCATATAATATATCATATAATATAATATATCATA 17310

Qy        438 AAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATG 497
              ||||  |   ||  | ||| | |||  |  ||   ||   | || | ||| ||| | |||
Db      17311 TAATATATTATATAATAAAATATATAATATAATATATAATATAATATATTATTATATATA 17370

Qy        498 TACTATTAAAACATTTACTTTCCAAATATTCTGTCATTCAGGAGTATGGAAGTATCGATG 557
              || |||    |     |  ||  | ||||  |   ||  |  | |||  ||      ||
Db      17371 TAATATATTATATAACATATTATATATATAATAATATATAATTATTATATAATATATTATA 17430

Qy        558 GCTTCTTTAAAATGAAGCAGGAGGGTCTGGCAGAGAGTATCTATGA 603
                | | ||| || |   |  |     |     | | ||| ||| |
Db      17431 TTATATATAATATAATATAATATATATTATATTATATTATATATTA 17476


RESULT 61
ABL33015/c
ID   ABL33015 standard; DNA; 17967 BP.
XX
AC   ABL33015;
```

XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 988.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 988; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 17967 BP; 5271 A; 247 C; 3443 G; 9006 T; 0 U; 0 Other;

Query Match 4.4%; Score 48.2; DB 6; Length 17967;
Best Local Similarity 46.5%; Pred. No. 1.1;
Matches 192; Conservative 0; Mismatches 218; Indels 3; Gaps 1;

```
Qy       128 CTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAA 187
            || |||| |   |  || |    ||||  |  || | ||  |  || ||    | || |
Db     10059 CTAATATAACAATAATTTATTTTTATTTTTATTTTCCAAATAAATATTAAACATATACAT 10000

Qy       188 GATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTT 247
              ||      |  |     |   ||     ||        |  |    | |||||  |  |
Db      9999 TATAATCAATACATTTACTAAACTTAAACTTAATAAATATAAAAAAATATAAAACCCAAT 9940

Qy       248 ACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCAT 307
              || |      |  ||  |  |  |||       ||  |||||    |||   |    |
Db      9939 TCATAACCTACGAAATCAAAATTATTTAAAAAAATTTTAAAAATACTTAATATATCTTAA 9880

Qy       308 TCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACA 367
               |   ||  |    | ||||   | |  |||||  ||    ||   | |  || || ||
Db      9879 ATAATAAATAACATAAAATTAAAAAATATAAAATCTAATTTTAATCAAATCAATCTTTCA 9820

Qy       368 TTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTT 427
              || | ||||    |  |||  |||||  | |  |||  |  |  |  |||  ||
Db      9819 ATATTAAATCAAATAAATAAACAAAAAAAAAAATCAAAAACAAAACCATCTAACTTAATT 9760

Qy       428 ATCGCACT---GCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAAC 484
```

```
            ||| | |   | || | ||| ||  || |  ||| || |   ||||||  || | |||
Db      9759 CTCGTAATATCACTAAAAAAAAAATAATACTACCTTTTTTCTTATAATAATAACAAATAAC 9700

Qy       485 ATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGTCATTCA 537
             || | ||    | |    || | ||  |||| | | || |   | |||||
Db      9699 ATTAATTAAACTCTTACTATATACCAAATACTCTACTAACAACTTATCATTTA 9647
```

RESULT 62
ABQ67014/c
ID ABQ67014 standard; DNA; 6676 BP.
XX
AC ABQ67014;
XX
DT 28-AUG-2002 (first entry)
XX
DE Human angiogenesis associated polynucleotide SEQ ID NO 44.
XX
KW Human; angiogenesis; methylation; eye disease; glaucoma; tumour;
KW inflammation; rheumatoid arthritis; diabetic retinopathy; antiulcerds;
KW macular degeneration; inflammatory bowel disease; Crohn's disease;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiarteriosclerotic; ds.
XX
OS Homo sapiens.
XX
PN WO200246454-A2.
XX
PD 13-JUN-2002.
XX
PF 06-DEC-2001; 2001WO-EP014320.
XX
PR 06-DEC-2000; 2000DE-01061338.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Schacht O;
XX
DR WPI; 2002-500450/53.
XX
PT New nucleic acid fragments from chemically treated angiogenesis-
PT associated genes, useful for determining methylation status, e.g. in
PT diagnosis or treatment of cancer.
XX
PS Claim 1; SEQ ID NO 44; 41pp + Sequence Listing; German.
XX
CC The invention relates to a nucleic acid (I) comprising a segment of 18
CC bases of chemically pretreated DNA of angiogenesis-associated genes (II)
CC having sequences (ABQ66971-ABQ67178) or their complements. (I), also
CC related oligomers, are used to evaluate the methylation status and/or
CC single-nucleotide polymorphisms, in angiogenesis-related genes, for
CC diagnosis and treatment of eye diseases, proliferative retinopathy,
CC neovascular glaucoma, solid tumours, inflammation, rheumatoid arthritis,
CC diabetic retinopathy, macular degeneration caused by neovascularisation,
CC psoriasis, arteriosclerosis, inflammatory bowel diseases, ulcers and
CC Crohn's disease. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 6676 BP; 1891 A; 55 C; 1377 G; 3353 T; 0 U; 0 Other;

Query Match 4.4%; Score 48; DB 6; Length 6676;
Best Local Similarity 46.0%; Pred. No. 0.93;
Matches 162; Conservative 0; Mismatches 190; Indels 0; Gaps 0;

```
Qy       171 TTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGA 230
             | || | || | ||    || | |  | ||| |||      |    |  ||  |  | | |
Db      1313 TAAAAAATTCTATATTTTATTTTAAATAAAACACAAATAATACAAATCATCTTATCATTA 1254

Qy       231 TTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAA 290
             | | ||  | | | ||  |    |    |   ||   | | |||  |  ||  || | ||
Db      1253 TATATTATATAAAACTAAATTATCATTATTCATAATTATTTTTCTTATTCTTTAACAATA 1194
```

```
Qy        291 GGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAA 350
              | |||| ||    ||  | | ||| | || | | ||  ||    ||||| | | | ||
Db       1193 TTTTTAAAAAACAAAATAAATATAAAAAACAAAAAATAATCAACCAAAAATAACAAAAAA 1134

Qy        351 AATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGT 410
              ||  |    | | | | |||     |  ||      |     |  | |   ||   |
Db       1133 AAATAAAAAAATTAATCTAACTACCCATACCAACTTTAAACTTACCCACTATAAAATCCT 1074

Qy        411 ATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAA 470
              | || |  | ||| | ||||  |    || |   |  || |  |  |  | |    |
Db       1073 AAACAATTTACTTACCTATCTAAACCTTAAAAATCAAAACTATCACCTAAAATAACATAT 1014

Qy        471 TAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAA 522
              ||||  |     | | ||  | |||| ||   || | | | |   || ||
Db       1013 TAATAACTACCTAAAAAAACTCTTATATAAATTTCACAAAATAAAAATCAAA 962
```

RESULT 63
ABL32226
ID ABL32226 standard; DNA; 6509 BP.
XX
AC ABL32226;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 199.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 199; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 6509 BP; 1585 A; 222 C; 1749 G; 2953 T; 0 U; 0 Other;

Query Match 4.4%; Score 47.8; DB 6; Length 6509;
Best Local Similarity 51.1%; Pred. No. 1;
Matches 112; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

```
Qy        285 AAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAA 344
              ||||||   || | ||||    || | ||||||| || |   |    | | || | ||
Db       2513 AAAAAAAAAAATGAGGAAAGAGTTTAATAGAAAGATTATTTGAAAAAGGGGTAAATAATAT 2572

Qy        345 ACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCT 404
              | |||||   |  |||| | |||  ||| | ||  |   ||   |||||| |
Db       2573 GAAGAAAATTTATTAATATAAGATTTATTATTTTTTAAAGAAAGATTAAAAAATATGATA 2632

Qy        405 AAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTT 464
              ||  || || |      | |  ||     |    | ||  || |     |  ||||||||
Db       2633 AATTGTTTAATTTTATTTATAATAAAAAAAGAAATATGAAAGTGTTTTCGGTTTTTATTT 2692

Qy        465 TGGAAATAATCTCAGAAAACATAATTTTTTATGTACTAT 503
              |||       || || ||| ||  | || || | |
Db       2693 ATTTAATTGGTAAGGATAAAATATTTGATAATATATTGT 2731
```

RESULT 64
ABL33943/c
ID ABL33943 standard; DNA; 12138 BP.
XX
AC ABL33943;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 1916.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 1916; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 12138 BP; 4076 A; 74 C; 2265 G; 5723 T; 0 U; 0 Other;

Query Match 4.4%; Score 47.8; DB 6; Length 12138;
Best Local Similarity 46.5%; Pred. No. 1.2;

```
  Matches  154;  Conservative    0;  Mismatches  177;  Indels    0;  Gaps    0;

Qy        197 TTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATA 256
              | ||| |||||   ||  | ||| ||| | || || || ||| | |  | | | | | |
Db       4766 TAAAATACATTACATTTATTATACACTTCATTTATATTATTATTATATATAATATATAAT 4707

Qy        257 CAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAA 316
              |   | | ||    ||     |  || ||||    || |||        ||
Db       4706 AAAACAATTATACAATTCACCATAATTTAAAATCAATAAAAACCCTAAACTTATTTTCCT 4647

Qy        317 GATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAAT 376
              |   ||||  | ||  |  |||| |||| |  |||| |    | |||| |||  |  |
Db       4646 AAAACTATACCATCCCATCTAAAAATAAAAAACAAATCATCAAACATTAAATTCTTATAA 4587

Qy        377 CTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTG 436
              ||  |||    ||    |||     |   | ||| |   | |||    ||  |
Db       4586 AAAACATACAACCTAAATCCCTCACGTATACAATTTACAATACGATTTACATTCCTATAA 4527

Qy        437 CAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTAT 496
              ||    | |||||  | ||  | |       |   || |||| | || |  |    | ||
Db       4526 AAATATAATATTACTTCTAACCTAACAAAAAACGAAAACTCAAACAATAATACCGATAAT 4467

Qy        497 GTACTATTAAAACATTTACTTTCCAAATATT 527
              |  || |  | |  |||     |||  ||
Db       4466 AAAAAATAACTATAAATACAAATAAAACTTT 4436


RESULT 65
ABK28336/c
ID   ABK28336 standard; DNA; 12138 BP.
XX
AC   ABK28336;
XX
DT   23-APR-2002  (first entry)
XX
DE   DNA transcription associated complementary genomic DNA #105.
XX
KW   DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;
KW   PNA; cytosine methylation state; SNP; retroviral infection; gene; ds;
KW   single nucleotide polymorphism; adenosine deaminase deficiency; cancer;
KW   viral infection; Sezary syndrome; haematological disorder; tuberculosis;
KW   immunological disorder; Werner syndrome; developmental disorder;
KW   psoriasis; Rieger's syndrome; neurological disorder; erythropoiesis;
KW   neurodegenerative disorder; Waardenburg syndrome; Niemann-Pick disease;
KW   myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;
KW   angiogenesis; congenital heart disease; HDR syndrome; gene therapy;
KW   polyglutamine disorder; solid tumour.
XX
OS   Unidentified.
XX
PN   WO200192565-A2.
XX
PD   06-DEC-2001.
XX
PF   06-APR-2001; 2001WO-EP003973.
XX
PR   06-APR-2000; 2000DE-01019058.
PR   07-APR-2000; 2000DE-01019173.
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-090046/12.
XX
PT   New nucleic acids or oligomers, useful for diagnosing or treating
PT   diseases associated with DNA transcription, e.g. immunological disorders,
PT   Werner syndrome, psoriasis, myocardial infarction, solid tumors or
PT   cancer.
XX
PS   Claim 1; SEQ ID NO 210; 32pp; English.
XX
CC   The invention relates to a nucleic acid, which comprises a segment of the
CC   chemically pretreated DNA of genes associated with DNA transcription from
CC   one of 346 sequences, and an oligomer, in particular an oligonucleotide
CC   or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical
CC   to the chemically pretreated DNA of genes associated with DNA
CC   transcription. The set of oligomer probes are useful for detecting the
CC   cytosine methylation state and/or single nucleotide polymorphisms (SNPs)
CC   in a chemically pretreated genomic DNA. The nucleic acids are useful for
CC   diagnosing or treating diseases associated with DNA transcription
CC   (particularly with the methylation status), e.g. adenosine deaminase
CC   deficiency, viral infection, retroviral infection, Sezary syndrome,
CC   haematological disorders, immunological disorders, Werner syndrome,
CC   tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,
CC   neurological disorders, neurodegenerative disorders, Waardenburg
CC   syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial
CC   infarction, hypertension, angiogenesis, erythropoiesis, congenital heart
CC   disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours
CC   or cancer. Sequences ABK28127-ABK28472 represent DNA transcription
CC   associated genomic DNA molecules of the invention. Note: The sequence
CC   data for this patent did not form part of the printed specification but
CC   was obtained in electronic format directly from the European Patent
CC   Office
XX
SQ   Sequence 12138 BP; 4076 A; 74 C; 2265 G; 5723 T; 0 U; 0 Other;

  Query Match             4.4%;  Score 47.8;  DB 6;  Length 12138;
  Best Local Similarity   46.5%;  Pred. No. 1.2;
  Matches  154;  Conservative    0;  Mismatches  177;  Indels    0;  Gaps    0;

Qy        197 TTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATA 256
              | ||| |||||   ||  | ||| ||| | || || || ||| | |  | | | | | |
Db       4766 TAAAATACATTACATTTATTATACACTTCATTTATATTATTATTATATATAATATATAAT 4707

Qy        257 CAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAA 316
              |   | | ||    ||     |  || ||||    || |||        ||
Db       4706 AAAACAATTATACAATTCACCATAATTTAAAATCAATAAAAACCCTAAACTTATTTTCCT 4647

Qy        317 GATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAAT 376
              |   ||||  | ||  |  |||| |||| |  |||| |    | |||| |||  |  |
Db       4646 AAAACTATACCATCCCATCTAAAAATAAAAAACAAATCATCAAACATTAAATTCTTATAA 4587

Qy        377 CTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTG 436
              ||  |||    ||    |||     |   | ||| |   | |||    ||  |
Db       4586 AAAACATACAACCTAAATCCCTCACGTATACAATTTACAATACGATTTACATTCCTATAA 4527

Qy        437 CAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTAT 496
              ||    | |||||  | ||  | |       |   || |||| | || |  |    | ||
Db       4526 AAATATAATATTACTTCTAACCTAACAAAAAACGAAAACTCAAACAATAATACCGATAAT 4467

Qy        497 GTACTATTAAAACATTTACTTTCCAAATATT 527
              |  || |  | |  |||     |||  ||
Db       4466 AAAAAATAACTATAAATACAAATAAAACTTT 4436


RESULT 66
ABQ67149/c
ID   ABQ67149 standard; DNA; 40324 BP.
XX
AC   ABQ67149;
XX
DT   28-AUG-2002  (first entry)
XX
DE   Human angiogenesis associated polynucleotide SEQ ID NO 179.
XX
KW   Human; angiogenesis; methylation; eye disease; glaucoma; tumour;
KW   inflammation; rheumatoid arthritis; diabetic retinopathy; antiulcerds;
KW   macular degeneration; inflammatory bowel disease; Crohn's disease;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiarteriosclerotic; ds.
XX
OS   Homo sapiens.
XX
```

PN WO200246454-A2.
XX
PD 13-JUN-2002.
XX
PF 06-DEC-2001; 2001WO-EP014320.
XX
PR 06-DEC-2000; 2000DE-01061338.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Schacht O;
XX
DR WPI; 2002-500450/53.
XX
PT New nucleic acid fragments from chemically treated angiogenesis-
PT associated genes, useful for determining methylation status, e.g. in
PT diagnosis or treatment of cancer.
XX
PS Claim 1; SEQ ID NO 179; 41pp + Sequence Listing; German.
XX
CC The invention relates to a nucleic acid (I) comprising a segment of 18
CC bases of chemically pretreated DNA of angiogenesis-associated genes (II)
CC having sequences (ABQ66971-ABQ67178) or their complements. (I), also
CC related oligomers, are used to evaluate the methylation status and/or
CC single-nucleotide polymorphisms, in angiogenesis-related genes, for
CC diagnosis and treatment of eye diseases, proliferative retinopathy,
CC neovascular glaucoma, solid tumours, inflammation, rheumatoid arthritis,
CC diabetic retinopathy, macular degeneration caused by neovascularisation,
CC psoriasis, arteriosclerosis, inflammatory bowel diseases, ulcers and
CC Crohn's disease. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 40324 BP; 12279 A; 436 C; 8127 G; 19482 T; 0 U; 0 Other;

Query Match 4.4%; Score 47.6; DB 6; Length 40324;
Best Local Similarity 47.5%; Pred. No. 1.8;
Matches 172; Conservative 0; Mismatches 189; Indels 1; Gaps 1;

```
Qy      149 AATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTG 208
            ||  ||  ||   | | || |   || | |  |  | || || |  | | ||  ||| |
Db    14565 AAATTTACTCTCAAAATATAACAAAAAATTAATAAAAAATATATATTTTAAATTACACTT 14506

Qy      209 TTTTGTTAATATACTCCCTTGATTTTTTAAAAGACCTTACAGACATACAGCTATTCATT 268
              |  |||| |  |      | | ||     ||| |   | |  |      |  ||  | ||
Db    14505 AATATTTAAAAACCAATTATTACTTAAAATAAACAAACTTCCAATTCCAACATAACCTTT 14446

Qy      269 TGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTA 328
            | |||||   ||    ||||    || | |  |||    ||  | ||| |   ||| |||
Db    14445 TATTTTTATATTACCAAAAATTTTTAAACAATAATAATATCTAATAAAAAATTTATCTTA 14386

Qy      329 GCCAGTTGAAAATTAAACACAAAATGAGTGCAT-ATTACATTACTTAATCTTGCAGTCAA 387
                |   |  ||     |  | | |  || |  |  ||  ||  |||   | | ||
Db    14385 TATCTATATTACACAACTTTTACCTAAATTAATAAACATTTTCTTTTTTCTAAAAATAAA 14326

Qy      388 AGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAAT 447
            |  |||| | |||| |      |    |  || ||  | | |  || |  |||| || |
Db    14325 ATATAAATACTCAAACCTCCTCTTCCATTTCTACTACCATTTTTCATTTTAAATCGATAA 14266

Qy      448 TACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAA 507
            | |  ||||| |   |   |||| |  |||  |||   |  |||  || |   || ||
Db    14265 TCCTTCAAATATCTATCATAAAATTAATTCACTAAAATAATATTTAAATTTTAAATAAAC 14206

Qy      508 AC 509
            ||
Db    14205 AC 14204
```

RESULT 67
AAC94812/c
ID AAC94812 standard; cDNA; 549 BP.
XX
AC AAC94812;
XX
DT 19-FEB-2001 (first entry)
XX
DE Cat flea hindgut and Malpighian tubule (HMT) cDNA, SEQ ID NO:1307.
XX
KW Cat flea; hindgut and Malpighian tubule nucleic acid; HMT;
KW flea infestation; vaccine; antiparasitic; therapeutic target; diagnosis;
KW detection; ss.
XX
OS Ctenocephalides felis.
XX
PN WO200061621-A2.
XX
PD 19-OCT-2000.
XX
PF 07-APR-2000; 2000WO-US009437.
XX
PR 09-APR-1999; 99US-0128704P.
XX
PA (HESK-) HESKA CORP.
XX
PI Brandt KS, Gaines PJ, Stinchcomb DT, Wisnewski N;
XX
DR WPI; 2000-656323/63.
XX
PT Flea Malpighian tubule and head and nerve cord tissue derived nucleic
PT acids useful for the prevention, diagnosis and treatment of flea
PT infestations.
XX
PS Claim 26; Page 666-667; 964pp; English.
XX
CC The invention relates to novel cat flea (Ctenocephalides felis) nucleic
CC acids which are expressed in hindgut and Malpighian tubule (HMT) tissue
CC or head and nerve cord (HNC) tissue. The invention also relates to the
CC encoded proteins. The invention additionally encompasses expression
CC constructs, recombinant viruses and recombinant cells comprising the
CC nucleic acids of the invention, recombinant production of the proteins,
CC antibodies against the proteins, a method of identifying inhibitors of
CC the proteins, and compositions comprising the inhibitors for
CC administration to an animal. The nucleic acids, and the proteins they
CC encode may be used in the prevention, treatment and diagnosis of diseases
CC associated with flea infestations. For example, the nucleic acids may be
CC used to produce an HMT or HNC protein according to standard recombinant
CC DNA methodology by inserting the nucleic acids into a host cell and
CC culturing the cell to express the protein. The HMT and HNC nucleic acids
CC may also be used as DNA probes in diagnostic assays (e.g., PCR) to detect
CC and quantitate the presence of cat flea or other homologous nucleic acid
CC sequences in samples. They may also be used to study the expression and
CC function of the proteins and their role in metabolism. The HMT and HNC
CC proteins may be used as antigens in the production of specific
CC antibodies, and in assays to identify modulators (agonists and
CC antagonists) of HMT and/or HNC protein expression and activity. The anti-
CC HMT/HNC protein antibodies and antagonists may also be used to
CC downregulate protein expression and activity. The antibodies may also be
CC used as diagnostic agents for detecting the presence of flea polypeptides
CC in samples (e.g., by enzyme linked immunosorbent assay (ELISA)). The
CC present sequence represents a cat flea HMT cDNA of the invention
XX
SQ Sequence 549 BP; 190 A; 67 C; 71 G; 219 T; 0 U; 2 Other;

Query Match 4.3%; Score 47.4; DB 3; Length 549;
Best Local Similarity 47.3%; Pred. No. 0.68;
Matches 175; Conservative 0; Mismatches 192; Indels 3; Gaps 1;

```
Qy      136 AGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCAT 195
            | ||| ||||| |   |   || ||    ||| | ||||||   ||    || | |
Db      451 ATTTTAGTTGAAATCAAAGTATTATACATAATTAATTAATAACATTTACTAATAAATGTA 392

Qy      196 CTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACAT 255
                ||      || | |   ||  ||      |||  | |||   | |     |||  ||
Db      391 GCAAAGTGTNGTGCTGTTAGAACCTAAAAATGTGAAATATTTTTTAAATAAACCAGTGAT 332

Qy      256 ACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAA 315
```

```
Db        331 TCAAAAATTAGTATTAATCAGATACCTGCAAAATATGTAAATAGATATTTATATACACCT 272

Qy        316 AGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAA 375

Db        271 AAAGCAATAGTATTTCAGATGGCGACAGATATAACAGTAAACG---AAAAAGTTTTTTAT 215

Qy        376 TCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACT 435

Db        214 TCATCCAATTATTTAAAAAAAATGAAATTATTTTTATTAAGTTTAATTGTTTAAAATAAA 155

Qy        436 GCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTA 495

Db        154 TAAATAATTCAATTTCACAAAAATTTGTCTGGTTTAATGCCGGCATAAAAAACTTTTTTG 95

Qy        496 TGTACTATTA 505

Db         94 TTTACTATTA 85
```

RESULT 68
AAS46514
ID AAS46514 standard; DNA; 5297 BP.
XX
AC AAS46514;
XX
DT 18-DEC-2001 (first entry)
XX
DE Tumour suppressor gene derived chemically modified sequence #236.
XX
KW Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer;
KW tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW cytosine methylation; ds.
XX
OS Homo sapiens.
XX
PN WO200168912-A2.
XX
PD 20-SEP-2001.
XX
PF 15-MAR-2001; 2001WO-EP002955.
XX
PR 15-MAR-2000; 2000DE-01013847.
PR 06-APR-2000; 2000DE-01019058.
PR 07-APR-2000; 2000DE-01019173.
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2001-602752/68.
XX
PT Fragments of chemically modified genes associated with tumor suppressor
PT genes and oncogenes, useful in designing primers and probes for analyzing
PT diseases associated with cytosine methylation state e.g. cancer.
XX
PS Claim 1; SEQ ID NO 236; 27pp; English.
XX
CC The invention relates to a nucleic acid comprising a sequence of 18
CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC bisulphite, of genes associated with tumour suppression and oncogenes
CC having a sequence taken from 536 (actually 533 since numbers 408, 458 and
CC 500 are missing from the sequence listing) sequences (Ss) and sequences
CC complementary to (Ss). The nucleic acid may be a peptide nucleic acid-
CC oligomer (PNA) of at least 9 nucleotides and may form part of a set of
CC probes for detecting the cytosine methylation state and/or single
CC nucleotide polymorphisms and also to be used in an array for analysing
CC diseases associated with CpG dinucleotides e.g. cancers and tumours. The
CC probes can also be used in a method for ascertaining genetic and/or
CC epigenetic parameters for the diagnosis and/or therapy of existing
CC diseases or the predisposition to specific diseases, by analysing
CC cytosine methylations. The parameters may be compared to another set of
CC genetic and/or epigenetic parameters, the differences serving as basis
CC for diagnosis and/or prognosis events which are disadvantageous to
CC patients. The present sequence is one of the 533 genomic sequences
CC derived from tumour suppressor genes and oncogenes. Sequences with even
CC numbered Seq ID numbers are the complementary sequence of the
CC corresponding odd numbered sequence (e.g. ID 2 and ID1, ID 536 and ID
CC 535, except for those whose partner sequence is missing). Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 5297 BP; 1353 A; 107 C; 1126 G; 2711 T; 0 U; 0 Other;

```
  Query Match             4.3%;  Score 47.4;  DB 4;  Length 5297;
  Best Local Similarity   44.7%;  Pred. No. 1.2;
  Matches  183;  Conservative    0;  Mismatches  226;  Indels    0;  Gaps    0;

Qy         87 TCTTCCTTATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGA 146

Db       1136 TGTTTTTTTATTTTTTGATTTAGTTTTAGTTATTTAGCGATATTTTGTGTTTTTTTTATT 1195

Qy        147 GAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACAT 206

Db       1196 AAAAATTTTGTTTTGAGAGATTATTTTATATTATGGAATAAAAATTGGTATTAGGTATTT 1255

Qy        207 TGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCA 266

Db       1256 TTTTTTTATTGGTTATTTAAATTGATGATAAGACGGATGTTATATATATGTATTTATATA 1315

Qy        267 TTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATAT 326

Db       1316 TTGTAGATTGTAGTAGGGATTAAAGTTTTAAGGAGATGTGGTAAAATAATGTAATTATAT 1375

Qy        327 TAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCA 386

Db       1376 TTTCGAATTAATTGGATAAAAAAAGGATTTGTATTTTTAAATTCGAAAGTAGGTATTTTT 1435

Qy        387 AAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAA 446

Db       1436 TGTTTAAAAAATATTTTAGATGAAAGAGTTATTTTTTTATTAGTTTTTTTTGATTTGGAA 1495

Qy        447 TTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTA 495

Db       1496 TCGTTTTTGGTTTTTATGTTTATTGAATTAGATATTATTTAAGAATATA 1544
```

RESULT 69
ABL33020/c
ID ABL33020 standard; DNA; 6092 BP.
XX
AC ABL33020;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 993.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.

XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 993; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 6092 BP; 1760 A; 110 C; 1203 G; 3019 T; 0 U; 0 Other;

Query Match 4.3%; Score 47.4; DB 6; Length 6092;
Best Local Similarity 49.2%; Pred. No. 1.2;
Matches 153; Conservative 0; Mismatches 156; Indels 2; Gaps 1;

```
Qy        215 TAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTT 274
              |||||  ||  |||  |||    |||   ||  || |   |||     ||  |||| |  |
Db       2524 TAATAATCTTCCTAAATTCAAATAATTAACAATAAAATTATAATTTAATAAATTTTTAAT 2465

Qy        275 TGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGT 334
                ||| ||||||||   ||| ||  ||   ||       || || |  || ||   | |
Db       2464 ATCTTTTTTCAAAAACAATATCAATTAAACAATCATCTTTAAAATAAATTACTAATAATT 2405

Qy        335 TGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAA 394
              |      |||     | |       | |  | |||||  | | || ||   ||   | ||
Db       2404 TATTTTCTAACTTTTATAATTACATAAAAAAAATTACAAATTATTTCATCTAA--TTTAA 2347

Qy        395 AAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACA 454
                | ||||  |  |    ||  ||||  ||| | ||   |||  || |  ||  | |||
Db       2346 TATTCAAAATCCATAACTATAACCTTATTTATAATTTTACTAAAATTTAAACCTCCCATC 2287

Qy        455 AATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTA 514
              ||    |  ||         |||   ||| |  ||    ||  ||   |    ||||||
Db       2286 AAAAACAAATTAATCCCCTACTCTTTAAAAAACATAACCTACCTAAATTCTCTACATTTC 2227

Qy        515 CTTTCCAAATA 525
              | ||| |||||
Db       2226 CATTCGAAATA 2216
```

RESULT 70
ADL44697/c
ID ADL44697 standard; DNA; 376 BP.
XX
AC ADL44697;
XX
DT 20-MAY-2004 (first entry)
XX
DE Human ovarian cancer DNA marker #18587.
XX
KW Human; ovarian cancer; ds; tumour; cytostatic; DNA marker.
XX
OS Homo sapiens.
XX
PN WO200170979-A2.
XX
PD 27-SEP-2001.

XX
PF 21-MAR-2001; 2001WO-US009126.
XX
PR 21-MAR-2000; 2000US-0191031P.
PR 25-MAY-2000; 2000US-0207124P.
PR 15-JUN-2000; 2000US-0211940P.
PR 07-JUL-2000; 2000US-0216820P.
PR 25-JUL-2000; 2000US-0220661P.
PR 21-DEC-2000; 2000US-0257672P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Lee J, Lillie J;
XX
DR WPI; 2001-611502/70.
XX
PT Novel isolated nucleic acid molecules (markers) overexpressed in ovarian
PT cancer cells as compared to their normal non-cancerous ovarian cells are
PT used to characterize stage, grade, histological type of ovarian cancer.
XX
PS Disclosure; SEQ ID NO 18587; 106pp; English.
XX
CC The invention relates to nucleic acid markers which are overexpressed in
CC ovarian cancer cells as compared to their expression in normal (i.e. non-
CC cancerous) ovarian cells. The invention also relates to polypeptides
CC encoded by the markers, antibodies that selectively bind to the
CC polypeptides, a method of inhibiting ovarian cancer in a patient at risk
CC of developing ovarian cancer involving inhibiting expression of a gene
CC corresponding to a marker of the invention and a method of treating a
CC patient afflicted with ovarian cancer comprising providing to cells of
CC the patient an antisense oligonucleotide complementary to a marker of the
CC invention. The markers are useful for assessing if a patient is afflicted
CC with ovarian cancer, which involves comparing the level of expression of
CC a marker in a patient sample and a normal level of expression of the
CC marker in a control non-ovarian cancer sample. A difference between the
CC expression levels indicates ovarian cancer. The level of expression of a
CC marker corresponds to a secreted protein or to a transcribed
CC polynucleotide or its portion. The level of expression of the marker is
CC assessed by detecting the presence in the sample, a protein or protein
CC fragment corresponding to the marker. The presence of protein or protein
CC fragment is detected using an antibody that specifically binds with the
CC protein or protein fragment. Alternatively, the level of expression of
CC the marker is assessed by detecting the presence of a transcribed
CC polynucleotide which anneals with the marker or anneals with a portion of
CC the polynucleotide comprising the marker, under stringent conditions. The
CC marker is also used for monitoring the progression of ovarian cancer in a
CC patient which involves detecting expression of the marker in a patient
CC sample at a first point in time, repeating the method at a subsequent
CC time and comparing the level of expression. The method is carried out
CC using an ovarian tissue sample. A composition comprising a marker,
CC polypeptide or antibody of the invention is used to treat ovarian cancer.
CC This sequence represents a human ovarian cancer DNA marker of the
CC invention. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 376 BP; 185 A; 38 C; 24 G; 128 T; 0 U; 1 Other;

Query Match 4.3%; Score 47.2; DB 5; Length 376;
Best Local Similarity 50.9%; Pred. No. 0.69;
Matches 137; Conservative 0; Mismatches 129; Indels 3; Gaps 1;

```
Qy         85 GATCTTCCTTATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTT 144
              | | ||  |||    | |     || | |  |  ||    | | |  | |||    |  ||
Db        328 GGTTTTTTTTAAGGTTTTTAAAAATTTTTTATCTTTTTTTTTTTTTTTTTTGAAAATTTTT 269

Qy        145 GAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGAC 204
                | |||||||  |  ||      ||  ||  |  |||  || || || | | ||||
Db        268 AAAAAATATTTTTTTTTTTTTTTCTTTTTTTTTTTTAATAAAATATTTTTTTTTAATTTN 209

Qy        205 ATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATT 264
              || | || || | |||    | || | |  ||    || |  ||     | ||  |  ||||
Db        208 TTTTTCTTTTTTAAATAAATCTTTAAATAATTATTAAAAGTTTTTGTATATTAAAGTATT 149
```

```
Qy        265 CATTTGTTTT---TGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCA 321
              |  | ||||   |  ||  || | || || |||| | ||||  | | | | ||| |  |
Db        148 TAAATATTTTATATATTTAATTAAGAATAGATATATATAAATAAAATAAAAAAAAAAAAA 89

Qy        322 TATATTAGCCAGTTGAAAATTAAACACAA 350
              | ||    ||    ||||  ||| | ||
Db         88 AAAATATATCAAAAAAAAAAAAAAAAAA 60
```

RESULT 71
ADL44888/c
ID ADL44888 standard; DNA; 457 BP.
XX
AC ADL44888;
XX
DT 20-MAY-2004 (first entry)
XX
DE Human ovarian cancer DNA marker #18778.
XX
KW Human; ovarian cancer; ds; tumour; cytostatic; DNA marker.
XX
OS Homo sapiens.
XX
PN WO200170979-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US009126.
XX
PR 21-MAR-2000; 2000US-0191031P.
PR 25-MAY-2000; 2000US-0207124P.
PR 15-JUN-2000; 2000US-0211940P.
PR 07-JUL-2000; 2000US-0216820P.
PR 25-JUL-2000; 2000US-0220661P.
PR 21-DEC-2000; 2000US-0257672P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Lee J, Lillie J;
XX
DR WPI; 2001-611502/70.
XX
PT Novel isolated nucleic acid molecules (markers) overexpressed in ovarian
PT cancer cells as compared to their normal non-cancerous ovarian cells are
PT used to characterize stage, grade, histological type of ovarian cancer.
XX
PS Disclosure; SEQ ID NO 18778; 106pp; English.
XX
CC The invention relates to nucleic acid markers which are overexpressed in
CC ovarian cancer cells as compared to their expression in normal (i.e. non-
CC cancerous) ovarian cells. The invention also relates to polypeptides
CC encoded by the markers, antibodies that selectively bind to the
CC polypeptides, a method of inhibiting ovarian cancer in a patient at risk
CC of developing ovarian cancer involving inhibiting expression of a gene
CC corresponding to a marker of the invention and a method of treating a
CC patient afflicted with ovarian cancer comprising providing to cells of
CC the patient an antisense oligonucleotide complementary to a marker of the
CC invention. The markers are useful for assessing if a patient is afflicted
CC with ovarian cancer, which involves comparing the level of expression of
CC a marker in a patient sample and a normal level of expression of the
CC marker in a control non-ovarian cancer sample. A difference between the
CC expression levels indicates ovarian cancer. The level of expression of a
CC marker corresponds to a secreted protein or to a transcribed
CC polynucleotide or its portion. The level of expression of the marker is
CC assessed by detecting the presence in the sample, a protein or protein
CC fragment corresponding to the marker. The presence of protein or protein
CC fragment is detected using an antibody that specifically binds with the
CC protein or protein fragment. Alternatively, the level of expression of
CC the marker is assessed by detecting the presence of a transcribed
CC polynucleotide which anneals with the marker or anneals with a portion of
CC the polynucleotide comprising the marker, under stringent conditions. The
CC marker is also used for monitoring the progression of ovarian cancer in a
CC patient which involves detecting expression of the marker in a patient
CC sample at a first point in time, repeating the method at a subsequent
CC time and comparing the level of expression. The method is carried out
CC using an ovarian tissue sample. A composition comprising a marker,
CC polypeptide or antibody of the invention is used to treat ovarian cancer.
CC This sequence represents a human ovarian cancer DNA marker of the
CC invention.
XX
SQ Sequence 457 BP; 166 A; 42 C; 57 G; 192 T; 0 U; 0 Other;

```
  Query Match             4.3%;  Score 47.2;  DB 5;  Length 457;
  Best Local Similarity   49.2%;  Pred. No. 0.73;
  Matches  124;  Conservative    0;  Mismatches  128;  Indels    0;  Gaps    0;

Qy        170 GTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTG 229
              |||  |  |||| | ||| |     | ||     |  || ||||  || |    |||
Db        328 GTTTTTTTTTTTTTTGAAAAAAAATTTTTTTTCAATTTTTTTTTATATTTATATCCAAAA 269

Qy        230 ATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAA 289
              ||||||||        ||| | |   | | |  ||  ||| |||   |||| |||  |||
Db        268 ATTTTTTTTTCCTCTTTTATATAATAAAACCCTTTTTTTTTTTTTGGTGTTTTTTCCCAAA 209

Qy        290 AGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACA 349
              |   | |||  ||   |   | | ||| |  | | |       |    ||||  ||| | |
Db        208 AAAAAAAAATTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAAAAAAAAAAAA 149

Qy        350 AAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGG 409
              |||  |    | | || |  |   ||       |   |||    ||||    |   |||
Db        148 AAAAAAAAAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAATCCGAAAAAAATAAAAAAAAA 89

Qy        410 TATACTACCTGC 421
              || | |||||||
Db         88 TAAAGTACCTGC 77
```

RESULT 72
ABL32693/c
ID ABL32693 standard; DNA; 6222 BP.
XX
AC ABL32693;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 666.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine

PT methylation.
XX
PS Claim 1; SEQ ID NO 666; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 6222 BP; 1890 A; 55 C; 1223 G; 3054 T; 0 U; 0 Other;

Query Match 4.3%; Score 47.2; DB 6; Length 6222;
Best Local Similarity 48.2%; Pred. No. 1.4;
Matches 133; Conservative 0; Mismatches 143; Indels 0; Gaps 0;

Qy 252 ACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAG 311
Db 3469 ATACACATATACCCTAAAACTTAAAATATAATAAAAAAAAAAAAAAAAATATATTAAATTA 3410

Qy 312 AGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTAC 371
Db 3409 ATAACTCCCAAATAATACAATTTTAATTCTTTAAAAAAAAATTAAAAAATCACAATATAC 3350

Qy 372 TTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCG 431
Db 3349 CTAATTTTAAAAATAAAACTAAATTATAAATATAAAAAACTTCCCATTATAACCTTTATC 3290

Qy 432 CACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTT 491
Db 3289 CCCAACTCATAAATATATATATATATTTTCTTTCCTTTTTTAACATTTAAAAAATTATTT 3230

Qy 492 TTTATGTACTATTAAAACATTTACTTTCCAAATATT 527
Db 3229 ATTTTTTAAAAATAAAATCTCTCTATTACCCAAACT 3194

RESULT 73
ABL92212/c
ID ABL92212 standard; DNA; 18060 BP.
XX
AC ABL92212;
XX
DT 01-JUL-2002 (first entry)
XX
DE Chemically treated DNA repair gene fragment#11.
XX
KW DNA repair; cytosine methylation; PMS2L1; PMS2L12; PMS2L2; PMS2L3; PMS2;
KW L4; PMS2L5; PMS2L6; MGMT; MSH2; NUDT1; TDG; INPPL1; RFC4; DDIT1L; FANCB;
KW XRCC8; ataxia telangiectasia; aging; Bloom's syndrome; Cockayne syndrome;
KW Nijmegen breakage syndrome; Werner syndrome; immunodeficiency;
KW trichthiodystrophy; Fanconi's anaemia; solid tumour; cancer; ds.
XX
OS Unidentified.
XX
PN WO200181622-A2.
XX
PD 01-NOV-2001.
XX
PF 06-APR-2001; 2001WO-EP003972.
XX
PR 06-APR-2000; 2000DE-01019058.
PR 07-APR-2000; 2000DE-01019173.
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-034446/04.
XX
PT New nucleic acid derived from genes associated with DNA repair, useful
PT for diagnosis, e.g. of ataxia telangiectasia, by determination of
PT cytosine methylation.
XX
PS Claim 1; SEQ ID NO 21; 25pp + Sequence Listing; English.
XX
CC The invention relates to nucleic acids containing a sequence of at least
CC 18 nucleotides of chemically treated DNA of genes associated with DNA
CC repair, and their complements. The invention also relates to nucleic
CC acids comprising at least 18 base pairs of the chemically pretreated DNA
CC of genes associated with DNA repair selected from PMS2L1, PMS2L12,
CC PMS2L2, PMS2L3, PMS2, L4, PMS2L5, PMS2L6, MGMT, MSH2, NUDT1, TDG, INPPL1,
CC RFC4, DDIT1L, FANCB, or XRCC8. Nucleic acids of the invention and related
CC oligomers, are useful for diagnosis of diseases associated with gene
CC repair, specifically ataxia telangiectasia, aging, Bloom's syndrome,
CC Cockayne syndrome, Nijmegen breakage syndrome or Werner syndrome,
CC immunodeficiency, trichthiodystrophy, Fanconi's anaemia, solid tumours
CC and cancer, particularly by determining status of cytosine methylation
CC and/or by detecting single-nucleotide polymorphisms. Determination of
CC individual methylation patterns may allow development of individualised
CC therapies. The sequences given in records ABL92192-ABL92335 represent
CC chemically pre-treated DNA fragments from genes associated with DNA
CC repair, and their complements. Note: The sequence data for this patent is
CC not represented in the specification, but is based on sequence
CC information supplied by the European Patent Office
XX
SQ Sequence 18060 BP; 4046 A; 384 C; 4257 G; 9373 T; 0 U; 0 Other;

Query Match 4.3%; Score 47.2; DB 6; Length 18060;
Best Local Similarity 50.4%; Pred. No. 1.8;
Matches 115; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

Qy 281 GTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAA 340
Db 2503 GTCTCAAAAACAAACAAAAAAAACAAATAATTTAAATACATACTTTCCCAAAAAAAATAT 2444

Qy 341 TTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCA 400
Db 2443 ACAAATAAAAAACAAACATATAAAAAAATACTCAACATCACTAATCATCAAAAAAATACA 2384

Qy 401 ACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTT 460
Db 2383 AATCAAAACCACAATAAAATATCCCTAAAACCCCTTAAAATAACCACTATCAAATAATAA 2324

Qy 461 ATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAA 508
Db 2323 ATATTAATAAAAATATAAAAAAACTAAAACCCTTATACACGATTAAAA 2276

RESULT 74
AAD22313/c
ID AAD22313 standard; DNA; 18060 BP.
XX
AC AAD22313;
XX
DT 12-FEB-2002 (first entry)
XX
DE Chemically treated human genomic DNA #3 associated with DNA adducts.
XX
KW DNA adduct; peptide nucleic acid; PNA; cytosine methylation;
KW gene therapy; tumour; cancer; human; ds.
XX
OS Homo sapiens.
XX
PN WO200177378-A2.
XX
PD 18-OCT-2001.
XX
PF 06-APR-2001; 2001WO-EP004015.
XX
PR 06-APR-2000; 2000DE-01019058.

PR 07-APR-2000; 2000DE-01019173.
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-010923/01.
XX
PT Novel nucleic acid comprising sequence of a segment of chemically
PT pretreated DNA of genes associated with DNA adduct, useful for diagnosis
PT and therapy of solid tumors and cancer.
XX
PS Claim 1; Page 31-36; 97pp; English.
XX
CC The invention relates to chemically modified DNA of genes associated with
CC DNA adducts, oligonucleotides and/or peptide nucleic acid (PNA) oligomers
CC for detecting cytosine methylations as well as method for ascertaining
CC genetic and/or epigenetic parameters of genes associated with DNA
CC adducts. Oligomers of the invention coupled to a solid phase is useful
CC for manufacturing an arrangement of different oligomers (array) fixed to
CC a carrier material for analysing diseases associated with the methylation
CC state of the CpG dinucleotides of chemically pretreated DNA of genes
CC associated with DNA adduct. They are also useful for ascertaining genetic
CC and/or epigenetic parameters for the diagnosis and/or therapy of existing
CC diseases or the predisposition to specific diseases by analysing cytosine
CC methylations. Sequences of the invention are useful for diagnosis and
CC therapy of solid tumours and cancers. They are also useful in gene
CC therapy. The present sequence is chemically pretreated human genomic DNA
CC associated DNA adducts
XX
SQ Sequence 18060 BP; 4046 A; 384 C; 4257 G; 9373 T; 0 U; 0 Other;

Query Match 4.3%; Score 47.2; DB 6; Length 18060;
Best Local Similarity 50.4%; Pred. No. 1.8;
Matches 115; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

Qy 281 GTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAA 340
Db 2503 GTCTCAAAAACAAACAAAAAAAAACAAATAATTTAAATACATACTTTCCCAAAAAAAATAT 2444

Qy 341 TTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCA 400
Db 2443 ACAAATAAAAAACAAACATATAAAAAAATACTCAACATCACTAATCATCAAAAAAATACA 2384

Qy 401 ACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTT 460
Db 2383 AATCAAAACCACAATAAAATATCCCTAAAACCCCTTAAAATAACCACTATCAAATAATAA 2324

Qy 461 ATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAA 508
Db 2323 ATATTAATAAAAATATAAAAAAACTAAAACCCTTATACACGATTAAAA 2276

RESULT 75
ABV08830/c
ID ABV08830 standard; cDNA; 420 BP.
XX
AC ABV08830;
XX
DT 13-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 8821.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US005171.
XX
PR 17-FEB-2000; 2000US-0183319P.
PR 16-MAR-2000; 2000US-0189862P.
PR 25-MAY-2000; 2000US-0207454P.
PR 09-JUN-2000; 2000US-0211314P.
PR 18-JUL-2000; 2000US-0219007P.
PR 13-DEC-2000; 2000US-0255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer.
XX
PS Claim 1; Page 1391; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for: (a) assessing whether
CC a patient is afflicted with prostate cancer; (b) monitoring the
CC progression of prostate cancer in a patient; (c) assessing the efficacy
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC determining whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or indolence of prostate cancer in a patient
CC ; (I) is also useful as a pharmacodyanamic or pharmacogenomic marker
XX
SQ Sequence 420 BP; 100 A; 27 C; 16 G; 216 T; 0 U; 61 Other;

Query Match 4.3%; Score 47; DB 5; Length 420;
Best Local Similarity 39.3%; Pred. No. 0.79;
Matches 149; Conservative 0; Mismatches 230; Indels 0; Gaps 0;

Qy 35 TCTATGTTCTTATAAGCAAATGTTACAATGAGATATGAGTTTTAAAGCCAGATCTTCCTT 94
Db 408 TATCTTNTNTNTTTTTCTNTTATAAAAANAAAATCATCTNTTTNNANGAAANTNATCANT 349

Qy 95 ATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATATT 154
Db 348 TTTTTNTNTTNTTNTGNAGGGNNTTTTTTTNTTTTTTATNGAAAGGNTGGNNANATATT 289

Qy 155 GATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGT 214
Db 288 TTNNANNANTNANNGGTTNTTTTNTTTTTTTAAAAAAAAAATAAANTNNNGAGNTTTTTN 229

Qy 215 TAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTT 274
Db 228 ANAAAAAANGNNTTTTNTTTTNTTNAATANATNATTTTNNNNNNGCGTAAAAAAAANAAT 169

Qy 275 TGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGT 334
Db 168 ANAAAAANTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 109

Qy 335 TGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAA 394
Db 108 AAAAAANAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAAAA 49

Qy 395 AAGTCAACCTAAAGGTATA 413
Db 48 AAAAAAAAAAAAAAAAANA 30

RESULT 76
ACN79585
ID ACN79585 standard; DNA; 804 BP.

XX
AC ACN79585;
XX
DT 02-DEC-2004 (first entry)
XX
DE Breast cancer related marker, seq id 735.
XX
KW Cancer; breast; tumour; cytostatic; marker; detection; therapy; ds.
XX
OS Homo sapiens.
XX
PN US2003099974-A1.
XX
PD 29-MAY-2003.
XX
PF 18-JUL-2002; 2002US-00198846.
XX
PR 18-JUL-2001; 2001US-0306220P.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX
DR WPI; 2003-787014/74.
XX
PT Novel isolated polypeptide associated with breast cancer, useful for
PT detecting presence of polypeptide in sample, as a marker for breast
PT cancer.
XX
PS Disclosure; SEQ ID NO 735; 36pp; English.
XX
CC The invention relates to an isolated polypeptide (I) associated with
CC breast cancer which is encoded by a nucleic acid molecule comprising a
CC nucleotide sequence (S1). Further disclosed is an antibody that binds to
CC the polypeptide of the invention. The activity of the polypeptide of the
CC invention may be described as cytostatic. The antibody is useful for
CC detecting the presence of (I) in a sample. Nucleic acid molecules of the
CC invention are useful in the detection of breast tumours. (I) is useful as
CC a marker for breast cancer and in breast cancer therapy. Sequences given
CC in records ACN78851-ACN92934 represent nucleic acid markers associated
CC with breast cancer. Note: The sequence listing does not form part of the
CC specification but may be obtained in electronic format from the USPTO web
CC site at seqdata.uspto.gov/sequence.html?DocID=20030099974
XX
SQ Sequence 804 BP; 255 A; 133 C; 145 G; 260 T; 0 U; 11 Other;

Query Match 4.3%; Score 47; DB 11; Length 804;
Best Local Similarity 49.8%; Pred. No. 0.93;
Matches 119; Conservative 0; Mismatches 120; Indels 0; Gaps 0;

```
Qy        250 AGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTC 309
              | |  ||   | |  | | || || |   | ||   | |    | |||    |||  ||
Db        424 ACAACTATCTCAAGCCCTGTGATTATAAATGGTGTTATAGGCTTGTAATCCTATGGGTTT 483

Qy        310 AGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATT 369
              |||  ||  |  | | ||    | ||||||  |||||||  | || | || |  |  ||
Db        484 AGACTAAAGTATTTTCTTCATAAATTGAAAGAGAAACACACTAGGAATTCAGAAGATTTT 543

Qy        370 ACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTAT 429
              | ||||  |||     |||  | ||  |  |   ||||||    ||  ||   |  || |
Db        544 AATTAAAGTTGTTCAAAAAATTTAATGGGAATTTTAAAGGCTGGCTCACTAAGTAGTTTT 603

Qy        430 CGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAA 488
              | || |  ||     ||  |  ||  || ||| || || | | |   ||  |  ||
Db        604 GATAATGAATTTAATTTATAGAAAGAAAGTTCTTTGGGCAAAATTTTGGTAATGCCAAA 662
```

RESULT 77
ABL32133/c
ID ABL32133 standard; DNA; 5421 BP.
XX
AC ABL32133;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 106.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 106; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 5421 BP; 1687 A; 57 C; 1123 G; 2544 T; 0 U; 10 Other;

Query Match 4.3%; Score 47; DB 6; Length 5421;
Best Local Similarity 45.9%; Pred. No. 1.5;
Matches 158; Conservative 0; Mismatches 186; Indels 0; Gaps 0;

```
Qy        209 TTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATT 268
              ||   |||| |   |  ||     ||||| |   |      |  | | | |  ||| | ||
Db       4113 TTCCTTTAAAACTTTTACTACTACTTTTTCATTAAAAAATCCAATAAAAACTTATCCTTT 4054

Qy        269 TGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTA 328
              |      ||  ||||  ||| |||||   ||    |  | |||  |   |  |||
Db       4053 CATACGACCTTCTTTCATCAAACGTATATCAAACCTAAAAAAAAAAAAACAATTTTAATTC 3994

Qy        329 GCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAA 388
              | |  ||||  |    |||| |    ||||| || |  |||  ||  | |
Db       3993 AAAAAATATCAATTTCAATATAAATAAAAATATATTAAATAATCTAACATTTTACTATTT 3934

Qy        389 GGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATT 448
              | |||   |||   |  | || ||    | |   |    || |      | ||  |
Db       3933 CAAATAAAAATAACTACACTATTTAATAAACACGTCAATTATTCAATTTTCTAAAAATAT 3874

Qy        449 ACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAA 508
              |   || ||| ||   |  || ||||  | |||  |  | | |||||  ||  ||  |
Db       3873 ATTTCATATTCTACACTAAAATTAATTCTACAAATAAATAATATTTATTTAAAATATACT 3814

Qy        509 CATTTACTTTCCAAATATTCTGTCATTCAGGAGTATGGAAGTAT 552
                  |  |||  | |||  |   | |   | | |  || |||
```

```
Db       3813 NTTAACTATTCTCACTATCATAAAAATATATAATTTAAAAATAT 3770

RESULT 78
ABL32487
ID   ABL32487 standard; DNA; 17183 BP.
XX
AC   ABL32487;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 460.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP007537.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful for
PT   diagnosis and treatment of diseases associated with abnormal cytosine
PT   methylation.
XX
PS   Claim 1; SEQ ID NO 460; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention
XX
SQ   Sequence 17183 BP; 5342 A; 149 C; 3418 G; 8274 T; 0 U; 0 Other;

  Query Match             4.3%;  Score 47;  DB 6;  Length 17183;
  Best Local Similarity   45.2%;  Pred. No. 2;
  Matches  213;  Conservative    0;  Mismatches  255;  Indels    3;  Gaps    1;

Qy        123 AGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTG 182
              ||| |   | || |   |   | ||| || |||   |     | |   |  ||  |||||
Db       5486 AGTATAAGAAATTAAAGTATATAAGAGATTTTGTAAAAGTTTAGTATATATATGATTTTG 5545

Qy        183 TAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAG 242
                | |  || | |    |   | ||  || | ||| |              | |||| |
Db       5546 AAAATTATTTTAATGAATTTATATAATTATTTTAGTTATTAAATCGTTAATGTTTAGAGA 5605

Qy        243 ACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAA 302
                  | | |  |  |    ||  | |  ||||  |||||| |   |    |  || |
Db       5606 GGAGAAAATATGTTGATTATTTTTTATTATTTTATTTTGTTAATTTAGTAGAGTAATATG 5665

Qy        303 TGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATA 362
              |   |  ||||||       |   |     | ||  ||||| |   ||| |  |   ||
Db       5666 TTATTAAAGAGAATTTAGTAAATATTTTTTGATGGGAATTAGA---TAAAGGGATAGTTA 5722

Qy        363 TTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCT 422
              ||  | ||   | | |||   |  || |||| ||    |      |   ||   |
Db       5723 TTTTAATATAGATTTTTGTGATTTAAAGTAATAATATTATTATTAATAATGGAATAGAAG 5782

Qy        423 TTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAA 482
              || || | | |        ||||||||    ||   |   ||| |  |||   |  ||||
Db       5783 TTATTTTAGGAGATAGGTAAGAAATTATAGTAATGATATATTTAGGGATATAGTTGGAAA 5842

Qy        483 ACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGTCATTCAGGAGT 542
              |  || ||||||||||||  |  |  | || | |||  |  | ||   |   | || ||
Db       5843 AGTTATTTTTTTATGTATGAGAACGAGTTTAAGTTTTAACGTTTTAAATTGATAAGTAGA 5902

Qy        543 ATGGAAGTATCGATGGCTTCTTTAAAATGAAGCAGGAGGGTCTGGCAGAGA 593
              |   ||| |   |    || ||| | || |    | ||  |  |  | | |
Db       5903 AAATAAGAAGTTAATTTTTTTTTTAGATTAGATGGTAGTTTAGGTAAAAAA 5953

RESULT 79
AAS46788/c
ID   AAS46788 standard; DNA; 61020 BP.
XX
AC   AAS46788;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Tumour suppressor gene derived chemically modified sequence #514.
XX
KW   Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer;
KW   tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW   cytosine methylation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200168912-A2.
XX
PD   20-SEP-2001.
XX
PF   15-MAR-2001; 2001WO-EP002955.
XX
PR   15-MAR-2000; 2000DE-01013847.
PR   06-APR-2000; 2000DE-01019058.
PR   07-APR-2000; 2000DE-01019173.
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2001-602752/68.
XX
PT   Fragments of chemically modified genes associated with tumor suppressor
PT   genes and oncogenes, useful in designing primers and probes for analyzing
PT   diseases associated with cytosine methylation state e.g. cancer.
XX
PS   Claim 1; SEQ ID NO 514; 27pp; English.
XX
CC   The invention relates to a nucleic acid comprising a sequence of 18
CC   bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC   bisulphite, of genes associated with tumour suppression and oncogenes
CC   having a sequence taken from 536 (actually 533 since numbers 408, 458 and
CC   500 are missing from the sequence listing) sequences (Ss) and sequences
CC   complementary to (Ss). The nucleic acid may be a peptide nucleic acid-
CC   oligomer (PNA) of at least 9 nucleotides and may form part of a set of
CC   probes for detecting the cytosine methylation state and/or single
CC   nucleotide polymorphisms and also to be used in an array for analysing
CC   diseases associated with CpG dinucleotides e.g. cancers and tumours. The
CC   probes can also be used in a method for ascertaining genetic and/or
CC   epigenetic parameters for the diagnosis and/or therapy of existing
CC   diseases or the predisposition to specific diseases, by analysing
```

```
CC   cytosine methylations. The parameters may be compared to another set of
CC   genetic and/or epigenetic parameters, the differences serving as basis
CC   for diagnosis and/or prognosis events which are disadvantageous to
CC   patients. The present sequence is one of the 533 genomic sequences
CC   derived from tumour suppressor genes and oncogenes. Sequences with even
CC   numbered Seq ID numbers are the complementary sequence of the
CC   corresponding odd numbered sequence (e.g. ID 2 and ID1, ID 536 and ID
CC   535, except for those whose partner sequence is missing). Note: The
CC   sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 61020 BP; 17884 A; 634 C; 12537 G; 29965 T; 0 U; 0 Other;

  Query Match             4.3%;  Score 47;  DB 4;  Length 61020;
  Best Local Similarity   43.2%;  Pred. No. 2.7;
  Matches  271;  Conservative    0;  Mismatches  355;  Indels    1;  Gaps    1;

Qy        136 AGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCAT 195
              | |||   | | ||  ||  |||  || ||||   | |||  |  |  | |     |  |
Db      34581 ACTTTAAATAATAATAATACATCGATATAAATTCATAAATTATAATAAATATACATTTCT 34522

Qy        196 CTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACAT 255
              ||||  ||||  |  |  ||| | |  |  |  || |   ||||  |   || | | |
Db      34521 AATAAA-ACATATTAATAATAAAAAATACTATACATATAACTAAATAAATATATATAAAA 34463

Qy        256 ACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAA 315
              |     |  ||| ||||     |     |||||||  |     | ||||   || | | ||
Db      34462 TAACTCTATATTTTATTTTACTATAAACCAAAAAATATCCTTAAAAATATCTTAAAAAAA 34403

Qy        316 AGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAA 375
              | |  | || |     | || |||||  | |   |||  ||    | ||||  |      ||
Db      34402 ATAAAAAATTTATTAAATTTAAAAATAATAATAAAATCGAAAAAACATTATAAACAAAAA 34343

Qy        376 TCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACT 435
                | |||   || ||     |     ||   | | ||     |     |       |
Db      34342 AAACAACTTAAAAAAAAATAACCATAAAATAAAACAAATTAAAACTTACAAAAAATATTT 34283

Qy        436 GCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTA 495
              |||| | |     |  |||||    |    |    || || |  | |||  ||||  | | |
Db      34282 ACAAAAAAACTAAAAAACAAAAACAAAAATATTAAAAAACGAATAAACAATAAAATATAA 34223

Qy        496 TGTACTATTAAAACATTTACTTTCCAAATATTCTGTCATTCAGGAGTATGGAAGTATCGA 555
              |   |  |||  |  ||    | ||  ||| | |    |     |       ||  | | |
Db      34222 AAAATAACAAAATAAAATAACAACAAATAATTATTTACATTTAAAAATAAAAAAAAACCA 34163

Qy        556 TGGCTTCTTTAAAATGAAGCAGGAGGGTCTGGCAGAGAGTATCTATGAAATAAGTTCCTC 615
              |   |       |||   |  |       |       | | ||   || | ||    |  |
Db      34162 TTAATCAAAATAAAATCAAAATTTAATTAATTTCAATAATAATAATAATATCCAATAATA 34103

Qy        616 TGACCTTCACGCTTAATTTTCTGAATGGAGTGGAGCAAATTACTTCAAGCTTCACTTAAC 675
              |     |       |||    ||   |     ||||  ||   |  | | | |||
Db      34102 ATAATAATAACTACCTTTTAACCAAACAATATATACAAAACACGATATTCATAATTTATA 34043

Qy        676 TTGCATATGAAATGAACCGTACAAAAATACAAGAGTGTCAGGAGAAAGTTATGCTCTGGT 735
              ||    ||   || ||    |||||| |  | | |   |  | |||  |||  | |
Db      34042 TTCATCATTTCATTAAAAACACAAAAAAAATAAAATCCTATAAAAAAAATATAATTTAAC 33983

Qy        736 AAATATTTTGCAAAACAGATAAAAGAT 762
              |||||       || |   |||| ||
Db      33982 TAATATAAAAACTAAAATTAAAAAAAT 33956


RESULT 80
ABA92787_5/c
Continuation (6 of 7) of ABA92787 from base 500001 (Buchnera sp. genomic DNA SEQ ID NO:1
WP Sequence split into 7 fragments  LOCUS ABA92787 Accession Aba92787
WP     Fragment Name      Begin     End
WP     ABA92787_0             1     110000
WP     ABA92787_1        100001     210000
WP     ABA92787_2        200001     310000
WP     ABA92787_3        300001     410000
WP     ABA92787_4        400001     510000
WP     ABA92787_5        500001     610000
WP     ABA92787_6        600001     640681

  Query Match             4.3%;  Score 47;  DB 6;  Length 110000;
  Best Local Similarity   46.7%;  Pred. No. 3.1;
  Matches  149;  Conservative    0;  Mismatches  170;  Indels    0;  Gaps    0;

Qy        190 TCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTAC 249
              |||  | |||      |   | | |  |||  |     | ||||||||||||     ||
Db      81884 TCTTTTTTTATTATTAGACTATAATGTATACAAATAATAGATTTTTTTAAATATTATTTT 81825

Qy        250 AGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTC 309
              |  || | |    |   ||| | | |  | | || | |         |  | ||    ||
Db      81824 AATCAGAAATAGCTATTTTTTTATATAATATTTTAATATTCATCTTTTAAAATATTATAT 81765

Qy        310 AGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATT 369
              | |    | || ||||    |   ||| || |     |||| | |  |  |   ||
Db      81764 TTATATTTTTGATTTATTTTAGAAAGGAATATAATTTTAAAAATTATTAAACTTATAATA 81705

Qy        370 ACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTAT 429
              | | ||  |  |   ||   ||||   |  | | |||  ||||     |  |||    |
Db      81704 ATGTTATTATTTAAAAAATTATAAATGTTTTAACAAAATTTATAGGTAATTTTTTACATT 81645

Qy        430 CGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAAT 489
              ||  | |||     | |  || ||| |||||||  ||||   | || ||| || ||
Db      81644 GAAACATCGAATTATTTTCAATACGAATATTATTTTTCAAATGTACGATGTAAAAATTAT 81585

Qy        490 TTTTTATGTACTATTAAAA 508
              | | |   || |||
Db      81584 AAAATTTTTTTATTTTAAA 81566


RESULT 81
ABL33761/c
ID   ABL33761 standard; DNA; 7851 BP.
XX
AC   ABL33761;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 1734.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP007537.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful for
PT   diagnosis and treatment of diseases associated with abnormal cytosine
PT   methylation.
XX
PS   Claim 1; SEQ ID NO 1734; 32pp + Sequence Listing; German.
```

```
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention
XX
SQ   Sequence 7851 BP; 1811 A; 284 C; 2095 G; 3661 T; 0 U; 0 Other;

  Query Match             4.3%;  Score 46.8;  DB 6;  Length 7851;
  Best Local Similarity   44.8%;  Pred. No. 1.8;
  Matches  180;  Conservative    0;  Mismatches  222;  Indels    0;  Gaps    0;

Qy        131 ATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGAT 190
              ||||   | |  || | | ||||  ||   |   |   | || |   | |       |
Db       7161 ATATATATATTTTTTAAATATATAAATACATCATATATATTTTAATATATAAATATATTA 7102

Qy        191 CTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACA 250
              || ||| | |  ||  | | | |||||| |   |  | | | ||  || |  | |
Db       7101 TATATATTATATATTTTAATATATAAATATATTATATATAATATATTTTAATATATAAAT 7042

Qy        251 GACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCA 310
              |||  |  |  |  |||  | |     | ||  | | |   || |   | |  | |
Db       7041 ATTATATATATTATATTTTAATATATAAATATTATATATATTAATTATATATTTTAATAT 6982

Qy        311 GAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTA 370
              | | ||  |||||||   | |   |||||| | |  | || |    ||| || ||||
Db       6981 ATAATATATATTATATTATATAATAATAAATTATATATTATATTATATAATAATAAATTA 6922

Qy        371 CTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATC 430
              || | |   |   ||   || ||  | |     ||  |||| ||     | | ||||
Db       6921 TATATTATACTATATAATAATATAATATAATATATAATATATATTATAATATATATTATA 6862

Qy        431 GCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATT 490
              |    | ||  | |    |  |    |    |   || | | | |     ||
Db       6861 TTATATTATATTATATTATATATTATAATATAAATTATATATATTTTATATATATATTTT 6802

Qy        491 TTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGTC 532
              | |||    ||||  || || ||      ||| | |   ||
Db       6801 ATATATTATATATTTTAATATATAAAACAAAAAAACTTCTTC 6760


RESULT 82
ABX46069
ID   ABX46069 standard; cDNA; 419 BP.
XX
AC   ABX46069;
XX
DT   21-FEB-2003  (first entry)
XX
DE   Bovine EST associated with lactation/muscle/fat deposition #11234.
XX
KW   Bovine; ss; EST; expressed sequence tag; lactation; LMFD;
KW   muscle deposition; fat deposition; genome mapping; gene identification;
KW   gene analysis; cattle breeding.
XX
OS   Bos Taurus.
XX
PN   US2002137139-A1.
XX
PD   26-SEP-2002.
XX
PF   24-SEP-2001; 2001US-00960352.
XX
PR   12-JAN-1999;   99US-0115707P.
PR   11-JAN-2000; 2000US-00480902.
XX
PA   (BYAT/) BYATT J C.
PA   (MATH/) MATHIALAGAN N.
PA   (TAON/) TAO N.
PA   (WARR/) WARREN W C.
XX
PI   Byatt JC,  Mathialagan N,  Tao N,  Warren WC;
XX
DR   WPI; 2003-110599/10.
XX
PT   New nucleic acid associated with lactation, and muscle and fat
PT   deposition, useful for genome mapping, gene identification and analysis,
PT   cattle breeding, or for genetically improving cattle.
XX
PS   Claim 2; SEQ ID NO 11234; 245pp; English.
XX
CC   The invention relates to a purified nucleic acid molecule associated with
CC   lactation or muscle and fat deposition (designated LMFD), derived from
CC   cattle, and the LMFD nucleic acid can specifically hybridise to a second
CC   nucleic acid molecule comprising any of 15112 nucleotide sequences,
CC   appearing as ABX34836-ABX49947, or complements of them. Also included are
CC   ; (1) a transformed cell having a nucleic acid comprising an LMFD nucleic
CC   acid linked to a promoter and a 3' non- translated sequence that
CC   functions in the cell to cause termination of transcription and addition
CC   of polyadenylated ribonucleotides to a 3' end of the mRNA molecule; and
CC   (2) determining a level or pattern of a molecule in a bovine cell or
CC   tissue comprising: (a) incubating a marker nucleic acid (comprising any
CC   of the 15112 nucleic acid sequences or its complement or fragment) with a
CC   complementary nucleic acid molecule obtained from the bovine cell or
CC   tissue, where hybridisation between the marker nucleic acid and the
CC   complementary nucleic acid permits the detection of the molecule; and (b)
CC   detecting the level or pattern of the complementary nucleic acid, where
CC   the detection of the complementary nucleic acid is predictive of the
CC   level or pattern of the molecule. The LMFD nucleic acid is used for
CC   determining a level or pattern of a molecule in a bovine cell or tissue.
CC   It is useful for genome mapping, gene identification and analysis, cattle
CC   breeding, preparation of constructs for use in cattle gene expression, or
CC   for genetically improving cattle. The present sequence is one of the
CC   15112 bovine LMFD EST (expressed sequence tag) nucleic acids. Note: The
CC   present sequence was not shown in the specification but was obtained in
CC   electronic format from the USPTO web site:
CC   seqdata.uspto.gov/sequence.html?DocID=20020137139
XX
SQ   Sequence 419 BP; 132 A; 8 C; 13 G; 266 T; 0 U; 0 Other;

  Query Match             4.3%;  Score 46.6;  DB 8;  Length 419;
  Best Local Similarity   49.2%;  Pred. No. 0.97;
  Matches  155;  Conservative    0;  Mismatches  154;  Indels    6;  Gaps    1;

Qy        206 TTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTC 265
              || |||| ||  | |  |   ||  |||||||      ||  |   |     | ||
Db         96 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTATTTTTTTTTTTTT 155

Qy        266 ATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATA 325
              |||  |||||  |||  ||   ||||   ||||  |||    || | | |   |  || ||
Db        156 TTTTTTTTTTTTTTTTTTTTAAAAATAATAATAAAAAAATTTTAAAAAGTTAAAATTTA 215

Qy        326 TTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTC 385
              | |     || || || ||| | ||  |    |  ||  ||  ||  | ||| |
Db        216 TAATTTTATTAAATATAAAAAAAAATTTTTTTATATTATATTTTTTTAAATTTAAATTAA 275

Qy        386 AAAGGTAAAAAGTC------AACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAA 439
              |    |||||  |       ||   ||  || |  |  |  |  |   | |     | ||
Db        276 ATTTTAAAATTTTTTAAAAAATTATAATTTAAATAAAATAATAAATAAAAAATTTTAAA 335

Qy        440 ATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTA 499
              | | |  ||   | |||| || || |  |||||||  |     | | |||||||| | | |
Db        336 AAAAATTTTTTAAAAAATATTTTTATAAAAATAATTTTTATATAAATAATTTTATTTTTT 395

Qy        500 CTATTAAAACATTTA 514
              ||   | | |||||
Db        396 TTAAATATATATTTA 410


RESULT 83
ABL32936/c
ID   ABL32936 standard; DNA; 6535 BP.
```

```
XX
AC   ABL32936;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 909.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP007537.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful for
PT   diagnosis and treatment of diseases associated with abnormal cytosine
PT   methylation.
XX
PS   Claim 1; SEQ ID NO 909; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention
XX
SQ   Sequence 6535 BP; 1651 A; 58 C; 1360 G; 3466 T; 0 U; 0 Other;

  Query Match             4.3%;  Score 46.6;  DB 6;  Length 6535;
  Best Local Similarity   47.2%;  Pred. No. 1.9;
  Matches  175;  Conservative    0;  Mismatches  194;  Indels    2;  Gaps    1;

Qy        144 TGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGA 203
              | | ||  |||   ||  |  | | | |||||  |       |  | |   ||| || |
Db       2362 TAATAATAATTACACAACATTACTAATATAATAAATACTACCGAATTATACACTTCAAAA 2303

Qy        204 CATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTAT 263
               | |       || ||  |   ||     | ||  || ||   | |  || | |   |
Db       2302 TAATTAAAATCGAAAATTTTATATTACAAATATTTTAACACAAAAAATTCAAAAAATAAA 2243

Qy        264 TCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATA 323
              | | |    ||   ||| || ||||||   | ||| |||   |   | | ||| |  | |
Db       2242 TAAATAAACTTATTTTTATTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2183

Qy        324 TATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAG 383
               |  |   |    ||||  ||| | ||||  |    | |  | | |    ||| |   |
Db       2182 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACTCAACAATATATTAT 2123

Qy        384 TCAAAGGTAAAAAGTCAACCT--AAAGGTATACTACCTGCTTTCTTATCGCACTGCAAAT 441
              |||||  |||| |   ||      |||  |||| |       |||||||   | |  |||
Db       2122 TCAAACATAAATATCAAAAAATCAAAACTATATTTTTAAATTTCTTATTAAAATAAAAAA 2063

Qy        442 AGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACT 501
              ||      | ||  ||   | |  ||| ||  | | |||| |  ||      |   ||
Db       2062 CCCAAAATATAAAATATTACATCTCAAAAAAAAATAAAAAAAAAACATAAAAAAAAAACA 2003

Qy        502 ATTAAAACATT 512
               |||||  | |
Db       2002 CTTAAATAAAT 1992


RESULT 84
ABL33983/c
ID   ABL33983 standard; DNA; 8592 BP.
XX
AC   ABL33983;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 1956.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP007537.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful for
PT   diagnosis and treatment of diseases associated with abnormal cytosine
PT   methylation.
XX
PS   Claim 1; SEQ ID NO 1956; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention
XX
SQ   Sequence 8592 BP; 2684 A; 132 C; 1778 G; 3998 T; 0 U; 0 Other;

  Query Match             4.3%;  Score 46.6;  DB 6;  Length 8592;
  Best Local Similarity   46.9%;  Pred. No. 2;
  Matches  177;  Conservative    0;  Mismatches  199;  Indels    1;  Gaps    1;

Qy        279 TTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAA 338
              || | || | ||  | ||||  | |       |   |  || | ||  | |  || || |
Db        542 TTTTACATATAACTTTTAAAATATTTATAAAAATAATCGAACCTAACTCCTTCATTTTAC 483

Qy        339 AATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGT 398
              || |||||| | ||| | | || |  | || |   |||  |  | |||||   | | |
```

```
Db        482 AAATAAACATATAATAAATCCA-AAAAAATAAAAAAATTATTTACTCAAAATCACACAAC 424

Qy        399 CAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATT 458

Db        423 TAAACTAAAACTAAAAACAAATCTCTTAATTCCTAATAATATATTCTTTTCACACTATTT 364

Qy        459 TTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTT 518

Db        363 AAACTTTCTAACCTTTACAAACAAATAAAATACGTTACAACTTTTCTAAAAATACAAATT 304

Qy        519 CCAAATATTCTGTCATTCAGGAGTATGGAAGTATCGATGGCTTCTTTAAAATGAAGCAGG 578

Db        303 TTTAAAATTTTTAAAAAAAATATTATAAAACTCAAATAATTTACCTTAAAATCATAATTC 244

Qy        579 AGGGTCTGGCAGAGAGTATCTATGAAATAAGTTCCTCTGACCTTCACGCTTAATTTTCTG 638

Db        243 CCTTTCTCTCACCCACTATCCCTATAATTTTTAATATATAAATATAACTTTAATTTTTTT 184

Qy        639 AATGGAGTGGAGCAAAT 655

Db        183 TATTTAATAAAAAAAAT 167


RESULT 85
ADB37663/c
ID   ADB37663 standard; DNA; 29993 BP.
XX
AC   ADB37663;
XX
DT   04-DEC-2003  (first entry)
XX
DE   Human chemically pretreated EYA4 gene SEQ ID NO:5.
XX
KW   colon cell proliferative disorder; EYA4; cytostatic; gene therapy; human;
KW   gene; ds; chromosome 6.
XX
OS   Synthetic.
OS   Homo sapiens.
XX
PN   WO2003072812-A2.
XX
PD   04-SEP-2003.
XX
PF   13-FEB-2003; 2003WO-EP001457.
XX
PR   27-FEB-2002; 2002EP-00004551.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Adorjan P,  Burger M,  Maier S,  Lesche R,  Cottrell S,  De Vos T;
XX
DR   WPI; 2003-731618/69.
XX
PT   Diagnosing a colon cell proliferative disorder in a subject comprises
PT   obtaining one or more samples from colon tissue or serum the subject, and
PT   detecting a decrease in the amount or expression of a polypeptide
PT   expressed from the EYA4 gene.
XX
PS   Claim 22; Page 91-99; 101pp; English.
XX
CC   The present invention describes a method for diagnosing a colon cell
CC   proliferative disorder in a subject. The method comprises obtaining one
CC   or more samples from colon tissue or serum or both of the subjects, and
CC   detecting a decrease in the amount or expression of a polypeptide
CC   expressed from the EYA4 gene. Also described: (1) repressing
CC   transformation in a colon cell by contacting the cell with an EYA4
CC   polypeptide to inhibit a transformed phenotype; (2) preventing or
CC   treating a colon cell proliferative disorder in a subject by
CC   administering to the subject a compound that agonises EYA4; (3) a nucleic
CC   acid comprising a sequence of at least 18 bases in length of a segment of
CC   the chemically pretreated genomic DNA or its complement; (4) an oligomer,
CC   in particular an oligonucleotide or peptide nucleic acid (PNA)-oligomer
CC   comprising in each case of at least one base sequence having a length of
CC   at least 9 nucleotides which is complementary to, or hybridises under
CC   moderately stringent or stringent conditions to a pretreated genomic DNA;
CC   (5) a set of oligomers comprising at least two of the oligomer described
CC   above; (6) manufacturing an arrangement of different oligomers (array)
CC   fixed to a carrier material; (7) an arrangement of different oligomers
CC   (array) obtainable in (6); (8) an array of different oligonucleotide-
CC   and/or PNA-oligomer sequences, which are arranged on a plane solid phase
CC   in the form of a rectangular or hexagonal lattice; (9) a composition of
CC   matter comprising the nucleic acid and a buffer comprising 1-5 mM
CC   magnesium chloride, 100-500 micromole dNTP, 0.5-5 units of taq
CC   polymerase, and the oligomer; (10) detecting, differentiating or
CC   distinguishing between colon cell proliferative disorders; (11) detecting
CC   a colon cell proliferative disorder; and (12) a kit useful for the method
CC   in (10) comprising a bisulfite reagent, and at least one of the nucleic
CC   acid molecule or peptide described above, or their complements. EYA4 has
CC   cytostatic activity, and can be used in gene therapy. The methods, vector
CC   and polypeptide from the present invention are useful for treating colon
CC   cell proliferative disorders. The EYA4 gene, the polypeptide expressed
CC   from the EYA4 gene and kit are useful for detecting, differentiating or
CC   distinguishing between colon cell proliferative disorders. The oligomers
CC   are useful for detecting the cytosine methylation state and/or single
CC   nucleotide polymorphisms (SNPs) within nucleic acid sequences. The array
CC   is useful for analysing diseases associated with the methylation state of
CC   the CpG dinucleotides. The present sequence represents a chemically
CC   pretreated human EYA4 gene, which is given in the exemplification of the
CC   present invention. Human EYA4 is mapped to chromosome 6, more
CC   specifically to 6q22.3.
XX
SQ   Sequence 29993 BP; 9861 A; 0 C; 5570 G; 14562 T; 0 U; 0 Other;

  Query Match             4.3%;  Score 46.6;  DB 10;  Length 29993;
  Best Local Similarity   46.6%;  Pred. No. 2.8;
  Matches  185;  Conservative    0;  Mismatches  209;  Indels    3;  Gaps    1;

Qy        134 TGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTC 193

Db      13597 TAATTTTCTAATAAAAAAACAAATATATCCAACATACATTAATATTTTCAATAAACTAAC 13538

Qy        194 ATCTTAAAGACA---TTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACA 250

Db      13537 ATTTAAAAAACAAATTTCTCATTAAAATTTATAATCTTCATAATTCTAATTTTTCTCCAAA 13478

Qy        251 GACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCA 310

Db      13477 AACAAACCTTAATCTCTTTAATTACCTTCTACTAAAATCAATTTCATATTATTTTAAAAA 13418

Qy        311 GAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTA 370

Db      13417 ATTTTTAACCCTCAAAATTATAAATTTCATTTAAAAAAAAACATATTCTATATTTATTTA 13358

Qy        371 CTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATC 430

Db      13357 TATCAATATTTCAAAATATAAAAACCAAAAATCCCAAATATATCTATTAAAAATATACAT 13298

Qy        431 GCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATT 490

Db      13297 AACCTTTTACTAAAAAAAATTACATCTTTTTTTTCAAACACAATTAAAAAAATAAAAAAA 13238

Qy        491 TTTTATGTACTATTAAAACATTTACTTTCCAAATATT 527

Db      13237 TTTTATTTAATTTTTAATAATTTATTTTTAAACATTT 13201


RESULT 86
AAL09384
ID   AAL09384 standard; cDNA; 639 BP.
XX
AC   AAL09384;
XX
DT   07-DEC-2001  (first entry)
XX
DE   Human breast cancer expressed polynucleotide 1841.
XX
```

KW Human; breast cancer; cell marker; cytostatic; ss.
XX
OS Homo sapiens.
XX
PN WO200151628-A2.
XX
PD 19-JUL-2001.
XX
PF 10-JAN-2001; 2001WO-US000798.
XX
PR 14-JAN-2000; 2000US-0176077P.
PR 14-MAR-2000; 2000US-0189167P.
PR 24-MAR-2000; 2000US-0192099P.
PR 29-MAR-2000; 2000US-0193480P.
PR 15-MAY-2000; 2000US-0205230P.
PR 09-JUN-2000; 2000US-0211315P.
PR 25-JUL-2000; 2000US-0220534P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX
DR WPI; 2001-451856/48.
XX
PT New peptide useful as a marker for the diagnosis of breast cancer.
XX
PS Claim 1; Page 363-364; 3695pp; English.
XX
CC The invention relates to human breast cancer expressed polynucleotides
CC (AAL07544-AAL26789) and methods of assessing whether a patient is
CC afflicted with breast cancer by examining the correlation between the
CC expression of certain markers and the cancerous state of breast cells.
CC The polynucleotides and encoded polypeptides are potential markers for
CC detecting, diagnosing, monitoring, characterising treating and
CC potentially preventing breast cancer. The polynucleotides and encoded
CC polypeptides are also useful for isolating compounds with cytostatic
CC activity
XX
SQ Sequence 639 BP; 211 A; 96 C; 126 G; 206 T; 0 U; 0 Other;

Query Match 4.3%; Score 46.4; DB 4; Length 639;
Best Local Similarity 50.0%; Pred. No. 1.2;
Matches 116; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

Qy 250 AGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTC 309
Db 404 ACAACTATCTCAAGCCCTGTGATTATAAATGGTGTTATAGGCTTGTAATCCTATGGGTTT 463

Qy 310 AGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATT 369
Db 464 AGACTAAAGTATTTTCTTCATAAATTGAAAGAGAAACACACTAGGAATTCAGAAGATTTT 523

Qy 370 ACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTAT 429
Db 524 AATTAAAGTTGTTCAAAAAATTTAATGGGAATTTTAAAGGCTGGCTCACTAAGTAGTTTT 583

Qy 430 CGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAA 481
Db 584 GATAATGAATTTAATTTATAGAAAGAAAGTTCTTTGGGCAAAATTTTGGTAA 635

RESULT 87
ABL32978/c
ID ABL32978 standard; DNA; 6963 BP.
XX
AC ABL32978;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 951.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 951; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 6963 BP; 2009 A; 47 C; 1548 G; 3359 T; 0 U; 0 Other;

Query Match 4.2%; Score 46.2; DB 6; Length 6963;
Best Local Similarity 48.9%; Pred. No. 2.4;
Matches 161; Conservative 0; Mismatches 158; Indels 10; Gaps 1;

Qy 206 TTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTC 265
Db 4889 TTACTTTCTTTATATACTCTCTCCATATATTTAAAAAATTTTTCATTTTTTTTTCATCATT 4830

Qy 266 ATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATA 325
Db 4829 ACCTATTTTTCACCTACACTCAAAACCTTTTATAATTTTTTTTTAATTAAAATCTTCATA 4770

Qy 326 TTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTC 385
Db 4769 AAATTTTAATAACACAAATTTAACATATCTATTTATATATATTACGTATACCTCTACTTC 4710

Qy 386 AAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAA 445
Db 4709 ACAAATAAAAACT----------ACAAAATTTACTCTTTTTATTCAAAATAAAATTAAAA 4660

Qy 446 ATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTA 505
Db 4659 ATTACAAAAAAATTAAACTAAAATCTAAATTTAAAAACTATTAACATTTATTTTAATTTA 4600

Qy 506 AAACATTTACTTTCCAAATATTCTGTCAT 534
Db 4599 ATAAATATATTTATAAAATAACTTTTAAT 4571

RESULT 88
AAS46287/c
ID AAS46287 standard; DNA; 8169 BP.

XX
AC AAS46287;
XX
DT 18-DEC-2001 (first entry)
XX
DE Tumour suppressor gene derived chemically modified sequence #9.
XX
KW Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer;
KW tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW cytosine methylation; ds.
XX
OS Homo sapiens.
XX
PN WO200168912-A2.
XX
PD 20-SEP-2001.
XX
PF 15-MAR-2001; 2001WO-EP002955.
XX
PR 15-MAR-2000; 2000DE-01013847.
PR 06-APR-2000; 2000DE-01019058.
PR 07-APR-2000; 2000DE-01019173.
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2001-602752/68.
XX
PT Fragments of chemically modified genes associated with tumor suppressor
PT genes and oncogenes, useful in designing primers and probes for analyzing
PT diseases associated with cytosine methylation state e.g. cancer.
XX
PS Claim 1; SEQ ID NO 9; 27pp; English.
XX
CC The invention relates to a nucleic acid comprising a sequence of 18
CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC bisulphite, of genes associated with tumour suppression and oncogenes
CC having a sequence taken from 536 (actually 533 since numbers 408, 458 and
CC 500 are missing from the sequence listing) sequences (Ss) and sequences
CC complementary to (Ss). The nucleic acid may be a peptide nucleic acid-
CC oligomer (PNA) of at least 9 nucleotides and may form part of a set of
CC probes for detecting the cytosine methylation state and/or single
CC nucleotide polymorphisms and also to be used in an array for analysing
CC diseases associated with CpG dinucleotides e.g. cancers and tumours. The
CC probes can also be used in a method for ascertaining genetic and/or
CC epigenetic parameters for the diagnosis and/or therapy of existing
CC diseases or the predisposition to specific diseases, by analysing
CC cytosine methylations. The parameters may be compared to another set of
CC genetic and/or epigenetic parameters, the differences serving as basis
CC for diagnosis and/or prognosis events which are disadvantageous to
CC patients. The present sequence is one of the 533 genomic sequences
CC derived from tumour suppressor genes and oncogenes. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 8169 BP; 2420 A; 116 C; 1665 G; 3960 T; 0 U; 8 Other;

Query Match 4.2%; Score 46.2; DB 4; Length 8169;
Best Local Similarity 49.3%; Pred. No. 2.5;
Matches 175; Conservative 0; Mismatches 178; Indels 2; Gaps 2;

```
Qy       175 TAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTT 234
Db      1043 TAACTTTATAACTCTTATATAAATAAACACCTCATATTAACAAACTATATAATTATTTAA 984

Qy       235 TTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTA 294
Db       983 TTTAAAAAAACTTAAAATAATAACCAACTCCATCTAATATCTCAATTTTTATAAAAAATA 924
```

```
Qy       295 TAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATG 354
Db       923 TAAATATAATTATTAAAAAATAAATTAACTATAAATAAAATAAATATTTATAAATAAACT 864

Qy       355 AGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATAC 414
Db       863 CATATAATTTAAAAT-CTTAAAATTTTATATTAAATCATATATACAAAATAAATATCTAT 805

Qy       415 TACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAAT 474
Db       804 ATCATAATCAAATAAAATAAAATAATTAATTATC-CCACATTTTTAAAAAAAAAAAAATT 746

Qy       475 CTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCT 529
Db       745 ATAAAACTAAAAAATAAAAATTAAAAAAATAAAAAATCTAAAATTAAAAAAAACT 691
```

RESULT 89
ADS89527/c
ID ADS89527 standard; DNA; 8169 BP.
XX
AC ADS89527;
XX
DT 18-NOV-2004 (first entry)
XX
DE Oligonucleotide of the invention SEQ ID NO:543.
XX
KW ss; cell proliferative disorder; breast; methylation; cytostatic;
KW gene therapy; single nucleotide polymorphism; SNP.
XX
OS Unidentified.
XX
PN WO2004035803-A2.
XX
PD 29-APR-2004.
XX
PF 01-OCT-2003; 2003WO-EP010881.
XX
PR 01-OCT-2002; 2002DE-01045779.
PR 07-JAN-2003; 2003DE-01000096.
PR 17-APR-2003; 2003DE-01017955.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Foekens J, Harbeck N, Koenig T, Maier S, Martens J, Model F;
PI Nimmrich I, Rujan T, Schmitt A, Schmitt M, Look MP, Marx A;
XX
DR WPI; 2004-348468/32.
XX
PT Predicting responsiveness of a subject with breast cell proliferative
PT disorder, useful for treating or differentiating breast cell
PT proliferative disorders comprises analyzing methylation pattern of a
PT genomic DNA from the subject.
XX
PS Disclosure; SEQ ID NO 543; 104pp; English.
XX
CC The invention relates to a novel method for predicting the responsiveness
CC of a subject with a cell proliferative disorder of the breast tissues to
CC a therapy comprising analysing the methylation pattern of a target
CC nucleic acid by contacting at least one of the target nucleic acids in a
CC biological sample obtained from the subject prior to or during treatment.
CC The method of the invention has cytostatic activity, and may have a use
CC in gene therapy. The set of oligonucleotides comprising at least two of
CC the oligomers are useful for detecting the cytosine methylation state
CC and/or single nucleotide polymorphisms (SNPs) within the sequences. The
CC methods, nucleic acid, oligonucleotide, and kit are useful for the
CC treatment, characterisation, classification and/or differentiation, of
CC breast cell proliferative disorders. The method is also useful for
CC predicting the responsiveness of a subject with a cell proliferative
CC disorder of the breast tissues to a therapy. The present sequence is used
CC in the exemplification of the invention.
XX
SQ Sequence 8169 BP; 2420 A; 0 C; 1665 G; 4076 T; 0 U; 8 Other;

```
  Query Match             4.2%;  Score 46.2;  DB 13;  Length 8169;
  Best Local Similarity   49.3%;  Pred. No. 2.5;
  Matches  175;  Conservative    0;  Mismatches  178;  Indels    2;  Gaps    2;

Qy        175 TAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTT 234
              ||  ||| ||     | |     |||| || |  | ||   ||  ||     ||  ||
Db       1043 TAACTTTATAACTCTTATATAAATAAACACCTCATATTAACAAACTATATAATTATTTAA 984

Qy        235 TTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTA 294
              ||||||| | |||| |   |||       | ||| |  | |     | || | ||||  ||
Db        983 TTTAAAAAAACTTAAAATAATAACCAACTCCATCTAATATCTCAATTTTTATAAAAAATA 924

Qy        295 TAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATG 354
              |||| | |   ||| | | | | || |  ||| |   |  | || ||| |  |  |||
Db        923 TAAATATAATTATTAAAAAATAAATTAACTATAAATAAAATAAATATTTATAAATAAACT 864

Qy        355 AGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATAC 414
                |  |  ||| | | |||||  ||  |    ||   | | |  |||  ||||  | ||
Db        863 CATATAATTTAAAAT-CTTAAAATTTTATATTAAATCATATATACAAAATAAATATCTAT 805

Qy        415 TACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAAT 474
              | | |   ||     |    || ||   ||  |||||  ||| |       ||| | |
Db        804 ATCATAATCAAATAAAATAAAATAATTAATTATC-CCACATTTTTAAAAAAAAAAAAATT 746

Qy        475 CTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCT 529
              | | |  | | |||      |  |  | ||||| || ||   |  ||| |  ||
Db        745 ATAAAACTAAAAAATAAAAATTAAAAAAATAAAAAATCTAAAATTAAAAAAAACT 691


RESULT 90
ADS89253/c
ID   ADS89253 standard; DNA; 8169 BP.
XX
AC   ADS89253;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Oligonucleotide of the invention SEQ ID NO:269.
XX
KW   ss; cell proliferative disorder; breast; methylation; cytostatic;
KW   gene therapy; single nucleotide polymorphism; SNP.
XX
OS   Unidentified.
XX
PN   WO2004035803-A2.
XX
PD   29-APR-2004.
XX
PF   01-OCT-2003; 2003WO-EP010881.
XX
PR   01-OCT-2002; 2002DE-01045779.
PR   07-JAN-2003; 2003DE-01000096.
PR   17-APR-2003; 2003DE-01017955.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Foekens J,  Harbeck N,  Koenig T,  Maier S,  Martens J,  Model F;
PI   Nimmrich I,  Rujan T,  Schmitt A,  Schmitt M,  Look MP,  Marx A;
XX
DR   WPI; 2004-348468/32.
XX
PT   Predicting responsiveness of a subject with breast cell proliferative
PT   disorder, useful for treating or differentiating breast cell
PT   proliferative disorders comprises analyzing methylation pattern of a
PT   genomic DNA from the subject.
XX
PS   Disclosure; SEQ ID NO 269; 104pp; English.
XX
CC   The invention relates to a novel method for predicting the responsiveness
CC   of a subject with a cell proliferative disorder of the breast tissues to
CC   a therapy comprising analysing the methylation pattern of a target
CC   nucleic acid by contacting at least one of the target nucleic acids in a
CC   biological sample obtained from the subject prior to or during treatment.
CC   The method of the invention has cytostatic activity, and may have a use
CC   in gene therapy. The set of oligonucleotides comprising at least two of
CC   the oligomers are useful for detecting the cytosine methylation state
CC   and/or single nucleotide polymorphisms (SNPs) within the sequences. The
CC   methods, nucleic acid, oligonucleotide, and kit are useful for the
CC   treatment, characterisation, classification and/or differentiation, of
CC   breast cell proliferative disorders. The method is also useful for
CC   predicting the responsiveness of a subject with a cell proliferative
CC   disorder of the breast tissues to a therapy. The present sequence is used
CC   in the exemplification of the invention.
XX
SQ   Sequence 8169 BP; 2420 A; 116 C; 1665 G; 3960 T; 0 U; 8 Other;

  Query Match             4.2%;  Score 46.2;  DB 13;  Length 8169;
  Best Local Similarity   49.3%;  Pred. No. 2.5;
  Matches  175;  Conservative    0;  Mismatches  178;  Indels    2;  Gaps    2;

Qy        175 TAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTT 234
              ||  ||| ||     | |     |||| || |  | ||   ||  ||     ||  ||
Db       1043 TAACTTTATAACTCTTATATAAATAAACACCTCATATTAACAAACTATATAATTATTTAA 984

Qy        235 TTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTA 294
              ||||||| | |||| |   |||       | ||| |  | |     | || | ||||  ||
Db        983 TTTAAAAAAACTTAAAATAATAACCAACTCCATCTAATATCTCAATTTTTATAAAAAATA 924

Qy        295 TAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATG 354
              |||| | |   ||| | | | | || |  ||| |   |  | || ||| |  |  |||
Db        923 TAAATATAATTATTAAAAAATAAATTAACTATAAATAAAATAAATATTTATAAATAAACT 864

Qy        355 AGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATAC 414
                |  |  ||| | | |||||  ||  |    ||   | | |  |||  ||||  | ||
Db        863 CATATAATTTAAAAT-CTTAAAATTTTATATTAAATCATATATACAAAATAAATATCTAT 805

Qy        415 TACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAAT 474
              | | |   ||     |    || ||   ||  |||||  ||| |       ||| | |
Db        804 ATCATAATCAAATAAAATAAAATAATTAATTATC-CCACATTTTTAAAAAAAAAAAAATT 746

Qy        475 CTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCT 529
              | | |  | | |||      |  |  | ||||| || ||   |  ||| |  ||
Db        745 ATAAAACTAAAAAATAAAAATTAAAAAAATAAAAAATCTAAAATTAAAAAAAACT 691


RESULT 91
ABL33064/c
ID   ABL33064 standard; DNA; 6074 BP.
XX
AC   ABL33064;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 1037.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP007537.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
```

```
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful for
PT   diagnosis and treatment of diseases associated with abnormal cytosine
PT   methylation.
XX
PS   Claim 1; SEQ ID NO 1037; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention
XX
SQ   Sequence 6074 BP; 1705 A; 114 C; 1315 G; 2940 T; 0 U; 0 Other;

  Query Match             4.2%;  Score 46;  DB 6;  Length 6074;
  Best Local Similarity   48.8%;  Pred. No. 2.6;
  Matches  124;  Conservative    0;  Mismatches  130;  Indels    0;  Gaps    0;

Qy        239 AAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAA 298
              | | |||     | |  || | ||  |||| | | |   || ||      |   | |
Db       4228 ATAAACCACCACGCCTAACCGAAATATATTTTTATATACATTATTTTCCTTAAAATTTTA 4169

Qy        299 GAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTG 358
                |||   || | | ||| || ||  | |         |||   ||| || |||
Db       4168 ACAATTTCTTAAAAAAAATATAATTAACTTTATCAAACAAATAAAAAAACTAAAACTCAA 4109

Qy        359 CATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACC 418
              ||| |  | || || ||| |  | ||| |    ||||   ||    ||   | ||||
Db       4108 AATAATTAAATAATTTATCCTCAAATCACACAACAAAAACAAAAAATAAAACACACTAAA 4049

Qy        419 TGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCA 478
              | |    || |   | |  ||| | ||    ||||||     |  |   ||| || || |
Db       4048 TTCAAAATTTTAAAATTCAAAAAATAATACCCCACAATAAAAACCTCAAAAACAACCTTA 3989

Qy        479 GAAAACATAATTTT 492
              |||| | ||||||
Db       3988 AAAAAAAAAATTTT 3975


RESULT 92
ABL33251/c
ID   ABL33251 standard; DNA; 6591 BP.
XX
AC   ABL33251;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 1224.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP007537.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful for
PT   diagnosis and treatment of diseases associated with abnormal cytosine
PT   methylation.
XX
PS   Claim 1; SEQ ID NO 1224; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention
XX
SQ   Sequence 6591 BP; 2053 A; 40 C; 1314 G; 3184 T; 0 U; 0 Other;

  Query Match             4.2%;  Score 46;  DB 6;  Length 6591;
  Best Local Similarity   47.0%;  Pred. No. 2.6;
  Matches  142;  Conservative    0;  Mismatches  160;  Indels    0;  Gaps    0;

Qy        225 CCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTC 284
              | |  | | |  ||||  | ||| |  || |  |  ||||||||  |  |    |  |||
Db       3329 CATACAATATAATAAACTAACTTTCGCACTTTAAATTATTCATTAATCATATAATAATTC 3270

Qy        285 AAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAA 344
              ||| ||   ||    ||   |  ||   |||        |  ||   | ||  || ||||
Db       3269 AAATAATTCATCTTCCAAAATAAACACTAAAACTAACACTCATAATAAATTATAATTTAA 3210

Qy        345 ACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCT 404
              | |  |||  |      | |    |  |||   |      ||    |   |   ||  |
Db       3209 AAAATAAAACAACCGCCAATCTCATTAATAAAAATTATAATAATTAAATTTATCAAAAAT 3150

Qy        405 AAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTT 464
              |||  || || | || ||||||  |           || |       || | | |  |||||
Db       3149 AAACATACACGATCTTCTTTCTAAAATTTTCTTTAAAATTTTCGCACAAATAATCTATTT 3090

Qy        465 TGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAAT 524
              ||||  |   | ||||  | |  ||||||  |    | |||  |  |  || |||||
Db       3089 CTTAAATTTTAAAAAAAAAACTTAAATTTTATTAATCCCTCAAAACTCAATCTTTCAAAT 3030

Qy        525 AT 526
              ||
Db       3029 TT 3028


RESULT 93
ABK28448/c
ID   ABK28448 standard; DNA; 9095 BP.
XX
AC   ABK28448;
XX
DT   23-APR-2002  (first entry)
XX
DE   DNA transcription associated complementary genomic DNA #161.
XX
KW   DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;
KW   PNA; cytosine methylation state; SNP; retroviral infection; gene; ds;
KW   single nucleotide polymorphism; adenosine deaminase deficiency; cancer;
KW   viral infection; Sezary syndrome; haematological disorder; tuberculosis;
KW   immunological disorder; Werner syndrome; developmental disorder;
KW   psoriasis; Rieger's syndrome; neurological disorder; erythropoiesis;
```

KW neurodegenerative disorder; Waardenburg syndrome; Niemann-Pick disease;
KW myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;
KW angiogenesis; congenital heart disease; HDR syndrome; gene therapy;
KW polyglutamine disorder; solid tumour.
XX
OS Unidentified.
XX
PN WO200192565-A2.
XX
PD 06-DEC-2001.
XX
PF 06-APR-2001; 2001WO-EP003973.
XX
PR 06-APR-2000; 2000DE-01019058.
PR 07-APR-2000; 2000DE-01019173.
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-090046/12.
XX
PT New nucleic acids or oligomers, useful for diagnosing or treating
PT diseases associated with DNA transcription, e.g. immunological disorders,
PT Werner syndrome, psoriasis, myocardial infarction, solid tumors or
PT cancer.
XX
PS Claim 1; SEQ ID NO 322; 32pp; English.
XX
CC The invention relates to a nucleic acid, which comprises a segment of the
CC chemically pretreated DNA of genes associated with DNA transcription from
CC one of 346 sequences, and an oligomer, in particular an oligonucleotide
CC or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical
CC to the chemically pretreated DNA of genes associated with DNA
CC transcription. The set of oligomer probes are useful for detecting the
CC cytosine methylation state and/or single nucleotide polymorphisms (SNPs)
CC in a chemically pretreated genomic DNA. The nucleic acids are useful for
CC diagnosing or treating diseases associated with DNA transcription
CC (particularly with the methylation status), e.g. adenosine deaminase
CC deficiency, viral infection, retroviral infection, Sezary syndrome,
CC haematological disorders, immunological disorders, Werner syndrome,
CC tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,
CC neurological disorders, neurodegenerative disorders, Waardenburg
CC syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial
CC infarction, hypertension, angiogenesis, erythropoiesis, congenital heart
CC disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours
CC or cancer. Sequences ABK28127-ABK28472 represent DNA transcription
CC associated genomic DNA molecules of the invention. Note: The sequence
CC data for this patent did not form part of the printed specification but
CC was obtained in electronic format directly from the European Patent
CC Office
XX
SQ Sequence 9095 BP; 2790 A; 109 C; 1749 G; 4447 T; 0 U; 0 Other;

Query Match 4.2%; Score 46; DB 6; Length 9095;
Best Local Similarity 44.6%; Pred. No. 2.8;
Matches 232; Conservative 0; Mismatches 280; Indels 8; Gaps 1;

```
Qy        33 TCTCTATGTTCTTATAAGCAAATGTTACAATGAGATATGAGTTTTAAAGCCAGATCTTCC 92
             ||| | | | |||  |   || |  |    | | |||  | |  ||||   | ||  |
Db      7037 TCTATTTATACTTTAATAAAATTTATCTTCTAATATAAAAATCATAAATAAATATAATTA 6978

Qy        93 TTATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATA 152
             | |      |      | |  |     |    | |   || | |   |   | | ||| |
Db      6977 TAACTATAATATAATTTTTTAAACCAATATTTTCTAAAATTTAAAAATAACTCAAAAAAA 6918

Qy       153 TTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTT 212
             |  |||| || |||| | |  ||| |||  || |  || |  ||  | | |||||    |
Db      6917 TCTATCAATAAAAATAAATACATACTTTATTACATCATTTTTTCACATAAACATTTCAAT 6858

Qy       213 GTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTT 272
```

```
             | ||||     ||  ||  |  | | | ||  |   | | | | ||||| |
Db      6857 AATTATATTAAAACTATCTTAAATCTATAAATTTTTTAATTAAAAATCAATTCATCTCAC 6798

Qy       273 TTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCA 332
             | |  |     |  || |  || |  ||||  | | | |         | |     |
Db      6797 TATATTAAACTATCAATATTAATTATCAAATTTAATTATATATTACATTACTTTACCTAA 6738

Qy       333 GTTGAAAATTAAACACAAAATGAGTGCATATTACATTACT--------TAATCTTGCAGT 384
              ||  |||| |||   ||  |      | |  |  ||| ||        |||  || || |
Db      6737 TTTTCAAATAAAAATAAACTTTTTATCTTTATTTATTTCTATAATAAATAAATTTACAAT 6678

Qy       385 CAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGA 444
              ||   ||  |   |   | |  | |  ||     |||  ||   |   || | |
Db      6677 AAATTACAATTAAATATTAAATAAATTTTATAATCAATTTAATACAAAAAATAAATTTTA 6618

Qy       445 AATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATT 504
             || || || ||   | | |   | |  |   | ||||  ||| ||  | || ||    |
Db      6617 AAATATCAAAATACATTATATATTACTTTTTCTAAAAAAAATATTTCCTCATTTATACAT 6558

Qy       505 AAAACATTTACTTTCCAAATATTCTGTCATTCAGGAGTAT 544
              ||| | | |       | || || | | |||    |||
Db      6557 TAAAAAATAAAAAAAACTAAATACTATAACTCATACCTAT 6518
```

RESULT 94
ABL32190/c
ID ABL32190 standard; DNA; 17137 BP.
XX
AC ABL32190;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 163.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 163; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines! The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel

```
CC   diseases. The present sequence is a gene of the invention
XX
SQ   Sequence 17137 BP; 4775 A; 240 C; 3303 G; 8819 T; 0 U; 0 Other;

  Query Match             4.2%;  Score 46;  DB 6;  Length 17137;
  Best Local Similarity   44.4%;  Pred. No. 3.3;
  Matches  184;  Conservative    0;  Mismatches  230;  Indels    0;  Gaps    0;

Qy        108 ATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAAT 167
              ||   ||  | |   |  | | |    |  |  ||  |  | | | |   || | || |  |
Db      11499 ATACATATATATATATATATATATAAATATATTTCTACACATAAAAATATAAATATACTT 11440

Qy        168 CAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCT 227
              |    |  | |||   | |   |    |  |||  |       |||    |||||    |
Db      11439 CCTAAACAATTTTCAAATACATTAATTTAATAATAAATACAATTTACATATATAAATACA 11380

Qy        228 TGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAA 287
              | || | | || |  || |   | | ||     ||| ||| | |   |   ||  | | |
Db      11379 TAATATATATATACTACATATTATATATTATTTTATACATATATACCTATATTTATAATA 11320

Qy        288 AAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACA 347
              | |  ||  |  ||||   |    | ||  |  | | | ||  |      ||||  || |
Db      11319 ATAACTACCATAAAATATTTAACAATAATTAAAAAAAAATACCACAAAAAAAAAAAATCT 11260

Qy        348 CAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAA 407
               |||||      |||||  |||| |||   ||         ||   | |||      ||| |
Db      11259 AAAAATCTAATAATATTTCATTTCTTCGCCTATATAATCAATACATAAACATCCACTACA 11200

Qy        408 GGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGG 467
              |     |   | | |     |       |      ||| | |  |  |  | | |||
Db      11199 TTACTCTCTACAATTATTTATAAACTTAAATATATACTATTTCAAATACATACAATTTTA 11140

Qy        468 AAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCA 521
              ||| |   |    ||| ||| |||| |    |  ||||||| | | ||| || |
Db      11139 AAAAACAATACACAAAAATATTTTTATTATAAAAATTAAAAAACTAACTCTCTA 11086


RESULT 95
ABL32191
ID   ABL32191 standard; DNA; 17137 BP.
XX
AC   ABL32191;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 164.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP007537.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful for
PT   diagnosis and treatment of diseases associated with abnormal cytosine
PT   methylation.
XX
PS   Claim 1; SEQ ID NO 164; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention
XX
SQ   Sequence 17137 BP; 5908 A; 240 C; 3151 G; 7838 T; 0 U; 0 Other;

  Query Match             4.2%;  Score 45.8;  DB 6;  Length 17137;
  Best Local Similarity   49.4%;  Pred. No. 3.7;
  Matches  119;  Conservative    0;  Mismatches  122;  Indels    0;  Gaps    0;

Qy        207 TGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCA 266
              |||||    |||  || |   ||  || ||||  | ||   ||   |  |    ||||
Db      14656 TGTTTAAGTAAAGTAATTATTTTTTTGTTTTTTATGAAAATAATTAGTTTTTTTTATTGG 14715

Qy        267 TTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATAT 326
              |     ||  ||  || || ||| ||||   |||||   ||   ||||   |   || |
Db      14716 ATATAAGTTTTTTATTTAAAGAAAAGTATTTTGAAATTAGTTTTTAGAATTTTTTAATGT 14775

Qy        327 TAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCA 386
              ||   | || |||  || | | | | |   | ||| | ||  || | ||     | |
Db      14776 TATTAAATTTAAATATATAAATTAGAAGGATTTGTTTTAAAATAAATAGTGTTATTATAA 14835

Qy        387 AAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAA 446
              | ||||||    |     || ||  ||  |  ||  |  | |    | || || | |
Db      14836 TATGTAAAATTGGATTAATTAGTTAATTTAAATTTTTGATAGTTGACGGGAAATTATATA 14895

Qy        447 T 447
              |
Db      14896 T 14896


RESULT 96
AAI61371
ID   AAI61371 standard; DNA; 335913 BP.
XX
AC   AAI61371;
XX
DT   16-OCT-2001  (first entry)
XX
DE   Soybean 240017 region G3, SEQ ID NO: 2.
XX
KW   Soybean; antihelmintic; gene therapy; soybean cyst nematode; SCN;
KW   SCN resistance; rhg1; Rhg4; SCN resistant allele; plant breeding;
KW   240017 region G3; 318013 region A3; 515002 region G2; ds.
XX
OS   Glycine max.
XX
PN   WO200151627-A2.
XX
PD   19-JUL-2001.
XX
PF   05-JAN-2001; 2001WO-US000552.
XX
PR   07-JAN-2000; 2000US-0174880P.
XX
PA   (MONS ) MONSANTO CO.
XX
PI   Hauge BM,  Wang ML,  Parsons JD,  Parnell LD;
XX
DR   WPI; 2001-425872/45.
DR   P-PSDB; AAM42214.
XX
```

PT New purified nucleic acid for producing a soybean plant having soybean
PT cyst nematode resistance and for use in plant breeding programs.
XX
PS Claim 2; Page 204-400; 1353pp; English.
XX
CC The invention relates to nucleic acid molecules from regions of the
CC soybean genome which are associated with soybean cyst nematode (SCN)
CC resistance. The nucleic acids are used to transform plants, and can
CC produce soybean plants having an rhg1 or an Rhg4 SCN resistant allele.
CC The nucleic acids can be used for investigating rhg1 or Rhg4 haplotypes
CC of soybean plants and for introgressing SCN resistance or partial SCN
CC resistance into soybean plants. They can also be used in plant breeding
CC programmes. The invention also relates to proteins encoded by such
CC nucleic acid molecules, as well as antibodies capable of recognising
CC these proteins. The present sequence is a nucleic acid molecule provided
CC in the specification
XX
SQ Sequence 335913 BP; 114579 A; 53403 C; 53026 G; 114905 T; 0 U; 0 Other;

Query Match 4.2%; Score 45.8; DB 5; Length 335913;
Best Local Similarity 49.4%; Pred. No. 7.7;
Matches 206; Conservative 0; Mismatches 202; Indels 9; Gaps 3;

Qy 119 TTGAAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGT 178
Db 98750 TTAAAATTATATATGTTTTATTTTTTTTTAACTTATGATAATTTTAAAGCTGTAATATTT 98809

Qy 179 TTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTA 238
Db 98810 TTTATTGTTTTTATGACTTTGAAATCATATTTTATTTTATGTTTAAAATTTTTTGTTTTA 98869

Qy 239 AAAGACCTTACAGACATACAGCTAT--TCATTTGTTTTTGGTTTGTTCAAAAAAGGTATA 296
Db 98870 TGATTTTTTATTTACTTGTATTTTTTATACTTTTATTTTCAATTCTTTTAAAAAATTATA 98929

Qy 297 AAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAG 356
Db 98930 AATAATTTTATTTATTTAATTATTAAATAA-----ATATTTTAATTTTACTAGTTATTAG 98984

Qy 357 TGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTA 416
Db 98985 TTTCATTTAATATATTTTGTATATTTAAAATTTATATAATTTTCTAATAATGTTATTGAA 99044

Qy 417 CCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAA--ATTTTATTTTGGAAATAAT 474
Db 99045 GTTGATTTGTTTTGTAAATAAAATAAAAAATTAACATAATGATATTTAATATTGTTTTTC 99104

Qy 475 CTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGT 531
Db 99105 TTTTAAATACTTAAATTTTAAAAAACTAAAATTTTAAAAAAAGTTAAAATATTTTGT 99161

RESULT 97
AAI61372
ID AAI61372 standard; DNA; 335913 BP.
XX
AC AAI61372;
XX
DT 16-OCT-2001 (first entry)
XX
DE Soybean 240017 region G3, SEQ ID NO: 3.
XX
KW Soybean; antihelmintic; gene therapy; soybean cyst nematode; SCN;
KW SCN resistance; rhg1; Rhg4; SCN resistant allele; plant breeding;
KW 240017 region G3; 318013 region A3; 515002 region G2; ds.
XX
OS Glycine max.
XX
PN WO200151627-A2.
XX
PD 19-JUL-2001.
XX
PF 05-JAN-2001; 2001WO-US000552.
XX
PR 07-JAN-2000; 2000US-0174880P.
XX
PA (MONS) MONSANTO CO.
XX
PI Hauge BM, Wang ML, Parsons JD, Parnell LD;
XX
DR WPI; 2001-425872/45.
DR P-PSDB; AAM42215.
XX
PT New purified nucleic acid for producing a soybean plant having soybean
PT cyst nematode resistance and for use in plant breeding programs.
XX
PS Claim 2; Page 400-595; 1353pp; English.
XX
CC The invention relates to nucleic acid molecules from regions of the
CC soybean genome which are associated with soybean cyst nematode (SCN)
CC resistance. The nucleic acids are used to transform plants, and can
CC produce soybean plants having an rhg1 or an Rhg4 SCN resistant allele.
CC The nucleic acids can be used for investigating rhg1 or Rhg4 haplotypes
CC of soybean plants and for introgressing SCN resistance or partial SCN
CC resistance into soybean plants. They can also be used in plant breeding
CC programmes. The invention also relates to proteins encoded by such
CC nucleic acid molecules, as well as antibodies capable of recognising
CC these proteins. The present sequence is a nucleic acid molecule provided
CC in the specification
XX
SQ Sequence 335913 BP; 114582 A; 53398 C; 53027 G; 114906 T; 0 U; 0 Other;

Query Match 4.2%; Score 45.8; DB 5; Length 335913;
Best Local Similarity 49.4%; Pred. No. 7.7;
Matches 206; Conservative 0; Mismatches 202; Indels 9; Gaps 3;

Qy 119 TTGAAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGT 178
Db 98750 TTAAAATTATATATGTTTTATTTTTTTTTAACTTATGATAATTTTAAAGCTGTAATATTT 98809

Qy 179 TTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTA 238
Db 98810 TTTATTGTTTTTATGACTTTGAAATCATATTTTATTTTATGTTTAAAATTTTTTGTTTTA 98869

Qy 239 AAAGACCTTACAGACATACAGCTAT--TCATTTGTTTTTGGTTTGTTCAAAAAAGGTATA 296
Db 98870 TGATTTTTTATTTACTTGTATTTTTTATACTTTTATTTTCAATTCTTTTAAAAAATTATA 98929

Qy 297 AAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAG 356
Db 98930 AATAATTTTATTTATTTAATTATTAAATAA-----ATATTTTAATTTTACTAGTTATTAG 98984

Qy 357 TGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTA 416
Db 98985 TTTCATTTAATATATTTTGTATATTTAAAATTTATATAATTTTCTAATAATGTTATTGAA 99044

Qy 417 CCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAA--ATTTTATTTTGGAAATAAT 474
Db 99045 GTTGATTTGTTTTGTAAATAAAATAAAAAATTAACATAATGATATTTAATATTGTTTTTC 99104

Qy 475 CTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGT 531
Db 99105 TTTTAAATACTTAAATTTTAAAAAACTAAAATTTTAAAAAAAGTTAAAATATTTTGT 99161

RESULT 98
ADA71938/c
ID ADA71938 standard; DNA; 2000 BP.
XX
AC ADA71938;
XX
DT 20-NOV-2003 (first entry)
XX
DE Rice gene, SEQ ID 5263.
XX
KW Plant; bacterial infection; fungal infection; viral infection; rice;

KW gene; ds.
XX
OS Oryza sativa.
XX
PN WO2003000898-A1.
XX
PD 03-JAN-2003.
XX
PF 22-JUN-2001; 2001WO-IB001105.
XX
PR 22-JUN-2001; 2001WO-IB001105.
XX
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
XX
DR WPI; 2003-175290/17.
XX
PT Identifying at least one gene involved in plant resistance or response to
PT pathogenic infection for conferring resistance or tolerance to a plant to
PT bacterial, fungal or viral infection by determining or detecting plant
PT gene expression.
XX
PS Claim 27; SEQ ID NO 5263; 899pp; English.
XX
CC The present invention relates to a method (M1) for identifying genes
CC involved in plant resistance or response to pathogenic infection. M1
CC comprises identifying a gene whose expression is significantly altered in
CC the incompatible interaction of plant gene expression relative to
CC expression of the gene in an uninfected plant, in a mutant plant that
CC does not express a gene associated with response to pathogenic infection,
CC or in a corresponding incompatible or compatible interaction. (M1) is
CC useful for conferring resistance to resistance or tolerance to a plant to
CC bacterial, fungal or viral infection. The present sequence was used to
CC illustrate the invention.
XX
SQ Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;

Query Match 4.2%; Score 45.6; DB 8; Length 2000;
Best Local Similarity 9.5%; Pred. No. 2.4;
Matches 66; Conservative 314; Mismatches 308; Indels 4; Gaps 2;

Qy 327 TAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCA 386
Db 1025 TAGGTGCTKMWRTYTRSMSWTYAMWKKYTKYMTAYSSTWKWYWAYKWRAYAWSRSRKTWW 966

Qy 387 AAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAA 446
Db 965 CTGGKRMA---TYCGTKMWAAGRWRWRMAWCWYCCMWKWWKMTSCMWWKYWRTWSCWYTM 909

Qy 447 TTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAA 506
Db 908 WWGAMRYAYYAMRRRRWTYKWSWRRMYWTMTKWAWTWMTCMCMAKWYMATGWATWMWWRY 849

Qy 507 AACATTTACTTTCCAAATATTCTGTCATTCAGGAGTATGGAAGTATCGATGGCTTCTTTA 566
Db 848 TMYTYCYAMTCAKCKYKMAMTKWWTTWACAWRATSWRWRAMAGMRWKRYKMKRAYWWRWW 789

Qy 567 AAATGAAGCAGGAGGGTCTGGCAGAGAGTATCTATGAAATAAGTTCCTCTGACCTTCACG 626
Db 788 RCWKAGWARWMKSRYRWKWKKYATRYYWKMWAMTWWWSWRRWKSYRMWSGMGRMRWSAWR 729

Qy 627 CTTAATTTTCTGAATGGAGTGGAGCAAATTACTTCAAGCTTCACTTAACTTGCATATGAA 686
Db 728 YCSRMKCAK-TKYASSARWTKRAKRSYRYRRRWYWKRKGWTYRYRYWRSCRMTRARMSKR 670

Qy 687 ATGAACCGTACAAAAATACAAGAGTGTCAGGAGAAAGTTATGCTCTGGTAAATATTTTGC 746
Db 669 RKWAGASMKSCWMYWRGARSMWYSKYSCSAKCCKKTRYMTSSYMSTGMYGMYSSYKSMSW 610

Qy 747 AAAACAGATAAAAGATAATACTAGAGCTCTGTCCTCAAAGAGTTAAGCAGCTAATCTAAG 806
Db 609 TSKMSYMGKMTCTMYTSMKGSTRRSKMGRWSGMSRMYMRWWKKMRKRKYMRYMKWKCTWR 550

Qy 807 GAGGTAAACTCTATGTCAGCAGGATGAACTGCTCTTCCCTTTCCTCCTCAATAAATTGCA 866
Db 549 RCMCYRWGYTMYTTSRSRMMYTGRYKARYTSKRRYMWYKYRKYCWYYYYGMYMKCSYMMR 490

Qy 867 AATCATCTAGTCCAACATCTTTACCACCAGTGCCTGAGGCTCCAGAGGAGCCATTGCCTT 926
Db 489 YGYCKACKKCCYAMCWKAAYSGMMMYWYRKYSKWMRMSTKYMWSMWYKKCRSMKYGAKGC 430

Qy 927 CTCAAGGTCACATAGGTGGTGGGTGAGTTAGGACCAAATCTAGAATTCCTGACTCCAGTA 986
Db 429 YGCKMWTYCSYGYMKWYTYMGSYKYSRCYKYMRMYMYKGWMYMMYYSAYSSMMTWYYYYA 370

Qy 987 ACTTCTGAAGTCATTTTGTTTTTTATTTTTAT 1018
Db 369 KYWKYWYKRRGTMSWYGKSYKKKYCTWWCYMK 338

RESULT 99
ABL33670/c
ID ABL33670 standard; DNA; 8446 BP.
XX
AC ABL33670;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 1643.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 1643; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 8446 BP; 2558 A; 50 C; 1741 G; 4097 T; 0 U; 0 Other;

Query Match 4.2%; Score 45.6; DB 6; Length 8446;

```
  Best Local Similarity   48.2%;  Pred. No. 3.4;
  Matches  160;  Conservative    0;  Mismatches  169;  Indels    3;  Gaps    1;

Qy        294 ATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAAT 353
              ||||| ||  ||   | | | | ||  || || | ||   | | |  ||||  | |||
Db       1896 ATAAAATAAACCAAATACAAACAAATACTACATAATCCCACTCACATATAAAATCTAAAC 1837

Qy        354 GAGTGCAT---ATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGT 410
               | |  ||   || | | |||  |||     | | |     ||||| | || | |
Db       1836 AAATTAATTCCATAAAAATACAAAATAAAATAATAATTACCAAAAATTAAAACGATTAAA 1777

Qy        411 ATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAA 470
              | ||          |||       |   |||  |||    || || ||| |||    |||
Db       1776 AAACGTAAAATAAAATTAAAAATATATTAATCAAAAAATACAAAATTTTAATTAAATAAA 1717

Qy        471 TAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTG 530
              ||  | |     | | ||| | ||||  |   ||| ||| | | |     | |||| |
Db       1716 AAAAATAATTTCAAAAAATCTATTATACAACGTTATAACTATAATTAAAAATATATTATA 1657

Qy        531 TCATTCAGGAGTATGGAAGTATCGATGGCTTCTTTAAAATGAAGCAGGAGGGTCTGGCAG 590
              |  || |  | ||   || ||   ||    || | | ||  ||  |  |   | |  |
Db       1656 TTCTTAAAAAATACTAAAATAAATATATTCTCCTCACAAAAAATAATAACTATATAAAAT 1597

Qy        591 AGAGTATCTATGAAATAAGTTCCTCTGACCTT 622
              |    || ||| || ||| |   | | | | |
Db       1596 AATACATATATTAATTAATTAAATTTAATCAT 1565


RESULT 100
ABL33680/c
ID   ABL33680 standard; DNA; 15951 BP.
XX
AC   ABL33680;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 1653.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP007537.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful for
PT   diagnosis and treatment of diseases associated with abnormal cytosine
PT   methylation.
XX
PS   Claim 1; SEQ ID NO 1653; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention
XX
SQ   Sequence 15951 BP; 4545 A; 97 C; 3159 G; 8150 T; 0 U; 0 Other;

  Query Match             4.2%;  Score 45.6;  DB 6;  Length 15951;
  Best Local Similarity   45.0%;  Pred. No. 4;
  Matches  171;  Conservative    0;  Mismatches  209;  Indels    0;  Gaps    0;

Qy        143 TTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAG 202
              || | ||| ||  ||  | || |         | | |    || ||   || | || |||
Db       8305 TTTATAAAAATAAATTCTCACTATATTACCCAAACTAACCTAAAACCCCTAAACTCAAAC 8246

Qy        203 ACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTA 262
              |     |||| || |     |       | |  | | | | |  | | |   ||||
Db       8245 AATACTTTTTCTTTAACCTCCAAAAATACTAATATTATAAAAATAAAAATAATACGATAC 8186

Qy        263 TTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCAT 322
              ||  ||| ||||   |   ||  ||  |   | |    | | ||   | | |||
Db       8185 TTTTTTTTTTTTAAATAAATTACAATCAAAAAAATTACACTACAAAAAAAAAAACTATTA 8126

Qy        323 ATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCA 382
               || | ||   | |||||||||| | |||   | |  || | |||| | | ||  |   |
Db       8125 CAATAATCCCAATAAAAATTAAAAAAAAACAAATTTAATCTAACATCAATAAAAATACAA 8066

Qy        383 GTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATA 442
               | |||   | |||   ||      |   | | |||     |    | |  | |   || |
Db       8065 ATAAAAAAAATAAATATAAAAATTATTAAAAATACAAAACTATACAACTTATTAATAACA 8006

Qy        443 GAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTA 502
               ||   |      |    | ||  || ||||     ||  |    | ||| |  || |
Db       8005 AAACCCAAATATTACAAACTATTACAACTAATACATTAATTCCCTTATATTTCTAAACGA 7946

Qy        503 TTAAAACATTTACTTTCCAA 522
              | | || ||| |   |  ||
Db       7945 TAATAATATTAATAATATAA 7926


Search completed: April 21, 2005, 00:49:19
Job time : 370.089 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 20, 2005, 22:49:10 ; Search time 3049.48 Seconds
(without alignments)
17319.707 Million cell updates/sec

Title: US-09-939-209A-3_COPY_11056_12145
Perfect score: 1090
Sequence: 1 tgctttccgtgtggtggtgc.....taactctaggatctcaggtg 1090

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1090	100.0	20300	6	AX451337	AX451337 Sequence
	2	1090	100.0	165329	9	AL583850	AL583850 Human DNA
c	3	1085.8	99.6	191699	2	AC031977	AC031977 Homo sapi
	4	65	6.0	248140	2	BX901898	BX901898 Danio rer
c	5	63	5.8	348034	3	CR382400	CR382400 Plasmodiu
	6	60.6	5.6	131459	9	AL391538	AL391538 Human DNA
c	7	60	5.5	216734	5	AB073376	AB073376 Oryzias l
	8	59.2	5.4	116696	3	PFMAL3P3	Z98547 Plasmodium
	9	59.2	5.4	214310	2	CR847982	CR847982 Danio rer
c	10	58.8	5.4	119227	5	BX901920	BX901920 Zebrafish
c	11	58.6	5.4	196788	2	AC021626	AC021626 Homo sapi
	12	58.4	5.4	2255	3	DDITUBA	L13999 Dictyosteli
	13	58.2	5.3	12469	3	AE001377	AE001377 Plasmodiu
	14	58.2	5.3	196722	2	CR352307	CR352307 Danio rer
c	15	57.8	5.3	6706	3	AF052006	AF052006 Dictyoste
	16	57.8	5.3	186431	2	AC022281	AC022281 Homo sapi
	17	56.8	5.2	2009	6	AX457067	AX457067 Sequence
	18	56.8	5.2	151143	2	AC141740	AC141740 Apis mell
	19	56.8	5.2	225600	10	AC115766	AC115766 Mus muscu
c	20	56.6	5.2	110000	2	PFMAL8P1_03	Continuation (4 of
c	21	56.6	5.2	176321	2	BX942829	BX942829 Danio rer
	22	56.6	5.2	176368	9	CNS01RGG	AL157971 Human chr
	23	56.4	5.2	348600	1	AB063521	AB063521 Wiggleswo
	24	56.2	5.2	88549	3	AC116924	AC116924 Dictyoste
	25	56	5.1	121958	2	CR759861	CR759861 Danio rer
c	26	56	5.1	185596	9	AC021553	AC021553 Homo sapi
	27	55.8	5.1	1141	6	AX083744	AX083744 Sequence
	28	55.8	5.1	4601	3	DMU11584	U11584 Drosophila
	29	55.8	5.1	19517	3	DMU37541	U37541 Drosophila
	30	55.8	5.1	36055	2	AP005815	AP005815 Homo sapi
	31	55.8	5.1	38849	9	AC024070	AC024070 Homo sapi
	32	55.8	5.1	110000	3	AC116984_1	Continuation (2 of
	33	55.8	5.1	160683	9	AP004607	AP004607 Homo sapi
	34	55.8	5.1	196762	2	CR759796	CR759796 Danio rer
c	35	55.6	5.1	172413	2	AC134388	AC134388 Pan trogl
c	36	55.6	5.1	172577	2	AL365174	AL365174 Homo sapi
	37	55.6	5.1	206038	2	CR847802	CR847802 Danio rer
	38	55.6	5.1	206464	2	CR847516	CR847516 Danio rer
c	39	55.6	5.1	225328	2	CR847887	CR847887 Danio rer
	40	55.6	5.1	232621	2	BX936449	BX936449 Danio rer
	41	55.2	5.1	532	3	AF527403	AF527403 Cydia pom
	42	55	5.0	110000	2	PFMAL8P1_07	Continuation (8 of
	43	55	5.0	110000	2	PFMAL8P1_08	Continuation (9 of
	44	55	5.0	146570	3	AC117076	AC117076 Dictyoste
c	45	55	5.0	169841	9	AL607043	AL607043 Human DNA

ALIGNMENTS

RESULT 1
AX451337
LOCUS AX451337 20300 bp DNA linear PAT 03-JUL-2002
DEFINITION Sequence 3 from Patent WO0216653.
ACCESSION AX451337
VERSION AX451337.1 GI:21698388
KEYWORDS .
SOURCE synthetic construct
ORGANISM synthetic construct
other sequences; artificial sequences.
REFERENCE 1
AUTHORS Levitt,P.R., Mirnics,K., Kodavali,V.C. and Nimgaonkar,V.L.
TITLE Methods and systems for facilitating the diagnosis and treatment of schizophrenia
JOURNAL Patent: WO 0216653-A 3 28-FEB-2002;
University of Pittsburgh (US)
FEATURES Location/Qualifiers
source 1. .20300
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="A genomic sequence containing RGS4 nucleic acid sequence and sequ ences upstream and downstream to the RGS4 nucleic acid sequence"
ORIGIN

Query Match 100.0%; Score 1090; DB 6; Length 20300;
Best Local Similarity 100.0%; Pred. No. 6.6e-198;
Matches 1090; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 TGCTTTCCGTGTGGTGGTGCAGTAACTTTCACTCTCTATGTTCTTATAAGCAAATGTTAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11056 TGCTTTCCGTGTGGTGGTGCAGTAACTTTCACTCTCTATGTTCTTATAAGCAAATGTTAC 11115

Qy         61 AATGAGATATGAGTTTTAAAGCCAGATCTTCCTTATCTCTCTGCCCCATCTCTAGTTCTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11116 AATGAGATATGAGTTTTAAAGCCAGATCTTCCTTATCTCTCTGCCCCATCTCTAGTTCTT 11175

Qy        121 GAAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11176 GAAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTT 11235
```

```
Qy        181 TGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11236 TGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAA 11295

Qy        241 AGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11296 AGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGA 11355

Qy        301 AATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11356 AATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCA 11415

Qy        361 TATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11416 TATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTG 11475

Qy        421 CTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11476 CTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGA 11535

Qy        481 AAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGTCATTCAGGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11536 AAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGTCATTCAGGA 11595

Qy        541 GTATGGAAGTATCGATGGCTTCTTTAAAATGAAGCAGGAGGGTCTGGCAGAGAGTATCTA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11596 GTATGGAAGTATCGATGGCTTCTTTAAAATGAAGCAGGAGGGTCTGGCAGAGAGTATCTA 11655

Qy        601 TGAAATAAGTTCCTCTGACCTTCACGCTTAATTTTCTGAATGGAGTGGAGCAAATTACTT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11656 TGAAATAAGTTCCTCTGACCTTCACGCTTAATTTTCTGAATGGAGTGGAGCAAATTACTT 11715

Qy        661 CAAGCTTCACTTAACTTGCATATGAAATGAACCGTACAAAAATACAAGAGTGTCAGGAGA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11716 CAAGCTTCACTTAACTTGCATATGAAATGAACCGTACAAAAATACAAGAGTGTCAGGAGA 11775

Qy        721 AAGTTATGCTCTGGTAAATATTTTGCAAAACAGATAAAAGATAATACTAGAGCTCTGTCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11776 AAGTTATGCTCTGGTAAATATTTTGCAAAACAGATAAAAGATAATACTAGAGCTCTGTCC 11835

Qy        781 TCAAAGAGTTAAGCAGCTAATCTAAGGAGGTAAACTCTATGTCAGCAGGATGAACTGCTC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11836 TCAAAGAGTTAAGCAGCTAATCTAAGGAGGTAAACTCTATGTCAGCAGGATGAACTGCTC 11895

Qy        841 TTCCCTTTCCTCCTCAATAAATTGCAAATCATCTAGTCCAACATCTTTACCACCAGTGCC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11896 TTCCCTTTCCTCCTCAATAAATTGCAAATCATCTAGTCCAACATCTTTACCACCAGTGCC 11955

Qy        901 TGAGGCTCCAGAGGAGCCATTGCCTTCTCAAGGTCACATAGGTGGTGGGTGAGTTAGGAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11956 TGAGGCTCCAGAGGAGCCATTGCCTTCTCAAGGTCACATAGGTGGTGGGTGAGTTAGGAC 12015

Qy        961 CAAATCTAGAATTCCTGACTCCAGTAACTTCTGAAGTCATTTTGTTTTTTATTTTTATGG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12016 CAAATCTAGAATTCCTGACTCCAGTAACTTCTGAAGTCATTTTGTTTTTTATTTTTATGG 12075

Qy       1021 TTTTATTATAAGAATACTTGCTAAGCACACTTACCCCCTGCATTGATTAATAACTCTAGG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12076 TTTTATTATAAGAATACTTGCTAAGCACACTTACCCCCTGCATTGATTAATAACTCTAGG 12135

Qy       1081 ATCTCAGGTG 1090
              ||||||||||
Db      12136 ATCTCAGGTG 12145
```

```
RESULT 2
AL583850
LOCUS       AL583850              165329 bp    DNA     linear   PRI 15-NOV-2001
DEFINITION  Human DNA sequence from clone RP11-430G6 on chromosome 1, complete
            sequence.
ACCESSION   AL583850
VERSION     AL583850.5  GI:16973044
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 165329)
  AUTHORS   Tracey,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Nov 16, 2001 this sequence version replaced gi:15020514.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            RP11-430G6 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-430G6 It may be shorter because we sequence overlapping
            sections only once, except for a short overlap.
            The true right end of clone RP11-430G6 is at 165329 in this
            sequence. The true right end of clone RP11-331H2 is at 2000 in this
            sequence.
FEATURES             Location/Qualifiers
     source          1. .165329
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-430G6"
                     /clone_lib="RPCI-11.2"
ORIGIN
```

```
  Query Match             100.0%;  Score 1090;  DB 9;  Length 165329;
  Best Local Similarity   100.0%;  Pred. No. 4.6e-198;
  Matches 1090;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TGCTTTCCGTGTGGTGGTGCAGTAACTTTCACTCTCTATGTTCTTATAAGCAAATGTTAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      82737 TGCTTTCCGTGTGGTGGTGCAGTAACTTTCACTCTCTATGTTCTTATAAGCAAATGTTAC 82796

Qy         61 AATGAGATATGAGTTTTAAAGCCAGATCTTCCTTATCTCTCTGCCCCATCTCTAGTTCTT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      82797 AATGAGATATGAGTTTTAAAGCCAGATCTTCCTTATCTCTCTGCCCCATCTCTAGTTCTT 82856

Qy        121 GAAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      82857 GAAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTT 82916

Qy        181 TGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      82917 TGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAA 82976
```

```
Qy        241 AGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      82977 AGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGA 83036

Qy        301 AATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83037 AATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCA 83096

Qy        361 TATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83097 TATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTG 83156

Qy        421 CTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83157 CTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGA 83216

Qy        481 AAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGTCATTCAGGA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83217 AAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGTCATTCAGGA 83276

Qy        541 GTATGGAAGTATCGATGGCTTCTTTAAAATGAAGCAGGAGGGTCTGGCAGAGAGTATCTA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83277 GTATGGAAGTATCGATGGCTTCTTTAAAATGAAGCAGGAGGGTCTGGCAGAGAGTATCTA 83336

Qy        601 TGAAATAAGTTCCTCTGACCTTCACGCTTAATTTTCTGAATGGAGTGGAGCAAATTACTT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83337 TGAAATAAGTTCCTCTGACCTTCACGCTTAATTTTCTGAATGGAGTGGAGCAAATTACTT 83396

Qy        661 CAAGCTTCACTTAACTTGCATATGAAATGAACCGTACAAAAATACAAGAGTGTCAGGAGA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83397 CAAGCTTCACTTAACTTGCATATGAAATGAACCGTACAAAAATACAAGAGTGTCAGGAGA 83456

Qy        721 AAGTTATGCTCTGGTAAATATTTTGCAAAACAGATAAAAGATAATACTAGAGCTCTGTCC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83457 AAGTTATGCTCTGGTAAATATTTTGCAAAACAGATAAAAGATAATACTAGAGCTCTGTCC 83516

Qy        781 TCAAAGAGTTAAGCAGCTAATCTAAGGAGGTAAACTCTATGTCAGCAGGATGAACTGCTC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83517 TCAAAGAGTTAAGCAGCTAATCTAAGGAGGTAAACTCTATGTCAGCAGGATGAACTGCTC 83576

Qy        841 TTCCCTTTCCTCCTCAATAAATTGCAAATCATCTAGTCCAACATCTTTACCACCAGTGCC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83577 TTCCCTTTCCTCCTCAATAAATTGCAAATCATCTAGTCCAACATCTTTACCACCAGTGCC 83636

Qy        901 TGAGGCTCCAGAGGAGCCATTGCCTTCTCAAGGTCACATAGGTGGTGGGTGAGTTAGGAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83637 TGAGGCTCCAGAGGAGCCATTGCCTTCTCAAGGTCACATAGGTGGTGGGTGAGTTAGGAC 83696

Qy        961 CAAATCTAGAATTCCTGACTCCAGTAACTTCTGAAGTCATTTTGTTTTTTATTTTTATGG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83697 CAAATCTAGAATTCCTGACTCCAGTAACTTCTGAAGTCATTTTGTTTTTTATTTTTATGG 83756

Qy       1021 TTTTATTATAAGAATACTTGCTAAGCACACTTACCCCCTGCATTGATTAATAACTCTAGG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83757 TTTTATTATAAGAATACTTGCTAAGCACACTTACCCCCTGCATTGATTAATAACTCTAGG 83816

Qy       1081 ATCTCAGGTG 1090
              ||||||||||
Db      83817 ATCTCAGGTG 83826
```

```
RESULT 3
AC031977/c
LOCUS       AC031977              191699 bp    DNA     linear   HTG 12-APR-2001
DEFINITION  Homo sapiens chromosome 1 clone RP11-288O18, WORKING DRAFT
            SEQUENCE, 3 unordered pieces.
ACCESSION   AC031977
VERSION     AC031977.7  GI:13194952
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 191699)
  AUTHORS   Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
            Federspiel,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,
            Mao,J., Komp,C., Kottler,S., Lam,B., Marathe,R., Miranda,M.,
            Morehouse,A.J., Nguyen,M., Oefner,P., Palm,C.J., Ramirez,D.,
            Southwick,A.M., Webb,C., Wilhelmy,J., Yu,S. and Davis,R.W.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 191699)
  AUTHORS   Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
            Federspiel,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,
            Mao,J., Marathe,R., Morehouse,A.J., Oefner,P., Palm,C.J.,
            Ramirez,D., Wilhelmy,J., Yu,S. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-APR-2000) DNA Sequencing and Technology Center,
            Stanford University, 855 California Avenue, Palo Alto, CA 94304,
            USA
COMMENT     On Mar 4, 2001 this sequence version replaced gi:9665085.
            -------------- Genome Center
                Center: Stanford DNA Sequencing and Technology Development
            Center
                Center code: SDSTDC
                Web site: http://sequence-www.stanford.edu/group/human/
                Contact: hum-info@sequence.stanford.edu
            -------------- Project Information
                Center project name: 880
                Center clone name: RP11-288O18
            -------------- Summary Statistics
                Sequencing Vector: M13mp18; X02513; 100% of reads
                Sequencing Vector: plasmid; plasmid_accession; 0% of reads
                Chemistry: Dye-primer; 1% of reads
                Chemistry: Dye-terminator Big Dye; 99% of reads
                Assembly program: Phrap; version 0.990319
                Consensus quality: 190680 bases at least Q40
                Consensus quality: 191287 bases at least Q30
                Consensus quality: 191336 bases at least Q20
                Insert size: 195548; agarose-fp
                Insert size: 191499; sum-of-contigs
                Quality coverage:  7.9x in Q20 bases; agarose-fp
                Quality coverage:  8.1x in Q20 bases; sum-of-contigs.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    12646: contig of 12646 bp in length
            *   12647    12746: gap of unknown length
            *   12747    94961: contig of 82215 bp in length
            *   94962    95061: gap of unknown length
            *   95062   191699: contig of 96638 bp in length.
FEATURES             Location/Qualifiers
     source          1. .191699
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-288O18"
                     /clone_lib="RPCI human BAC library 11"
     misc_feature    1. .12646
                     /note="assembly_name:Contig5"
     misc_feature    12747. .94961
                     /note="assembly_name:Contig6
                     clone_end:T7"
     misc_feature    95062. .191699
                     /note="assembly_name:Contig7
                     clone_end:SP6"
ORIGIN

  Query Match             99.6%;  Score 1085.8;  DB 2;  Length 191699;
  Best Local Similarity   99.8%;  Pred. No. 2.8e-197;
  Matches 1087;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;
```

```
Qy          2 GCTTTCCGTGTGGTGGTGCAGTAACTTTCACTCTCTATGTTCTTATAAGCAAATGTTACA 61
              |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db       1880 GCTTTCCGTGTGGTGGTGCAGTAACTTTCACTCTCTATGTTCTTATAAGAAAATGTTACA 1821

Qy         62 ATGAGATATGAGTTTTAAAGCCAGATCTTCCTTATCTCTCTGCCCCATCTCTAGTTCTTG 121
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1820 ATGAGATATGAGTTTTAAAGCCAGATCTTCCTTATCTCTCTGCCCCATCTCTAGTTCTTG 1761

Qy        122 AAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTT 181
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1760 AAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTT 1701

Qy        182 GTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAA 241
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1700 GTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAA 1641

Qy        242 GACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAA 301
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1640 GACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAA 1581

Qy        302 ATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCAT 361
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1580 ATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCAT 1521

Qy        362 ATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGC 421
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1520 ATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGC 1461

Qy        422 TTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAA 481
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1460 TTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAA 1401

Qy        482 AACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGTCATTCAGGAG 541
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1400 AACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGTCATTCAGGAG 1341

Qy        542 TATGGAAGTATCGATGGCTTCTTTAAAATGAAGCAGGAGGGTCTGGCAGAGAGTATCTAT 601
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1340 TATGGAAGTATCGATGGCTTCTTTAAAATGAAGCAGGAGGGTCTGGCAGAGAGTATCTAT 1281

Qy        602 GAAATAAGTTCCTCTGACCTTCACGCTTAATTTTCTGAATGGAGTGGAGCAAATTACTTC 661
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1280 GAAATAAGTTCCTCTGACCTTCACGCTTAATTTTCTGAATGGAGTGGAGCAAATTACTTC 1221

Qy        662 AAGCTTCACTTAACTTGCATATGAAATGAACCGTACAAAAATACAAGAGTGTCAGGAGAA 721
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db       1220 AAGCTTCACTTAACTTGCATATGAAATGAACCGTACAAAAATACAAGAGTGTCAGGATAA 1161

Qy        722 AGTTATGCTCTGGTAAATATTTTGCAAAACAGATAAAAGATAATACTAGAGCTCTGTCCT 781
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1160 AGTTATGCTCTGGTAAATATTTTGCAAAACAGATAAAAGATAATACTAGAGCTCTGTCCT 1101

Qy        782 CAAAGAGTTAAGCAGCTAATCTAAGGAGGTAAACTCTATGTCAGCAGGATGAACTGCTCT 841
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1100 CAAAGAGTTAAGCAGCTAATCTAAGGAGGTAAACTCTATGTCAGCAGGATGAACTGCTCT 1041

Qy        842 TCCCTTTCCTCCTCAATAAATTGCAAATCATCTAGTCCAACATCTTTACCACCAGTGCCT 901
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1040 TCCCTTTCCTCCTCAATAAATTGCAAATCATCTAGTCCAACATCTTTACCACCAGTGCCT 981

Qy        902 GAGGCTCCAGAGGAGCCATTGCCTTCTCAAGGTCACATAGGTGGTGGGTGAGTTAGGACC 961
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        980 GAGGCTCCAGAGGAGCCATTGCCTTCTCAAGGTCACATAGGTGGTGGGTGAGTTAGGACC 921

Qy        962 AAATCTAGAATTCCTGACTCCAGTAACTTCTGAAGTCATTTTGTTTTTTATTTTTATGGT 1021
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        920 AAATCTAGAATTCCTGACTCCAGTAACTTCTGAAGTCATTTTGTTTTTTATTTTTATGGT 861

Qy       1022 TTTATTATAAGAATACTTGCTAAGCACACTTACCCCCTGCATTGATTAATAACTCTAGGA 1081
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        860 TTTATTATAAGAATACTTGCTAAGCACACTTACCCCCTGCATTGATTAATAACTCTAGGA 801

Qy       1082 TCTCAGGTG 1090
              |||||||||
Db        800 TCTCAGGTG 792
```

RESULT 4
BX901898

```
LOCUS       BX901898              248140 bp    DNA     linear   HTG 29-JAN-2004
DEFINITION  Danio rerio clone DKEY-148E3, *** SEQUENCING IN PROGRESS ***, 12
            unordered pieces.
ACCESSION   BX901898
VERSION     BX901898.2  GI:41392721
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 248140)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JAN-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jan 29, 2004 this sequence version replaced gi:40556530.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK148E3
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 243513 bases at least Q40
            Consensus quality: 244233 bases at least Q30
            Consensus quality: 245148 bases at least Q20
            Insert size: 247040; sum-of-contigs
            Insert size: 230827; 3.5% error; agarose-fp
            Quality coverage: 6.60x in Q20 bases; sum-of-contigs Quality
            coverage: 7.30x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 12 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       -1     3342: contig of 3342 bp in length
            *     3343     3442: gap of 100 bp
            *     3443   100473: contig of 97031 bp in length
            *   100474   100573: gap of 100 bp
            *   100574   113486: contig of 12913 bp in length
            *   113487   113586: gap of 100 bp
            *   113587   116700: contig of 3114 bp in length
            *   116701   116800: gap of 100 bp
            *   116801   159022: contig of 42222 bp in length
            *   159023   159122: gap of 100 bp
            *   159123   199472: contig of 40350 bp in length
            *   199473   199572: gap of 100 bp
            *   199573   204017: contig of 4445 bp in length
            *   204018   204117: gap of 100 bp
            *   204118   223587: contig of 19470 bp in length
            *   223588   223687: gap of 100 bp
            *   223688   231390: contig of 7703 bp in length
            *   231391   231490: gap of 100 bp
            *   231491   233892: contig of 2402 bp in length
            *   233893   233992: gap of 100 bp
            *   233993   237252: contig of 3260 bp in length
            *   237253   237352: gap of 100 bp
```

```
          *   237353   248140: contig of 10788 bp in length.
FEATURES             Location/Qualifiers
     source          1. .248140
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-148E3"
                     /clone_lib="DanioKey"
     misc_feature    1. .3342
                     /note="assembly_fragment:00099
                     fragment_chain:1"
     misc_feature    3443. .100473
                     /note="assembly_fragment:01752
                     fragment_chain:1"
     misc_feature    100574. .113486
                     /note="assembly_fragment:00427
                     fragment_chain:1"
     misc_feature    113587. .116700
                     /note="assembly_fragment:00184
                     fragment_chain:1"
     misc_feature    116801. .159022
                     /note="assembly_fragment:01226
                     fragment_chain:1"
     misc_feature    159123. .199472
                     /note="assembly_fragment:00736
                     fragment_chain:1"
     misc_feature    199573. .204017
                     /note="assembly_fragment:00124
                     fragment_chain:1"
     misc_feature    204118. .223587
                     /note="assembly_fragment:00558
                     fragment_chain:1"
     misc_feature    223688. .231390
                     /note="assembly_fragment:00226
                     fragment_chain:1"
     misc_feature    231491. .233892
                     /note="assembly_fragment:00059"
     misc_feature    233993. .237252
                     /note="assembly_fragment:00152"
     misc_feature    237353. .248140
                     /note="assembly_fragment:00314.0"
ORIGIN

  Query Match             6.0%;  Score 65;  DB 2;  Length 248140;
  Best Local Similarity   47.9%;  Pred. No. 0.0072;
  Matches  182;  Conservative    0;  Mismatches  198;  Indels    0;  Gaps    0;

Qy        146 AGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACA 205
              | | ||||  ||   || | ||   |  |||  | | || |         | || |
Db     236808 ATATATATATATATATATATATATATATATATATATATATATATATATTTTTTTATTTTT 236867

Qy        206 TTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTC 265
              | |||| || || ||      ||  |||||||| |       || | | ||| |  |||
Db     236868 CTTTTTTTTTTATTTATAATTTTTTTTTTTTTATATATATATATATATATATATATAT 236927

Qy        266 ATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATA 325
              |||| |  ||   | |      |   |   ||  |||  | | | | ||| |  | | |
Db     236928 ATTTTTAATTATATATTATTTTTATATTTATAAATAATTAAATAAAAAAAAAAAAAAAAA 236987

Qy        326 TTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTC 385
              ||   |    |||| ||||   | ||  |    ||  | |  |  ||||  |    |
Db     236988 ATAAAAAAAAAAAAAATAAAATAATAAAAAAATAATTAAAAAAAAAAATAATTATTTTTTT 237047

Qy        386 AAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAA 445
              |   |||| | | ||   |||   | |  |     ||| |||   | |   || | |
Db     237048 TATTTTAAATATTTAAAAAAAAAAAAAAAAAATAATATTTTTTATATAAATAATAAAAAAT 237107

Qy        446 ATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTA 505
              |||  | | | ||   | |  ||| ||  | | |||| |   |||||| | |  | ||
Db     237108 TTTAATATATAATTATATATTTAAAAAAAATAAAAAAAAANNNTTTTTTTTATTTTTTT 237167

Qy        506 AAACATTTACTTTCCAAATA 525
                |  |||  |    ||| |
```

```
Db     237168 TTATTTTTTATAAAAAAAAA 237187


RESULT 5
CR382400/c
LOCUS       CR382400               348034 bp    DNA     linear   INV 26-AUG-2004
DEFINITION  Plasmodium falciparum chromosome 6, complete sequence; segment 3/5.
ACCESSION   CR382400 AL844505
VERSION     CR382400.1  GI:46361120
KEYWORDS    HTG.
SOURCE      Plasmodium falciparum 3D7
  ORGANISM  Plasmodium falciparum 3D7
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 348034)
  AUTHORS   Hall,N., Pain,A., Berriman,M., Churcher,C., Harris,B., Harris,D.,
            Mungall,K., Bowman,S., Atkin,R., Baker,S., Barron,A., Brooks,K.,
            Buckee,C.O., Burrows,C., Cherevach,I., Chillingworth,C.,
            Chillingworth,T., Christodoulou,Z., Clark,L., Clark,R., Corton,C.,
            Cronin,A., Davies,R., Davis,P., Dear,P., Dearden,F., Doggett,J.,
            Feltwell,T., Goble,A., Goodhead,I., Gwilliam,R., Hamlin,N.,
            Hance,Z., Harper,D., Hauser,H., Hornsby,T., Holroyd,S.,
            Horrocks,P., Humphray,S., Jagels,K., James,K.D., Johnson,D.,
            Kerhornou,A., Knights,A., Konfortov,B., Kyes,S., Larke,N.,
            Lawson,D., Lennard,N., Line,A., Maddison,M., McLean,J., Mooney,P.,
            Moule,S., Murphy,L., Oliver,K., Ormond,D., Price,C., Quail,M.A.,
            Rabbinowitsch,E., Rajandream,M.A., Rutter,S., Rutherford,K.M.,
            Sanders,M., Simmonds,M., Seeger,K., Sharp,S., Smith,R., Squares,R.,
            Squares,S., Stevens,K., Taylor,K., Tivey,A., Unwin,L.,
            Whitehead,S., Woodward,J., Sulston,J.E., Craig,A., Newbold,C. and
            Barrell,B.G.
  TITLE     Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13
  JOURNAL   Nature 419 (6906), 527-531 (2002)
  MEDLINE   22255708
   PUBMED   12368867
REFERENCE   2  (bases 1 to 348034)
  AUTHORS   Cherevach,I., Davis,P., Goodhead,I., Stevens,K., Mungall,K.,
            Berriman,M., Pain,A., Hall,N., Atkin,R., Chillingworth,C.,
            Doggett,J., Ormond,D., Sanders,M., Hayes,R., Hall,S., Quail,M. and
            Barrell,B.G.
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 348034)
  AUTHORS   Cherevach,I., Davis,P., Goodhead,I., Stevens,K., Mungall,K.,
            Berriman,M., Pain,A., Hall,N., Atkin,R., Chillingworth,C.,
            Doggett,J., Ormond,D., Sanders,M., Hayes,R., Hall,S., Quail,M. and
            Barrell,B.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-SEP-2002) P.falciparum Genome Sequencing Consortium,
            The Welcome Trust Sanger Institute, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA, UK
REFERENCE   4  (bases 1 to 348034)
  AUTHORS   Cherevach,I., Davis,P., Goodhead,I., Stevens,K., Mungall,K.,
            Berry,A.E., Berriman,M., RA Pain,A., Hall,N., Atkin,R.,
            Chillingworth,C., Doggett,J., Ormond,D., Sanders,M., Hayes,R.,
            Hall,S., Quail,M. and Barrell,B.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-MAR-2004) P.falciparum Genome Sequencing Consortium,
            The Welcome Trust Sanger Institute, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA, UK
COMMENT     For more information about this sequence or the Malaria Project,
            see http://www.sanger.ac.uk/Projects/P_falciparum.
FEATURES             Location/Qualifiers
     source          1. .348034
                     /organism="Plasmodium falciparum 3D7"
                     /mol_type="genomic DNA"
                     /isolate="3D7"
                     /db_xref="taxon:36329"
                     /chromosome="6"
     gene            1757. .5722
                     /locus_tag="PFF0820w"
     CDS             1757. .5722
                     /locus_tag="PFF0820w"
                     /note="MAL6P1.257"
                     /codon_start=1
```

/product="hypothetical protein"
/protein_id="CAG24985.1"
/db_xref="GI:46361121"
/db_xref="UniProt/TrEMBL:Q6LF47"
/translation="MFSKRKIIKKNIKKSFEDNDEEEQIDKLNAHTNYKNEKEFIDTH
DDIYQLKFKETNQDNNATFSASQNVKNNYLNMKEIKEKEEEKKNVQNGQNEYGLTCTF
IQSNDEKKKEEKIIKKKSNKMSCIERKKNEEEKRVNKMNTSFHIYSEDDNDSYITLKK
KKKKKSSSVKMYKEEKNNKIDNIHNNNYNNSDNIFKENQSSDNETFNNAYNNLSIYNS
TQGEKHIYNKDHMKIKKDKNLSYNNHNIKKNEIYLSEDNHTNYGDHTSDNDEENDNKD
IIYDDTKFYVNDSTSYKLKQRNISSYEVGNDYSNNDNENVVNLGSDNSGVVSKGLKDN
NQIFNNSEDVFIIDDKCDNYQSININLQDEYDDKDNDEEKKLIERIKLKKNILRKKKE
INNINYNNNDNNLFLDDDYDDNFLFPMEKKNFFFEIKKNKNVTYEDDLDVDDMYNYET
INEMKNRILIKKNRNEEIKNLYEERNIEENVVDTNDLNIDNNYEDYDLYEDEKINQIV
DNKMLVELKKQNDFLYDSKKKKKKKEEEEEENKRERVKDKDLCFKEDNINAYIKSNNN
EPKEEEHVKNKEYSDDDLFSLSREKDKELYSLDNNNYYNDKNKLLFEKIQNAYKNKGV
CNLEIIKMYEHIQNLFSEYKNKIQESNEIKELENSNQVEYNKYEIICKKGKKEIIICR
VFLQFLQILINLICEKYEPVEDALNGLYKLESTFHLIYHNLKLYIYKEYYEKYKLTFI
NDYIFNSKYYKNLKEKKKKIQLQYLIDNNIIQNNGLNDNVTQINEYNILNEICDVKNM
NNSEFYYMFDGFSSNYSTDTLSSTSLLLENDLNGDNKNFIQRKENKDILKSDALKSDA
LKSDELKSDALKNDMSKKNLDKQKYFQTLKMKDLKKKFLFSINNIFKNVNMYFFNFKN
VIKYFGLLKLYNYHFYVNNKCSSYFDDVLFFFVKCELLYWDPLYQFFMSKKKKKRILN
YFDEDINSSSSMKKNESNKNNLNILDSNKMDNNDKHFNIDEDNKKFFSLYSSPSPQYD
NDEKNIYNNTNERTIEYDMDMSGSSSSSSNFSLSHTSSCNENNLNNENILDREEENVKK
ICSNIFKNKLFKKKKYKLRKTFFHNPSIKSFEWYKFMDELILIYDNIEEEKDILKKLY
EHIFNKKIYELINVWNPLSLKQSFHLSRIINDFILYNDNINELRDKIKDKIQTYITTF
LQCYKNITSQKKKNIFLMRCLKFLKSMKNILQLLCNEELYDIVKNVFNNYVLPNCDDG
NKLHNLIVLKIIQVIHSLDNIPKEDSFFHKTTEIMSKISDCLKTGYFDMLKVENGCMI
K"

gene complement(6540. .7217)
/locus_tag="PFF0825c"

CDS complement(6540. .7217)
/locus_tag="PFF0825c"
/note="MAL6P1.256"
/codon_start=1
/product="mitochondrial import receptor subunit tom40"
/protein_id="CAG24986.1"
/db_xref="GI:46361122"
/db_xref="UniProt/TrEMBL:Q6LF46"
/translation="MISRINIDGSVNGRFCKKINNNIDCKLNFNTYAKNDTRNMYEMS
LEVNKPLYTYNFKSIWQGAWIFNTSYTQLLTKKLQAGVDLTYIASNCASIGSFGLRYN
HKNNVLTMQIVRQPNFKSPEFMLNQTHLYKIQYAKKISDRLSLGTELEITPQTKESAM
RLGWDYSFRHAKVQGSIDTSGKISVFTQDYSGFGVSGYIDYLNNDYKFGFMMHISPSQ
EQPQTAA"

gene join(12564. .15172,15385. .15927,16064. .16232,16353. .16553,
16658. .16699,16811. .16869,16994. .17081)
/locus_tag="PFF0830w"

CDS join(12564. .15172,15385. .15927,16064. .16232,16353. .16553,
16658. .16699,16811. .16869,16994. .17081)
/locus_tag="PFF0830w"
/note="Strong Pfam hit (2e-22;codon 25-837) to Adaptin N
terminal region
MAL6P1.255; putative alpha adaptin-like protein"
/codon_start=1
/product="alpha adaptin-like protein, putative"
/protein_id="CAG24987.1"
/db_xref="GI:46361123"
/db_xref="UniProt/TrEMBL:Q6LF45"
/translation="MIKHSIKGLYCFIDEVRNCRCKEDEEKKVLQEIIKIKKKFNEKN
ITNYKRKKYIWKLIYCHILGYGISLSLYLDIIKLISSNNFSDKYCGYTALSLLVDENNE
MLNIMVSTIKADIKNNDEKINFLAFHFASHKINGLLIENLYEDILHIVTSNYIYKPNI
RKKAFLCLANIYKKRHDLLLRGKNDFEIFKFLDQNLNEINMFNIFAYLNLLYVIILIF
QKYESMYYYIKKEEKKKNSLEEKNDDILKREELKFDNSSSSSGSSNHSRRISLHIKNE
LKKKRSKSNIYLNEQNFLYNDLKKVDKILEIESECTFNIEEIDFYEIKKYINKYIHFI
LNVLYLILDENLKIDEGFYYSHVRYPFLLIKCLQMIQLYDVHNLSQTTINNINDIMYK
IISKPYRKLQGRLGNENNPLNNNIPNNNRHTIANHFNSTNNNNNNNNNNNNNDYNNNN
KSSNNNNSSNIFKKSFSKNSIYSHSKKNKDILRNEKSAIYIEYGIIYECSNLYNFLDD
KIEDRNRDIIICLITSFDTKKSNINYVILNSLSKLKMNIKIYNMLETHIMHLINLLNS
DDITIKLQTFNILFNMCNYNNWKLLINVFLHHLPFIDPYIQNEVIIKISILAENFSPD
LSWYIDVIFKMIEITHKSIFQEVWFRLVQIITGFVEGDGKNSSKLSKGENNALNNINN
NSNESNNSNNNSNNNNNNNNNNNNNNNNNNNNNNNNNKNKDDQKVQTYAAMKCYRYLSDHF
TRIELLVDLCTYIIGSFGYLIKDKVPMKEQLKILEKYFAFASSNTKCVILMAIMKMVF
YDNKLMNSVKKILNTCINHSDLELQTRACEFLNLCNMNNMSMLNYLLKNMPLYNMKKV
HENFLIKRLLETNKHAVIDFNEKDNSLKFDKKHDNKKSGKHNKHNDEDYNDTNSDYSH
NKVSSHSSCNSSSSEDEDEDEDEDEEEENEEKYSKKRSRNLSDSSHNSKKSSSSVSSH
KSKNSSDNSDDNEKLFKILCLQKANNINDLWFNACLLDKCLFFKNNMVSILMKQKYQE
NKGILIFYIKNISKAALNLSSINIEECEQINIKEQKIVNNEKIDSDGMYVHKIILTVS
DIFYNIPSIYFNIMTPNAANASFSSKLPLLLTRFIKENKMNETTFKNYWKALTQVHNE
DVIMGVPKFSKKYFLNYLINAFNFYILKIGMLICASGYIHFPDLNMENKILILAKIRH
EVKGCQISVVSSIKHLNNYLNKIFEIYLIKNN"

gene join(19555. .19659,19825. .20227,20451. .21203,21343. .21422)
/locus_tag="PFF0835w"

CDS join(19555. .19659,19825. .20227,20451. .21203,21343. .21422)
/locus_tag="PFF0835w"
/note="Although it is moderately similar to yeast nuclear
segregation protein bfr1, the evidence is not strong
enough to call this protein a bfr1 homologue
MAL6P1.254"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAG24988.1"
/db_xref="GI:46361124"
/db_xref="UniProt/TrEMBL:Q6LF44"
/translation="MTEAENIKIEKPDFDAYNEKLGSISQSIDEIKKKIDNLQKEIKV
ANKGKEEYNKKKKDIVTKIDCFQSDIDKLENERRSILDDIEKKQKHKKELKVNAQNMK
KQIGFENEEDIEKKIREIENKLMTSTISIKEEKLLINQIQALNKNKPLLSSYSKIENE
ASKYDDETIVPLKSKMDSIREKINKLRNEKKMERNKLKELQNSYQEKNNKLNELNNLR
DNYSKKMNHYFMERRSITVEMEEKKQQYRSFKLNQLQAKQQKIKEDRERKNLELEKQS
LEKKLEDIDVLPYREELALIENMLAYLKKLQEEFKLEESKKQQNAAKKINGEVETNEN
KDDNKTKNENNNKKAKSKKDKQKTFKLDMNILCYFVTAGINPPVSFDEIDSCIQKLHE
KKDMYEQKRDESVKNVETRREDLTNKIKEIDEKLTTFKSGEHHNKGTKANKVKA"

gene join(23040. .23150,23307. .23486,23689. .23817)
/locus_tag="PFF0840w"

CDS join(23040. .23150,23307. .23486,23689. .23817)
/locus_tag="PFF0840w"
/note="MAL6P1.253"
/codon_start=1
/product="hypothetical protein, conserved"
/protein_id="CAG24989.1"
/db_xref="GI:46361125"
/db_xref="UniProt/TrEMBL:Q6LF43"
/translation="MDSSYRKLRTVELQKRKVTQSKNAPSSIERKLGNIYAPKDFDED
STIVVCNFPQKTTISECKNFMQWIGPVIKVEKIPSFMQENNYLVVFCNPYFAKAALEI
PLFYENKTKLFTRAVEKRETLWQNINDMITTNMNFFS"

gene complement(join(25622. .26665,27855. .33995))
/gene="VAR"
/locus_tag="PFF0845c"

CDS complement(join(25622. .26665,27855. .33995))
/gene="VAR"
/locus_tag="PFF0845c"
/note="MAL6P1.252; pfEMP1"
/codon_start=1
/product="erythrocyte membrane protein 1 (PfEMP1)"
/protein_id="CAG24990.1"
/db_xref="GI:46361126"
/db_xref="UniProt/TrEMBL:Q6LF42"
/translation="MGPAPRTATTTYSSAKDLLEDIGESVQKEAKKQALGRSESVLHG
LLSNATIKGVKNKATKPIQLEYEYHTNVTGGFDKNNPCANRLDVRFSDIYGGQCTDNK
INGNDDETGGTCAPLRRLFLCDQHLSHMKEGNINNTDNLLLEVSLAAKYEGDSIINNY
PDNRDKKEGICTALARSFADIGDIIRGKDLFLGYTKKDEKEKEKVQKNLKRIFNEIYK
KMQDPAKSHYSGDSSDFYKLREDWWALNRKEVWKAITCKAKNDAEYFRKKDSDGKHCS
VQNCKCVDGDPPTNLDYVPQHLRWFDEWSEEFCRKRKDQLKISLEKCRGKNKDKYCSL
NGCNCKTTVRGKKKFDYQQECNDCLVACDPFVHWIDNEKLEFLKQKKKYKNEIKEREP

Query Match 5.8%; Score 63; DB 3; Length 348034;
Best Local Similarity 45.3%; Pred. No. 0.016;
Matches 228; Conservative 0; Mismatches 275; Indels 0; Gaps 0;

Qy 23 TAACTTTCACTCTCTATGTTCTTATAAGCAAATGTTACAATGAGATATGAGTTTTAAAGC 82
Db 22774 TATATATATATATATATATATATATATATATATATTCCATTTACATAATTATATAATATC 22715

Qy 83 CAGATCTTCCTTATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGG 142
Db 22714 GAATTTTTTTTTTTTATTCTTAAGATATATTCATTTTTATATTTATTCACTTGAAAATAT 22655

Qy 143 TTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAG 202

```
Db      22654 TATACATATATATATGTACTTGTGAAAAGTTATAATTATATACGTGTAAATTTTTTATGA 22595

Qy        203 ACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTA 262

Db      22594 AAAATAAAAAATTGTCTTTCAATATACAAAGCTTTTCAAATATATATATATATATATATA 22535

Qy        263 TTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCAT 322

Db      22534 TTTATATATAATATTGAAGAAAAAAAAAAATAATAAATAACCCTTTTTAACATATAAAAAT 22475

Qy        323 ATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCA 382

Db      22474 TTTATGAGACTCTTTTAAAAAAAAAATTTTTGAATTCGTATATTTAATAATTTTATTTCA 22415

Qy        383 GTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATA 442

Db      22414 GTTTAAATAAAGATGTGCACCAACACATAAAATATATATATATATATATATATATATA 22355

Qy        443 GAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTA 502

Db      22354 TATATATACATATATACAATTCTTTCATTCATTAAAAAAAATAAATTCTTTTTTATTATA 22295

Qy        503 TTAAAACATTTACTTTCCAAATA 525

Db      22294 TTAAATTAAATAGACTTGAAACA 22272


RESULT 6
AL391538
LOCUS       AL391538              131459 bp    DNA     linear   PRI 16-MAY-2001
DEFINITION  Human DNA sequence from clone RP11-334H12 on chromosome 6, complete
            sequence.
ACCESSION   AL391538
VERSION     AL391538.11  GI:14141384
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 131459)
  AUTHORS   Martin,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On May 17, 2001 this sequence version replaced gi:13872373.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            RP11-334H12 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-334H12 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true left end of clone RP11-334H12 is at 1 in this sequence.
            The true left end of clone RP1-57E3 is at 131360 in this sequence.
            The true right end of clone RP11-193P4 is at 87630 in this
            sequence.
FEATURES             Location/Qualifiers
     source          1. .131459
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP11-334H12"
                     /clone_lib="RPCI-11.2"
     repeat_region   413. .469
                     /note="L1MA6 repeat: matches 6233. .6291 of consensus"
     repeat_region   1530. .1587
                     /note="29 copies 2 mer aa 75% conserved"
     repeat_region   2469. .2512
                     /note="22 copies 2 mer tg 95% conserved"
     repeat_region   3034. .3429
                     /note="MSTD repeat: matches 1. .394 of consensus"
     repeat_region   3433. .4372
                     /note="L2 repeat: matches 1534. .2497 of consensus"
     repeat_region   6544. .6840
                     /note="AluSx repeat: matches 1. .309 of consensus"
     repeat_region   7836. .7909
                     /note="MLT1I repeat: matches 308. .390 of consensus"
     repeat_region   8148. .9562
                     /note="L1PA16 repeat: matches 4706. .6157 of consensus"
     repeat_region   10096. .11255
                     /note="L1PA5 repeat: matches 5153. .6143 of consensus"
     repeat_region   11247. .11331
                     /note="L1MA8 repeat: matches 6013. .6097 of consensus"
     repeat_region   12374. .12629
                     /note="AluSx repeat: matches 49. .297 of consensus"
     repeat_region   16692. .16787
                     /note="L1M1 repeat: matches 5691. .5786 of consensus"
     repeat_region   17050. .17155
                     /note="AluJb repeat: matches 134. .232 of consensus"
     repeat_region   17627. .17784
                     /note="MIR repeat: matches 61. .231 of consensus"
     repeat_region   19053. .19088
                     /note="18 copies 2 mer ac 94% conserved"
     repeat_region   20718. .20811
                     /note="47 copies 2 mer aa 70% conserved"
     repeat_region   23097. .23388
                     /note="AluSx repeat: matches 1. .288 of consensus"
     repeat_region   23644. .23715
                     /note="L2 repeat: matches 2678. .2750 of consensus"
     repeat_region   23726. .24090
                     /note="MLT1I repeat: matches 27. .390 of consensus"
     repeat_region   24858. .25289
                     /note="L1M4c repeat: matches 1284. .1711 of consensus"
     repeat_region   25577. .25730
                     /note="MER86 repeat: matches 2. .163 of consensus"
     repeat_region   26145. .26539
                     /note="MER95 repeat: matches 9. .430 of consensus"
     repeat_region   26883. .27067
                     /note="MIR repeat: matches 67. .257 of consensus"
     repeat_region   28242. .29600
                     /note="L1MA2 repeat: matches -1389. .-42 of consensus"
     repeat_region   29601. .29895
                     /note="AluY repeat: matches 1. .296 of consensus"
     repeat_region   29896. .33311
                     /note="L1MA2 repeat: matches -42. .3427 of consensus"
     repeat_region   33529. .34003
                     /note="LTR28 repeat: matches 373. .941 of consensus"
     repeat_region   34022. .34164
                     /note="LTR28 repeat: matches 275. .414 of consensus"
     repeat_region   34197. .34297
                     /note="LTR28 repeat: matches 1. .110 of consensus"
```

```
repeat_region   34299. .34547
                /note="L1MA2 repeat: matches 3422. .3669 of consensus"
repeat_region   34675. .37318
                /note="L1MA2 repeat: matches 3669. .6308 of consensus"
repeat_region   37342. .37475
                /note="L1ME1 repeat: matches 5470. .5605 of consensus"
repeat_region   37494. .38480
                /note="L1MCb repeat: matches 102. .1028 of consensus"
repeat_region   38487. .38911
                /note="L1ME1 repeat: matches 5722. .6163 of consensus"
repeat_region   40273. .40461
                /note="MER8 repeat: matches 1. .180 of consensus"
repeat_region   40495. .40554
                /note="TIGGER2 repeat: matches 5. .64 of consensus"
repeat_region   40949. .41199
                /note="MIR repeat: matches 1. .248 of consensus"
repeat_region   43783. .44083
                /note="AluSx repeat: matches 1. .311 of consensus"
repeat_region   45424. .45681
                /note="AluSg repeat: matches 3. .259 of consensus"
repeat_region   45967. .46060
                /note="MLT2D repeat: matches 1. .94 of consensus"
repeat_region   46061. .46426
                /note="MLT1A1 repeat: matches 1. .365 of consensus"
repeat_region   46427. .46890
                /note="MLT2D repeat: matches 94. .553 of consensus"
repeat_region   48332. .48971
                /note="LTR13 repeat: matches 311. .945 of consensus"
repeat_region   48972. .49337
                /note="LTR13 repeat: matches 2. .364 of consensus"
repeat_region   49354. .49490
                /note="L1P4 repeat: matches 5497. .5636 of consensus"
repeat_region   50194. .50675
                /note="L1 repeat: matches 4175. .4668 of consensus"
repeat_region   50993. .51612
                /note="L1M4 repeat: matches 4447. .5061 of consensus"
repeat_region   51640. .51925
                /note="L1ME2 repeat: matches 5779. .6066 of consensus"
repeat_region   53644. .53751
                /note="L2 repeat: matches 2607. .2710 of consensus"
repeat_region   56448. .56576
                /note="MER5A repeat: matches 64. .184 of consensus"
repeat_region   56828. .56976
                /note="HAL1 repeat: matches 377. .537 of consensus"
repeat_region   57045. .57256
                /note="106 copies 2 mer tt 60% conserved"
repeat_region   61093. .61419
                /note="AluSx repeat: matches 1. .308 of consensus"
repeat_region   62068. .62290
                /note="L1MA8 repeat: matches 6050. .6289 of consensus"
repeat_region   63592. .63653
                /note="31 copies 2 mer aa 67% conserved"
repeat_region   63747. .64054
                /note="AluSx repeat: matches 4. .300 of consensus"
repeat_region   64246. .65557
                /note="HSMAR2 repeat: matches 6. .1300 of consensus"
repeat_region   65562. .65692
                /note="L1MA6 repeat: matches 6098. .6231 of consensus"
repeat_region   65693. .65757
                /note="MADE1 repeat: matches 1. .77 of consensus"
repeat_region   65758. .66770
                /note="L1MA6 repeat: matches 5118. .6098 of consensus"
repeat_region   67555. .67688
                /note="AluJb repeat: matches 1. .134 of consensus"
repeat_region   70009. .70319
                /note="AluY repeat: matches 1. .311 of consensus"
repeat_region   71532. .71828
                /note="AluY repeat: matches 1. .298 of consensus"
repeat_region   72226. .72529
                /note="AluJo repeat: matches 1. .298 of consensus"
repeat_region   73182. .73511
                /note="L1MC4 repeat: matches 7362. .7692 of consensus"
repeat_region   73518. .73987
                /note="LTR12 repeat: matches 38. .437 of consensus"
repeat_region   74302. .74398
                /note="L1MC4 repeat: matches 7861. .7960 of consensus"
repeat_region   74937. .75237
                /note="AluY repeat: matches 3. .306 of consensus"
repeat_region   75521. .75824
                /note="AluSp repeat: matches 3. .298 of consensus"
repeat_region   76247. .76542
                /note="AluSq repeat: matches 1. .296 of consensus"
repeat_region   76685. .77130
                /note="L1MB8 repeat: matches 5701. .6173 of consensus"

  Query Match            5.6%;  Score 60.6;  DB 9;  Length 131459;
  Best Local Similarity  47.5%;  Pred. No. 0.055;
  Matches  180;  Conservative    0;  Mismatches  199;  Indels    0;  Gaps    0;

Qy        157 TCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTA 216
              | |||| | || |  |||   || ||||| | ||   |  || ||     | | |  |||
Db      83049 TTATTATATATTATATAAGTATTATGTAGTATATACAAATTTTAATTATATTTATATTTA 83108

Qy        217 ATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTG 276
              |||   |  || ||  ||| | || |  | |   |||   |  |||  |||| | | |
Db      83109 ATAATTTATATTTATAATTTAATAATATATAATTAACAGGTAAGTATATATTTATATATA 83168

Qy        277 GTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTG 336
              |   || | |  |||| | | ||  ||  |   | | |   |  |   |  | ||
Db      83169 AAATAAACATATATTTTATATAAATATAAATATACTTATATAAATTTATATGTATATATG 83228

Qy        337 AAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAA 396
              ||| ||||| | |      | | | ||| |   ||  ||      | |   | |    |||
Db      83229 AAATTTAAATATATTTATATTTCCTATAAATATAAATATATAAGTATATATATAATTAAA 83288

Qy        397 GTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAA 456
              ||  |  ||    ||  ||  |   | |     |    ||||| | | ||    ||
Db      83289 GTATATATATTTTAATTTTATATATAATATATAAATATATTAAATATATAATATATAAAT 83348

Qy        457 TTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACT 516
              | |    |  | |  || | |  | |   ||| | | || ||  ||  ||| || |  |
Db      83349 ATATTAAATATATAATATATAAATATATTAAATATATAATATATAATATAAATATATTAT 83408

Qy        517 TTCCAAATATTCTGTCATT 535
              |  | ||||  | | |||
Db      83409 ATTAATATATAATATAATT 83427


RESULT 7
AB073376/c
LOCUS       AB073376               216734 bp    DNA     linear   VRT 02-MAY-2002
DEFINITION  Oryzias latipes DNA, MHC Class I Region, complete cds, section 1/2.
ACCESSION   AB073376 BA000027
VERSION     AB073376.1  GI:18157518
KEYWORDS    .
SOURCE      Oryzias latipes (Japanese medaka)
  ORGANISM  Oryzias latipes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
            Beloniformes; Adrianichthyidae; Oryziinae; Oryzias.
REFERENCE   1
  AUTHORS   Matsuo,M.Y., Asakawa,S., Shimizu,N., Kimura,H. and Nonaka,M.
  TITLE     Nucleotide sequence of the MHC class I genomic region of a teleost,
            the medaka (Oryzias latipes)
  JOURNAL   Immunogenetics 53 (10-11), 930-940 (2002)
  MEDLINE   21850510
   PUBMED   11862394
REFERENCE   2  (bases 1 to 216734)
  AUTHORS   Matsuo,M.Y., Asakawa,S., Shimizu,N., Kimura,H. and Nonaka,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-OCT-2001) Megumi Y Matsuo, University of Tokyo,
            Department of Biological Sciences, Graduate School of Science;
            7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
            (E-mail:meg@biol.s.u-tokyo.ac.jp, Tel:81-3-5841-4064,
            Fax:81-3-5841-4064)
```

```
FEATURES             Location/Qualifiers
     source          1. .216734
                     /organism="Oryzias latipes"
                     /mol_type="genomic DNA"
                     /strain="Hd-rR"
                     /db_xref="taxon:8090"
                     /clone="188P8"
                     /clone_lib="medaka-fish sperm DNA BAC library"
                     /note="Linkage Group 11"
     gene            complement(join(314. .459,819. .887))
                     /gene="GNBP"
     CDS             complement(join(<314. .459,819. .887))
                     /gene="GNBP"
                     /note="supported by GENSCAN and homologous to guanine
                     nucleotide binding protein (gi:121013)"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="BAB83834.1"
                     /db_xref="GI:18157519"
                     /translation="MAAEKAEMDALKKECDALRTQIEAARKAANDGSMSAAAGGVASV
                     GRVQLKLRKNLKGHLAKIYAMHWSADS"
     gene            complement(join(6794. .6953,9685. .9800,11390. .11574,
                     11734. .11895,14226. .14313))
                     /gene="CLEP"
     CDS             complement(join(6794. .6953,9685. .9800,11390. .11574,
                     11734. .11895,14226. .14313))
                     /gene="CLEP"
                     /note="supported by GENSCAN and partially homologous to
                     the c-type lectin domain"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="BAB83835.1"
                     /db_xref="GI:18157520"
                     /translation="MSNSVAFSKLVDDEGNRDTLRHRADVQHSDKNRSQRGDEGTTLM
                     QTDVSALTALISKLQQENKQLQEEKRALLAKLDLKLTTKAPPVVSLMPSMTGVPVECP
                     TNWHLFNSSCYFISTQMKPWRDSQTYCQRQGAHLAIIHTAEEQTFLWDLLPRAHWNAY
                     WFGISDRQKEDEWKWVDGTSVEKSFWEEGEPNNHINEDCGYIVKTQVLERVAIRSWYD
                     APCEMSIKFICEKEMKTQ"
     LTR             17152. .17844
                     /note="LTR of LReO_3"
     gene            17886. .22382
                     /gene="LReO_3"
     CDS             17886. .22382
                     /gene="LReO_3"
                     /note="LTR retrotransposable element
                     partially supported by GENSCAN"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="BAB83836.1"
                     /db_xref="GI:18157521"
                     /translation="MTRGKKTETPDLPLDDDKATSSHQEGAAAPDEPDKLEELTTLVK
                     TLVQSQASRDKLVEKESARQDQRWKNLQHQFQMLQTQVKELKDDFETRAEQEDDDDDD
                     VGEGADDVQLAMPTSCGPVPPTQRGPILQKQPKILPLSPEDDIEHYLMTFERIATVCR
                     WPKEEWAIQLIPLLTGKARSAYVLMDFDDSEDYEKVKEVILAKYEITAETYRKRFRAL
                     DINPGETPRELYVRLRELFNKWVKPEACTVKEISEKLILEQFLRMMNPEMEIWIRERD
                     PQSAEEASRLAEVFLSARSGRGRPNFGRDSYFSGRSKSYAGERGGQTHSRMHNSPRQF
                     SAKAKESRRSFSSSKQNVRCYQCNEIGHTQYTCPATRQNPPSLLCTVPRPSQPVQVQK
                     EVVTVPVLINGQRKEALLDSGCFKSVVLESLVPRDLWNEATATIGCVHGDEKAYRTAE
                     VYLTIDGQTYLMPVVLVPELPFSVILGRDVPLLFDLIQQAHAKMDGGVCETSVQAQRK
                     PVGSDHQHAVSLCNIVTTRANAARNLLEELPFHNTELEISPVKVRKSRAQRRREKFLA
                     TVETQCDDEKQPSHVLDFQVPADIAALQRADPRLKVWFDKVSEIEGQKTGRVECLTDK
                     IYLIKNGILYQRHGTVEALALPQCFSKKVMGIGHSIPWAGHLAFQKSLNRIANRFVWP
                     AMYTQLKEFCASCEICQLTAPQGVARAPLQPLPLIETPFDRIGMDIVGPLEKSSSGHK
                     YILVICDYATRYPEAFPLRSVKARQIANCLIQLFSRVGIPKEILTDCGTNFLSKLLKQ
                     VYSLLGVKGIKTTPYHPQTDGLVERFNQTLKTMLRKFVSETGADWDQWLPYLLFAYRE
                     VPQASTGFSPFELLYGRQVRGPLDLLKDYWEKPVTDKDNVVSYVLKMRERLERMSTLA
                     QEHMRSAQAQQKTWYDKKARDRIFQVGQQVLLLLPTSDNKLLARWQGPYSITKRLGEV
                     TYELYMPEKLKKHQRFHVNLLKEFQLPASNKQECNQTLFIRVADDDEEEREKLFPHEE
                     MASSPVDVSHLQPDRQREVGRLLDAELFKETPGFTTLVQHKVHLKEDAVPRRRYYRIP
                     ERLVVKLEKEIELMLKLGVIEPSTSEWCSPVVLVPKKDGSLRFCIDFRYLNAVSKIQS
                     YPMPRIDELLERVGKSKFITTLDLSKGYWQVALAQETKELTAFTTPYGKFQFKVMPFG
                     LQGAPATFQRLMDEILRDFPQFAAAYLDDVIIFSHSWRDHMSHLRHVLHLIKAAGLTI
                     NPGKCVVAQQQVEYLGHVVGQGLVKPRVGKVEAIQEYQIPTTKKKVRAFVGLVGWYSK
                     FIPHFADRAAVLTDLTRASAPNKVVWTEDCDAAFRDLKGAITSESVLYSPDFTRPFIL
                     QTDASAVGLGAVLVQEAEGERHPVLFLSRKLLDRETRYSTVEKECLAMKWAIDSLRYY
                     LLGRHFCLETDHRALQWLRRMKDSNTRLTAWYLSLQAYDFTVQYRAGKTNCVADCLSR
                     VHEN"
     LTR             22415. .23107
                     /note="LTR of LReO_3"
     gene            complement(join(23657. .23844,24033. .24365,24473. .24751,
                     26771. .27052,27490. .27568))
                     /gene="OrlacI-S4"
     CDS             complement(join(23657. .23844,24033. .24365,24473. .24751,
                     26771. .27052,27490. .27568))
                     /gene="OrlacI-S4"
                     /note="non-classical class Ib
                     partially supported by GENSCAN and GrailEXP"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="BAB83837.1"
                     /db_xref="GI:18157522"
                     /translation="MKCMLLKCDFCLNYYIQKSKKQFNLMTKHSLEYFCTAASGVKEF
                     PEFVGVATVDKDFVGSCSMIIHRTEPKSTVIKKFIDENSKQIWWYYTTCKEKERFYAS
                     IIDGIKKHLNQNTVKYVFSKHVFQQHSGCEWDDETGETKSFNRYAYDGEDFLKLDLLK
                     HEWTALTEKAYATKLTWENNKVSLKHSEYFYSIQCPDWLKKYVTFGREYLQRAVLPLV
                     SLLQRSNSSLVSCHATGFFPDIADLFWMKDGEEPHGDVEKGEILPNNDETFQMSIYLN
                     TSSLPLTDWQKFSCVFQLFSSQKNITVSLDKTKIRTNDATHVMCSGNTDSFTKRHWGA
                     GTALFILVAMAVSGSNMFKRLLNLQYCLCSSDKENKDVNGSICPQVDVSECN"
     gene            join(52614. .52763,53361. .53514,53616. .53795,57879. .58039,
                     58223. .58346,60583. .60715,66607. .66642,68636. .68727,
                     68889. .68994,69164. .69311)
                     /gene="RXRB"
     CDS             join(52614. .52763,53361. .53514,53616. .53795,57879. .58039,
                     58223. .58346,60583. .60715,66607. .66642,68636. .68727,
                     68889. .68994,69164. .69311)
                     /gene="RXRB"
                     /note="partially supported by GENSCAN and GrailEXP
                     retinoic-acid receptor B"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="BAB83838.1"
                     /db_xref="GI:18157523"
                     /translation="MTSLPPISTSGVNSPVSSIGSPFSVISSSLGSPCLPGTPSVGYG
                     PISSPQINSTVSMSGLHTVSSSDDVKPPLGLKQLSSHSPGPMLSQKRLCSICGDRSSG
                     KHYGVYSCEGCKGFFKRTVRKDLTYTCRDNKDCTVDKRQRNRCQYCRYQKCLAMGMKR
                     EAVQEERQRNKEREGEVESTSAVNEEMPVEKILEAEMAVEQKTELHSDGGSGGSSPND
                     PVTNISQAADKQLFTLVEWAKRIPHFSELPLDDQVILLRWNELLIASFSHRSISVKDG
                     ILLATGLHVHRNSAHSAGVGAIFDRAHNAEVGAIFERVLTELVSKMRDMQMDKTELGC
                     LRAIILFNPDAKGLSNSSEVELLRERVYASLETYCKHKYPDQQGRFAKLLLRLPALRS
                     IGLKCLEHLFFFKLIGDTPIDTFLMEMLEAPHQLT"
     gene            join(100457. .100538,103270. .103431,104635. .104845,
                     105409. .105571,105691. .105813,109660. .110049,
                     110248. .110403,110560. .110637,115214. .115267,
                     116168. .116209,116401. .116463,116661. .116735,
                     116910. .116996,117362. .117418,117612. .117665,
                     117858. .117911,118128. .118181,118484. .118537,
                     118670. .118723,118800. .118844,119050. .119103,
                     121738. .121782,121933. .121986,123392. .123436,
                     124339. .124392,124502. .124546,124763. .124816,
                     124954. .124998,125091. .125144,125308. .125415,
                     127044. .127097,127222. .127275,127428. .127472,
                     127866. .127919,128078. .128122,128244. .128297,
                     128741. .128794,129460. .129513,129679. .129876,
                     130300. .130353,130559. .130666,130761. .130872,
                     131008. .131048,131203. .131256,131363. .131470,
                     131591. .131644,131728. .131781,132041. .132094,
                     132175. .132228,132324. .132431,132521. .132574,
                     132692. .132745,132830. .132883,133086. .133139,
                     133571. .133678,133778. .133831,133996. .134157,
                     134251. .134304,134385. .134420,134521. .134574,
                     134667. .134952,135086. .135195,136338. .136574,
                     137391. .137528)
                     /gene="COL11A2"
     CDS             join(100457. .100538,103270. .103431,104635. .104845,
                     105409. .105571,105691. .105813,109660. .110049,
                     110248. .110403,110560. .110637,115214. .115267,
```

```
                     116168. .116209,116401. .116463,116661. .116735,
                     116910. .116996,117362. .117418,117612. .117665,
                     117858. .117911,118128. .118181,118484. .118537,
                     118670. .118723,118800. .118844,119050. .119103,
                     121738. .121782,121933. .121986,123392. .123436,
                     124339. .124392,124502. .124546,124763. .124816,
                     124954. .124998,125091. .125144,125308. .125415,
                     127044. .127097,127222. .127275,127428. .127472,
                     127866. .127919,128078. .128122,128244. .128297,
                     128741. .128794,129460. .129513,129679. .129876,
                     130300. .130353,130559. .130666,130761. .130872,
                     131008. .131048,131203. .131256,131363. .131470,
                     131591. .131644,131728. .131781,132041. .132094,
                     132175. .132228,132324. .132431,132521. .132574,
                     132692. .132745,132830. .132883,133086. .133139,
                     133571. .133678,133778. .133831,133996. .134157,
                     134251. .134304,134385. .134420,134521. .134574,
                     134667. .134952,135086. .135195,136338. .136574,
                     137391. .137528)
                     /gene="COL11A2"
                     /note="collagen type XI alpha 2
                     partially supported by GENSCAN and GraiEXP"
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="BAB83839.1"
                     /db_xref="GI:18157524"
                     /translation="MLYLLLGLGNVLQEENCSQSKWPWADSYPVDVLRALQVPSLPEG
                     VKKVPGFCTSRRSSSPDHAYRITKKAQISAPTKQLFSGRFPENFSIMALVKAQAGLQA

  Query Match             5.5%;  Score 60;  DB 5;  Length 216734;
  Best Local Similarity   49.7%;  Pred. No. 0.066;
  Matches  180;  Conservative    0;  Mismatches  180;  Indels    2;  Gaps    1;

Qy        166 ATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATAC--T 223
              | |||  | || ||||  |||||| ||  | |||     ||| |||| | |   | |  |
Db     188031 ACCAGCAAGTAATTTTTCAGAAGAACTTTTATTATTATTATTATTTTTTAATGCTTCTTT 187972

Qy        224 CCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTT 283
              ||||| || ||  | || | |    | |||| ||   ||| |  | | |     |   ||
Db     187971 CCCTTACTTATTAAAGAAAAACAATAAAACATTCAATAATTAACGTTTCTGATATAATTT 187912

Qy        284 CAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTA 343
                  |||   || ||| |||||    |  | |  |||  | | ||     |     || |  |
Db     187911 TGTAAATACTAAAAAAAAATGTCAGCTTATATTGATGTTTTTTTTTAAAACACAACAAAA 187852

Qy        344 AACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACC 403
              ||   || |  |    | |      ||||||  | |  | || || | |     | |  |
Db     187851 AAAGAAAGACAAACTTAAAGACATTTACTTGTTTTGCCGGTAAATGTTTGTGGTTAATTC 187792

Qy        404 TAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATT 463
              | ||     | ||    |   ||  |   |      | |  |||||||   |||  || ||
Db     187791 TTAATAATAATTATAGACAGACTGCTGCGAAGATCACTTCAAATTACATTAAACATTTTT 187732

Qy        464 TTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAA 523
              ||  |   |  |||  | || |||  ||||  |   |   | ||||  |||||  ||
Db     187731 TTTTATTAATCCTCCAATAAAATATGTTTTCTTTAGCCTGTCAAACCGTTACTCACCTGT 187672

Qy        524 TA 525
              ||
Db     187671 TA 187670


RESULT 8
PFMAL3P3
LOCUS       PFMAL3P3               116696 bp    DNA     linear   INV 29-JAN-2003
DEFINITION  Plasmodium falciparum MAL3P3, complete sequence.
ACCESSION   Z98547 AL008977 AL010164 AL139179 AL844502 Z98548 Z98549 Z98550
VERSION     Z98547.2  GI:15383894
KEYWORDS    HTG; 60S Acidic ribosomal protein P2; asparagine synthetase;
            calcium-dependent protein kinase; DNA polymerase delta; dual
            specificity protein phosphatase; guanine nucleotide-binding
            protein; PDZ domain; rat BRAIN-like protein; serine/threonine
            protein kinases; T-complex protein eta; T08A11.2-like protein; WD
            domain; Y48E1C.2-like protein.
SOURCE      Plasmodium falciparum 3D7
  ORGANISM  Plasmodium falciparum 3D7
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 116696)
  AUTHORS   Bowman,S., Lawson,D., Basham,D., Brown,D., Chillingworth,T.,
            Churcher,C.M., Craig,A., Davies,R.M., Devlin,K., Feltwell,T.,
            Gentles,S., Gwilliam,R., Hamlin,N., Harris,D., Holroyd,S.,
            Hornsby,T., Horrocks,P., Jagels,K., Jassal,B., Kyes,S., McLean,J.,
            Moule,S., Mungall,K., Murphy,L., Oliver,K., Quail,M.A.,
            Rajandream,M.-A., Rutter,S., Skelton,J., Squares,R., Squares,S.,
            Sulston,J.E., Whitehead,S., Woodward,J.R., Newbold,C. and
            Barrell,B.G.
  TITLE     The complete nucleotide sequence of chromosome 3 of Plasmodium
            falciparum
  JOURNAL   Nature 400 (6744), 532-538 (1999)
  MEDLINE   99376085
   PUBMED   10448855
REFERENCE   2
  AUTHORS   Hall,N., Pain,A., Berriman,M., Churcher,C., Harris,B., Harris,D.,
            Mungall,K., Bowman,S., Atkin,R., Baker,S., Barron,A., Brooks,K.,
            Buckee,C.O., Burrows,C., Cherevach,I., Chillingworth,C.,
            Chillingworth,T., Christodoulou,Z., Clark,L., Clark,R., Corton,C.,
            Cronin,A., Davies,R., Davis,P., Dear,P., Dearden,F., Doggett,J.,
            Feltwell,T., Goble,A., Goodhead,I., Gwilliam,R., Hamlin,N.,
            Hance,Z., Harper,D., Hauser,H., Hornsby,T., Holroyd,S.,
            Horrocks,P., Humphray,S., Jagels,K., James,K.D., Johnson,D.,
            Kerhornou,A., Knights,A., Konfortov,B., Kyes,S., Larke,N.,
            Lawson,D., Lennard,N., Line,A., Maddison,M., Mclean,J., Mooney,P.,
            Moule,S., Murphy,L., Oliver,K., Ormond,D., Price,C., Quail,M.A.,
            Rabbinowitsch,E., Rajandream,M.A., Rutter,S., Rutherford,K.M.,
            Sanders,M., Simmonds,M., Seeger,K., Sharp,S., Smith,R., Squares,R.,
            Squares,S., Stevens,K., Taylor,K., Tivey,A., Unwin,L.,
            Whitehead,S., Woodward,J., Sulston,J.E., Craig,A., Newbold,C. and
            Barrell,B.G.
  TITLE     Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13
  JOURNAL   Nature 419 (6906), 527-531 (2002)
  MEDLINE   22255708
   PUBMED   12368867
REFERENCE   3  (bases 1 to 116696)
  AUTHORS   Lawson,D., Bowman,S. and Barrell,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-AUG-1997) P.falciparum Genome Sequencing Consortium,
            The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
            CB10 1SA, UK
COMMENT     On or before Aug 31, 2001 this sequence version replaced
            gi:2894463, gi:2894495, gi:2894369, gi:2982521, gi:2665376,
            gi:3649752.
            For more information about this sequence or the Malaria Project,
            see http://www.sanger.ac.uk/Projects/P_falciparum. The following
            numbers may have changed slightly:
            The true left end of clone MY434-3D7 is at 1 in this sequence. The
            true right end of clone MY549-3D7 is at 79511 in this sequence. The
            true right end of clone MY434-3D7 is at 114736 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .116696
                     /organism="Plasmodium falciparum 3D7"
                     /mol_type="genomic DNA"
                     /isolate="3D7"
                     /db_xref="taxon:36329"
                     /chromosome="3"
                     /clone="MAL3P3"
     gene            complement(join(1. .4302,4520. .4587,4804. .4830))
                     /gene="MAL3P3.1"
                     /note="synonym: PFC0325c"
     CDS             complement(join(<1. .4302,4520. .4587,4804. .4830))
                     /gene="MAL3P3.1"
                     /note="cDNA EST EMBL:T18200 comes from this gene, revised:
                     added two 5' and two 3' exons; splicing conserved in P.
                     yoelii"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAD49084.1"
```

```
                     /db_xref="GI:23476979"
                     /db_xref="UniProt/TrEMBL:Q8I223"
                     /translation="MNNEKKTEFRLDQSTASNGLFTVIERKPLSSLENNESLNFSNFN
                     EQLNPSVDDTLKKKEDYNNDNNNVDNMNNSNNQSDSICGIYRLNSSQLLNENLDNTLS
                     RISYVDEDVSSFFDSDKEEELYFDNDISTLGSCYSINTPNKIYKKKKTRNSINGIHAH
                     GKEHINLLLEKNEEKKNSPFIHNKKEEKNNIYYKPEEGLNEDYYDIMNKKIKKKKKSA
                     FLKDHYFKENVNVTNDDINIHSMVSNHHSDHSDHSNKSEKLLNEKIVPIKNNNEEKAI
                     RLNKNIYRVVNINKEEEKNNKESFIRRITQIRRSMINEKNFQRNEDDHHHKYDKKYEN
                     DDDHHHKYDKKYENDDDHHHKYDKKYDHDDGHHHKYDKKYDHSDGHHHKYDKKYDHSD
                     GHHHKYDKKYDHGDGHHHKYDKKYDHGDGHHHKYDKKYDHGDGHHHKYDKKYDHGDGH
                     HHKYDKKYDHSDGYHDHNHLYKYDKKYDHNDKDKIEHHNDVVDQGEHVQHEETSDKNS
                     PSNNYSNYNNRFSEGYILKRDNKLARYNSDKRVKYKQLNSDEGYENVLKYTNKEKYSN
                     IDSDIYYREEDYPEYYNDVPSRYNDYYNRVSNRNDKKNELNENSYYYKNVKYNEKNMF
                     MENNDDKENEPYAEVHYNDDKYIDNQYDDRYYNGNRDQYHFYDDEENVHKSKYMNKNY
                     YNDYKDYYNKGNAVNNYYEKDKLKRYSSFDQKVDIKRGDENDYIYNNNNNNMDEKIHT
                     LDKDKHNLDKNNIYDDKNNIYDDKNNMYDDKNNIYVDKNNIYVDKNKMYDDKNNYGYN
                     HHIVGGTNTTTYLKNLPKKSYKRREESGEINNESMNLESGEEEYRMMQKKKNTTFKHI
                     DENNMYYKHNELLNEYNTIDDPDNLNNNINRNRNIYMHKNRISNEQNNKFNLNNNELE
                     ADEKKNKMDNMFYNKERNYLINNQEYNKNYIYRNEYYKNEKNVYLDSYPYEMVEHKNK
                     YTSNMNNQYDNKDSSYLYDDVSSSFNNKKYTKLIIKDNLCNNKYKHNNYDNHNNIDGN
                     DVYYKHPYGIEEGRAYDKNYYHDNELNDHHLNYNDNYMNTKVNIHEDHNNNMDDKNIS
                     YNNMNLTDENNNNNNDNNNEKYNDDNNNNNNRYHVMNQRNIYPNNIPMNHVGDINEDK
                     NNYNMSDKQNIPSIDEKREENSRNFINETNKDDKENVNNMNNLNMVNNNMKNFIVPNM
                     NGMVLQNTSVPANVVPANVVPANVVPANVVPANVVPANVIPANVVPANVVPANVVPSNVVPAN
                     VAPANVVPANVVPANVVPANVVPANVVPANIIPANIVPTNIVPNVPITNNNIQ
                     NNNNLENINKTNDNPAPPVKRKRGRPKMKKNTPEETQPTDLASNSNASSSYQATNTNI
                     QNMNLNNLNHNLVNDMNQPLYTNLNTIINPPFNNQLANAVINLPLDQQVVQPLGQQVG
                     QPLGQAINYQVIPPPQTIGQQLAEPINQQ"
gene                 6740. .8839
                     /gene="MAL3P3.2"
                     /note="synonym: PFC0330w"
CDS                  6740. .8839
                     /gene="MAL3P3.2"
                     /note="Weak match (E-value 0.29) to PF00595 PDZ, PDZ
                     domain (Also known as DHR or GLGF); signal anchor
                     predicted by SignalP 2.0 HMM (Signal peptide probabilty
                     0.293, signal anchor probability 0.631) with cleavage site
                     probability 0.142 between residues 23 and 24"
                     /codon_start=1
                     /product="PDZ domain protein, putative"
                     /protein_id="CAB11102.1"
                     /db_xref="GI:3649753"
                     /db_xref="GOA:O77318"
                     /db_xref="UniProt/TrEMBL:O77318"
                     /translation="MYLLMSKFLVWIIIFGNFFVVHFFLCFKIPNNKITNVLTYDNSK
                     KQRINSKHNTFLYKNDTLNKLLNRKKNKHTLFYRSKKEKDIQESGYLYPFDHLEKKKT
                     KFPPDPYIPSEDEKESSIDIFHDKINDDDLMIYEDDDDAKKYFGKTTFNKPNIIQKDE
                     NSEDDQDDDLNDQNDDPINNPTNNLTDNPTNNSTDNPTNNSIDNPTNNSIDNPTNNSI
                     DNPTNNSIDNPTNNSIDNPTNNSIDNPTNNSTDNPINSPNYQQVTQSNIILEEENNTN
                     KTDIEKDIININFYSSEEKKIACEMIKDQIYYLLHSSMDKTFIETFNINIESHIVPSE
                     KRICVSYNMKSHTFKRSISNLSDLIEFSNKLKYVLDERIHKLYYEDKNNENAYLNKTL
                     RKELEIGKKERDVTNLSKKNLQMSIESSLKLIEKCMQNMYKQNYEITLHDLKLAYNML
                     PCKYYGFFLSNLCSTISTLSYYVNDLQGTFNFALKSIKYEPKYNVAWKCLGDAYRSFR
                     KFWQAKNFYDIAIYLGYKDEKGERFEDIVQELNNLTENALKNVDLLKEHMNNSTNIII
                     KKNIGIVLNKNPDSIGGCYVSYIIEGSKASKKNLHHGDQIVALNNHVTYGKPIDFCLK
                     AFQKNDGTYDIIFFKGNIIELYGLKAYKYLMKNDLFYALFENEQIVSFKRNETILFDM
                     KGFGTFSEYGMSKAEFV"
gene                 complement(join(9589. .16129,16327. .20041,20295. .21213))
                     /gene="MAL3P3.3"
                     /note="synonym: PFC0335c"
CDS                  complement(join(9589. .16129,16327. .20041,20295. .21213))
                     /gene="MAL3P3.3"
                     /note="Splicing confirmed in P. yoelii"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAB11104.1"
                     /db_xref="GI:3649755"
                     /db_xref="UniProt/TrEMBL:O77320"
                     /translation="MKEKEVDNLFPGDKDVTNLKQYFEGGDIKFENAELKKEREEIRK
                     NKVFIECVQNIWMNILNKKMDDKFSKVEYFKIMLRICKVLIPQFDIKEIIKIVNDEWI
                     NDSEGKKYLNFSSFFNAFFELSDIWTPTINAYDYAHFLKSLFYRITYIEIKTKDGRII
                     KKKPMIMINFKRTAENKKLSLYNDTNIIEKHVIKEKNKNTLKNLQSFKRSSMQIQNLK
                     KKHFIKRFSENYKLENSFLYDMDDDGDNSLIPKNLFENLLNNKQHNDYLQRNIILMDV
                     NDINPLEHPDEQNQSLNKNKCLTGTNKKEKYMIPKGTEYQDKEKESSILTINQNDKKY
                     KKKKKKYSELQDSNISSNNTLTSKRYTYTCGMDKETKETDEQNKRRNTEVASICNVCN
                     VETNEKQRKKTNKKKSVQKDDLVEENALLDNLIIIDGINFDDVTKECKIIDDNNNENI
                     DNNIYVENNKLKDQKSYDLFSSEGKNLILGVNEGEEFNEEVFENIEKELQERKDEKNK
                     NDKTIYNNNEEQTDDLNRNINKIESINNNNDNNNNINNKKEFNKIRTEHILNKESISK
                     HIKGSPSRDKEIKKLYTTNKNEDSTFELKKELEIITNNKVNVYEEDIIGSNEDDEYIH
                     VLKENLKEDANEYNNDKENKNNKTKEILKSKNYLENEKRTLEELKLRGKNNIFKKDEK
                     YNSLGEVIINEIQINEENKINDIQDGNISKQKIIQSSSRTNDTFNIKDISLNDDLEKE
                     KRKKKSQHFIDNLVKADKNEISENINKICDNNINNIYDESINNIYDESINNIYDESIN
                     NIYDESINNIYDENINNIYDENINNIYDEGINKICDDNILENKNIKTTND
                     IVQVEENNESIEKNELMISLNKDINNTYNMFKENVDIFINKIKRESLLKIDKNIKDNN
                     NDDEYIMDNFYENDFIINHKMEITNKELDPLEINTQNEFIENLDIKKKKYTNDHFFND
                     ADKMFYEMNKILNKDMKKNKEQEFFKTDETFGSLQSHKIKKYNKGEEKHDKNNEEEKN
                     ILYDENQVYSVLYSDHKIEQDIQDIHSIQTNICDENNIEQINEENSKKGVRISGTDME
                     NKNDMENKNDMEKKNDMEKKNDIEKKNDMEKKNDMEKKNDMEKKNDMEKNDMENENN
                     MENKSDIENENKMNEYKNNMKNENKMIENNTIENNMIENNMIENNMIENNMIENNMIE
                     NNMIENNMIENNMIENNTKIAYNMNENKNNIKIEDDAKNQINVQNNIQNKVHVKDKKN
                     DLINNVDIINDVLKSDDKFENMINSKEINIKDFKCVDKNNDICLNLNELNKQSYNSNT
                     TEKCIDIHDFDYVENKDVHDKIHEDRKEYCDESKLKYLPADDIKKMRSFIKINKKSKR
                     ENFLIISYMNEEIYFLIKKLKVEIIEPSIILNKNMHIEKQFLSPLGKYINDNLKEGKC
                     ISSKIIIDLMLEYMNTIISKNNGYVLYFNNYNILQHLDYFLHKLKGLSTIKYINYVHF
                     LNINFLKNVNDQKDVNIYDYLLDDQKIIVSNPFKGMFPKFCLILKKHNIVSIHSGKHF
                     NPQKIYRRLLSKFNDKQIDSSNNVKEEVEEKHSELFAWDKNEEKYKNVKENQEENDHK
                     SVCSDEYYLSISSEGKYNYDGEDEQDEQDEQDEQDEQDEQEEEDEQEEEDEQEEEDEQ
                     EEEDDEDDEDDEEEDDEDDNDNDDDDEDDNDNDDDDDDEYDDHYDKPSDDATQNISI
                     NINKSIIEEINNYENTNKNDKNHTDESYYNLMNKEINFIYKCTSPDKYYKNYYNNLK
                     KYCDDMEHEKDLSMLIKHLSVLLKDEKENIKNNEGGEESIDNILIDKNMKSDQEYNEN
                     EKDHLNNIEYIENLSEHLRHKYDIEMIPIDIVGNIRLYKLFYINMIKNNYSYYLKNNF
                     FPYKYKGKRGKAKKGMLMINIKRKHEYDNLMKKKNGNEKTGNNNKKKNNKEDKNDYIN
                     NNNNNNNNNNNDNNINVGYKRMQNNINPNNLYEIKNNNFNLYHRYWSLFYKYCPVSYI
                     NDNKIIEGKKKFCVIYKNKIYMLNNHDYMIEFLKHPKFYASKKVKESVCFSLVIYFPN
                     MDIQKKIIYLLETYEKLIHIDIGKYFKQYFFLSHQDKEQHKNNMKSNNNDDDVDNKIE
                     DAHIFEIHRIIEGKTISPLFIVRYICIHMKKIHQFNIYIIKEIIKSIIKLKHNFNLLI
                     MKKKNILINQCEIKKKMKRKKKMKKKKKKKIPEKKKLVNNEKKNLLFLFSSIMKIKFL
                     YNLIIIKQKNIKNVSFKNVLDKLSHITNIKDMAKDINLLLVLIEDMFFREKRKYLNKV
                     ERRDNILLSVKKSGSTCHTDYNNNEDNEDNENNEDNENNENKENNEDNKMRDDSYYVG
                     ERESRQRYIITGLPLHSKIFFYLKCMNISIDHVISFDIIKKEMNKDEENEDKGDGDDF
                     YGDDFYGDEKESGKRNTNFKNLSQYYYNDKMNEEDTYELYKKYVSYNTYNDNIYHMNI
                     LNDYKKKKYDIKIIPFDEDEKKSIYFYLNEIRKIINPYVHSYYNYNDILKNRICCKQL
                     NNSYINVYCPYVLKKHNWLYKSIDNIKLVYDEKVYYLSNLYHAFIFSKNIKYYQHNFN
                     IPPLRLLFLYDVPYKFNKDIKKNMVRKFFFKYINLEKECIIYIKYVFIYLLKYIFLFK
                     IYKYDYKKKTNIDIKNFSINSIPIKNLLEKKYIYYLYREDIIYLENNINQINVYEQNN
                     HFNEFHIYDLLFQIIKPFNVFLNIIKILTIRYILHEHLQNCIITYNTFQHSIFNITDV
                     NSLKYVIQKYRQKKYNKDHSSKQKEDIFKGSNQNELLTNEKDTGEKKQNLIHQDDHLK
                     NRKNKNHVDEYINNYNINEKDKQDPNKNTSNQNDTYLVTNQITINSFGYDDDNNSDSI
```

```
Query Match              5.4%;  Score 59.2;  DB 3;  Length 116696;
Best Local Similarity   47.6%;  Pred. No. 0.1;
Matches  175;  Conservative    0;  Mismatches  193;  Indels    0;  Gaps    0;

Qy       162 ACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATA 221
             |  |||  |    |||  ||||  |      ||  |  |  ||   |  |  ||  |||  |  |||||
Db     76696 AAAAAAAAAAAGATAATTTTTTTTTTTATTTTTTGTTGCAAATTTTAATTTATATATATA 76755

Qy       222 CTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTG 281
             |  || || |  |  | ||  |  | ||  ||  |  |  ||  |||||  |||||  ||
Db     76756 ATATATTTATATAATAATAATATTTCACTGATTTTTATTTAAACATTTTTTTTTTTTTAT 76815

Qy       282 TTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAAT 341
             |  | | |  ||| |  ||||  |  | | ||| |  |  | | |     |||||||
Db     76816 ATATATATATATATTATAAAATAAAAAAAAAAAAAAAAAAAACTTTAAAATTATTGAAAAA 76875

Qy       342 TAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAA 401
             | ||  || | | | | | ||| | | | | | |  ||        ||    ||
Db     76876 TTAAAGAAAGAATATTTCTTTTTATTTAATTAATTTTAAATTTCCCCATATTAATAAAAA 76935

Qy       402 CCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTA 461
             ||     |  |  |  |  |  | | | |  |||| | ||  |  | | || ||
Db     76936 AATATTAACCTTAAAAAAAGTAAGATTACAAAAAGAGAATATATATATATATTATTTATA 76995

Qy       462 TTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCA 521
             || |  || | |  | | |  | |||  |||||||      |  | |  || |   |
Db     76996 TTATACAATTGAATTTATATTATATATATTTTTATTATTGTTATATAATTTGATACTTAT 77055
```

```
Qy        522 AATATTCT 529
              | |||| |
Db      77056 ATTATTAT 77063


RESULT 9
CR847982
LOCUS       CR847982              214310 bp    DNA     linear   HTG 11-OCT-2004
DEFINITION  Danio rerio clone DKEY-5F13, *** SEQUENCING IN PROGRESS ***, 13
            unordered pieces.
ACCESSION   CR847982
VERSION     CR847982.2  GI:54034980
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 214310)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 11, 2004 this sequence version replaced gi:54021686.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK5F13
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 209425 bases at least Q40
            Consensus quality: 210080 bases at least Q30
            Consensus quality: 210572 bases at least Q20
            Insert size: 213110; sum-of-contigs
            Insert size: 206305; 5.0% error; agarose-fp
            Quality coverage: 8.23x in Q20 bases; sum-of-contigs Quality
            coverage: 8.51x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 13 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     3687: contig of 3687 bp in length
            *     3688     3787: gap of 100 bp
            *     3788    17738: contig of 13951 bp in length
            *    17739    17838: gap of 100 bp
            *    17839    31263: contig of 13425 bp in length
            *    31264    31363: gap of 100 bp
            *    31364    50563: contig of 19200 bp in length
            *    50564    50663: gap of 100 bp
            *    50664    56490: contig of 5827 bp in length
            *    56491    56590: gap of 100 bp
            *    56591    67651: contig of 11061 bp in length
            *    67652    67751: gap of 100 bp
            *    67752    79277: contig of 11526 bp in length
            *    79278    79377: gap of 100 bp
            *    79378    97587: contig of 18210 bp in length
            *    97588    97687: gap of 100 bp
            *    97688   124183: contig of 26496 bp in length
            *   124184   124283: gap of 100 bp
            *   124284   136752: contig of 12469 bp in length
            *   136753   136852: gap of 100 bp
            *   136853   173914: contig of 37062 bp in length
            *   173915   174014: gap of 100 bp
            *   174015   203520: contig of 29506 bp in length
            *   203521   203620: gap of 100 bp
            *   203621   214310: contig of 10690 bp in length.
FEATURES             Location/Qualifiers
     source          1. .214310
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-5F13"
                     /clone_lib="DanioKey"
     misc_feature    1. .3687
                     /note="assembly_fragment:00003
                     fragment_chain:1"
     misc_feature    3788. .17738
                     /note="assembly_fragment:00683
                     fragment_chain:1"
     misc_feature    17839. .31263
                     /note="assembly_fragment:00528
                     fragment_chain:1"
     misc_feature    31364. .50563
                     /note="assembly_fragment:01026
                     fragment_chain:1"
     misc_feature    50664. .56490
                     /note="assembly_fragment:00030
                     fragment_chain:1"
     misc_feature    56591. .67651
                     /note="assembly_fragment:00381
                     fragment_chain:1"
     misc_feature    67752. .79277
                     /note="assembly_fragment:00100
                     fragment_chain:1"
     misc_feature    79378. .97587
                     /note="assembly_fragment:01259
                     fragment_chain:1"
     misc_feature    97688. .124183
                     /note="assembly_fragment:01494
                     fragment_chain:1"
     misc_feature    124284. .136752
                     /note="assembly_fragment:00841
                     fragment_chain:1"
     misc_feature    136853. .173914
                     /note="assembly_fragment:02209
                     fragment_chain:1"
     misc_feature    174015. .203520
                     /note="assembly_fragment:01845
                     fragment_chain:1"
     misc_feature    203621. .214310
                     /note="assembly_fragment:00235.0"
ORIGIN

  Query Match             5.4%;  Score 59.2;  DB 2;  Length 214310;
  Best Local Similarity   39.6%;  Pred. No. 0.094;
  Matches  151;  Conservative    0;  Mismatches  230;  Indels    0;  Gaps    0;

Qy        138 TTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCT 197
              |||   | | ||| ||  || |  | |||  |   |       |   |        | |
Db       3819 TTTTTATTATAAAAATATATTAAAATAAAAAANAAANNNNANTNNNNTANTTNNNTNTTT 3878

Qy        198 TAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATAC 257
              |     || |||| ||  | |  |   |           |   | | | | ||
Db       3879 TTTTTTTTTTTTTTTNTTTTTTTTTTTTTNNTTTNNNANAANNNAANANNAAAAAAAAAAA 3938

Qy        258 AGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAG 317
              | |  | | |  |||| ||| ||   |||   ||       |||   ||
Db       3939 AANTNNTNTTNTNNTTTTTTTTTTTTTTTAAANTTTANTNNATAATTNTTTTTTTATTTT 3998

Qy        318 ATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATC 377
              | | | ||       ||   | | || | ||||  | |  || ||   ||  ||  |
Db       3999 TTNNTTTTTTTTNNNTTTNNATATTAAAANAAAAAAAATATATTTTTATTTTTTTTTTT 4058

Qy        378 TTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGC 437
              ||    |      ||| |    ||  | ||  |             | || |
```

```
Db       4059 TTTTTTTTTTNTTTAATATAAAAATTTTAANTTTNNNNNNNNNTNNTTTTTTTNTNNNNNN 4118

Qy        438 AAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATG 497
              ||||| |  ||   | |||| || ||||  |       |    | | |||  | | |
Db       4119 AAATATATTTTTTTATAAATATTTTTTTTTANNNTNNATNTANATATATATATATANANA 4178

Qy        498 TACTATTAAAACATTTACTTT 518
              |  | || ||  |||  |||
Db       4179 AATAAATATAATTTTTTATTT 4199


RESULT 10
BX901920/c
LOCUS       BX901920              119227 bp    DNA     linear   VRT 18-MAR-2004
DEFINITION  Zebrafish DNA sequence from clone CH211-184M19 in linkage group 20,
            complete sequence.
ACCESSION   BX901920
VERSION     BX901920.9  GI:45581114
KEYWORDS    HTG.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 119227)
  AUTHORS   Barker,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-MAR-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Mar 18, 2004 this sequence version replaced gi:44903871.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep Clone-derived
            Zebrafish pUC subclones occasionally display inconsistency over the
            length of mononucleotide A/T runs and conserved TA repeats. Where
            this is found the longest good quality representation will be
            submitted.
            Repeat names beginning 'Dr' were identified by the Recon repeat
            discovery system (Zhirong Bao and Sean Eddy, submitted), and those
            beginning 'drr' were identified by Rick Waterman (Stephen Johnson
            lab, WashU).  For further information see
            http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml
            CH211-184M19 is from a CHORI-211 BAC library
            VECTOR: pTARBAC2.1.
FEATURES             Location/Qualifiers
     source          1. .119227
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-184M19"
                     /clone_lib="CHORI-211"
ORIGIN

  Query Match             5.4%;  Score 58.8;  DB 5;  Length 119227;
  Best Local Similarity   49.2%;  Pred. No. 0.12;
  Matches  182;  Conservative    0;  Mismatches  187;  Indels    1;  Gaps    1;

Qy        197 TTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATA 256
              ||||   | ||  |       ||||  |       |||| | | |  | ||||   |
Db      58492 TTAATTTCCTTTATCCAAGCCTATAAGCATAATCTTTTATAATATAATCTTAATAATCAT 58433

Qy        257 CAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAA 316
              |   ||||   || || || | || || || |||| ||||| || |     |  | | | |
Db      58432 CTTTTATTTGGTTTTTATTAGGTTTTTGAATAAAGCTATAAGGACAGATGTATAAAAATA 58373

Qy        317 GATCATATATTAGCCA-GTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAA 375
              | || | || |   |   |  | | || |||  |||  | |  | || |||    | | ||
Db      58372 GTGCACAGATAAATAATAATTTATAATACACAACAAAGTAATTAAAATAAATACATTAAA 58313

Qy        376 TCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACT 435
              | |       ||  |||| |||||  |     | |   |  || |||    |
Db      58312 ACATTTTAAATAACAAAAAATGTCAATTTTTTTATGTTGCATTTGATTTAAATTTTTTGC 58253

Qy        436 GCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTA 495
              |  ||  |||  | |||| | | |  |  |||||||||    |||| |||| |||||
Db      58252 AGTGTTTGAGTTTATAAAAAATCTAAGTATTTAAATAATCTATTAAAAAATAATATTTTA 58193

Qy        496 TGTACTATTAAAACATTTACTTTCCAAATATTCTGTCATTCAGGAGTATGGAAGTATCGA 555
              |  |  |  | |  |  | | |    |||| | | ||  |    |||  |  |||  |
Db      58192 GGCTCCCTGTATATTTAAAATATATGTATATATTTTTATGTATATATATATATATATATA 58133

Qy        556 TGGCTTCTTT 565
              |   || | |
Db      58132 TACTTTTTAT 58123


RESULT 11
AC021626/c
LOCUS       AC021626              196788 bp    DNA     linear   HTG 12-MAR-2000
DEFINITION  Homo sapiens clone RP11-391I2, WORKING DRAFT SEQUENCE, 28 unordered
            pieces.
ACCESSION   AC021626
VERSION     AC021626.3  GI:7229839
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 196788)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-391I2
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 196788)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
```

```
COMMENT     On Mar 12, 2000 this sequence version replaced gi:6978227.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L5883
                Center clone name: 391_I_2
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 181094 bases at least Q40
                Consensus quality: 189581 bases at least Q30
                Consensus quality: 192532 bases at least Q20
                Insert size: 208000; agarose-fp
                Insert size: 194088; sum-of-contigs
                Quality coverage: 3.7 in Q20 bases; agarose-fp
                Quality coverage: 4.0 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 28 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1281: contig of 1281 bp in length
            *     1282     1381: gap of 100 bp
            *     1382     2899: contig of 1518 bp in length
            *     2900     2999: gap of 100 bp
            *     3000     5681: contig of 2682 bp in length
            *     5682     5781: gap of 100 bp
            *     5782     8351: contig of 2570 bp in length
            *     8352     8451: gap of 100 bp
            *     8452    10639: contig of 2188 bp in length
            *    10640    10739: gap of 100 bp
            *    10740    13655: contig of 2916 bp in length
            *    13656    13755: gap of 100 bp
            *    13756    15927: contig of 2172 bp in length
            *    15928    16027: gap of 100 bp
            *    16028    19028: contig of 3001 bp in length
            *    19029    19128: gap of 100 bp
            *    19129    22669: contig of 3541 bp in length
            *    22670    22769: gap of 100 bp
            *    22770    26791: contig of 4022 bp in length
            *    26792    26891: gap of 100 bp
            *    26892    30845: contig of 3954 bp in length
            *    30846    30945: gap of 100 bp
            *    30946    35594: contig of 4649 bp in length
            *    35595    35694: gap of 100 bp
            *    35695    40148: contig of 4454 bp in length
            *    40149    40248: gap of 100 bp
            *    40249    45042: contig of 4794 bp in length
            *    45043    45142: gap of 100 bp
            *    45143    49477: contig of 4335 bp in length
            *    49478    49577: gap of 100 bp
            *    49578    54518: contig of 4941 bp in length
            *    54519    54618: gap of 100 bp
            *    54619    61982: contig of 7364 bp in length
            *    61983    62082: gap of 100 bp
            *    62083    67811: contig of 5729 bp in length
            *    67812    67911: gap of 100 bp
            *    67912    74070: contig of 6159 bp in length
            *    74071    74170: gap of 100 bp
            *    74171    80685: contig of 6515 bp in length
            *    80686    80785: gap of 100 bp
            *    80786    89056: contig of 8271 bp in length
            *    89057    89156: gap of 100 bp
            *    89157    99104: contig of 9948 bp in length
            *    99105    99204: gap of 100 bp
            *    99205   109014: contig of 9810 bp in length
            *   109015   109114: gap of 100 bp
            *   109115   122471: contig of 13357 bp in length
            *   122472   122571: gap of 100 bp
            *   122572   138753: contig of 16182 bp in length
            *   138754   138853: gap of 100 bp
            *   138854   156691: contig of 17838 bp in length
            *   156692   156791: gap of 100 bp
            *   156792   173620: contig of 16829 bp in length
            *   173621   173720: gap of 100 bp
            *   173721   196788: contig of 23068 bp in length.
FEATURES             Location/Qualifiers
     source          1. .196788
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="RP11-391I2"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1281
                     /note="assembly_fragment"
     misc_feature    1382. .2899
                     /note="assembly_fragment"
     misc_feature    3000. .5681
                     /note="assembly_fragment"
     misc_feature    5782. .8351
                     /note="assembly_fragment"
     misc_feature    8452. .10639
                     /note="assembly_fragment"
     misc_feature    10740. .13655
                     /note="assembly_fragment"
     misc_feature    13756. .15927
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    16028. .19028
                     /note="assembly_fragment"
     misc_feature    19129. .22669
                     /note="assembly_fragment"
     misc_feature    22770. .26791
                     /note="assembly_fragment"
     misc_feature    26892. .30845
                     /note="assembly_fragment"
     misc_feature    30946. .35594
                     /note="assembly_fragment"
     misc_feature    35695. .40148
                     /note="assembly_fragment"
     misc_feature    40249. .45042
                     /note="assembly_fragment"
     misc_feature    45143. .49477
                     /note="assembly_fragment"
     misc_feature    49578. .54518
                     /note="assembly_fragment"
     misc_feature    54619. .61982
                     /note="assembly_fragment"
     misc_feature    62083. .67811
                     /note="assembly_fragment"
     misc_feature    67912. .74070
                     /note="assembly_fragment"
     misc_feature    74171. .80685
                     /note="assembly_fragment"
     misc_feature    80786. .89056
                     /note="assembly_fragment"
     misc_feature    89157. .99104
                     /note="assembly_fragment"
     misc_feature    99205. .109014
                     /note="assembly_fragment"
     misc_feature    109115. .122471
                     /note="assembly_fragment"
     misc_feature    122572. .138753
                     /note="assembly_fragment"
```

```
     misc_feature    138854. .156691
                     /note="assembly_fragment"
     misc_feature    156792. .173620
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    173721. .196788
                     /note="assembly_fragment"
ORIGIN

  Query Match             5.4%;  Score 58.6;  DB 2;  Length 196788;
  Best Local Similarity   46.4%;  Pred. No. 0.12;
  Matches  226;  Conservative    0;  Mismatches  259;  Indels    2;  Gaps    1;

Qy         58 TACAATGAGATATGAGTTTTAAAGCCAGATCTTCCTTATCTCTCTGCCCCATCTCTAGTT 117
              ||  |  | |||| |  | || ||  | ||  |   ||| | | |     || | ||||
Db      65644 TATTACTATATATTAAATATATAGTAATATTATATATATTTATATATTTAATATATAGTA 65585

Qy        118 CTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAG 177
                |    | | |  |   |  ||   |    | |||  || ||    | |   |  ||
Db      65584 TATATTATATATATTTATATATTTAATATATAGTATATATTATATATATTTTATATATTT 65525

Qy        178 TTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTT 237
                | || |  ||   || | || | ||| | |   | |||||      |  ||| | |
Db      65524 AAATATATATAATTATATATATAATAAATTTTATATATTATATAAAAATATATTTTATAT 65465

Qy        238 AAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAA 297
              |||| |  ||| | |||||||  |||  || | |  |   || | ||| | | |
Db      65464 AAAATATTTTATACACATACATATATGAATATATCATATATGCATTTATAAATGTTTTTT 65405

Qy        298 AGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGT 357
              | |  | | || ||||| | ||  | |  ||  ||  |  ||| || | |  | ||| ||
Db      65404 ATATGTACTTTAAGAGATATAT--TTTTGTATGCATATTTAAAATATATATTACATGTGT 65347

Qy        358 GCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTAC 417
                | | |  ||  |    |  | || | |  | |  |  |    |
Db      65346 ATTTTTAAATATATACATATAATACATATGATATATACATATATACACATAAATATATTT 65287

Qy        418 CTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTC 477
                |  | | |||     ||   ||| | ||         |||||||| || | ||   | |
Db      65286 ATATATACATATAAGTGTGTGTATATATATATATTACTATTTTATTATGTATATTCACAC 65227

Qy        478 AGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGTCATTCA 537
              | |    | || | |  |||| |    |||| |    |  |  || ||    | ||||
Db      65226 ACACTCTAGAAATGTGACTGTATGAAGGGAACAATCCAATAACCCATTTTGATGCCTTCA 65167

Qy        538 GGAGTAT 544
                | | |
Db      65166 TGTCTCT 65160


RESULT 12
DDITUBA
LOCUS       DDITUBA                 2255 bp    DNA     linear   INV 21-SEP-1993
DEFINITION  Dictyostelium discoideum alpha tubulin (TUBA) gene, complete cds.
ACCESSION   L13999
VERSION     L13999.1  GI:290058
KEYWORDS    alpha-tubulin.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 2255)
  AUTHORS   Trivinos-Lagos,L., Ohmachi,T., Albrightson,C., Burns,R.G.,
            Ennis,H.L. and Chisholm,R.L.
  TITLE     The highly divergent alpha- and beta-tubulins from Dictyostelium
            discoideum are encoded by single genes
  JOURNAL   J. Cell. Sci. 105 (Pt 4), 903-911 (1993)
  MEDLINE   94043557
   PUBMED   8227212
COMMENT     Original source text: Dictyostelium discoideum (strain AX-3)
            (library: Clontech) DNA.
FEATURES             Location/Qualifiers
     source          1. .2255
                     /organism="Dictyostelium discoideum"
                     /mol_type="unassigned DNA"
                     /strain="AX-3"
                     /db_xref="taxon:44689"
                     /tissue_lib="Clontech"
     gene            178. .1662
                     /gene="TUBA"
     CDS             join(178. .805,917. .1662)
                     /gene="TUBA"
                     /note="putative"
                     /codon_start=1
                     /product="alpha tubulin"
                     /protein_id="AAC37343.1"
                     /db_xref="GI:290059"
                     /translation="MGREIISIHIGQAGVQVGNSCWELYCLEHGIERDGSIPADRKQS
                     SDVNNLNKDLGTFFSESTNGKKVVPRAIFLDLEPTVIDEIRTGDYKNLFHPEQLITGK
                     EDAANNYARGHYTVGKELIDVCVDRIRRLADQCDGLQGFLVFHSVGGGTGSGFGSLLL
                     QKLALDYGGKKSKLDFCVYPSPQVSTSVVEPYNSVLSTHSLLEHTDVSFMLDNEAIYN
                     ICKNSLDIEKPTYTNLNRLIAQVISSLTSSLRFPGQLNLDINDIQTNLVPFPRLHFVL
                     CSYAPVISREKAHHETITVDNITSAVFSEKNIMAKCQPNLGKYMACCLMYRGDIVPKE
                     AQKAVQNIRSEKSRNVSFVDWSPTGFKCGINNQAPVSTKDSEMAEVKKSVCMLSNTTA
                     ISQVFSRINHKFDLMFVKRAFVHWYVGEGMEEGEFAEARDDLLALEKDYESVSASTEG
                     EEQEEEY"
     exon            <178. .805
                     /gene="TUBA"
     intron          806. .916
                     /gene="TUBA"
     exon            917. .>1662
                     /gene="TUBA"
ORIGIN

  Query Match             5.4%;  Score 58.4;  DB 3;  Length 2255;
  Best Local Similarity   46.2%;  Pred. No. 0.29;
  Matches  194;  Conservative    0;  Mismatches  226;  Indels    0;  Gaps    0;

Qy        118 CTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAG 177
              ||| ||      |  |||   ||       | ||||||   | | | ||||| |  |||  |
Db      1691 CTTTAAACTGTTGTTATTCTTTCATAATAAAAATATAATTTAATTCAAATAAAATAAATG 1750

Qy        178 TTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTT 237
              ||||  |   | |||  |  |||||   || |||| ||  | |  |   ||  ||    |
Db      1751 TTTTATTTTATATCAAACTTTAAAATTTTTTTTTTTTTTTTTTTTTTTTTTTTAGAAT 1810

Qy        238 AAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAA 297
                |  |||| |  ||     |  ||    |  ||||  ||   |  ||||  |  ||
Db      1811 TTGAACACTTAAAACCAATTTCCATTTTTAATTATTTTTTTTCTATTTTAAAAATTCAAA 1870

Qy        298 AGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGT 357
              |  ||    |  | |  |  || | ||  ||    |  | ||| | |||   |||| |
Db      1871 ATTAAAATAAATAAATTATAATTAAATTGTAATTAAATTAAATTAAAATTAAAAATAAAA 1930

Qy        358 GCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTAC 417
                ||   | || |  ||| ||  | | |||   |||||    |   |||   | | |
Db      1931 AAATTAAAAATAAAAAAATTTTAAAATAAAAAAAAAAAAAAATTAAAAAAAAAAAAAAAA 1990

Qy        418 CTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTC 477
                ||  ||       |        || |||   |||||   ||  | |        |
Db      1991 AAAAATTAGTAAAAACTTTTTTTCTAAATTTAAACCAAATAAAATATAGATCTATTAATT 2050

Qy        478 AGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGTCATTCA 537
                ||      | |||| | || | ||       |||  |||   ||| || | |  || |
Db      2051 TTTAATATCTTTATTTTTTTTAATTTTTTTTTTTTTTTTTTTTTTAATTTTTTTTTTTTAA 2110


RESULT 13
AE001377
LOCUS       AE001377               12469 bp    DNA     linear   INV 04-OCT-2002
DEFINITION  Plasmodium falciparum 3D7 chromosome 2 section 14 of 73 of the
            complete sequence.
ACCESSION   AE001377 AE001362
VERSION     AE001377.2  GI:23503349
KEYWORDS    .
SOURCE      Plasmodium falciparum 3D7
```

```
  ORGANISM  Plasmodium falciparum 3D7
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 12469)
  AUTHORS   Gardner,M.J., Tettelin,H., Carucci,D.J., Cummings,L.M., Aravind,L.,
            Koonin,E.V., Shallom,S., Mason,T., Yu,K., Fujii,C., Pederson,J.,
            Shen,K., Jing,J., Aston,C., Lai,Z., Schwartz,D.C., Pertea,M.,
            Salzberg,S., Zhou,L., Sutton,G.G., Clayton,R., White,O.,
            Smith,H.O., Fraser,C.M., Adams,M.D., Venter,J.C. and Hoffman,S.L.
  TITLE     Chromosome 2 sequence of the human malaria parasite Plasmodium
            falciparum
  JOURNAL   Science 282 (5391), 1126-1132 (1998)
  MEDLINE   99021743
   PUBMED   9804551
REFERENCE   2  (bases 1 to 12469)
  AUTHORS   Gardner,M.J., Hall,N., Fung,E., White,O., Berriman,M., Hyman,R.W.,
            Carlton,J.M., Pain,A., Nelson,K.E., Bowman,S., Paulsen,I.T.,
            James,K., Eisen,J.A., Rutherford,K., Salzberg,S.L., Craig,A.,
            Kyes,S., Chan,M.-S., Nene,V., Shallom,S.J., Suh,B., Peterson,J.,
            Angiuoli,S., Pertea,M., Allen,J., Selengut,J., Haft,D.,
            Mather,M.W., Vaidya,A.B., Martin,D.M.A., Fairlamb,A.H.,
            Fraunholz,M.J., Roos,D.S., Ralph,S.A., McFadden,G.I.,
            Cummings,L.M., Subramanian,G.M., Mungall,C., Venter,J.C.,
            Carucci,D.J., Hoffman,S.L., Newbold,C., Davis,R.W., Fraser,C.M. and
            Barrell,B.
  TITLE     Genome sequence of the human malaria parasite Plasmodium falciparum
  JOURNAL   Nature 419, 498-511 (2002)
REFERENCE   3  (bases 1 to 12469)
  AUTHORS   Gardner,M.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-OCT-2002) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
COMMENT     On Oct 4, 2002 this sequence version replaced gi:3845111.
FEATURES             Location/Qualifiers
     source          1. .12469
                     /organism="Plasmodium falciparum 3D7"
                     /mol_type="genomic DNA"
                     /isolate="3D7"
                     /db_xref="taxon:36329"
                     /chromosome="2"
     repeat_region   <1. .23
                     /rpt_type=tandem
     repeat_region   complement(<1. .23)
                     /rpt_type=tandem
                     /rpt_unit="(ta)n"
     repeat_region   88. .118
                     /rpt_type=tandem
     repeat_region   complement(88. .118)
                     /rpt_type=tandem
                     /rpt_unit="(a)n"
     repeat_region   complement(170. .206)
                     /rpt_type=tandem
                     /rpt_unit="(ta)n"
     gene            <232. .>960
                     /locus_tag="PFB0160w"
                     /note="rad10"
     mRNA            <232. .>960
                     /locus_tag="PFB0160w"
     CDS             232. .960
                     /locus_tag="PFB0160w"
                     /codon_start=1
                     /product="ERCC1 nucleotide excision repair protein,
                     putative"
                     /protein_id="AAC71822.1"
                     /db_xref="GI:3845112"
                     /translation="MVSDKEDQCNKINNNDNVTSLESINEEKKNNTNEGGESFFDNNA
                     EQYLIISLRQKLNPVIKKIKRVRYKFNNIIPDFLVGKNNACLFISMKYHRLRSNYLKA
                     RIETLSNKYNNRILLCLVDMENIENSLGEINQLSFSFNMTLILCWSNEECARVIEDFR
                     IYEKKISYIIKKKISSSNQEEKIHELLKKIRCIHTTDCITLTTKFKNFKNIIQAKKED
                     LISCSGLGIKKIQALMATFNDPFF"
     repeat_region   complement(432. .453)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   complement(711. .754)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   complement(841. .867)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   complement(968. .1003)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   1031. .1163
                     /rpt_type=tandem
     repeat_region   complement(1031. .1163)
                     /rpt_type=tandem
                     /rpt_unit="(ta)n"
     repeat_region   1036. .1174
                     /rpt_type=tandem
     repeat_region   1177. .1212
                     /rpt_type=tandem
     repeat_region   complement(1177. .1212)
                     /rpt_type=tandem
                     /rpt_unit="(a)n"
     repeat_region   complement(1275. .1310)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   1357. .1399
                     /rpt_type=tandem
     repeat_region   complement(1357. .1399)
                     /rpt_type=tandem
                     /rpt_unit="(ta)n"
     repeat_region   complement(1414. .1450)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   1479. .1499
                     /rpt_type=tandem
                     /rpt_unit="(a)n"
     repeat_region   1540. .1595
                     /rpt_type=tandem
     repeat_region   complement(1540. .1591)
                     /rpt_type=tandem
                     /rpt_unit="(ta)n"
     repeat_region   1656. .1683
                     /rpt_type=tandem
                     /rpt_unit="(a)n"
     repeat_region   complement(1693. .1738)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   complement(1808. .1834)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   complement(1835. .1891)
                     /rpt_type=tandem
                     /rpt_unit="(ta)n"
     repeat_region   1839. .1891
                     /rpt_type=tandem
     gene            complement(<1911. .>2971)
                     /locus_tag="PFB0161c"
     mRNA            complement(join(<1911. .2092,2415. .2504,2941. .>2971))
                     /locus_tag="PFB0161c"
     CDS             complement(join(1911. .2092,2415. .2504,2941. .2971))
                     /locus_tag="PFB0161c"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAN37604.1"
                     /db_xref="GI:23503350"
                     /translation="MMYFLCIHVKYVNDNPFIKNRSSTRITNAPGGNSSVSFGNYVDT
                     ESKKVSNKNEKSQTTIKENTKAAGNLDVNNENKKSDRRTNVKVNQPPGGASSSNKT"
     repeat_region   complement(2098. .2120)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   complement(2121. .2286)
                     /rpt_type=tandem
                     /rpt_unit="(ta)n"
     repeat_region   2121. .2158
                     /rpt_type=tandem
```

```
     repeat_region   2211. .2242
                     /rpt_type=tandem
     repeat_region   complement(2385. .2409)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   complement(2506. .2543)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   complement(2554. .2587)
                     /rpt_type=tandem
                     /rpt_unit="(ta)n"
     repeat_region   complement(2676. .2717)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   2736. .2768
                     /rpt_type=tandem
                     /rpt_unit="(a)n"
     repeat_region   2740. .2764
                     /rpt_type=tandem
     repeat_region   complement(2774. .2884)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   complement(2897. .3070)
                     /rpt_type=tandem
                     /rpt_unit="(ta)n"
     repeat_region   3008. .3064
                     /rpt_type=tandem
     repeat_region   complement(3088. .3140)
                     /rpt_type=tandem
                     /rpt_unit="(ta)n"
     repeat_region   complement(3146. .3166)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   complement(3177. .3234)
                     /rpt_type=tandem
                     /rpt_unit="(a)n"
     repeat_region   complement(3257. .3280)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   complement(3300. .3323)
                     /rpt_type=tandem
                     /rpt_unit="(ta)n"
     repeat_region   complement(3325. .3348)
                     /note="AT_rich"
                     /rpt_type=tandem
     repeat_region   complement(3367. .3434)
                     /rpt_type=tandem

  Query Match             5.3%;  Score 58.2;  DB 3;  Length 12469;
  Best Local Similarity   48.4%;  Pred. No. 0.24;
  Matches  192;  Conservative    0;  Mismatches  203;  Indels    2;  Gaps    1;

Qy        131 ATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGAT 190
              |||| | | |  || | ||    |      |   |||||  |||   |||     ||
Db      10243 ATATAATTATATTTAAAAATATATATATGATCATAATCATATAAATATTTCTCTAAAATG 10302

Qy        191 CTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACA 250
              |  || |||| | ||   | | |  |||   |   ||  |||| ||| || |    | |
Db      10303 ATTTTCCTAAATATATATATATATATATAATATATTTTTTTTTTCTTATAAAATAGCAAA 10362

Qy        251 GACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCA 310
              ||     |  | | |  | ||      |     ||||||   ||||  | |   |   |
Db      10363 GAACAGTAAGAAATAAAATATTAAGTAATAAAAAAAAAAAATAATAATAATAAATAAATA 10422

Qy        311 GAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTA 370
              | |||  |  |||||     | |   ||||  | |  | |    |  || ||| ||||
Db      10423 AATAAATTTTCTATATATATAAATCATAAATATTATAATATAAATATATATTTTATATTA 10482

Qy        371 CTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATC 430
                ||| ||  |  || |  || | | |  |  || |  |||| ||     | | ||||
Db      10483 AGAAATTTTCTA--CATACATATATAATATATATATATATATATTATAAATTATATTATA 10540

Qy        431 GCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATT 490
              |  || ||| | |||  |  ||  ||  ||||    | ||||| |    |      |||
Db      10541 ACTTTGAAAAAAAAAAAAAATTACCCATAATATTAAATATATAATATATATTATATATT 10600

Qy        491 TTTTATGTACTATTAAAACATTTACTTTCCAAATATT 527
              | ||||||  |||       ||| | |     |||||
Db      10601 ATGTATGTATAATTTTTTTTTTTCCATATGGTATATT 10637


RESULT 14
CR352307
LOCUS       CR352307              196722 bp    DNA     linear   HTG 16-MAR-2004
DEFINITION  Danio rerio clone DKEY-81P21, *** SEQUENCING IN PROGRESS ***, 6
            unordered pieces.
ACCESSION   CR352307
VERSION     CR352307.4  GI:45504093
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 196722)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Mar 16, 2004 this sequence version replaced gi:45433936.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK81P21
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 194096 bases at least Q40
            Consensus quality: 194526 bases at least Q30
            Consensus quality: 194947 bases at least Q20
            Insert size: 196222; sum-of-contigs
            Insert size: 201559; 2.1% error; agarose-fp
            Quality coverage: 10.40x in Q20 bases; sum-of-contigs Quality
            coverage: 10.30x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    64612: contig of 64612 bp in length
            *    64613    64712: gap of 100 bp
            *    64713    70583: contig of 5871 bp in length
            *    70584    70683: gap of 100 bp
            *    70684    82793: contig of 12110 bp in length
            *    82794    82893: gap of 100 bp
            *    82894   147353: contig of 64460 bp in length
            *   147354   147453: gap of 100 bp
            *   147454   188859: contig of 41406 bp in length
            *   188860   188959: gap of 100 bp
            *   188960   196722: contig of 7763 bp in length.
FEATURES             Location/Qualifiers
     source          1. .196722
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-81P21"
                     /clone_lib="DanioKey"
     misc_feature    1. .64612
```

```
                     /note="assembly_fragment:02309
                     fragment_chain:1"
     misc_feature    64713. .70583
                     /note="assembly_fragment:00022
                     fragment_chain:1"
     misc_feature    70684. .82793
                     /note="assembly_fragment:00283
                     fragment_chain:1"
     misc_feature    82894. .147353
                     /note="assembly_fragment:01182
                     fragment_chain:1"
     misc_feature    147454. .188859
                     /note="assembly_fragment:00466
                     fragment_chain:1"
     misc_feature    188960. .196722
                     /note="assembly_fragment:00123.0"
ORIGIN

  Query Match             5.3%;  Score 58.2;  DB 2;  Length 196722;
  Best Local Similarity   36.5%;  Pred. No. 0.15;
  Matches  201;  Conservative    0;  Mismatches  349;  Indels    0;  Gaps    0;

Qy         23 TAACTTTCACTCTCTATGTTCTTATAAGCAAATGTTACAATGAGATATGAGTTTTAAAGC 82
Db      64295 TATATATTTATATATATATAAATATATATAAATATATTTATTATAAAAATTTTTTTATTT 64354

Qy         83 CAGATCTTCCTTATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGG 142
Db      64355 TTTTTTTATTTATATTTAANTATNTTATTATTATATATATTTTTTTTAAATAAAATATAA 64414

Qy        143 TTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAG 202
Db      64415 NAAAANATAATAAAAAAAAATANATATATAAAAAAAATATTANAAAATNTTAAATATTAA 64474

Qy        203 ACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTA 262
Db      64475 ATAATTTTTTAATTATATTAAATAATTATTATTANAAAATTTTATTAATAAATAATAATA 64534

Qy        263 TTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCAT 322
Db      64535 TTTTATTTTATNTTAAATTTAAATAAAAAATATAAATAAAATAAATTNTAANATTTTATA 64594

Qy        323 ATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCA 382
Db      64595 TATAATTAAATTATAAAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 64654

Qy        383 GTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATA 442
Db      64655 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNAA 64714

Qy        443 GAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTA 502
Db      64715 TACCCTTTTACTATTTTTTTTTTTTAAATTTATATATATATATATATAAAAAATTTTATT 64774

Qy        503 TTAAAACATTTACTTTCCAAATATTCTGTCATTCAGGAGTATGGAAGTATCGATGGCTTC 562
Db      64775 TTTATTTAAATTTTTTAAAAAAATTATAATATATATATATATATAAATATATATATATAT 64834

Qy        563 TTTAAAATGA 572
Db      64835 ATTATATTTA 64844


RESULT 15
AF052006/c
LOCUS       AF052006               6706 bp    DNA     linear   INV 26-JUN-2002
DEFINITION  Dictyostelium discoideum C-module-binding factor (cbfA) gene,
            complete cds.
ACCESSION   AF052006
VERSION     AF052006.2  GI:21592302
KEYWORDS    .
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 6706)
  AUTHORS   Horn,J., Dietz-Schmidt,A., Zundorf,I., Garin,J., Dingermann,T. and
            Winckler,T.
  TITLE     A Dictyostelium protein binds to distinct oligo(dA) x oligo(dT) DNA
            sequences in the C-module of the retrotransposable element DRE
  JOURNAL   Eur. J. Biochem. 265 (1), 441-448 (1999)
  MEDLINE   99421761
   PUBMED   10491202
REFERENCE   2  (bases 1 to 6706)
  AUTHORS   Winckler,T. and Dingermann,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-MAR-1998) Pharmaceutical Biology, University of
            Frankfurt, Marie-Curie-Strasse 9, Frankfurt 60439, Germany
REFERENCE   3  (bases 1 to 6706)
  AUTHORS   Winckler,T. and Dingermann,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUN-2002) Pharmaceutical Biology, University of
            Frankfurt, Marie-Curie-Strasse 9, Frankfurt 60439, Germany
  REMARK    Sequence update by submitter
COMMENT     On Jun 26, 2002 this sequence version replaced gi:4335861.
FEATURES             Location/Qualifiers
     source          1. .6706
                     /organism="Dictyostelium discoideum"
                     /mol_type="genomic DNA"
                     /strain="AX2"
                     /db_xref="taxon:44689"
     gene            <114. .>3525
                     /gene="cbfA"
     mRNA            join(<114. .293,703. .>3525)
                     /gene="cbfA"
                     /product="C-module-binding factor"
     CDS             join(114. .293,703. .3525)
                     /gene="cbfA"
                     /note="DNA-binding protein; CMBF"
                     /codon_start=1
                     /product="C-module-binding factor"
                     /protein_id="AAD17489.2"
                     /db_xref="GI:21592303"
                     /translation="MEELIKAPNSNFIIMSNQPYQTTSPEIVEDYIIKRGQPFVLTGT
                     TQGWSRSNMFTLDFLSERYSEMELINSPRNNETHTDLQGWRMKDFISYLQVSPEERNP
                     KHLYGKDIACPREWQEYLSHKLQPQYSYKSRFDLVSHLPDYLQPETLLVYIGSNGTYT
                     PGHIDMCGSLSQNLMVSSDQDAFAWWFIVPTEYKDEALKFWGDKGGDVYNESRFIRPI
                     DLLGAPFPIYVFKQRPGDFIFVPPDSVHQVVNCGPGISTKVAWNSISLKSLPISYFSS
                     LPHTRRMAKPELFRIKAIAYYTLRKIMGDVENTNFNTIDVNDVIDIIAPLLEIFHNIL
                     QTESILIPKPNYPYCNGETIPFLQPFKYFNGDRIQDRRCDHCNSDIFNRCYHCETCKT
                     DDGQQGKDFCFDCVSSGIGCEFHFKVMVLKEFISHSKLKKELSSFYEIYKNLLAHSGRR
                     PKEVDDIITKSTDRVSDECGFLTTATVAYHVVFYSSQKKIKCHRCEKRFKKFSIIFCT
                     NCNARFCEQCVVNTFGQNFQVLMKRNEWECFCCKGLCDCSNCTSNSNSSNHPRILNNN
                     QQLGLPYNNNNNSNNNNNNNINNNNNNNNNNNINNNNNNMNNNNSINNNNNNNNNNIN
                     NNNINNNNHHNNNGNNNLNSSYSSLNALSSLSQQQSYGSYDNYNNNNNNNNYNNNNNN
                     NGHIQILKSGRQYDDEQSSSSGSGSSNSTPTKPRPRNGGDDGLMSHFSGNNNNNNNHH
                     NNNNNNNNNHHMMSHHHHNNNNNNNNNNPTTSSLSSLSTSLSSSSTSTQKPMDVHSK
                     KRPIVLDNDKPKGRPPKNLKEWTSTHKFIISLIELFRSSNNAILGKPNPHYKPIENLP
                     PLVQLYLSQRKAFGGVLWAKTNSCPLLPCIWVKDLSVIPPNTKLLPSLIQGKKIVVLF
                     FGDQDQEEYVGIVGKKSIFSFDEVNQTLLLKCGEVPLAQLEDLFNTTEPEIAMKKDIA
                     AFNYKNQIEEKEEGLYVKQELYNNKKII"
ORIGIN

  Query Match             5.3%;  Score 57.8;  DB 3;  Length 6706;
  Best Local Similarity   50.1%;  Pred. No. 0.31;
  Matches  196;  Conservative    0;  Mismatches  192;  Indels    3;  Gaps    2;

Qy        146 AGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACA 205
Db        649 AAAAAAAGTTAGTATTTAAAACATTTTTTTTTTCTTTTAAATTTAACCATATAGTGATA 590

Qy        206 TTGTTTTGTTAATATACTCCCTTGATTTTTTAAAAGACCTTACAGACATACAGCTATTC 265
Db        589 TTGTTTATAATTTATTATTTATTTATTTTTTTATTATTATTTGATA-AGATGATTTTTT 531

Qy        266 ATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATA 325
```

```
Db        530 TTTTTTTTTTTTTTTTTTCAATTGAATTAAAAAAAATTTAATTTAAAAAAAAAAAAAAAA 471

Qy        326 TTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACT--TAATCTTGCAG 383

Db        470 TTAATTTGGAAAAAAAAAAAAAAAATAAAAATTCTGATACGAAATAGTGATATTTTTAAAT 411

Qy        384 TCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAG 443

Db        410 TGAAATTAAAAAACAAAAAATAAAATAAAATAAAATATTTTTTAATTTAAAAAAAAATAA 351

Qy        444 AAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTAT 503

Db        350 TAATAATAATAATAATAATAATAAAAATGAATAATAATGATGATAATAATAACTTACAGA 291

Qy        504 TAAAACATTTACTTTCCAAATATTCTGTCAT 534

Db        290 TAAAAAGTCAAGAGTAAACATATTACTTCTT 260


RESULT 16
AC022281
LOCUS       AC022281              186431 bp    DNA     linear   HTG 24-JAN-2002
DEFINITION  Homo sapiens chromosome 10 clone RP11-10E13, WORKING DRAFT
            SEQUENCE, 9 unordered pieces.
ACCESSION   AC022281
VERSION     AC022281.3  GI:13399348
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN;
            HTGS_CANCELLED.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 186431)
  AUTHORS   Smith,D.R.
  TITLE     Genome Therapeutics Corporation Sequencing Center: Human Genome
            Sequence Data
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 186431)
  AUTHORS   Smith,D.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JAN-2000) Genome Therapeutics Corporation, 100 Beaver
            Street, Waltham, MA 02453, USA
COMMENT     On Mar 21, 2001 this sequence version replaced gi:9929643.
            -------------- Genome Center
                           Center: Genome Therapeutics Corporation
                           Center code: GTC
                           Web site: http://www.genomecorp.com/
                           Contact: gtc-seqcenter@genomecorp.com
                       -------------- Project Information
                           Center project name: hg143
                       -------------- Summary Statistics
                           Sequencing vector: N/A
                           Chemistry: Dye-terminator Big Dye; 100% of reads
                           Assembly program: Phrap; version 990315
                           Consensus quality: 395389 bases at least Q40
                           Consensus quality: 405067 bases at least Q30
                           Consensus quality: 410525 bases at least Q20
                           Insert size: 185631; sum-of-contigs
                           Quality coverage: 6.4x in Q20 bases; sum-of-contigs
                       --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1   124725: contig of 124725 bp in length
            *   124726   124825: gap of unknown length
            *   124826   125924: contig of 1099 bp in length
            *   125925   126024: gap of unknown length
            *   126025   128493: contig of 2469 bp in length
            *   128494   128593: gap of unknown length
            *   128594   156515: contig of 27922 bp in length
            *   156516   156615: gap of unknown length
            *   156616   170011: contig of 13396 bp in length
            *   170012   170111: gap of unknown length
            *   170112   176430: contig of 6319 bp in length
            *   176431   176530: gap of unknown length
            *   176531   179019: contig of 2489 bp in length
            *   179020   179119: gap of unknown length
            *   179120   181965: contig of 2846 bp in length
            *   181966   182065: gap of unknown length
            *   182066   186431: contig of 4366 bp in length.
FEATURES             Location/Qualifiers
     source          1. .186431
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP11-10E13"
                     /clone_lib="RPCI-11"
     misc_feature    1. .124725
                     /note="assembly_name:Contig243
                     clone_end:SP6"
     misc_feature    124826. .125924
                     /note="assembly_name:Contig204"
     misc_feature    126025. .128493
                     /note="assembly_name:Contig233"
     misc_feature    128594. .156515
                     /note="assembly_name:Contig240"
     misc_feature    156616. .170011
                     /note="assembly_name:Contig239"
     misc_feature    170112. .176430
                     /note="assembly_name:Contig235"
     misc_feature    176531. .179019
                     /note="assembly_name:Contig231"
     misc_feature    179120. .181965
                     /note="assembly_name:Contig232"
     misc_feature    182066. .186431
                     /note="assembly_name:Contig236
                     clone_end:T7"
ORIGIN

  Query Match             5.3%;  Score 57.8;  DB 2;  Length 186431;
  Best Local Similarity   45.8%;  Pred. No. 0.18;
  Matches  170;  Conservative    0;  Mismatches  201;  Indels    0;  Gaps    0;

Qy        138 TTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCT 197

Db     175849 TTTATTATTNAATTTTTATTAATTATATATATATTATATTATAANNNTATTTTTTATTAT 175908

Qy        198 TAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATAC 257

Db     175909 TTTTATAATTTTATTTATATTTATATATATTAATATTATTATATTAANATATATATTTAT 175968

Qy        258 AGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAG 317

Db     175969 ATTTTTAATTAAATTTTATTTATATATANAATTTTAATAATANTTTAAATTATATTTTTA 176028

Qy        318 ATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATC 377

Db     176029 AATTNATATTATATTTTATAAAATTATNTATTTATTTATAAATTAATANTTTATTTAAAT 176088

Qy        378 TTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGC 437

Db     176089 ATAAATTTATNANTATATATATAATTAAAAATTATATTAATATATTATTAAAATAANTTT 176148

Qy        438 AAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATG 497

Db     176149 AATTTAAAATAAAATTATATAATTNTAATTATATAATTTTTTATAATAAAATATTATATA 176208

Qy        498 TACTATTAAAA 508

Db     176209 TAAAATATAAA 176219
```

```
RESULT 17
AX457067
LOCUS       AX457067               2009 bp    DNA     linear   PAT 06-JUL-2002
DEFINITION  Sequence 28 from Patent WO0231186.
ACCESSION   AX457067
VERSION     AX457067.1  GI:21715849
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Berlin,K.
  TITLE     Method for the detection of cytosine methylations
  JOURNAL   Patent: WO 0231186-A 28 18-APR-2002;
            Epigenomics AG (DE)
FEATURES             Location/Qualifiers
     source          1. .2009
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="lower strand of DNA sequence 25; after bisulphite
                     treatment and P CR"
ORIGIN

  Query Match             5.2%;  Score 56.8;  DB 6;  Length 2009;
  Best Local Similarity   44.7%;  Pred. No. 0.59;
  Matches  220;  Conservative    0;  Mismatches  272;  Indels    0;  Gaps    0;

Qy          46 ATAAGCAAATGTTACAATGAGATATGAGTTTTAAAGCCAGATCTTCCTTATCTCTCTGCC 105
               |||| | |  | | |||    |||   |||||||   |         ||     | |
Db         348 ATAATTATAATTAATAATATATTATATATTTTAAAATAATTAAAAAATTTAAAATATTAT 407

Qy         106 CCATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAA 165
                 || | ||    |  ||     | | |  | | |  ||   |||||||     |||
Db         408 TAATATATAAAAATAAAATATATTTAAAATAATATATTTTTTAAATATTTGAATTTAATT 467

Qy         166 ATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCC 225
               || |  || |    | || || |  |  | ||| | |  || |||| | ||||||
Db         468 ATTATATATTTTGAATATATAAAAAATGTTTTTATATATTTTATTTTATAAATATATAAA 527

Qy         226 CTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCA 285
               |    | | |||  | |    | |   | | |  ||   |    | |   ||   |
Db         528 ATATTATATATTATTAAAAAAAAAAATAAAAAAAATAAAAAAAATATATATATTATATTT 587

Qy         286 AAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAA 345
               |     |  | | |  |  |||   |  || |  ||  |||     |  | ||  |
Db         588 TATTTAATTAATAAATTTTTATTTTAAATAAAAATATTAATTTTAAAAATTAATTTATGT 647

Qy         346 CACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTA 405
                  ||  |    |  | | || |  ||  |  | |  ||  | |||| | ||
Db         648 TTTTTAAAAATAATTAAAAATAATAATATATTATTTATTTTAAAATTAAAAATTAAAAAT 707

Qy         406 AAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTT 465
               ||  |||   |  |   |  ||||   | |   || |  ||     |  |||| | ||||
Db         708 AAATTATTTAAATTTTGTAATTATATAAATTATAAAAATAAAATAAATTAATTATTTTTT 767

Qy         466 GGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATA 525
                 | | |  | | ||||  |  |||||| | |  | |||  |   ||  |||  ||||
Db         768 ATTATTTAATTTAAAAAAATTTTTTTTTTTTATTTTTTTATTATGATTTTTTTTAAAATT 827

Qy         526 TTCTGTCATTCA 537
               |  | | ||| |
Db         828 TATTTTGATTTA 839


RESULT 18
AC141740
LOCUS       AC141740              151143 bp    DNA     linear   HTG 19-MAR-2003
DEFINITION  Apis mellifera clone CH224-58F11, WORKING DRAFT SEQUENCE, 10
            unordered pieces.
ACCESSION   AC141740
VERSION     AC141740.1  GI:29123924
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Apis mellifera (honey bee)
  ORGANISM  Apis mellifera
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Apoidea;
            Apidae; Apis.
REFERENCE   1  (bases 1 to 151143)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 151143)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-MAR-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: AMEK
                Center clone name: CH224-58F11
            -------------- Summary Statistics
                Sequencing vector: Plasmid;
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 148444 bases at least Q40
                Consensus quality: 149298 bases at least Q30
                Consensus quality: 149814 bases at least Q20
                Estimated insert size: 147256; sum-of-contigs estimation
                Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
```

```
             --------------
             * NOTE: Estimated insert size may differ from sequence length
             *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
             * NOTE: This is a 'working draft' sequence. It currently
             * consists of 10 contigs. The true order of the pieces
             * is not known and their order in this sequence record is
             * arbitrary. Gaps between the contigs are represented as
             * runs of N, but the exact sizes of the gaps are unknown.
             * This record will be updated with the finished sequence
             * as soon as it is available and the accession number will
             * be preserved.
             *        1     1058: contig of 1058 bp in length
             *     1059     1158: gap of unknown length
             *     1159     2581: contig of 1423 bp in length
             *     2582     2681: gap of unknown length
             *     2682     4880: contig of 2199 bp in length
             *     4881     4980: gap of unknown length
             *     4981     9305: contig of 4325 bp in length
             *     9306     9405: gap of unknown length
             *     9406    12997: contig of 3592 bp in length
             *    12998    13097: gap of unknown length
             *    13098    22656: contig of 9559 bp in length
             *    22657    22756: gap of unknown length
             *    22757    36494: contig of 13738 bp in length
             *    36495    36594: gap of unknown length
             *    36595    53260: contig of 16666 bp in length
             *    53261    53360: gap of unknown length
             *    53361    96147: contig of 42787 bp in length
             *    96148    96247: gap of unknown length
             *    96248   151143: contig of 54896 bp in length.
FEATURES             Location/Qualifiers
     source          1. .151143
                     /organism="Apis mellifera"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7460"
                     /clone="CH224-58F11"
ORIGIN

  Query Match             5.2%;  Score 56.8;  DB 2;  Length 151143;
  Best Local Similarity   47.7%;  Pred. No. 0.29;
  Matches  229;  Conservative    0;  Mismatches  247;  Indels    4;  Gaps    2;

Qy          61 AATGAGATATGAGTTTTAAAGCCAGATCTTCCTTATCTCTCTGCCCCATCTCTAGTTCTT 120
               ||||  | |  | || || |     |  |  | ||   | ||     ||   ||  | ||
Db       96713 AATGTTAAAAAAATTATATATTATATTAATTATAATTCATATGAATTATAATTATATTTT 96772

Qy         121 GAAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTT 180
               | ||   |  ||  |       |      ||||| || || |  |   | | |      |
Db       96773 AATGTAAATGTTAAAAAAAATTATATATTATATTAATTATAATTATATATTAATATAAAT 96832

Qy         181 TGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAA 240
               | ||  | || | || |||| | | | | |    | |||| |   |  |||  | ||||
Db       96833 TATATTATATGTTATATTAATTATAATTTATATGAATTATAAT-TATATATTAATATAAA 96891

Qy         241 AGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGA 300
               |  | |  ||| |     | | || |  ||  | ||    | ||  |||    |
Db       96892 TGTTAAGAAAATTATATATTATATTAATTATAATTTATATGAATTATAATTATATTTTAA 96951

Qy         301 AATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCA 360
               ||  ||    | ||| || |||| |||      | | ||||||    ||    | |  |
Db       96952 TATAAATGTTAAAAAAAATTATATGTTATATTAATTATAATTAATATAAATTATATTATA 97011

Qy         361 TATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAG---TCAACCTAAAGGTATACTAC 417
               | ||| |||| |||    ||     | ||   |||  |     | ||  ||||  |  | |
Db       97012 TGTTATATTAATTATAATTTATATGAATTATAATTATATATTAATATAAATATTAAGAAA 97071

Qy         418 CTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTC 477
               |  | | ||||   | |   |||  | || |     |||| ||||||   |  ||| |
Db       97072 ATTATATATTATATTAATTATAATTTATATAAATTATAATTATATTTTAATATAAATGTT 97131

Qy         478 AGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGTCATTCA 537
               | ||||  ||  | ||    || |  |||   || |   ||  || |||    | || ||
Db       97132 AAAAAAATTATATATTATATTAATTATAATTTATATGAATTATAATTATATATTAATACA 97191
```

```
RESULT 19
AC115766
LOCUS       AC115766              225600 bp    DNA     linear   ROD 31-JAN-2004
DEFINITION  Mus musculus chromosome 1, clone RP23-82I24, complete sequence.
ACCESSION   AC115766
VERSION     AC115766.9  GI:41470217
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 225600)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus chromosome 1, clone RP23-82I24
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 225600)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 225600)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-DEC-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 225600)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
```

```
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-JAN-2004) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jan 31, 2004 this sequence version replaced gi:40445419.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L23317
                Center clone name: 82_I_24
            --------------.
FEATURES             Location/Qualifiers
     source          1. .225600
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="1"
                     /map="1"
                     /clone="RP23-82I24"
                     /clone_lib="RPCI-23 Female Mouse BAC"
     misc_feature    1. .10056
                     /note="wgs_end_extension
                     clone_end:SP6"
     repeat_region   2. .58
                     /rpt_family="(TCC)n"
     repeat_region   complement(61. .205)
                     /rpt_family="B1_MM"
     repeat_region   complement(210. .297)
                     /rpt_family="MIR"
     repeat_region   1115. .1173
                     /rpt_family="(TTTA)n"
     repeat_region   1653. .1689
                     /rpt_family="(TG)n"
     repeat_region   3086. .3215
                     /rpt_family="B1_MM"
     repeat_region   3216. .3263
                     /rpt_family="(CAA)n"
     repeat_region   4885. .4984
                     /rpt_family="PB1D9"
     repeat_region   4993. .5023
                     /rpt_family="(CACCAT)n"
     repeat_region   5024. .5064
                     /rpt_family="(CAA)n"
     repeat_region   7312. .7368
                     /rpt_family="GA-rich"
     repeat_region   complement(8192. .8990)
                     /rpt_family="RMER13A"
     unsure          complement(9153. .9174)
                     /note="single clone coverage"
     repeat_region   9731. .9796
                     /rpt_family="(TG)n"
     repeat_region   9802. .9981
                     /rpt_family="B2_Mm2"
     misc_feature    10057. .10062
                     /note="clone_boundary
                     clone_end:SP6.
                     site:EcoRI"
     repeat_region   complement(11356. .11527)
                     /rpt_family="L2"
     repeat_region   11528. .11613
                     /rpt_family="(TTATA)n"
     repeat_region   complement(11614. .11763)
                     /rpt_family="L2"
     repeat_region   complement(11832. .12036)
                     /rpt_family="MSTD"
     repeat_region   complement(12839. .16086)
                     /rpt_family="Lx2"
     repeat_region   17100. .17149
                     /rpt_family="(TG)n"
     repeat_region   complement(18967. .19081)
                     /rpt_family="ID_B1"
     repeat_region   19301. .19350
                     /rpt_family="(TAA)n"
     repeat_region   complement(20733. .21013)
                     /rpt_family="B4A"
     repeat_region   complement(22840. .23104)
                     /rpt_family="MLT1J"
     repeat_region   26941. .26982
                     /rpt_family="(TTTTG)n"
     repeat_region   complement(28528. .29162)
                     /rpt_family="Lx"
     repeat_region   29150. .29445
                     /rpt_family="L1_MM"
     repeat_region   complement(30205. .30499)
                     /rpt_family="B4"
     repeat_region   32068. .32318
                     /rpt_family="ORR1D"
     repeat_region   32329. .32734
                     /rpt_family="ORR1B-int"
     repeat_region   33123. .33759
                     /rpt_family="Lx"
     repeat_region   33793. .34312
                     /rpt_family="ORR1B-int"
     repeat_region   34313. .34627
                     /rpt_family="ORR1D"
     repeat_region   34628. .34805
                     /rpt_family="(TTTC)n"
     repeat_region   complement(34808. .34947)
                     /rpt_family="B1_MM"
     repeat_region   complement(35046. .35421)
                     /rpt_family="L1MA6"
     repeat_region   35468. .35533
                     /rpt_family="L2"
     repeat_region   complement(35744. .35834)
                     /rpt_family="B1_MM"
     repeat_region   35841. .35977
                     /rpt_family="B3"
     repeat_region   35979. .36012
                     /rpt_family="(CA)n"
     repeat_region   complement(36706. .36743)
                     /rpt_family="U2"

  Query Match             5.2%;  Score 56.8;  DB 10;  Length 225600;
  Best Local Similarity   65.2%;  Pred. No. 0.27;
  Matches  116;  Conservative    0;  Mismatches   57;  Indels    5;  Gaps    2;

Qy        887 TTACCACCAGTGCCTGAGGCTCCAGAGGAGCCATTGCCTTCTCAAGGTCACA---TAGGT 943
              | | |  ||||||  | | |||||||||  |  ||  || ||||||||||||   | |||
Db     132992 TCATCTTCAGTGCTAGGGTCTCCAGAGGGTC--TTATCTGCTCAAGGTCACATGGTTGGT 133049

Qy        944 GGTGGGTGAGTTAGGACCAAATCTAGAATTCCTGACTCCAGTAACTTCTGAAGTCATTTT 1003
              |||||  |||  |||||||  || ||  | |||| ||||||  ||||||  | ||  | |||
Db     133050 GGTGGTAGAGGAAGGACCATGTCCAGGTTACCTGTCTCCAGGGACTTCTAACGTAACTTT 133109

Qy       1004 GTTTTTTATTTTTATGGTTTTATTATAAGAATACTTGCTAAGCACACTTACCCCCTGC 1061
              | ||||||||||||  | || | || |   |  |  |  |   || ||   |  || ||
```

```
Db     133110 GGTTTTTATTTTTCGGCTTCTGTTGTGTTACAAGATATTTGCCAAACACGCTTCCAGC 133167


RESULT 20
PFMAL8P1_03/c
WPCOMMENT
  Sequence split into 14 fragments  LOCUS PFMAL8P1 Accession AL844507
     Fragment Name      Begin     End
     PFMAL8P1_00            1    110000
     PFMAL8P1_01       100001    210000
     PFMAL8P1_02       200001    310000
     PFMAL8P1_03       300001    410000
     PFMAL8P1_04       400001    510000
     PFMAL8P1_05       500001    610000
     PFMAL8P1_06       600001    710000
     PFMAL8P1_07       700001    810000
     PFMAL8P1_08       800001    910000
     PFMAL8P1_09       900001   1010000
     PFMAL8P1_10      1000001   1110000
     PFMAL8P1_11      1100001   1210000
     PFMAL8P1_12      1200001   1310000
     PFMAL8P1_13      1300001   1325595
Continuation (4 of 14) of PFMAL8P1 from base 300001 (AL844507 Plasmodium falciparum 3D7

  Query Match             5.2%;  Score 56.6;  DB 2;  Length 110000;
  Best Local Similarity   46.9%;  Pred. No. 0.33;
  Matches  176;  Conservative    0;  Mismatches  199;  Indels    0;  Gaps    0;

Qy        157 TCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTA 216
              | |  | | | || ||  |  || | |  || |     |    ||| | |   | || |
Db      72138 TAAAAATATACCATTTTTTTTTTCTATTAAAAAAAAAAAAAAAAAAAAAAAATCTTTTCC 72079

Qy        217 ATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTG 276
              | |  ||  ||  | |  ||  ||||  |   | || |  |  |||  || |  || |
Db      72078 TTTTTTTCTTTTATTATAATTTTAAGAAATATTAAACTTTTATTTATATATATTATTCTT 72019

Qy        277 GTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTG 336
              |||  || | || |  || | | ||||  |   | | | | |        ||   |  |
Db      72018 ATTTTTTTATAATATATAAATATAAATATAAATATAAATATAAATATATATATATATATA 71959

Qy        337 AAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAA 396
              | |||||| |  ||||   |  ||  | |  ||  | ||           |   |||
Db      71958 TATATTAAAAAGTAAATATATTGATTGTTCTATAACTTATTAATACCAATTTGAAGAAAT 71899

Qy        397 GTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAA 456
              ||||  ||||      |  |  |  ||| ||||      |   || | |||      | | |
Db      71898 TTCAAAATAAAAAATAAAGATATAATTTATTATTATTTTAAGAAAAAAAAAATATATATA 71839

Qy        457 TTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACT 516
              | |   | |  | |||     | | || |||  ||| ||| ||    |  ||||| || |
Db      71838 TATATATATATATATATATATATATAATATATTTATATATATATATGTAACATATATT 71779

Qy        517 TTCCAAATATTCTGT 531
              |    || || | |
Db      71778 ATATTGATCTTTTAT 71764


RESULT 21
BX942829/c
LOCUS       BX942829              176321 bp    DNA     linear   HTG 03-OCT-2004
DEFINITION  Danio rerio clone CH211-95H12, WORKING DRAFT SEQUENCE, 12 unordered
            pieces.
ACCESSION   BX942829
VERSION     BX942829.4  GI:53755952
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 176321)
  AUTHORS   McLaren,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 3, 2004 this sequence version replaced gi:46019758.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zC95H12
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 173270 bases at least Q40
            Consensus quality: 173761 bases at least Q30
            Consensus quality: 174259 bases at least Q20
            Insert size: 175221; sum-of-contigs
            Insert size: 183561; 3.8% error; agarose-fp
            Quality coverage: 7.74x in Q20 bases; sum-of-contigs Quality
            coverage: 7.39x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 12 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     7786: contig of 7786 bp in length
            *     7787     7886: gap of 100 bp
            *     7887    14905: contig of 7019 bp in length
            *    14906    15005: gap of 100 bp
            *    15006    44564: contig of 29559 bp in length
            *    44565    44664: gap of 100 bp
            *    44665    59129: contig of 14465 bp in length
            *    59130    59229: gap of 100 bp
            *    59230    64301: contig of 5072 bp in length
            *    64302    64401: gap of 100 bp
            *    64402    76420: contig of 12019 bp in length
            *    76421    76520: gap of 100 bp
            *    76521    86373: contig of 9853 bp in length
            *    86374    86473: gap of 100 bp
            *    86474    96468: contig of 9995 bp in length
            *    96469    96568: gap of 100 bp
            *    96569   123997: contig of 27429 bp in length
            *   123998   124097: gap of 100 bp
            *   124098   145113: contig of 21016 bp in length
            *   145114   145213: gap of 100 bp
            *   145214   151421: contig of 6208 bp in length
            *   151422   151521: gap of 100 bp
            *   151522   176321: contig of 24800 bp in length.
FEATURES             Location/Qualifiers
     source          1. .176321
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-95H12"
                     /clone_lib="CHORI-211"
     misc_feature    1. .7786
                     /note="assembly_fragment:00181
                     fragment_chain:1
                     clone_end:T7
                     vector_side:left"
     misc_feature    7887. .14905
                     /note="assembly_fragment:00055
                     fragment_chain:1"
     misc_feature    15006. .44564
                     /note="assembly_fragment:01666
                     fragment_chain:1"
     misc_feature    44665. .59129
                     /note="assembly_fragment:00681
```

```
                     fragment_chain:1"
     misc_feature    59230. .64301
                     /note="assembly_fragment:00012
                     fragment_chain:1"
     misc_feature    64402. .76420
                     /note="assembly_fragment:00402
                     fragment_chain:1"
     misc_feature    76521. .86373
                     /note="assembly_fragment:00284
                     fragment_chain:1"
     misc_feature    86474. .96468
                     /note="assembly_fragment:00541
                     fragment_chain:1"
     misc_feature    96569. .123997
                     /note="assembly_fragment:01078
                     fragment_chain:1"
     misc_feature    124098. .145113
                     /note="assembly_fragment:00834
                     fragment_chain:1"
     misc_feature    145214. .151421
                     /note="assembly_fragment:00109
                     fragment_chain:1"
     misc_feature    151522. .176321
                     /note="assembly_fragment:01370
                     fragment_chain:1
                     clone_end:SP6
                     vector_side:right"
ORIGIN

  Query Match             5.2%;  Score 56.6;  DB 2;  Length 176321;
  Best Local Similarity   38.4%;  Pred. No. 0.3;
  Matches  200;  Conservative    0;  Mismatches  319;  Indels    2;  Gaps    1;

Qy         24 AACTTTCACTCTCTATGTTCTTATAAGCAAATGTTACAATGAGATATGAGTTTTAAAGCC 83
              ||   | | | | | |  ||  |||  | |||   ||||||  |    |  || || ||
Db      59547 AAAACTAATTTTGTTTATAATTTAAAGTGATTGTCTAAATGAGCTTGTAAATTAAAGGCT 59488

Qy         84 AGATCTTCCTTATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGT 143
              | ||  |  |||  |              ||  | | ||  | | | | |||| |  ||
Db      59487 AAATAGT--TTAGTTTATATATTATAGTCATGATATTATATTATATTATATTATATTATA 59430

Qy        144 TGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGA 203
              | | |    | |   ||  | || |  | ||| | |  || |  || | || ||| |
Db      59429 TTATATTATATTAATTTTTTATATTATATTATATTATATTATATTATTTTATATTATATT 59370

Qy        204 CATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTAT 263
                |  | |  || | || |   ||  || | ||| |   |  ||| |        |  | |
Db      59369 ATATTATATTATATTTTATTTTATTTTTTATATTATATTATATTAGATTTTATTAAATTT 59310

Qy        264 TCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATA 323
              |  | | || |    |  || |  |||  |    | |   |  |  |    |  ||  ||
Db      59309 TATTATATTATATTATATTTTATTAAATTTTGTTATATTAGATTATATGATATTATATTA 59250

Qy        324 TATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAG 383
              |||||    |  | ||||
Db      59249 TATTATATTATATGATATTANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 59190

Qy        384 TCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAG 443

Db      59189 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 59130

Qy        444 AAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTAT 503
                ||||    | | ||||||||     | || | | |  | || || ||||||  | | |
Db      59129 CTATTATATTATACTTTATTTTATTTTTTATATTATATTATATTATATTTTATTAAATTT 59070

Qy        504 TAAAACATTTACTTTCCAAATATTCTGTCATTCAGGAGTAT 544
              ||  | |||   ||      ||||   |  ||   | |||
Db      59069 TATTATATTATATTATATTTTATTAAATTTTTTTATATTAT 59029


RESULT 22
CNS01RGG
LOCUS     CNS01RGG              176368 bp    DNA     linear   PRI 13-NOV-2001
DEFINITION  Human chromosome 14 DNA sequence BAC R-933I1 of library RPCI-11
            from chromosome 14 of Homo sapiens (Human), complete sequence.
ACCESSION   AL157971
VERSION     AL157971.4  GI:16944346
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 176368)
  AUTHORS   Heilig,R., Petit,J.L., Vico,V., Dasilva,C., Robert,C., Wincker,P.,
            Brottier,P., Cattolico,L., Barbe,V., Pelletier,E., Artiguenave,F.,
            Levy,M., Eckenberg,R., Bruls,T., deBerardinis,V., Cruaud,C.,
            Gyapay,G., Saurin,W. and Weissenbach,J.
  TITLE     Sequencing of the human chromosome 14
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 176368)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-NOV-2001) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     On Nov 15, 2001 this sequence version replaced gi:14715170.
            -------------- Genome Center
            Center: Genoscope / Centre National de Sequencage
            Center code: GS
            Web site: http://www.genoscope.cns.fr/
            Contact: SeqRef@genoscope.cns.fr
            --------------
            The following BAC sequence is oriented from the T7 to the SP6 end.
            -------------- Summary Statistics
            Assembly program: Phrap; version 2.0
            Quality coverage: 7.53x in Q20 bases; sum-of-contigs
            --------------
            ----------------------
            Overall quality chart :
            Range     :   bases
            0         :
            1  -  9   :       5
            10 - 19   :      25
            20 - 29   :      67
            30 - 39   :     185
            40 - 49   :    2535
            50 - 59   :    5835
            60 - 69   :    7981
            70 - 79   :   20287
            80 - 89   :   49616
            90 - 99   :   89832
            ----------------------
            Percentage of bases with a quality value >= 40 : 99 %.
FEATURES              Location/Qualifiers
     source           1. .176368
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:9606"
                      /chromosome="14"
                      /clone="R-933I1"
                      /clone_lib="RPCI-11"
ORIGIN

  Query Match             5.2%;  Score 56.6;  DB 9;  Length 176368;
  Best Local Similarity   48.5%;  Pred. No. 0.3;
  Matches  184;  Conservative    0;  Mismatches  194;  Indels    1;  Gaps    1;

Qy        144 TGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGA 203
              | | || |||  || |||| | || |  | |||   || || | || | ||  || | |
Db      56337 TAATAATTATATATTATTATATATAATATTATATAATTATATATTATATTATTATATATA 56396

Qy        204 CATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTAT 263
                |  | |  |||||||      |     || | || || |   || | |   | |  ||
Db      56397 ATATTATATAATAATATATTATTATTATTATATATAATATTATATATAATAATATATAA 56456

Qy        264 TCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATA 323
```

```
Db      56457 TATATATTATTATATAATAATATATTATATATTATTATATATAATATTATATAATAATTA 56516

Qy        324 TATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAG 383

Db      56517 TATATTATATTATAATATTATATATAATATTATATAATAATATATATTATATTATAATAT 56576

Qy        384 TCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAG 443

Db      56577 ATAATTATATATAATATAAAAATAATTATATATTATATATTAATATTATATATAATAATT 56636

Qy        444 AAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTAT 503

Db      56637 ATATATTAAATAATAATATATATTATATATAATAT-ATATTATATATTATTATATGTAATAT 56695

Qy        504 TAAAACATTTACTTTCCAA 522

Db      56696 TATATATTATTCTGTATAA 56714


RESULT 23
AB063521
LOCUS       AB063521              348600 bp    DNA     linear   BCT 20-FEB-2004
DEFINITION  Wigglesworthia glossinidia endosymbiont of Glossina brevipalpis
            DNA, complete genome, section 1/2.
ACCESSION   AB063521 BA000021
VERSION     AB063521.2  GI:25165955
KEYWORDS    .
SOURCE      Wigglesworthia glossinidia endosymbiont of Glossina brevipalpis
  ORGANISM  Wigglesworthia glossinidia endosymbiont of Glossina brevipalpis
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Wigglesworthia.
REFERENCE   1
  AUTHORS   Akman,L., Yamashita,A., Watanabe,H., Oshima,K., Shiba,T.,
            Hattori,M. and Aksoy,S.
  TITLE     Genome sequence of the endocellular obligate symbiont of tsetse
            flies, Wigglesworthia glossinidia
  JOURNAL   Nat. Genet. 32 (3), 402-407 (2002)
  MEDLINE   22297718
   PUBMED   12219091
REFERENCE   2  (bases 1 to 348600)
  AUTHORS   Yamashita,A., Oshima,K., Furuya,K., Yoshino,C., Shiba,T. and
            Hattori,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JUN-2001) Masahira Hattori, Kitasato University,
            School of Science; Kitasato 1-15-1, Sagamihara, Kanagawa 228-8555,
            Japan (E-mail:hattori@genome.sci.kitasato-u.ac.jp,
            Tel:81-42-778-9926, Fax:81-42-778-9926)
COMMENT     On Nov 21, 2002 this sequence version replaced gi:19071885.
FEATURES             Location/Qualifiers
     source          1. .348600
                     /organism="Wigglesworthia glossinidia endosymbiont of
                     Glossina brevipalpis"
                     /mol_type="genomic DNA"
                     /specific_host="Glossina brevipalpis"
                     /db_xref="taxon:36870"
                     /note="equivalent name: Wigglesworthia brevipalpis"
     gene            182. .2068
                     /gene="gidA"
     CDS             182. .2068
                     /gene="gidA"
                     /note="Unknown function"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAC24147.1"
                     /db_xref="GI:25165956"
                     /translation="MCSCYKEIDIIIVGGGHAGTEAAAVAAKMKCNTMLITQNIDTIG
                     QMSCNPSIGGIGKGHLVKEIDAMGGIMGIAADNSGIQFRTLNSSKGPAVRATRAQIDR
                     KLYSSLIQKILSNQKNLYLLQQEVIDILVEKNKIKGVVTNAGLIIKSKIVIITTGTFL
                     NGKIHIGNKQYVSGRSGEFSSIKLSQILKEKFCFRINRLKTGTPPRIDLKSVNFSKLE
                     SQYGDTPIPQFSFMKFVGIRPKQLPCYITYTNKNTHNIIIDNIKHSPIYNGSIKGIGP
                     RYCPSIEDKVVRFSKKESHQIFLEPEGLNNYEIYPNGISTSLPLNLQIKIIRSIKGLE
                     SANITRPGYSIEYDFLDPRDLKLTLESKIISGLFFAGQINGTTGYEEAAAQGLLSGIN
                     AALMLKNGSFWFPKRSQAYLGVLIDDLCTLGTKEPYRMFTSRSEHRLIIREDNADLRL
                     TEIGRKLGIIEDDRWNNFYNKKNIIYKEIKNLKNTKILFNKEEKNKINLNINKSGSIT
                     AEEILRRPEITYNDLICYNFIKRKLKDEEIIKQIEIQIKYAGYIKRQEKEINKKTYYD
                     NIILPNEINFKKISGLSKEVIEKLNIFKPTSIGQAIKISGITPAAISIILIWLKKNNL
                     MYNK"
     gene            2230. .3066
                     /gene="atpB"
     CDS             2230. .3066
                     /gene="atpB"
                     /note="Energy metabolism"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAC24148.1"
                     /db_xref="GI:25165957"
                     /translation="MTIYNQSLSSKEYISHHLKNLQIDLRTFKIVQENDSSSFWFLNI
                     DSLFFSFFLGIIFLLFFNYISKKFTIGTPGRIQASIEILVEFINSNVKDIFGHNTNKI
                     IPPLSLTIFVWLFLMNLMDLIPVDFVPYIANFIFNISELRIVPSSDINITLSMSLGVF
                     LLILYFGIKNKGIVVFLKEFFFQPFNNYLLIPVNLVLELISLLSKPVSLSLRLFGNIY
                     SGELIFILISGLLPWWLQWILSVPWAIFHILIIILQAFIFMILTIIYLEMSLEKPDKN
                     TI"
     gene            3076. .3315
                     /gene="atpE"
     CDS             3076. .3315
                     /gene="atpE"
                     /note="Energy metabolism"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAC24149.1"
                     /db_xref="GI:25165958"
                     /translation="MNSINADLLYISAAIIMGFASIGAAIGIGILGGKFLEGAARQPD
                     LIPTLRTQFFIVMGLVDAIPMISVGLGLYLIFAAS"
     gene            3328. .3813
                     /gene="atpF"
     CDS             3328. .3813
                     /gene="atpF"
                     /note="Energy metabolism"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAC24150.1"
                     /db_xref="GI:25165959"
                     /translation="MNINATIFGQTIAFFLFVFFCMKYIWPNLISLVEKRRENIAQAL
                     NEAKQAKLNLKISKEKAKKRIESAQIKCKNIINEANETKKLLIEEAKKEAIKIKEHII
                     SQGRLDILDEKKRMCEDLKTKISEIIVMSVEKIIESSINKKISDNIIERSISKINKIK
                     G"
     gene            3817. .4347
                     /gene="atpH"
     CDS             3817. .4347
                     /gene="atpH"
                     /note="Energy metabolism"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAC24151.1"
                     /db_xref="GI:25165960"
                     /translation="MTIYEIKSYLSRPYAEAAFKFANENNIVDDWIYMIEEICKILKN
                     IKIYSLSSFILKKNKNFLFESIKNNFDVYFNNFVKIIIENNRLIIFPEILNQFINLKN
                     DQNNIENITIISKYKLDKKILNKIEKKIKSLIFKEIKINIKIDKSIIGGYIIKSKNFL
                     IDNSIKNKLYRFSEYL"
     gene            4368. .5900
                     /gene="atpA"
     CDS             4368. .5900
                     /gene="atpA"
                     /note="Energy metabolism"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAC24152.1"
                     /db_xref="GI:25165961"
                     /translation="MQLRSSEISELIKYRISKFNAESEMQDEGIIISISDGIIKIYGL
                     NNIMQNEMIKLPEKTFAIALNLEKEIVGAVVMGSYFHLSEGMKVYSTGRILEVPTGNN
                     LLGRVLNALGEPLDSKGDILHDCFSPIETIAPGVIDRLSINEPLQTGYKSIDSMVPIG
                     KGQRELIIGDRQTGKSSLAIDTIINQKDSNIKCIYVSIGQKFSTVVKTVEKLKEHDAL
                     SNTIIVCATASDSAVLQYLAPYSGCTMGEYFRDRGEDSLIVYDDLSKQAIAYRQISLL
                     LRRPPGREAYPGDIFYLHSRLLERSARVNEIYVENFTKGKVKNKSGSLTALPIVETQG
                     GDISSFVPTNLISITDGQIFLESNLFNSGIRPAINPGISVSRVGGSAQTEIIRKLSGN
```

```
                IRTSLAQYNELSSFAQFSSDLDDSTKKQLQHGEKIIEILKQKQYSPMSLSKQALLLFA
                VQNKYLDKIKSSEVEKYENSLLKHAEKKYSNYMEKINKSYIYDDFVIKKLKKIMDSFE
                FF"
     gene       5954. .6817
                /gene="atpG"
     CDS        5954. .6817
                /gene="atpG"
                /note="Energy metabolism"
                /codon_start=1
                /transl_table=11
                /protein_id="BAC24153.1"
                /db_xref="GI:25165962"
                /translation="MYNIKDIRQKIIGIKKTQSITSAMEKISAIKMKKSQNLLNSTKP
                YYEKISLLLNKLLFQNIKYKHPYLKKRQIKCIGYIVVSTDRGLAGSLNINLFKKLLYE
                INKSKEKKINTKLVLIGSKAISFFRFFEDEIISTISGIGDTPKISELIKPVQIMLKEY
                DSNIIDKIYIISNKFINTMTYVPDIKKVLPISIKKSSIKFKNLNYLYEPDFSSLFESI
                LPRYIESLIYQSIVENISSEQSARMVAMKSAMDNSKNLIEELSLIYNKARQSNITKEL
                TEIIAGSNNIY"
     gene       6842. .8230
                /gene="atpD"
     CDS        6842. .8230
                /gene="atpD"
                /note="Energy metabolism"
                /codon_start=1
                /transl_table=11
                /protein_id="BAC24154.1"
                /db_xref="GI:25165963"
                /translation="MDTGKIVQIIGPVIDVKFEISFIPKINFALEIKIKNKKLIMEVM
                QQIGDGIVRCILMGSSDGLKRGLKVFNLNKGIEVPVGKPTLGRIMNVLGDPIDRKGKI
                LSSEYRPIHNKSPSYNEISSSDELLETGIKIIDLICPFAKGGKIGLFGGAGVGKTVNM
                MELIRNIAIEHSGYSVFSGVGERTREGNDFYNEMISSNVIDKVSLVYGQMNEPPGNRL
                RVALTGLTLAEKFRDEGNDVLFFVDNIYRYTLAGTEVSALLGRIPSAVGYQPTLSEEM
                GILQERITSTKSGSITSVQAIYVPADDLTDPSPATTFVHLDATVVLSRQIASLGIYPA
                IDPLDSTSRQLDPLIVGEDHYLVARKIQSILQRYKDLKDIIAILGIDELSEEDKILVS
                RARKIQRFLSQPFFVAEIFTGFSGKYVTLKDTIDGFNSIINGEYDYIPEQAFYMTGNI
                SDVLDKYKKFYS"
     gene       8242. .8658
                /gene="atpC"
     CDS        8242. .8658
                /gene="atpC"
                /note="Energy metabolism"
                /codon_start=1
                /transl_table=11
                /protein_id="BAC24155.1"
                /db_xref="GI:25165964"
                /translation="MNNKFYQLTVISAENFIFSNLVKKSNITGSEGNLGILPGHAPLI
                TKIKPGLIHIISSKNIEEYIYLSGGILEINSNVVTVLADTAIRGKEIDEKKAIESKKN
                AEELILKSKERHEYIRAVIELSKAIAKLRVCKLSNK"
     gene       8848. .10233
                /gene="glmU"
     CDS        8848. .10233
                /gene="glmU"
                /note="Cell envelope"
                /codon_start=1
                /transl_table=11
                /protein_id="BAC24156.1"
                /db_xref="GI:25165965"
                /translation="MNVLLQELFIIILSAGKGKRMFTSIPKVLHKLAGKPILQHIIDK
                VLYLKAKKINIVYGYEGKLLRKKIISRGFSLNWTLQSEQNGTGHAVQQVIFKEIGNDN
                DKILILYGDVPLISINTLRKLLYSHSGFNISLLTAVIDSPDGYGRIIRKNGKISKIIE
                HEDALSIDKNIREINTGVMVVNRYYLRFWLNKLLNKRNKDKEIYLTDIISIAYKSGNI
                INSVQPEKVFEIFGINNRFQLMKLEKIYQIEQAKKLLLNGVTLSDYNRFDLRGTLKHG
                KDIFIDINVVLKGSVIIGDRVKIGNGCVLKNVIINNDVIIHPYSIIEDACLDSNSVIG
                PFAHIHSKSKIKKNVHVGNFVEIKNTIFGKNSKVGHLSYLGDSDIGKNVNIGAGTITC
                NFDGKKKNKTIIKNNVFIGANSELIAPVIINSGAVVGAGTTVTKNINRKDKIISRIRQ
                FSILRKENNSK"
     gene       10399. .12237
                /gene="glmS"
     CDS        10399. .12237
                /gene="glmS"
                /note="Central intermediary metabolism"
                /codon_start=1
                /transl_table=11
                /protein_id="BAC24157.1"
                /db_xref="GI:25165966"
                /translation="MCGIVGIVSQRNILKFLLTGLNHLEYRGYDSSGLAVIDNNNKLR
```

```
  Query Match             5.2%;  Score 56.4;  DB 1;  Length 348600;
  Best Local Similarity   44.7%;  Pred. No. 0.3;
  Matches  219;  Conservative    0;  Mismatches  271;  Indels    0;  Gaps    0;

Qy         46 ATAAGCAAATGTTACAATGAGATATGAGTTTTAAAGCCAGATCTTCCTTATCTCTCTGCC 105
              ||||  |||| ||| ||  |   || |  |||       | |  ||  ||||   ||  |
Db     221936 ATAAATAAATATTAAAAATATGAATAAAATTTTTTTAAAAACTTTTTTTATAGATCGTTC 221995

Qy        106 CCATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAA 165
              | | | ||     | |  |   | | || |       || | |||  |  ||| ||   |
Db     221996 ACGTTTTTAAAAAATTATTTTATTTAAATAATACATATTAAAAAAATTACATCTTTTGTA 222055

Qy        166 ATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCC 225
              ||   | | ||      |       | |      | | |     || |    |  |
Db     222056 ATTGATAATTATAAAAATTATTTTAATAAATGCTTAAAAAAATAATTTTATTAAATATTT 222115

Qy        226 CTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCA 285
                ||  | || | | |   || | | |   |    || ||| | |  ||| || |
Db     222116 TAGAATGATATTTATATAAAATAAATAAAAGAATTATTTTATTATTAATAACTTTTTTAA 222175

Qy        286 AAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAA 345
              |||||   || || ||||  |      || | |    | ||||        | | |||
Db     222176 AAAAATAAATTAAAAAATAGAAGTTTTGATAAAAACAAAATTATAATTAATTGATTAAAA 222235

Qy        346 CACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTA 405
                | | | | |||| |   ||   ||| ||    | ||     |||| |  |   |
Db     222236 ATTATATTTAATATATATAAATATATAAAATATTTATATAAATATCTAAAAATTTAGAAA 222295

Qy        406 AAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTT 465
              |    | | |  |   |  | | | | ||   |||| |  |  | |||| |||
Db     222296 ATCTAGTTCAAATTAATAAAATGTTGTAATTTAATATAAAATAAAAAATATTTTTTATTT 222355

Qy        466 GGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATA 525
              ||||||  |   ||| || | ||||      |  || | ||  | | ||   ||| |
Db     222356 TTAAATAAAATTTAAAATCAAATTTTTAATAAAAAAATAATAAATATAAATTAATAAAAA 222415

Qy        526 TTCTGTCATT 535
              ||     |||
Db     222416 TTAATAAATT 222425
```

```
RESULT 24
AC116924
LOCUS       AC116924               88549 bp    DNA     linear   INV 21-FEB-2004
DEFINITION  Dictyostelium discoideum chromosome 2 map 6357117-6445670 strain
            AX4, complete sequence.
ACCESSION   AC116924
VERSION     AC116924.3  GI:42733690
KEYWORDS    HTG.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 88549)
  AUTHORS   Gloeckner,G., Eichinger,L., Szafranski,K., Pachebat,J., Dear,P.,
            Lehmann,R., Baumgart,C., Parra,G., April,J.F., Guigo,R., Kumpf,K.,
            Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and
            Noegel,A.A.
  TITLE     Sequence and analysis of chromosome 2 of Dictyostelium discoideum
  JOURNAL   Nature 418 (6893), 79-85 (2002)
  MEDLINE   22092622
   PUBMED   12097910
  REMARK    The Dictyostelium Genome Sequencing Consortium
REFERENCE   2  (bases 1 to 88549)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-APR-2002) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   3  (bases 1 to 88549)
  AUTHORS   Baumgart,C.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   4  (bases 1 to 88549)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   5  (bases 1 to 88549)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-FEB-2004) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
COMMENT     On Feb 21, 2004 this sequence version replaced gi:28829074.
            CDS predictions from GeneID do not necessarily reflect true genes.
            Further Information is available from IMB Jena, Department of
            Genome Analysis
            (http://genome.imb-jena.de/dictyostelium/)
            and the Univerity Colonge, Institute for Biochemistry I
            (http://www.uni-koeln.de/dictyostelium/project.shtml
            Funding
            Agency : Deutsche Forschungsgemeinschaft (DFG).
FEATURES              Location/Qualifiers
     source           1. .88549
                      /organism="Dictyostelium discoideum"
                      /mol_type="genomic DNA"
                      /strain="AX4"
                      /db_xref="taxon:44689"
                      /chromosome="2"
                      /map="6357117-6445670"
     CDS              1864. .2241
                      /note="GeneID exon scores (in order of location ranges):
                      27.07 - GSCJ_ID dd_01259"
                      /codon_start=1
                      /product="hypothetical protein"
                      /protein_id="AAO51640.2"
                      /db_xref="GI:42733691"
                      /translation="MDQFKVPPPTYAPVAGQTIYGAPSYNNYYMRTPQVSTAPMIYPT
                      PMMAPPIMTPPIMTQPIMTPPMMYPPIIPSQPFMGPSMVSPIMSPPMIPSQPFMGPSM
                      VSPGYGVTPNLNVPFSTNVKYYD"
     CDS              2839. .4479
                      /note="GeneID exon scores (in order of location ranges):
                      88.56 - GSCJ_ID dd_01260"
                      /codon_start=1
                      /product="hypothetical protein"
                      /protein_id="AAO51641.2"
                      /db_xref="GI:42733692"
                      /translation="MDCKIHNEKNIMWCLNCTTLCCVDCIEFDHFQHNLKSFKNIDKF
                      HCPTHDEPIIKYCILCNCVICNLCEKTTHFHNNNNNQINITQFPWIPKKKSKSIELEL
                      NEIDEYNNNYNNNINNNNNNNNKNNVNNVNNVNNENKEINENHNNKNQRSNSPPPLPK
                      IKLLNIENITLIKHGSYRNLNKIEEENENDEKEILDRKQYNKENKEFLKKQSEKGNSI
                      AQCLLGSSYYYGMNGYSVNFEKSLKYLTKSAEQNNAKAYFELFNYYKQQKKDLIKSKY
                      YLLKSAEFIYEDSIILLGDYYYNGFKTIDNHDDHDDGSGENEELNLICQDYFKAFNLF
                      KIGASLNVDSNSLTKLGICFYYGRGVTIDYNEAYRLFYQSSICDHHQEHNPEITFYYL
                      GLCLFYGKGVLKNQCKGFEYFMKSASLNHYSPALEAVGRCFLNGEGISQNFMQAKLYF
                      TTAKSQGSNQDRNIQDVNKIIELNKSIKIKSPNILIEIEKIINENKSFNLEQSCLNYI
                      NHQKNNYKQEIDNTNIKNIKCKQDSLIDKTFNKVNYLN"
     CDS              complement(join(4673. .5239,5354. .6118,6205. .6621,
                      6720. .7982))
                      /note="GeneID exon scores (in order of location ranges):
                      34.62, 63.42, 31.78, 100.55 - GSCJ_ID dd_01261"
                      /codon_start=1
                      /product="hypothetical protein"
                      /protein_id="AAO51642.2"
                      /db_xref="GI:42733693"
                      /translation="MEIDKNVEQSSNKVEYNQPKSPITSILVNNNSSNNNNNSNNKNN
                      NNEEEFDISSPLTLSPIANCGASGSNIDSSIDNTPHLLSQISTPPISSTSSITLLNDK
                      KKGIYSTFLDLDYYNNTYLKNKQKILQPKLNQQKQQQQQQQQQPKTLTFSLDSQQQQQ
                      QQPQLQSPQLIDQQQQQQQQQQQQQQQQQQQQQQQQIPYHQRKHHISIRDGDDNDNEVDS
                      FNILSQNQLIHQTFNNPQNHHYFSSTSPSLLYHHQQQTQLEQKPIYSSSSSSSILSSL
                      SSSTISFYNSITSLFLTFPQTISTDPIIIREYERLKMLQEQQVGPRHTEILLRKEDGV
                      GGRKALKEKFVEIYESFFNGEDPSENQPIFWEQLFLIKVNIPPLERCIILTSEENLLA
                      LKQNINKIFTQSCFWLKDSNSLRLTHIIETLSILLKSIFRKKFNNFGFDILNILCGIE
                      NSDQVFTDMIKDLEQLLKNSSIDKKVKIGVINLFTIIATATEYINQNTLLQYFMSVNI
                      FDTLVTFLGEDDIPKQTQLNILTLLCYFSNFQKYEITNPYLKDLSNLSDPIKILKILN
                      LITQNLKDHNEFFEQQYRDPNLGMVSRFSGYLSSWVYSPKVVQPLSFYETGSSLMILY
                      ELFYQNDHFVNMLVKGGLPFANGGENGSTCPDIIKQFFSFCAYLASDAQKDLKTIYSR
                      MALLILLCICEKRELEDFFHDIKSTCYIFIYSKKSLFSDPKETEKHPLSCWTLDILSQ
                      FIKCNLKGKISMDVHQSAVDCIHRIVSYKKKTLNRLPYKWSELWSALFSLISMVSTLN
                      ISNDKIQLNQSIHVGISTVNIFNLFINYGDSFLPSPNDYDELFYEIIRSGQVIENFYQ
                      FILNQTTTGGQSVTSPSAAASAFNPTTPTPLSPQQLPTITTTTTTTITSNNESNNPLLN
                      QLLNIRSIVQHFTSKLEEWSANNAEVALTAPQVSKIIKDNYDTLRLKLQENLDQFEPY
                      VENPKEVLFFKHIIKELINDLRKQINVTNII"
     CDS              complement(9007. .10431)
                      /note="GeneID exon scores (in order of location ranges):
                      117.93 - GSCJ_ID dd_03414"
                      /codon_start=1
                      /product="similar to Plasmodium falciparum. Hypothetical
                      protein"
                      /protein_id="AAS38634.1"
                      /db_xref="GI:42733694"
                      /translation="MESYSPSIQKLCVFLKPESDISNTGISYLNNFLADMSNAIIVES
                      IHSSKSHKHKTILIEDIESALNVVFPSEIADSIINQSKGRIQSTSLKNLMQVKTFNNG
                      GNNNNNNEQDNDLQEKEQNELSDDSDVEGFEIIPNNNNIYSHDVSLCFPIDLVCRVTL
                      DYICRKKVIHQNPSDFYKVIEQQQSNHVIYTYISYALEVITTKILIKISSSTDLIDED
                      IIRKVIQDNHHLNSVRHKNIQASNIAGKLLDSSSSSNSNLQTSLSSSFIFSSSSNNTT
                      TFTQQEQQEQFNDDKISGDEIEFEINNSKYLEKEDKKDQLIQKDEEIILKTFNDNNNN
                      NNNNNNNNNNNNNNNNNNININNDNDNNNNNNKNNDNKNYDNNNNNNADKIEKLSPVN
                      NNVKNSEIEKNIEKLKNINNNNNVEINEALISIIKETCKNIPSSSFSSVKNLKRSPPS
                      LFKFKEPIDLVQVLAEGGIHMEAT"
     CDS              complement(11849. .12583)
                      /note="GeneID exon scores (in order of location ranges):
                      50.22 - GSCJ_ID dd_02577"
                      /codon_start=1
                      /product="hypothetical protein"
                      /protein_id="AAS38635.1"
                      /db_xref="GI:42733695"
                      /translation="MKLLFGLILLIVNYCYLINANKLFVNFVGTYNSESCSESPIGTG
                      FSMALGQCVQMQGIFSYVTSSSPEATIWMATLGGDGNSFTLKQFHSSDKLCTEDPINT
                      VKYNKFDTCVQQPSFLTSNNSLLANNNNNNNNNNNNNNNNNNNNNNLVYSKLTLSQNAPS
                      YAPNSVIIARYNSSDSVCNYNNQNYGISYSTGSVIYNENGVSVYVVCDQSNQSILFHC
                      QNDQCVQSFTDTSCDMPPNDLQIYCN"
     CDS              complement(13151. .13840)
                      /note="GeneID exon scores (in order of location ranges):
                      57.23 - GSCJ_ID dd_02576"
                      /codon_start=1
                      /product="hypothetical protein"
                      /protein_id="AAS38636.1"
                      /db_xref="GI:42733696"
                      /translation="MKLLLTLILIIVNYCCFINANKLFVNFVGSYSSNTCDGNPSGIG
                      FSLAVGQCLPIEGVLSYVINDMPPTTNWFITMGGNGNSFTLKQYDSQDQWCKQDPINT
                      LQFDNFDTCVEQPAFINVNNSLLADTPVVYSKLTLSQNVPTYAPDSVIIGQYNTNTNV
                      CTYNDQQFGVSYTSGTVIYNENGISIHVLCNQVNQSMMFKCKESKCGETFIDTYCQMS
                      GDVQIYCNSNN"
     CDS              join(20802. .21095,21199. .21978)
                      /note="GeneID exon scores (in order of location ranges):
                      23.32, 97.21 - GSCJ_ID dd_01856"
                      /codon_start=1
                      /product="similar to Dictyostelium discoideum (Slime
                      mold). Guanine nucleotide-binding protein alpha-2 subunit
                      (G alpha 2)"
                      /protein_id="AAS38637.1"
                      /db_xref="GI:42733697"
                      /translation="MGICASSMEGEKTNTDINLSIEKERKKKHNEVKLLLLGAGESGK
                      STISKQMKIIHQSGYSNEERKEFKPIITRNILDNMRVLLDGMGRLGMTIDPSNSDAAV
                      MIKELTSLQASIVTDCWGELNEDQGKKIKALWTDPGVKQAMRRANEFSTLPDSAPYFF
                      DSIDRMTSPVYIPTDQDILHTRVMTRGVHETNFEIGKIKFRLVDVGGQRSERKKWLSC
                      FDDVTAVVFCVALSEYDLLLYEDNSTNRMLESLRVFSDVCNSWFVNTPIILFLNKSDL
                      FREKIKHVDLSETFPEYKGGRDYERASNYIKERFWQINKTEQKAIYSHITCATDTNNI
                      RVVFEAVKDIIFTQCVMKAGLYS"
     CDS              join(22686. .22743,22820. .22926,23004. .23139,23232. .23558,
                      23639. .27180)
                      /note="GeneID exon scores (in order of location ranges):
                      -4.93, 3.57, -1.17, 18.69, 174.64 - GSCJ_ID dd_01857"
                      /codon_start=1
                      /product="hypothetical protein"
```

```
                     /protein_id="AAS38638.1"
                     /db_xref="GI:42733698"
                     /translation="MNRNIHLYIFVSILLSLINGIKSIKVGNGDFSVIDITDSFDDLV
                     TLYNDQYSIGGLDENIFWLKVINTVQIGENQKFNLLLTLDNSTTIQFDDFFTLTCKER
                     PVNIERELVGQKLFIDKQYKGRYGILVHNIGLEQYLSVPGLENNFISNTFTQGYYDDS
                     IFYTDVSKNYTIFKPVIKRSASVISQLSIDTIFGSVSTSTKLYLTSFLNSTYFKNPTI
                     APLQHDNNCSSLVYDQCSLLFVVQADFQRPGIYFLMNMDSNSNLPELVYGNKSYGHWL
                     FTIKNIAERQGSYIYPISSFDSTLGFSTIFSRNISYFGPTGVNLPVIQQSYFGTQIQN
                     SIFISFSKFVIPTTVNYGDIFIDNGFNTLSHFKYPYFINTINAQGGATNTLWVPSSPY
                     QKSIALEITLQFYFIKYPSPILVGGISGIASFKESIPPVFENIEYFDSVSDDCIVRIS
                     ASDEGGSCIQSIAISFIGAYEPSKVQSIFRITNRNLIEGDCFKGIYQISIPKYSIRRN
                     YFIEMNDYANNLVKLYPNNFYSGKRYLGSFPYQVLTLEDIESIGFRNNNIDLSNSNSS
                     NVLYIKLKDQSISNLVFQFIPKLGIDTYEPNNKISGTPRFFKSKWNSASSIYEVSFDL
                     PARLPTGPISYILNPLEIDSSILYQYFGSPSELNVYSQIADIMGPIALSYNTIGNTTI
                     PSINEPPNTIGFQILFGDYLNGISQVNITVTSDKTYKQMFKNFTFTEPSLGKWIQFTW
                     EIKYGDIPQTFSISELSAIDSNGYITEFPNPTKVNPLMLLVELYPSFSIQTSGETPTD
                     VTPPYMQVYHVNQLNIENYLDRTLEINFTVYDGESGIYIEEPPTIYFINDNVEIISKQ
                     TSYISGTSSGLGFFENAIFGVNITLPYLFGYPNGFLTQIHGIYDNKFNVLGATPDIIN
                     QYNRGNSFINVTIGQSQPFIESFVMDIDGDLLTVNGLKFGILPTGKLRINGAIENLNI
                     SLPFLNNTYIVFDLKQYRNINDTNCQIYITNQLGFSSNEINFVLKAHTNVTMPPTTTP
                     HPTSNPTATPSLTANPTPLPTNSPSQCKNDCGGIAQGQCSNLGCICISPYIGIDCSSK
                     VIVDIDPIVNNSQPEVITPYQHENSGNKFSTLVSIYSLRELGFDGKVENDYIFEMWNY
                     QYYNNDNHTYFTSIKNKNDLSKRTNITVSIKMYKQLTTIQFAGENITIQPSSIKYFIS
                     LSDYQFKSNLNTLDLVMYTSIETEKIDDACFSSEFGDSSSQLSSQYLKIQINDVSLYG
                     EFIKRAIVDTHIVGVLNKVRGNNINSTSNGSSSSQYISIQIPNYKTSVELDPNFSVLL
                     ESSKIENSTCNKNQSGLTKTQLIGIIIGSTAGFILILIIVISIVYKKSTTLRININRA
                     KKFRMKKLSFK"
     CDS             complement(join(27340. .29056,29330. .29490))

  Query Match             5.2%;  Score 56.2;  DB 3;  Length 88549;
  Best Local Similarity   48.1%;  Pred. No. 0.41;
  Matches  189;  Conservative    0;  Mismatches  203;  Indels    1;  Gaps    1;

Qy        131 ATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGAT 190
              |||| |  |   |  | | |||| ||  | ||    | |   ||||  |||  |  | ||
Db      74583 ATATAAAGTATATATATATATATAGAAGAATAACTTTTAAGCAATAAATTTTAATCAAAT 74642

Qy        191 CTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACA 250
              | |    | | ||  |  ||| |||| ||  |   ||  ||||| |           |
Db      74643 ATAAATAAATAAACTATAGTTTATTAAAATCGTTTTTTTTTTTTTCTTTCTTTTTGAAAT 74702

Qy        251 GACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCA 310
                | ||    | | || |||||     | ||||||  |||||| || | .||  |
Db      74703 ATTTAATAGAACATTACTTTTTTTTAAAAGAATAAAAAAAAATATAAACAATTCAATATA 74762

Qy        311 GAGAAAGATCATATATTAGCCAGTTG-AAAATTAAACACAAAATGAGTGCATATTACATT 369
              |  |  |||||| ||||   | ||  |||| || | |   ||| |    | ||   | |
Db      74763 AAACACAATCATACATTATAAACTTTAAAAAATATAAAATTAATTAAAATAAATACAAAT 74822

Qy        370 ACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTAT 429
              |   ||   | |||  |||  || |||         ||||  |    | |    ||  |
Db      74823 AAACAAATATACAGATAAATATATAAAAAATGATGAAAGTAAATAGATCACAATTTAAAG 74882

Qy        430 CGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAAT 489
                |||  | | | ||| ||  |  |||  || |    ||| ||         | | ||
Db      74883 AAAACTTTATAAAAAAAATAATAATAATAATAATAAAAAAACAAATAATATTGATACAAA 74942

Qy        490 TTTTTATGTACTATTAAAACATTTACTTTCCAA 522
              | ||||| ||   |  |     |   | || |||
Db      74943 TATTTATTTATATTATAGTATTCCCCATTTCAA 74975


RESULT 25
CR759861
LOCUS       CR759861              121958 bp    DNA     linear   HTG 13-SEP-2004
DEFINITION  Danio rerio clone DKEY-193P20, *** SEQUENCING IN PROGRESS ***, 6
            unordered pieces.
ACCESSION   CR759861
VERSION     CR759861.3  GI:52000372
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 121958)
  AUTHORS   McLay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 13, 2004 this sequence version replaced gi:51988222.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK193P20
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 119459 bases at least Q40
            Consensus quality: 119870 bases at least Q30
            Consensus quality: 120219 bases at least Q20
            Insert size: 121458; sum-of-contigs
            Insert size: 173520; 3.2% error; agarose-fp
            Quality coverage: 9.53x in Q20 bases; sum-of-contigs Quality
            coverage: 6.73x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    27224: contig of 27224 bp in length
            *    27225    27324: gap of 100 bp
            *    27325    44196: contig of 16872 bp in length
            *    44197    44296: gap of 100 bp
            *    44297    95722: contig of 51426 bp in length
            *    95723    95822: gap of 100 bp
            *    95823    98782: contig of 2960 bp in length
            *    98783    98882: gap of 100 bp
            *    98883   108140: contig of 9258 bp in length
            *   108141   108240: gap of 100 bp
            *   108241   121958: contig of 13718 bp in length.
FEATURES             Location/Qualifiers
     source          1. .121958
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-193P20"
                     /clone_lib="DanioKey"
     misc_feature    1. .27224
                     /note="assembly_fragment:00620
                     fragment_chain:1"
     misc_feature    27325. .44196
                     /note="assembly_fragment:00346
                     fragment_chain:1"
     misc_feature    44297. .95722
                     /note="assembly_fragment:01086
                     fragment_chain:1"
     misc_feature    95823. .98782
                     /note="assembly_fragment:00015
                     fragment_chain:1"
     misc_feature    98883. .108140
                     /note="assembly_fragment:00042.0"
     misc_feature    108241. .121958
                     /note="assembly_fragment:00174"
ORIGIN

  Query Match             5.1%;  Score 56;  DB 2;  Length 121958;
  Best Local Similarity   41.2%;  Pred. No. 0.42;
```

```
  Matches  167;  Conservative    0;  Mismatches  238;  Indels    0;  Gaps    0;

Qy        146 AGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACA 205
              | | ||||  ||   || | ||   |  ||| | | || |        || | |
Db     107653 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 107712

Qy        206 TTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTC 265
              | | |   ||       |   |  || | | || |       || | | ||| |  | |
Db     107713 ATATATATATATATATATATATATATATATATATATATATATATATATATATATNTAT 107772

Qy        266 ATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATA 325
              || | ||| |   ||  |  | | |  |||||| |||   |   | |   | || | |
Db     107773 ATATATTTATATATTTATATATATATATATAAAAAAAAAAAAAAAAAAAANTATATAAAAAT 107832

Qy        326 TTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTC 385
              |    |  |    | ||      |||      | |    | ||    ||       |
Db     107833 ATNAANANATNNNNANAAANNNNNNAATNNANAAAAANANAANTAAANNATAANNNNNTN 107892

Qy        386 AAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAA 445
              ||    |||    |    |||  |||   |      |  ||          ||      |
Db     107893 TAAAAATAAATANTATAAAAAAAATATNNNAATAAAATAATTTATNNTNAAAAANATNNA 107952

Qy        446 ATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTA 505
              |||  |  |||  | | |   | ||||    | |||   | |||| |||  || | |
Db     107953 TTTATTANTAATNNTTTATAANANATAANTNNATAAATNTTTATTTATTAAATATTTTAT 108012

Qy        506 AAACATTTACTTTCCAAATATTCTGTCATTCAGGAGTATGGAAGT 550
              ||  ||   ||    ||| |  |  ||| |      ||  || |
Db     108013 TAATTTTNTNTTATATAATTTNATAATATTTANTTAAATATAATT 108057


RESULT 26
AC021553/c
LOCUS       AC021553               185596 bp    DNA     linear   PRI 07-MAY-2002
DEFINITION  Homo sapiens chromosome 15, clone RP11-709B3, complete sequence.
ACCESSION   AC021553
VERSION     AC021553.14  GI:20429558
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 185596)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 15, clone RP11-709B3
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 185596)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 185596)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-MAY-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 185596)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAY-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 3, 2002 this sequence version replaced gi:15706183.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L5662
                Center clone name: 709_B_3
            --------------.
FEATURES             Location/Qualifiers
     source          1. .185596
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA" .
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15"
                     /clone="RP11-709B3"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   1. .381
                     /rpt_family="L1MA4"
     repeat_region   397. .418
                     /rpt_family="(TAAA)n"
     repeat_region   422. .1327
```

```
                     /rpt_family="MER83B-int"
     repeat_region   1328. .1636
                     /rpt_family="AluSx"
     repeat_region   1641. .1953
                     /rpt_family="AluY"
     repeat_region   1954. .1989
                     /rpt_family="MER83B-int"
     repeat_region   1990. .2009
                     /rpt_family="(TAA)n"
     repeat_region   2010. .2635
                     /rpt_family="L1MA4"
     repeat_region   2636. .2905
                     /rpt_family="AluSx"
     repeat_region   2906. .3135
                     /rpt_family="L1MA4"
     repeat_region   3136. .3451
                     /rpt_family="AluJb"
     repeat_region   3452. .3622
                     /rpt_family="L1MA4"
     repeat_region   3670. .4728
                     /rpt_family="L1MA4"
     repeat_region   4729. .4943
                     /rpt_family="AluSx"
     repeat_region   4945. .5282
                     /rpt_family="L1MA4"
     repeat_region   5283. .5302
                     /rpt_family="(CAA)n"
     repeat_region   5303. .5601
                     /rpt_family="L1MA4"
     repeat_region   5602. .5906
                     /rpt_family="AluSq"
     repeat_region   5907. .6220
                     /rpt_family="L1MA4"
     repeat_region   6221. .6526
                     /rpt_family="AluY"
     repeat_region   6527. .7623
                     /rpt_family="L1MA4"
     repeat_region   complement(7624. .7924)
                     /rpt_family="AluSg"
     repeat_region   7925. .8104
                     /rpt_family="L1MA4"
     unsure          8062. .8068
                     /note="<30 qual SNGL region"
     repeat_region   8196. .8501
                     /rpt_family="AluSc"
     repeat_region   complement(8556. .8630)
                     /rpt_family="Tigger5"
     repeat_region   8732. .9247
                     /rpt_family="L1M1"
     repeat_region   9248. .9551
                     /rpt_family="AluJo"
     repeat_region   9552. .9716
                     /rpt_family="L1M1"
     repeat_region   9736. .9786
                     /rpt_family="L1ME"
     repeat_region   9787. .9818
                     /rpt_family="(TG)n"
     repeat_region   complement(9823. .10128)
                     /rpt_family="AluSx"
     repeat_region   10129. .10442
                     /rpt_family="L1ME"
     repeat_region   10469. .10580
                     /rpt_family="L1ME"
     repeat_region   complement(11726. .12033)
                     /rpt_family="AluY"
     repeat_region   12220. .12397
                     /rpt_family="MER33"
     repeat_region   13586. .13638
                     /rpt_family="MER33"
     repeat_region   complement(13639. .13943)
                     /rpt_family="AluJb"
     repeat_region   complement(13953. .14263)
                     /rpt_family="AluSx"
     repeat_region   14264. .14346
                     /rpt_family="MER33"
     repeat_region   14443. .14832
                     /rpt_family="MER45C"
     repeat_region   15220. .15468
                     /rpt_family="MER45B"
     repeat_region   15469. .15608
                     /rpt_family="L2"
     repeat_region   15609. .15984
                     /rpt_family="(TA)n"
     repeat_region   16034. .16224
                     /rpt_family="(TA)n"
     repeat_region   16248. .16441
                     /rpt_family="(TA)n"

  Query Match          5.1%;  Score 56;  DB 9;  Length 185596;
  Best Local Similarity   46.5%;  Pred. No. 0.39;
  Matches  214;  Conservative    0;  Mismatches  245;  Indels    1;  Gaps    1;

Qy        109 TCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAATC 168
              | | || |  || || | |   || | |  |   ||   | || || |  ||     |
Db     155419 TTTATAATATTTTAAATATAATATTTTAAATATTTTATAATATTTTAAATATATTTTAAA 155360

Qy        169 AGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTT 228
              |  |  || |||   | |   | | || ||    | ||  |  |  |  |||| |  ||
Db     155359 ATATTTTATTTTAAAATATTTTATAATATTTTTAAAATATTAATAATTTTATATTTTATT 155300

Qy        229 GATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAA 288
              ||||  ||   |       ||   | ||| |   | | || | |  |    ||   | | ||
Db     155299 AATTTATTATTATATTAATATTTATATATAATAAATAATATATAATATATTATATAATAA 155240

Qy        289 AAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACAC 348
              |   |||| |  |     ||  |   ||  ||     ||      |  | | || |||
Db     155239 ATAATATATAATATATTATATAATAAATAATATATAATATATTATATAATAAATAATATA 155180

Qy        349 AAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAG 408
              ||   | |  ||| || || |  |||| ||    | |||  |  ||  | |   |
Db     155179 TAATATATTATATATAATAAATAATATAATATTATTTTAAAATTTTTAATTTAATATTTTTA 155120

Qy        409 GTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTTATTTTGGA 468
              |||  |   |    | | ||   |     |||| | | ||    | || |||| |   |
Db     155119 TAATATTTTATAAAATATAATATAATACATAATATATAATATATAATATATTATATAATA 155060

Qy        469 AATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTC 528
              ||||| | | || | || || |  ||| ||  ||| |||   ||| |  |   || || |
Db     155059 TATAATAT-ATAATATATTATATAATATATAAAATTTAAAATATTAAATAATAATTAATA 155001

Qy        529 TGTCATTCAGGAGTATGGAAGTATCGATGGCTTCTTTAAA 568
              | | ||    | |||  ||  |   ||   |  ||| |
Db     155000 TATAATAATTTATTATATATAATAAATAATTTATATATTATA 154961


RESULT 27
AX083744
LOCUS       AX083744               1141 bp    DNA     linear   PAT 28-FEB-2001
DEFINITION  Sequence 22 from Patent WO0111061.
ACCESSION   AX083744
VERSION     AX083744.1  GI:13185472
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Kunst,L. and Clemens,S.
  TITLE     Regulation of embryonic transcription in plants
  JOURNAL   Patent: WO 0111061-A 22 15-FEB-2001;
            UNIVERSITY OF BRITISH COLUMBIA (CA)
FEATURES             Location/Qualifiers
     source          1. .1141
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
     promoter        1. .1141
```

/note="consensus sequence of A.t., L.a., and B.n. FAE1 promoters"
ORIGIN

```
  Query Match             5.1%;  Score 55.8;  DB 6;  Length 1141;
  Best Local Similarity   11.2%;  Pred. No. 1;
  Matches   72;  Conservative  231;  Mismatches  342;  Indels    0;  Gaps    0;

Qy          1 TGCTTTCCGTGTGGTGGTGCAGTAACTTTCACTCTCTATGTTCTTATAAGCAAATGTTAC 60
              :::: |:: ::::: : ::::::: :: ::  : | :   :   : :|:|:: : :
Db        219 WSYDVTYYWWVWDDMCKRKVRRWVRTRGRMRNYMVAWBTAHRRRYNNGWTBAMAYRRWTM 278

Qy         61 AATGAGATATGAGTTTTAAAGCCAGATCTTCCTTATCTCTCTGCCCCATCTCTAGTTCTT 120
                 |:|   : ::: : : ::  |||| : |   |:    |:  :   ::: ::
Db        279 NNNNNNAKAMCKRAKYWGWNRABVNSTCTTWKSKTTKVRTSCWANNCRAGDANKDHKWWK 338

Qy        121 GAAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTT 180
               |  | ::        |:::  ::|:::: ::  :: ::   |:    ::|  :::|
Db        339 WSAAMGVYWNNNNNNNWTYKKARHBARWDWVWHSAWKKWHANAAHYSRKKWTBYKRKTMV 398

Qy        181 TGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAA 240
                 |   ::   |:::::::::: |:      |    : :     ::|  :::     ::
Db        399 NNNNGTTMWKRMWAWYWKMDMDWBGTYNNNNNGGRTYYGWTKNKKMWTYYKWKANNCKWR 458

Qy        241 AGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGA 300
               | :: | :   :: : ::   :::     : :::   ::|  :::::: :     :
Db        459 AWDHKTCTHNNTTWWKMKTYWNNCYWKSMTNGKSHRBAAAVYTWYMWWWRRYAHANNNNW 518

Qy        301 AATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCA 360
               : :  | |  : ::  :::  :| ::| ::   | :  :::|    : : :  ::|
Db        519 DYWWKACTWYKYBVCSKWWNNYAAWYTKSSWNYTSRYYRWKTNNSWRWRSDTRSMGRANN 578

Qy        361 TATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTG 420
               :|   :  :   |:::::   :::::|   : :: : ::| |: : |  :|: :     :
Db        579 YARABHYGYKWNTRWWBWSHTWBHBRAGAAHYWMBMMYBAKCHCMKAWYKAKKYAGAGGS 638

Qy        421 CTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGA 480
                        | ::  |  : :|:| |:::: ::     |  |:|  :   :
Db        639 NNNNNNNNNNNNNNNNATCARDDYYAASRWYAMANAKWYYYKBAANNAYYTHANNWWGCW 698

Qy        481 AAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGTCATTCAGGA 540
               | :|::| ::       | |::       |:     : : |||      : : :   :  :
Db        699 NNATDTRRTMWKNNNNNNAGTWKNNNNNNAKNASAAKNYAAAAVKAAKKHWRWANKWAMR 758

Qy        541 GTATGGAAGTATCGATGGCTTCTTTAAAATGAAGCAGGAGGGTCTGGCAGAGAGTATCTA 600
               |:: : ||:| : :  | :|::|||     | : ::   | ::
Db        759 GWHADAAABTTDKRNNGAYTKYTTTNNNNTYRGVVTNTAARDGWANNNNNNNNNNNNNNG 818

Qy        601 TGAAATAAGTTCCTCTGACCTTCACGCTTAATTTTCTGAATGGAG 645
               :::::  :  :  : |              : ::| ::::  ::: ::
Db        819 WSDMWVTWWAYANYGTNNNNNNNNNNAYAWWTNKWYYTTDDRWRB 863
```

```
RESULT 28
DMU11584
LOCUS       DMU11584                4601 bp    DNA     linear   INV 23-JUL-1994
DEFINITION  Drosophila melanogaster Oregon-R mitochondrial A+T region.
ACCESSION   U11584
VERSION     U11584.1  GI:508826
KEYWORDS    mitochondrial DNA; A+T region; tandem repeats.
SOURCE      mitochondrion Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 4601)
  AUTHORS   Lewis,D.L., Farr,C.L., Farquhar,A.L. and Kaguni,L.S.
  TITLE     Sequence, organization, and evolution of the A+T region of
            Drosophila melanogaster mitochondrial DNA
  JOURNAL   Mol. Biol. Evol. 11 (3), 523-538 (1994)
  MEDLINE   94285822
   PUBMED   8015445
REFERENCE   2  (bases 1 to 4601)
  AUTHORS   Kaguni,L.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUN-1994) Laurie S. Kaguni Ph.D, Dept. of
            Biochemistry, Michigan State University, East Lansing, MI,
            48824-1318, USA
FEATURES             Location/Qualifiers
     source          1. .4601
                     /organism="Drosophila melanogaster"
                     /organelle="mitochondrion"
                     /mol_type="genomic DNA"
                     /strain="Oregon-R"
                     /db_xref="taxon:7227"
                     /dev_stage="embryo"
     repeat_unit     650. .1022
                     /note="repeat I-A"
                     /rpt_type=tandem
     repeat_unit     1023. .1360
                     /note="repeat I-B1"
                     /rpt_type=tandem
     repeat_unit     1361. .1705
                     /note="repeat I-C/A"
                     /rpt_type=tandem
     repeat_unit     1706. .2043
                     /note="repeat I-B2"
                     /rpt_type=tandem
     repeat_unit     2044. .2388
                     /note="repeat I-C"
                     /rpt_type=tandem
     misc_feature    2491. .2511
                     /note="deoxythymidylate stretch"
     repeat_unit     2512. .2648
                     /rpt_type=tandem
     repeat_unit     2649. .3112
                     /note="repeat II-A"
                     /rpt_type=tandem
     repeat_unit     3113. .3576
                     /note="repeat II-B1"
                     /rpt_type=tandem
     repeat_unit     3577. .4040
                     /note="repeat II-B2"
                     /rpt_type=tandem
     repeat_unit     4041. .4504
                     /note="repeat II-C"
                     /rpt_type=tandem
     misc_feature    complement(4565. .4585)
                     /note="deoxythymidylate stretch"
ORIGIN
```

```
  Query Match             5.1%;  Score 55.8;  DB 3;  Length 4601;
  Best Local Similarity   46.0%;  Pred. No. 0.8;
  Matches  225;  Conservative    0;  Mismatches  262;  Indels    2;  Gaps    1;

Qy         45 TATAAGCAAATGTTACAATGAGATATGAGTTTTAAAGCCAGATCTTCCTT--ATCTCTCT 102
              |||||  |||  ||| ||  | || |   ||| |||   |  | ||  ||  ||    ||
Db       2177 TATAATTAAAACTTAAAAAAATATTTTTTTTTAAAAAAAAATTATTTATTAAATTATACT 2236

Qy        103 GCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTA 162
                 | || |  || |   |   | |  ||| | |    ||    || ||   | || |
Db       2237 TAATAAACTATTTTTATAATAAATTATTTTATAAATAAAATTATTTAAAATAATTAATAA 2296

Qy        163 CAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATAC 222
               || |   |  |||  | | || |     | |  | |||   |||  ||   ||  |||
Db       2297 AAATTATATATATATATATATATATATTAAAATGAAAATAATTTTTAAATTTTAATAA 2356

Qy        223 TCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGT 282
               |   |  ||||  | | |     ||| | | |   |   | | |||  | |||  ||  |
Db       2357 TAAATAAATTTAATAATTAATAATTAAATAAAATCTATTCATTATTAATATTTAATTAAT 2416

Qy        283 TCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATT 342
                 ||| |  | |||  |     | | | | |||  | ||  |||||    |  |  || |
Db       2417 AATAAATAAATTTAATAACTAATAATTAAATAAAATTTATTTATTACTAATATTTAATTA 2476

Qy        343 AAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAAC 402
```

```
                | |   ||||    |   | ||   ||  ||  |     | | ||      |   |  |
Db       2477 ATAATAAAAAATTATTTTTTTTTTTTTTTTTTTTTTAATAATTTAATTAATTATTATATAT 2536

Qy        403 CTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTAT 462
              || |  | ||     |   |   |  |   | | | ||| | ||    | |||||  ||
Db       2537 TTATAAATTTATATATTATTGAATATTTATAATATATATATATATATAGAAAAATTAAAT 2596

Qy        463 TTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAA 522
              | |  ||||||| | | | ||  |   ||    || |  ||     |   ||  | |  ||
Db       2597 TATTTAAATAATTTAATATAAATTTTTTAAAAATTTCTTAAATGTATTATTTTTATAAAA 2656

Qy        523 ATATTCTGT 531
              |   | | |
Db       2657 AATATTTAT 2665
```

RESULT 29
DMU37541
LOCUS DMU37541 19517 bp DNA circular INV 02-MAR-2001
DEFINITION Drosophila melanogaster complete mitochondrial genome.
ACCESSION U37541
VERSION U37541.1 GI:1166529
KEYWORDS .
SOURCE mitochondrion Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 12511 to 12682)
AUTHORS Clary,D.O., Goddard,J.M., Martin,S.C., Fauron,C.M. and Wolstenholme,D.R.
TITLE Drosophila mitochondrial DNA: a novel gene order
JOURNAL Nucleic Acids Res. 10 (21), 6619-6637 (1982)
MEDLINE 83090428
PUBMED 6294611
REFERENCE 2 (bases 5269 to 5695)
AUTHORS Clary,D.O., Wahleithner,J.A. and Wolstenholme,D.R.
TITLE Transfer RNA genes in Drosophila mitochondrial DNA: related 5' flanking sequences and comparisons to mammalian mitochondrial tRNA genes
JOURNAL Nucleic Acids Res. 11 (8), 2411-2425 (1983)
MEDLINE 83220794
PUBMED 6304652
REFERENCE 3 (bases 404 to 5272)
AUTHORS de Bruijn,M.H.
TITLE Drosophila melanogaster mitochondrial DNA, a novel organization and genetic code
JOURNAL Nature 304 (5923), 234-241 (1983)
MEDLINE 83245048
PUBMED 6408489
REFERENCE 4 (bases 804 to 1778)
AUTHORS Satta,Y., Ishiwa,H. and Chigusa,S.I.
TITLE Analysis of nucleotide substitutions of mitochondrial DNAs in Drosophila melanogaster and its sibling species
JOURNAL Mol. Biol. Evol. 4 (6), 638-650 (1987)
MEDLINE 88174373
PUBMED 2832697
REFERENCE 5 (bases 5268 to 13619)
AUTHORS Garesse,R.
TITLE Drosophila melanogaster mitochondrial DNA: gene organization and evolutionary considerations
JOURNAL Genetics 118 (4), 649-663 (1988)
MEDLINE 88212147
PUBMED 3130291
REFERENCE 6 (bases 441 to 2967)
AUTHORS Satta,Y. and Takahata,N.
TITLE Evolution of Drosophila mitochondrial DNA and the history of the melanogaster subgroup
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 87 (24), 9558-9562 (1990)
MEDLINE 91088557
PUBMED 2124697
REFERENCE 7 (bases 14215 to 14512)
AUTHORS Ballard,J.W., Olsen,G.J., Faith,D.P., Odgers,W.A., Rowell,D.M. and Atkinson,P.W.
TITLE Evidence from 12S ribosomal RNA sequences that onychophorans are modified arthropods
JOURNAL Science 258 (5086), 1345-1348 (1992)
MEDLINE 93088057
PUBMED 1455227
REFERENCE 8 (bases 14917 to 19517)
AUTHORS Lewis,D.L., Farr,C.L., Farquhar,A.L. and Kaguni,L.S.
TITLE Sequence, organization, and evolution of the A+T region of Drosophila melanogaster mitochondrial DNA
JOURNAL Mol. Biol. Evol. 11 (3), 523-538 (1994)
MEDLINE 94285822
PUBMED 8015445
REFERENCE 9 (bases 1 to 408; 13319 to 19517)
AUTHORS Lewis,D.L., Farr,C.L. and Kaguni,L.S.
TITLE Drosophila melanogaster mitochondrial DNA: completion of the nucleotide sequence and evolutionary comparisons
JOURNAL Insect Mol. Biol. 4 (4), 263-278 (1995)
MEDLINE 96423163
PUBMED 8825764
REFERENCE 10 (bases 1 to 19517)
AUTHORS Lewis,D.L., Farr,C.L. and Kaguni,L.S.
TITLE Direct Submission
JOURNAL Submitted (03-OCT-1995) Laurie S. Kaguni, Biochemistry Department, Michigan State University, East Lansing, MI 48824-1319, USA
FEATURES Location/Qualifiers
source 1. .19517
/organism="Drosophila melanogaster"
/organelle="mitochondrion"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/note="derived from new and previously submitted sequences; sequence is a composite containing sequences obtained from different Drosophila melanogaster strains"
tRNA 1. .65
/product="tRNA-Ile"
tRNA complement(97. .165)
/product="tRNA-Gln"
tRNA 171. .239
/product="tRNA-Phe"
CDS 240. .1265
/codon_start=1
/transl_table=5
/product="NADH dehydrogenase subunit 2"
/protein_id="AAC47811.1"
/db_xref="GI:1166530"
/translation="MFNNSSKILFITIMIIGTLITVTSNSWLGAWMGLEINLLSFIPL
LSDNNNLMSTEASLKYFLTQVLASTVLLFSSILLMLKNNMNNEINESFTSMIIMSALL
LKSGAAPFHFWFPNMMEGLTWMNALMLMTWQKIAPLMLISYLNIKYLLLISVILSVII
GAIGGLNQTSLRKLMAFSSINHLGWMLSSLMISESIWLILFFFYSFLSFVLTFMFNIF
KLFHLNQLFSWFVNSKILKFTLFMNFLSLGGLPPFLGFLPKWLVIQQLTLCNQYFMLT
IMMMSTLITLFFYLRICYSAFMMNYFENNWIMKMNMNSINYNMYMIMTFFSIFGLFLI
SLFYFMF"
tRNA 1264. .1329
/product="tRNA-Trp"
tRNA complement(1322. .1383)
/product="tRNA-Cys"
tRNA complement(1403. .1468)
/product="tRNA-Tyr"
CDS join(1470. .1472,1474. .3009)
/note="mechanism underlying reading frame shift after first codon uncertain"
/codon_start=1
/transl_table=5
/product="cytochrome c oxidase subunit I"
/protein_id="AAC47812.2"
/db_xref="GI:7412849"
/translation="MSRQWLFSTNHKDIGTLYFIFGAWAGMVGTSLSILIRAELGHPG
ALIGDDQIYNVIVTAHAFIMIFFMVMPIMIGGFGNWLVPLMLGAPDMAFPRMNNMSFW
LLPPALSLLLVSSMVENGAGTGWTVYPPLSAGIAHGGASVDLAIFSLHLAGISSILGA
VNFITTVINMRSTGISLDRMPLFVWSVVITALLLLLSLPVLAGAITMLLTDRNLNTSF
FDPAGGGDPILYQHLFWFFGHPEVYILILPGFGMISHIISQESGKKETFGSLGMIYAM
LAIGLLGFIVWAHHMFTVGMDVDTRAYFTSATMIIAVPTGIKIFSWLATLHGTQLSYS

```
                PAILWALGFVFLFTVGGLTGVVLANSSVDIILHDTYYVVAHFHYVLSMGAVFAIMAGF
                IHWYPLFTGLTLNNKWLKSHFIIMFIGVNLTFFPQHFLGLAGMPRRYSDYPDAYTTWN
                IVSTIGSTISLLGILFFFFIIWESLVSQRQVIYPIQLNSSIEWYQNTPPAEHSYSELP
                LLTN"
tRNA            3012. .3077
                /product="tRNA-Leu"
CDS             3083. .3767
                /note="TAA stop codon is completed by the addition of 3' A
                residues to the mRNA"
                /codon_start=1
                /transl_except=(pos:3767,aa:TERM)
                /transl_table=5
                /product="cytochrome c oxidase subunit II"
                /protein_id="AAC47813.1"
                /db_xref="GI:1166532"
                /translation="MSTWANLGLQDSASPLMEQLIFFHDHALLILVMITVLVGYLMFM
                LFFNNYVNRFLLHGQLIEMIWTILPAIILLFIALPSLRLLYLLDEINEPSVTLKSIGH
                QWYWSYEYSDFNNIEFDSYMIPTNELMTDGFRLLDVDNRVVLPMNSQIRILVTAADVI
                HSWTVPALGVKVDGTPGRLNQTNFFINRPGLFYGQCSEICGANHSFMPIVIESVPVNY
                FIKWISSNNS"
tRNA            3768. .3838
                /product="tRNA-Lys"
tRNA            3840. .3906
                /product="tRNA-Asp"
CDS             3907. .4068
                /codon_start=1
                /transl_table=5
                /product="ATPase 8"
                /protein_id="AAC47814.1"
                /db_xref="GI:1166533"
                /translation="MPQMAPISWLLLFIIFSITFILFCSINYYSYMPNSPKSNELKNI
                NLNSMNWKW"
CDS             4062. .4736
                /codon_start=1
                /transl_table=5
                /product="ATPase 6"
                /protein_id="AAC47815.1"
                /db_xref="GI:1166534"
                /translation="MMTNLFSVFDPLAIFNFSLNWLSTFLGLLMIPSIYWLMPSRYNI
                MWNSILLTLHKEFKTLLGPSGHNGSTFIFISLFSLILFNNFMGLFPYIFTSTSHLTLT
                LSLALPLWLCFMLYGWINHTQHMFAHLVPQGTPAILMPFMVCIETISNIIRPGTLAVR
                LTANMIAGHLLLTLLGNTGSSMSYMLMTFLLMAQIALLVLESAVAMIQSYVFAVLSTL
                YSSEVN"
CDS             4736. .5524
                /codon_start=1
                /transl_table=5
                /product="cytochrome c oxidase subunit III"
                /protein_id="AAC47816.1"
                /db_xref="GI:1166535"
                /translation="MSTHSNHPFHLVDYSPWPLTGAIGAMTTVSGMVKWFHQYDISLF
                VLGNIITILTVYQWWRDVSREGTYQGLHTYAVTIGLRWGMILFILSEVLFFVSFFWAF
                FHSSLSPAIELGASWPPMGIISFNPFQIPLLNTAILLASGVTVTWAHHSLMENNHSQT
                TQGLFFTVLLGIYFTILQAYEYIEAPFTIADSIYGSTFFMATGFHGIHVLIGTTFLLV
                CLLRHLNNHFSKNHHFGFEAAAWYWHFVDVVWLFLYITIYWWGG"
tRNA            5543. .5607
                /product="tRNA-Gly"
CDS             5608. .5961
                /codon_start=1
                /transl_table=5
                /product="NADH dehydrogenase subunit 3"
                /protein_id="AAC47817.1"
                /db_xref="GI:1166536"
                /translation="MFSIIFIALLILLITTIVMFLASILSKKALIDREKSSPFECGFD
                PKSSSRLPFSLRFFLITIIFLIFDVEIALILPMIIIMKYSNIMIWTITSIIFILILLI
                GLYHEWNQGMLNWSN"
tRNA            5981. .6045
                /product="tRNA-Ala"
tRNA            6055. .6118
                /product="tRNA-Arg"
tRNA            6119. .6183
                /product="tRNA-Asn"
tRNA            6184. .6251
```

```
  Query Match             5.1%;  Score 55.8;  DB 3;  Length 19517;
  Best Local Similarity   46.0%;  Pred. No. 0.63;
  Matches  225;  Conservative    0;  Mismatches  262;  Indels    2;  Gaps    1;

Qy         45 TATAAGCAAATGTTACAATGAGATATGAGTTTTAAAGCCAGATCTTCCTT--ATCTCTCT 102
              |||||  |||  ||| ||  | || |   ||| |||   |  | ||  ||  ||    ||
Db      17093 TATAATTAAAACTTAAAAAAATATTTTTTTTTAAAAAAAAATTATTTATTAAATTATACT 17152

Qy        103 GCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTA 162
                | || |  || |    |   | |  ||| | |    ||    || ||   | || |
Db      17153 TAATAAACTATTTTTATAATAAATTATTTTATAAATAAAATTATTTAAAATAATTAATAA 17212

Qy        163 CAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATAC 222
               || |   |  |||  | | || |      | |  | |||   ||| ||   ||   |||
Db      17213 AAATTATATATATATATATATATATATTAAAATGAAAATAATTTTTAAATTTTAATAA 17272

Qy        223 TCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGT 282
              |   |  ||||  | | |   |  ||| | | |  |   |  | |||  | |||  ||  |
Db      17273 TAAATAAATTTAATAATTAATAATTAAATAAAATCTATTCATTATTAATATTTAATTAAT 17332

Qy        283 TCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATT 342
                ||| |  | |||  |      | | | | |||  | || |||||   |  |  || |
Db      17333 AATAAATAAATTTAATAACTAATAATTAAATAAAATTTATTTATTACTAATATTTAATTA 17392

Qy        343 AAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAAC 402
              | |   ||||    |   | ||   ||  ||  |    | | ||      |   |  |
Db      17393 ATAATAAAAAATTATTTTTTTTTTTTTTTTTTTTAATAATTTAATTAATTATTATATAT 17452

Qy        403 CTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTAT 462
              || |  | ||     |  |   |  |   | |  | ||| | ||    | |||||  ||
Db      17453 TTATAAATTTATATATTATTGAATATTTATAATATATATATATATATAGAAAAATTAAAT 17512

Qy        463 TTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAA 522
              | |  |||||||| | | | ||  |  ||    || |  ||      |  ||  | |  ||
Db      17513 TATTTAAATAATTTAATATAAATTTTTTAAAAATTTCTTAAATGTATTATTTTTATAAAA 17572

Qy        523 ATATTCTGT 531
              |   | | |
Db      17573 AATATTTAT 17581


RESULT 30
AP005815
LOCUS       AP005815               36055 bp    DNA     linear   HTG 11-OCT-2002
DEFINITION  Homo sapiens chromosome 11 clone CMF9-380I9 map 11q, WORKING DRAFT
            SEQUENCE, 3 unordered pieces.
ACCESSION   AP005815
VERSION     AP005815.1  GI:23821400
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens genomic DNA of 11q
  JOURNAL   Published Only in Database (2002)
REFERENCE   2  (bases 1 to 36055)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-OCT-2002) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     -------------- Genome Center
                Center: RIKEN Genomic Sciences Center(GSC)
                Center code: RIKEN
                Web site: http://hgp.gsc.riken.go.jp/
                Contact: hattori@gsc.riken.go.jp
            -------------- Project Information
```

```
            Center project name: HumDraft11
            Center clone name: CMF9-380I9
          -------------- Summary Statistics
            Sequencing vector: PCR products; 100% of reads
            Chemistry: Dye-terminator ET-amersham; 100% of reads
            Assembly program: Phrap; version 0.990329
            Consensus quality: 35497 bases at least Q40
            Consensus quality: 35693 bases at least Q30
            Consensus quality: 35797 bases at least Q20
            Insert size: 35855; sum-of-contigs
            Quality coverage: 15.63x in Q20 bases; sum-of-contigs
          --------------.
          NOTE: This is a 'working draft' sequence. It currently consists of
          3 contigs. The true order of the pieces is not known and their
          order in this sequence record is arbitrary. Gaps between the
          contigs are represented as runs N, but the exact sizes of the gaps
          are unknown. This record will be updated with the finished sequence
          as soon as it is available and the accession number will be
          preserved
                1     15568 contig of  15568 bp in length
             15669     34838 contig of  19170 bp in length
             34939     36055 contig of   1117 bp in length.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 3 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *      1    15568: contig of 15568 bp in length
          *  15569    15668: gap of 100 bp
          *  15669    34838: contig of 19170 bp in length
          *  34839    34938: gap of 100 bp
          *  34939    36055: contig of 1117 bp in length.
FEATURES             Location/Qualifiers
     source          1. .36055
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q"
                     /clone="CMF9-380I9"
     misc_feature    1. .15568
                     /note="assembly_fragment"
     misc_feature    15669. .34838
                     /note="assembly_fragment"
     misc_feature    34939. .36055
                     /note="assembly_fragment"
ORIGIN

  Query Match             5.1%;  Score 55.8;  DB 2;  Length 36055;
  Best Local Similarity   48.6%;  Pred. No. 0.56;
  Matches  186;  Conservative    0;  Mismatches  192;  Indels    5;  Gaps    1;

Qy        150 ATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGT 209
              ||||  ||  |      ||  || |||| || | ||       ||      ||||     |
Db      15935 ATATGTATTTTGTATTATAGGTAAATATTTATATATTTATACTATATATAAAATACATTA 15994

Qy        210 TTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTT 269
              ||  |||||        | | || || | |  |     | |||||  |||| || |
Db      15995 TATACAAATATTTATATATTTATATTATATATAAAATACATAATATACAAATATTTATAT 16054

Qy        270 GTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAG 329
              ||| |  ||| |  ||| |    ||| | |||    |  | |   | ||  ||||| |
Db      16055 ATTTATATTTTATATAAATATATAATATACAAATATTTATATATTTATATTTTATATAAA 16114

Qy        330 CCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAG 389
                |   | |  ||       ||   |  || ||| || | |  ||          ||
Db      16115 TATATAATATACAAATATTTGTATATTTATATTTTATATAAATATATAAATATACAAATA 16174

Qy        390 GTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCT-----TATCGCACTGCAAATAGA 444
              .      |||  | ||    |  ||||| |     | | |     |||  |||   ||||| |
Db      16175 TACAAATATAAAATATATAATATACAAATATTTATATATTTATATTTCACATAAAATATA 16234

Qy        445 AATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATT 504
              | ||     ||  ||||| ||  | ||||  |   || | | || | |||||  |   ||
Db      16235 TAATATACAAATATTTATATTTCATATAAAATATAAATATACAAATATTTATATTTCATA 16294

Qy        505 AAAACATTTACTTTCCAAATATT 527
              ||||   | | | | ||||||||
Db      16295 AAAAATATAAATATACAAATATT 16317


RESULT 31
AC024070
LOCUS       AC024070               38849 bp    DNA     linear   PRI 09-APR-2003
DEFINITION  Homo sapiens chromosome 22 clone chk89 map 22q11, complete
            sequence.
ACCESSION   AC024070
VERSION     AC024070.2  GI:7712132
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 38849)
  AUTHORS   Hu,P., Wu,H., Prescot,A., Emanuel,B., Budarf,M. and Roe,B.A.
  TITLE     Homo sapiens Chromosome 22q11 Cosmid Clone chk89
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 38849)
  AUTHORS   Hu,P., Wu,H., Prescot,A., Emanuel,B., Budarf,M. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-FEB-2000) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   3  (bases 1 to 38849)
  AUTHORS   Hu,P., Wu,H., Prescot,A., Emanuel,B., Budarf,M. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAY-2000) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   4  (bases 1 to 38849)
  AUTHORS   Hu,P., Wu,H., Prescot,A., Emanuel,B., Budarf,M. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-APR-2003) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On May 6, 2000 this sequence version replaced gi:7019305.
            -------------- Genome Center
            Center: Department Of Chemistry And Biochemistry
            The University Of Oklahoma
            Center code:UOKNOR
            ----------------------.
FEATURES             Location/Qualifiers
     source          1. .38849
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="22"
                     /map="22q11"
                     /clone="chk89"
ORIGIN

  Query Match             5.1%;  Score 55.8;  DB 9;  Length 38849;
  Best Local Similarity   47.5%;  Pred. No. 0.56;
  Matches  199;  Conservative    0;  Mismatches  217;  Indels    3;  Gaps    1;

Qy        117 TCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATA 176
              | |  || | | | |||| |  ||    | | | |||||     |||| |    |  ||||
Db      18626 TATATAATTATATTATTATTATATTATATTATATATATTTTATATTATATTATATATAATT 18685

Qy        177 GTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTT 236
              | ||.||  | ||   || ||| | | | | | |  | |||||  |   ||   | |
Db      18686 ATATTATATTATATTATATATTATATATAATTTATGATATATATAATGTATTATATATAA 18745
```

```
Qy        237 TAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATA 296
              |  | ||  |     | |||     | |||| | |  |  | |  |   | |   ||||
Db      18746 TTTATGATATATATTATATATTATAAATCATATATCATATATATTATGCTATATTATATA 18805

Qy        297 AAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAG 356
              |  |  ||||       | ||  ||||| |   |    | |||  || | || || | |
Db      18806 TAATATAGCATAATATATCATATATTATATAATATATAATATAATATAATATAATATAAT 18865

Qy        357 TGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTA 416
              ||  || || |  ||    |       |     ||| |     |  | | | |||||||
Db      18866 GTAATCATATATAATATATAATATATTATATTATAAAATATATTATATTATGATATACTA 18925

Qy        417 CCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCT 476
              | |   ||   ||   | |||  | ||   | ||  | || | |  |     || |
Db      18926 ---TATATTACATATTACATTAGATAATATTTTATATAATATATAATATATATTATATAT 18982

Qy        477 CAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGTCATT 535
              | |    || || ||  || ||  |||| |   | || ||   | ||| | | | |||
Db      18983 TATACTTTATTATATTACATATATAATTATACATTATATTTCATATATAATTTATAATT 19041


RESULT 32
AC116984_1
WPCOMMENT
  Sequence split into 6 fragments  LOCUS AC116984 Accession AC116984
     Fragment Name      Begin     End
     AC116984_0             1     110000
     AC116984_1        100001     210000
     AC116984_2        200001     310000
     AC116984_3        300001     410000
     AC116984_4        400001     510000
     AC116984_5        500001     541399
Continuation (2 of 6) of AC116984 from base 100001 (AC116984 Dictyostelium discoideum ch

  Query Match             5.1%;  Score 55.8;  DB 3;  Length 110000;
  Best Local Similarity   48.1%;  Pred. No. 0.47;
  Matches  217;  Conservative    0;  Mismatches  232;  Indels    2;  Gaps    2;

Qy        122 AAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTT 181
              || ||  |  | |  |  | ||  ||       | || ||  ||    ||  ||  | |||||
Db      79445 AATTGAATTTTTTTTTTATTTTTTTTTTTTTTTTATTTTTTTTTATTTTTTTTTGTTTT 79504

Qy        182 GTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAA 241
              | | ||||| |  | || |     || ||||  || | | ||   |  |||  |||  ||
Db      79505 GAAAAAGATTTTTTTTTTATTTTTTTATTTTCATATTTTTCTTTTTCTTTTTCTTTTAA 79564

Qy        242 GACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAA 301
              |   | ||   ||      | ||  ||| |||||  ||| |||  || |  || ||
Db      79565 AAAAATGCATTGATTTTTTTTTTTTTTTTTTTTTTTTTTTTTCTTAAGATTTAAAATTTC 79624

Qy        302 ATGCATTCAGAGA-AAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCA 360
              ||       | | | ||    |||| | |   | |  ||||   || |      |  ||
Db      79625 ATATTCAAATAAATAATGAAATATTTGAAAAAATCAAAAAAGCAAAATTTTTTTTGTTAC 79684

Qy        361 TATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTG 420
              || |||   |  | ||  | | |    |      ||  | ||  ||    |  |
Db      79685 ACCTAAATTTTATGTTATTTTATTTATTTTTTTTTTTAAAAAATTAATTTAATTAATTTT 79744

Qy        421 CTTTCTTATCGCACTGCA-AATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAG 479
              |||  || |   | |||| || | |||  |  | |||   ||       ||| ||    |
Db      79745 TTTTTTTTTTAAGTGCACAACAAAAAAAAAAATAAAAAATAAAACAAAAAAAAAGAAAT 79804

Qy        480 AAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGTCATTCAGG 539
              | |  | |||  || |  | |   |||| ||||  |||  |||||| ||  |  |  ||
Db      79805 CACATTTTATTCATTTTTAAATTAAAAAATATTTCTTTTAAAAATATACTACCTGTGTGG 79864

Qy        540 AGTATGGAAGTATCGATGGCTTCTTTAAAAT 570
              | ||  |  ||   |   || ||| || |
Db      79865 CGAATCAATTTAAATTTTATTTTTTTTAATT 79895


RESULT 33
AP004607
LOCUS       AP004607              160683 bp    DNA     linear   PRI 25-JUN-2003
DEFINITION  Homo sapiens genomic DNA, chromosome 11 clone:RP11-529A4, complete
            sequence.
ACCESSION   AP004607
VERSION     AP004607.4  GI:32188043
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens genomic DNA
  JOURNAL   Published Only in Database (2002)
REFERENCE   2  (bases 1 to 160683)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JAN-2002) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     On Jun 24, 2003 this sequence version replaced gi:19879829.
FEATURES             Location/Qualifiers
     source          1. .160683
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q"
                     /clone="RP11-529A4"
ORIGIN

  Query Match             5.1%;  Score 55.8;  DB 9;  Length 160683;
  Best Local Similarity   48.6%;  Pred. No. 0.44;
  Matches  186;  Conservative    0;  Mismatches  192;  Indels    5;  Gaps    1;

Qy        150 ATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGT 209
              ||||  ||  |    ||  || |||| || | ||      ||    ||||      |
Db      21515 ATATGTATTTTGTATTATAGGTAAATATTTATATATTTATACTATATATAAAATACATTA 21574

Qy        210 TTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTT 269
              | |  |||||       |   | || || | |  |     | |||||  |||| || |
Db      21575 TATACAAATATTTATATATTTATATTATATATAAAATACATAATATACAAATATTTATAT 21634

Qy        270 GTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAG 329
              ||| |  ||| |  ||| |     ||| | ||||   |  | |   | ||  |||||  |
Db      21635 ATTTATATTTTATATAAATATATAATATACAAATATTTATATATTTATATTTTATATAAA 21694

Qy        330 CCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAG 389
              |   | |  ||       ||   |  || ||| || | |  ||        ||
Db      21695 TATATAATATACAAATATTTGTATATTTATATTTTATATAAATATATAAATATACAAATA 21754

Qy        390 GTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCT-----TATCGCACTGCAAATAGA 444
              |||  | ||    |  |||||  |     | | |     |||  |||   ||||| |
Db      21755 TACAAATATAAAATATATAATATACAAATATTTATATATTTATATTTCACATAAAATATA 21814

Qy        445 AATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATT 504
              | ||    ||  ||||| ||  | ||||  |   || | | || | |||||  |   ||
Db      21815 TAATATACAAATATTTATATTTCATATAAAATATAAATATACAAATATTTATATTTCATA 21874

Qy        505 AAAACATTTACTTTCCAAATATT 527
              ||||  | | | | ||||||||
Db      21875 AAAAATATAAATATACAAATATT 21897


RESULT 34
CR759796
LOCUS       CR759796              196762 bp    DNA     linear   HTG 24-SEP-2004
DEFINITION  Danio rerio clone DKEY-18A10, WORKING DRAFT SEQUENCE, 10 unordered
            pieces.
```

```
ACCESSION   CR759796
VERSION     CR759796.4  GI:52673634
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 196762)
  AUTHORS   Burton,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 24, 2004 this sequence version replaced gi:52626701.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK18A10
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 193114 bases at least Q40
            Consensus quality: 193781 bases at least Q30
            Consensus quality: 194189 bases at least Q20
            Insert size: 195862; sum-of-contigs
            Insert size: 192525; 5.1% error; agarose-fp
            Quality coverage: 6.57x in Q20 bases; sum-of-contigs Quality
            coverage: 6.68x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     5760: contig of 5760 bp in length
            *     5761     5860: gap of 100 bp
            *     5861    20214: contig of 14354 bp in length
            *    20215    20314: gap of 100 bp
            *    20315    34680: contig of 14366 bp in length
            *    34681    34780: gap of 100 bp
            *    34781    43318: contig of 8538 bp in length
            *    43319    43418: gap of 100 bp
            *    43419    73325: contig of 29907 bp in length
            *    73326    73425: gap of 100 bp
            *    73426   109883: contig of 36458 bp in length
            *   109884   109983: gap of 100 bp
            *   109984   118156: contig of 8173 bp in length
            *   118157   118256: gap of 100 bp
            *   118257   129072: contig of 10816 bp in length
            *   129073   129172: gap of 100 bp
            *   129173   182677: contig of 53505 bp in length
            *   182678   182777: gap of 100 bp
            *   182778   196762: contig of 13985 bp in length.
FEATURES             Location/Qualifiers
     source          1. .196762
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-18A10"
                     /clone_lib="DanioKey"
     misc_feature    1. .5760
                     /note="assembly_fragment:00009
                     fragment_chain:1"
     misc_feature    5861. .20214
                     /note="assembly_fragment:00500
                     fragment_chain:1"
     misc_feature    20315. .34680
                     /note="assembly_fragment:00347
                     fragment_chain:1"
     misc_feature    34781. .43318
                     /note="assembly_fragment:00123
                     fragment_chain:1"
     misc_feature    43419. .73325
                     /note="assembly_fragment:00837
                     fragment_chain:1"
     misc_feature    73426. .109883
                     /note="assembly_fragment:01226
                     fragment_chain:1"
     misc_feature    109984. .118156
                     /note="assembly_fragment:00051
                     fragment_chain:1"
     misc_feature    118257. .129072
                     /note="assembly_fragment:00216
                     fragment_chain:2"
     misc_feature    129173. .182677
                     /note="assembly_fragment:01642
                     fragment_chain:2"
     misc_feature    182778. .196762
                     /note="assembly_fragment:00664.0"
ORIGIN

  Query Match             5.1%;  Score 55.8;  DB 2;  Length 196762;
  Best Local Similarity   47.5%;  Pred. No. 0.42;
  Matches  199;  Conservative    0;  Mismatches  217;  Indels    3;  Gaps    1;

Qy        127 TCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGA 186
              | | | || | | |  ||    |  ||  || ||   |||        ||  ||||    |
Db      73755 TTTAAAATAAATATATTTTTATATAATATATTATATAAAAAATAAATATTATTTTAATAA 73814

Qy        187 AGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAA---AAGA 243
              | |    || ||| | | |   || |  | ||||| |   || ||| || |||    | |
Db      73815 ATAATAAATTTTATATAAAAAATTATAATTATATAATTATTTAATTATTATAATTTTATA 73874

Qy        244 CCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAAT 303
              | | | |         |||| || |  |||| ||| |   | ||    | |   | |
Db      73875 TTTAATAAATTATATATTATTTATATTATTTTATTTTTTATTATAATAAAAAATTTATAA 73934

Qy        304 GCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATAT 363
                |  | |  || ||  ||  ||     || | | | | |    |||| | |  ||
Db      73935 ATTTATATAATAATATATTAATATAATATATTAATATTAATAATATAAATTATTATATTA 73994

Qy        364 TACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTT 423
              |  | | |||     |  |   ||    || | | ||   | |  |||| ||     ||
Db      73995 AATTTAATTTATATATATATAAAATATAAATTATTAAATTAATATTTATAATATAATTTT 74054

Qy        424 TCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAA 483
              | | ||   | |  ||||| | ||||  | |||||||| ||||    ||  | |    ||
Db      74055 TTTAATAATAATATAAATAAATATTAAAAAAAATTTTTTTTTATTTATTTTTTATATAAT 74114

Qy        484 CATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGTCATTCAGGAGT 542
                |  |||||||  ||  |  ||||    |  ||      | || | |  || |  | |
Db      74115 TTTTTTTTTTTAAATATAAAAAAAAATAAAATATTTTTTTTTTTTATATTTAATTATT 74173


RESULT 35
AC134388/c
LOCUS       AC134388              172413 bp    DNA     linear   HTG 19-FEB-2004
DEFINITION  Pan troglodytes clone rp43-137d12, WORKING DRAFT SEQUENCE, 6
            ordered pieces.
ACCESSION   AC134388
VERSION     AC134388.9  GI:42628006
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 172413)
  AUTHORS   Oommen,S., Shaikh,T. and Roe,B.A.
  TITLE     Pan troglodytes BAC Clone rp43-137d12
```

```
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 172413)
  AUTHORS   Oommen,S. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-SEP-2002) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   3  (bases 1 to 172413)
  AUTHORS   Oommen,S., Shaikh,T. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-FEB-2004) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On Feb 19, 2004 this sequence version replaced gi:42539055.
            -------------- Genome Center
            Center: Department Of Chemistry And Biochemistry
            The University Of Oklahoma
            Center code:UOKNOR
            ----------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *      1    85914: contig of 85914 bp in length
            *  85915    86014: gap of unknown length
            *  86015    93633: contig of 7619 bp in length
            *  93634    93733: gap of unknown length
            *  93734   104748: contig of 11015 bp in length
            * 104749   104848: gap of unknown length
            * 104849   161223: contig of 56375 bp in length
            * 161224   161323: gap of unknown length
            * 161324   166338: contig of 5015 bp in length
            * 166339   166438: gap of unknown length
            * 166439   172413: contig of 5975 bp in length.
FEATURES             Location/Qualifiers
     source          1. .172413
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="rp43-137d12"
                     /clone_lib="RPCI - 43 Male Chimpanzee BAC Library"
ORIGIN

  Query Match             5.1%;  Score 55.6;  DB 2;  Length 172413;
  Best Local Similarity   47.4%;  Pred. No. 0.47;
  Matches  166;  Conservative    0;  Mismatches  184;  Indels    0;  Gaps    0;

Qy        174 ATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTT 233
              ||| | || || || || | ||    | | ||  | |     |||||| | | |    |
Db      93606 ATAATATTATATAATATATAATACATATTATATATTATATAATATATAATACATATTATA 93547

Qy        234 TTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGT 293
              |  ||||  |  |   | | |||   |   | || |  |  |   |  | ||    | |
Db      93546 TAATAAATAATATATAATATATAATACATATTATATAATAATATATAATACATGTTATAT 93487

Qy        294 ATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAAT 353
              |  |  | ||    |   | | |  | ||||| ||     |   | || | ||| || ||
Db      93486 AATATTATATAATATATAATACATGTTATATAATATTATATAATATATAATACATAATAT 93427

Qy        354 GAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATA 413
              |    |||| | ||||  |||| |   |  || |    || | |  |     |   ||||
Db      93426 AATATTATATAATATTATATAATATATAATACATAATATAATATTATATAATATTATATA 93367

Qy        414 CTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAA 473
              ||  |  |   | ||   |    | |||  |  ||  ||| | | |||  |   |   |
Db      93366 ATATATAATACATAATATAATATTATATAATATATAATACATAATATATAATATTATATA 93307

Qy        474 TCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAA 523
```

```
              | || ||||| | | | || |  |||| |  ||  |  ||  | |
Db      93306 ATATATAATACATATTATATAATATTATATTATATAATACATATTATATA 93257


RESULT 36
AL365174/c
LOCUS       AL365174               172577 bp    DNA     linear   HTG 10-JUL-2001
DEFINITION  Homo sapiens chromosome 1 clone RP11-435N22, 25 unordered pieces.
ACCESSION   AL365174
VERSION     AL365174.3  GI:12331089
KEYWORDS    HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Burton,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jan 22, 2001 this sequence version replaced gi:9798209.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bA435N22
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 161378 bases at least Q40
            Consensus quality: 165923 bases at least Q30
            Consensus quality: 168179 bases at least Q20
            Insert size: 170177; sum-of-contigs
            Insert size: 185612; 7.2% error; agarose-fp
            Quality coverage: 3.99x in Q20 bases; sum-of-contigs Quality
            coverage: 4.26x in Q20 bases; agarose-fp
            ----------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 25 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *      1     6277: contig of 6277 bp in length
            *   6278     6377: gap of 100 bp
            *   6378     9441: contig of 3064 bp in length
            *   9442     9541: gap of 100 bp
            *   9542    24962: contig of 15421 bp in length
            *  24963    25062: gap of 100 bp
            *  25063    29276: contig of 4214 bp in length
            *  29277    29376: gap of 100 bp
            *  29377    37516: contig of 8140 bp in length
            *  37517    37616: gap of 100 bp
            *  37617    48139: contig of 10523 bp in length
            *  48140    48239: gap of 100 bp
            *  48240    53407: contig of 5168 bp in length
            *  53408    53507: gap of 100 bp
            *  53508    74246: contig of 20739 bp in length
            *  74247    74346: gap of 100 bp
            *  74347    77364: contig of 3018 bp in length
            *  77365    77464: gap of 100 bp
            *  77465    81839: contig of 4375 bp in length
            *  81840    81939: gap of 100 bp
            *  81940    84471: contig of 2532 bp in length
            *  84472    84571: gap of 100 bp
            *  84572    87957: contig of 3386 bp in length
            *  87958    88057: gap of 100 bp
```

```
            *    88058    91000: contig of 2943 bp in length
            *    91001    91100: gap of 100 bp
            *    91101   103888: contig of 12788 bp in length
            *   103889   103988: gap of 100 bp
            *   103989   118772: contig of 14784 bp in length
            *   118773   118872: gap of 100 bp
            *   118873   121121: contig of 2249 bp in length
            *   121122   121221: gap of 100 bp
            *   121222   126906: contig of 5685 bp in length
            *   126907   127006: gap of 100 bp
            *   127007   130692: contig of 3686 bp in length
            *   130693   130792: gap of 100 bp
            *   130793   134245: contig of 3453 bp in length
            *   134246   134345: gap of 100 bp
            *   134346   137280: contig of 2935 bp in length
            *   137281   137380: gap of 100 bp
            *   137381   144713: contig of 7333 bp in length
            *   144714   144813: gap of 100 bp
            *   144814   147044: contig of 2231 bp in length
            *   147045   147144: gap of 100 bp
            *   147145   154661: contig of 7517 bp in length
            *   154662   154761: gap of 100 bp
            *   154762   166849: contig of 12088 bp in length
            *   166850   166949: gap of 100 bp
            *   166950   172577: contig of 5628 bp in length.
FEATURES             Location/Qualifiers
     source          1. .172577
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-435N22"
                     /clone_lib="RPCI-11.2"
     misc_feature    1. .6277
                     /note="assembly_fragment:00258
                     fragment_chain:1"
     misc_feature    6378. .9441
                     /note="assembly_fragment:01194
                     fragment_chain:1"
     misc_feature    9542. .24962
                     /note="assembly_fragment:01531
                     fragment_chain:2"
     misc_feature    25063. .29276
                     /note="assembly_fragment:00634
                     fragment_chain:2"
     misc_feature    29377. .37516
                     /note="assembly_fragment:01777
                     fragment_chain:3"
     misc_feature    37617. .48139
                     /note="assembly_fragment:00172
                     fragment_chain:3"
     misc_feature    48240. .53407
                     /note="assembly_fragment:00476"
     misc_feature    53508. .74246
                     /note="assembly_fragment:00575"
     misc_feature    74347. .77364
                     /note="assembly_fragment:00599"
     misc_feature    77465. .81839
                     /note="assembly_fragment:00625"
     misc_feature    81940. .84471
                     /note="assembly_fragment:00833"
     misc_feature    84572. .87957
                     /note="assembly_fragment:01082"
     misc_feature    88058. .91000
                     /note="assembly_fragment:01216"
     misc_feature    91101. .103888
                     /note="assembly_fragment:01263"
     misc_feature    103989. .118772
                     /note="assembly_fragment:01287"
     misc_feature    118873. .121121
                     /note="assembly_fragment:01388"
     misc_feature    121222. .126906
                     /note="assembly_fragment:01454"
     misc_feature    127007. .130692
                     /note="assembly_fragment:01462"
     misc_feature    130793. .134245
                     /note="assembly_fragment:01596"
     misc_feature    134346. .137280
                     /note="assembly_fragment:01622"
     misc_feature    137381. .144713
                     /note="assembly_fragment:01718"
     misc_feature    144814. .147044
                     /note="assembly_fragment:01746"
     misc_feature    147145. .154661
                     /note="assembly_fragment:01761"
     misc_feature    154762. .166849
                     /note="assembly_fragment:01771"
     misc_feature    166950. .172577
                     /note="assembly_fragment:00064
                     clone_end:T7
                     vector_side:right"
ORIGIN

  Query Match             5.1%;  Score 55.6;  DB 2;  Length 172577;
  Best Local Similarity   51.3%;  Pred. No. 0.47;
  Matches  153;  Conservative    0;  Mismatches  144;  Indels    1;  Gaps    1;

Qy        228 TGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAA 287
              | || || | | | | | | | | ||      ||| |||| ||| |   || | |||
Db     126887 TTATATTATAATTAAAATTAATAATTATTATAATATATTTTTTTTATTTATTATATAAA 126828

Qy        288 AAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACA 347
              |||  |||||  |  |  |   | | ||| || |||   ||   | || | ||| | |
Db     126827 AAAAATATAATAATTTAAAAAAAAATAAAAATAATAATATAATAATTTTATAATAATAAT 126768

Qy        348 CAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAA 407
               || |  |    ||  || | || ||||  ||  | | |||   | ||| | ||  |||
Db     126767 AAATAATAAATTATTATATAATATTTAAAATTA-ATTAAAAATAATAAAATAAATTAAAA 126709

Qy        408 GGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGG 467
              | | ||     | |  | |   | |    | |  ||| |    |   | || |
Db     126708 ATAAAAATATAATTTATTATTTAATATTAATTAAAATAATAAAATAATTAATAATATAAT 126649

Qy        468 AAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATA 525
              | ||||| | | |||| | ||| ||| ||  |  |  |||| | |||  ||  ||| |
Db     126648 ATATAATATAATAAAATAAAATATTTAATAAATAAAAAAAATAATTAAATTTAAAAAA 126591


RESULT 37
CR847802
LOCUS       CR847802              206038 bp    DNA     linear   HTG 05-OCT-2004
DEFINITION  Danio rerio clone DKEY-48N1, *** SEQUENCING IN PROGRESS ***, 14
            unordered pieces.
ACCESSION   CR847802
VERSION     CR847802.2  GI:53828252
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 206038)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 5, 2004 this sequence version replaced gi:53749620.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK48N1
            -------------- Summary Statistics
```

```
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 201054 bases at least Q40
            Consensus quality: 201870 bases at least Q30
            Consensus quality: 202463 bases at least Q20
            Insert size: 204738; sum-of-contigs
            Insert size: 207929; 5.7% error; agarose-fp
            Quality coverage: 7.77x in Q20 bases; sum-of-contigs Quality
            coverage: 7.88x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 14 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     3987: contig of 3987 bp in length
            *     3988     4087: gap of 100 bp
            *     4088     6329: contig of 2242 bp in length
            *     6330     6429: gap of 100 bp
            *     6430    11109: contig of 4680 bp in length
            *    11110    11209: gap of 100 bp
            *    11210    15846: contig of 4637 bp in length
            *    15847    15946: gap of 100 bp
            *    15947    18312: contig of 2366 bp in length
            *    18313    18412: gap of 100 bp
            *    18413    28830: contig of 10418 bp in length
            *    28831    28930: gap of 100 bp
            *    28931    37979: contig of 9049 bp in length
            *    37980    38079: gap of 100 bp
            *    38080    94886: contig of 56807 bp in length
            *    94887    94986: gap of 100 bp
            *    94987   104471: contig of 9485 bp in length
            *   104472   104571: gap of 100 bp
            *   104572   142831: contig of 38260 bp in length
            *   142832   142931: gap of 100 bp
            *   142932   148148: contig of 5217 bp in length
            *   148149   148248: gap of 100 bp
            *   148249   183568: contig of 35320 bp in length
            *   183569   183668: gap of 100 bp
            *   183669   189062: contig of 5394 bp in length
            *   189063   189162: gap of 100 bp
            *   189163   206038: contig of 16876 bp in length.
FEATURES             Location/Qualifiers
     source          1. .206038
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-48N1"
                     /clone_lib="DanioKey"
     misc_feature    1. .3987
                     /note="assembly_fragment:00082
                     fragment_chain:1"
     misc_feature    4088. .6329
                     /note="assembly_fragment:00047
                     fragment_chain:1"
     misc_feature    6430. .11109
                     /note="assembly_fragment:00264
                     fragment_chain:1"
     misc_feature    11210. .15846
                     /note="assembly_fragment:00103
                     fragment_chain:1"
     misc_feature    15947. .18312
                     /note="assembly_fragment:00030
                     fragment_chain:1"
     misc_feature    18413. .28830
                     /note="assembly_fragment:00409
                     fragment_chain:1"
     misc_feature    28931. .37979
                     /note="assembly_fragment:00321
                     fragment_chain:1"
     misc_feature    38080. .94886
                     /note="assembly_fragment:01954
                     fragment_chain:1"
     misc_feature    94987. .104471
                     /note="assembly_fragment:00544
                     fragment_chain:1"
     misc_feature    104572. .142831
                     /note="assembly_fragment:01458
                     fragment_chain:1"
     misc_feature    142932. .148148
                     /note="assembly_fragment:00154
                     fragment_chain:1"
     misc_feature    148249. .183568
                     /note="assembly_fragment:00983
                     fragment_chain:1"
     misc_feature    183669. .189062
                     /note="assembly_fragment:00207
                     fragment_chain:1"
     misc_feature    189163. .206038
                     /note="assembly_fragment:00723.0"
ORIGIN

  Query Match             5.1%;  Score 55.6;  DB 2;  Length 206038;
  Best Local Similarity   38.1%;  Pred. No. 0.46;
  Matches  229;  Conservative    0;  Mismatches  372;  Indels    0;  Gaps    0;

Qy        165 AATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTC 224
              | | | ||   | |||| ||||| |         | || || |      |   | | | |
Db     141746 ATTTATTTTTCATTTTTTTAGAAAACAAACAATAAACGAGACATGATCAACATTTAAGTG 141805

Qy        225 CCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTC 284
              ||  || | | || |     || | | ||| |  |||  || | | | |   | |
Db     141806 TCTATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 141865

Qy        285 AAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAA 344
              | | |  |||| | | ||  ||  | | | | ||        ||   |    ||||  ||
Db     141866 TATATATATATATATATATATATATATATATATATATATATATATATATAAAAAAAAAAA 141925

Qy        345 ACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCT 404
              | | ||||  |    | |  | |  |   ||       |   ||    || ||   |
Db     141926 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAANAAAAANAAAAAANNNN 141985

Qy        405 AAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTT 464
              |   || | ||  |  ||| || | |   |     || | | |    | |||    |  |
Db     141986 NANNTTAAANTAAATTATTTTTTTTTTTATATTTAAAAAATAAANAAAAAAAAAAAAAAT 142045

Qy        465 TGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAAT 524
                | ||||  | | || |  |||| |||  | |    ||       |||        | |
Db     142046 AAAATATAAATTAATAATAATTAATATTTNNTTTTTNTTTTTTNNTTTTNANNAAAANNT 142105

Qy        525 ATTCTGTCATTCAGGAGTATGGAAGTATCGATGGCTTCTTTAAAATGAAGCAGGAGGGTC 584
              ||       ||          | |||   |  || |  ||||  ||      |
Db     142106 TTTTANNTTTTNNTTTAATATATAATATTATTTAATTATAAAAAAAAAAAANAAAANNNA 142165

Qy        585 TGGCAGAGAGTATCTATGAAATAAGTTCCTCTGACCTTCACGCTTAATTTTCTGAATGGA 644
                |     |                                  |     |     |
Db     142166 ANNAATTNNNNANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNANATNANNNNAATNNNA 142225

Qy        645 GTGGAGCAAATTACTTCAAGCTTCACTTAACTTGCATATGAAATGAACCGTACAAAAATA 704
              |       |     | ||  || |   || ||  |||  |||   |      | ||| |
Db     142226 NTATTNANNAAANAATNAAAATTTATAAAANTTAAATAAAAAAAATATATATAATAAAAA 142285

Qy        705 CAAGAGTGTCAGGAGAAAGTTATGCTCTGGTAAATATTTTGCAAAACAGATAAAAGATAA 764
              |  |           ||| | |          |||||     |||  | | |||| | ||
Db     142286 TATAAAAAAATATTAAAAATAAAATAANATATAATATAAAAAAAATTATAAAAAAAAAAA 142345

Qy        765 T 765
              |
Db     142346 T 142346


RESULT 38
CR847516
```

```
LOCUS       CR847516              206464 bp    DNA     linear   HTG 01-OCT-2004
DEFINITION  Danio rerio clone DKEY-74J13, *** SEQUENCING IN PROGRESS ***, 11
            unordered pieces.
ACCESSION   CR847516
VERSION     CR847516.2  GI:53748724
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 206464)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 1, 2004 this sequence version replaced gi:52851336.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK74J13
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 204746 bases at least Q40
            Consensus quality: 205167 bases at least Q30
            Consensus quality: 205285 bases at least Q20
            Insert size: 205464; sum-of-contigs
            Insert size: 188759; 6.6% error; agarose-fp
            Quality coverage: 8.18x in Q20 bases; sum-of-contigs Quality
            coverage: 9.05x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    15848: contig of 15848 bp in length
            *    15849    15948: gap of 100 bp
            *    15949    48299: contig of 32351 bp in length
            *    48300    48399: gap of 100 bp
            *    48400    70853: contig of 22454 bp in length
            *    70854    70953: gap of 100 bp
            *    70954    90664: contig of 19711 bp in length
            *    90665    90764: gap of 100 bp
            *    90765    97350: contig of 6586 bp in length
            *    97351    97450: gap of 100 bp
            *    97451   169433: contig of 71983 bp in length
            *   169434   169533: gap of 100 bp
            *   169534   174455: contig of 4922 bp in length
            *   174456   174555: gap of 100 bp
            *   174556   183558: contig of 9003 bp in length
            *   183559   183658: gap of 100 bp
            *   183659   193915: contig of 10257 bp in length
            *   193916   194015: gap of 100 bp
            *   194016   196128: contig of 2113 bp in length
            *   196129   196228: gap of 100 bp
            *   196229   206464: contig of 10236 bp in length.
FEATURES             Location/Qualifiers
     source          1. .206464
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-74J13"
                     /clone_lib="DanioKey"
     misc_feature    1. .15848
                     /note="assembly_fragment:00542
                     fragment_chain:1"
     misc_feature    15949. .48299
                     /note="assembly_fragment:01258
                     fragment_chain:1"
     misc_feature    48400. .70853
                     /note="assembly_fragment:00958
                     fragment_chain:1"
     misc_feature    70954. .90664
                     /note="assembly_fragment:00695
                     fragment_chain:1"
     misc_feature    90765. .97350
                     /note="assembly_fragment:00090
                     fragment_chain:2"
     misc_feature    97451. .169433
                     /note="assembly_fragment:01703
                     fragment_chain:2"
     misc_feature    169534. .174455
                     /note="assembly_fragment:00046
                     fragment_chain:2"
     misc_feature    174556. .183558
                     /note="assembly_fragment:00248
                     fragment_chain:3"
     misc_feature    183659. .193915
                     /note="assembly_fragment:00393
                     fragment_chain:3"
     misc_feature    194016. .196128
                     /note="assembly_fragment:00030"
     misc_feature    196229. .206464
                     /note="assembly_fragment:00138.0"
ORIGIN

  Query Match             5.1%;  Score 55.6;  DB 2;  Length 206464;
  Best Local Similarity   50.8%;  Pred. No. 0.46;
  Matches  133;  Conservative    0;  Mismatches  129;  Indels    0;  Gaps    0;

Qy        117 TCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATA 176
              | || ||  || | |||    |||   |   | || ||  ||| || | |||| ||| |
Db     169686 TTTTTAACAGTTTTATACCCTTTTAACTACAATATTTTACTCAATAAACATCAATTATTG 169745

Qy        177 GTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTT 236
              | |  |   |||| |   ||  || | || || | || |  |  |   ||   || |
Db     169746 GCTAAGATTAAGAGATGCACTACAAAATTTTCTTATTTTGAGGTTTTGTTTTTTATTCTA 169805

Qy        237 TAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATA 296
              | || |   ||  | | |    | |  |||| ||||    || || ||||| || ||
Db     169806 AATAATATTCTAAGTAAAAAATATTGTCAATTTTTTTTAAAGTTTTTAAAAAAGTGTCTA 169865

Qy        297 AAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAG 356
                |||   |   |   ||     ||   |||||  |||| ||   ||| | |  ||
Db     169866 CTTAAAACAAGCAAACCAATTCAGATTATTTAGCTTGTTGTAAGGGAAAAAAATTATTTA 169925

Qy        357 TGCATATTACATTACTTAATCT 378
              |  ||||   | | ||| | |
Db     169926 AGTTTATTCATTCATTTATTTT 169947


RESULT 39
CR847887/c
LOCUS       CR847887              225328 bp    DNA     linear   HTG 08-OCT-2004
DEFINITION  Danio rerio clone DKEY-178N6, *** SEQUENCING IN PROGRESS ***, 7
            unordered pieces.
ACCESSION   CR847887
VERSION     CR847887.2  GI:54019808
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 225328)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
```

```
  JOURNAL   Submitted (07-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 9, 2004 this sequence version replaced gi:53850225.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK178N6
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 222988 bases at least Q40
            Consensus quality: 223337 bases at least Q30
            Consensus quality: 223737 bases at least Q20
            Insert size: 224728; sum-of-contigs
            Insert size: 220016; 4.5% error; agarose-fp
            Quality coverage: 9.09x in Q20 bases; sum-of-contigs Quality
            coverage: 9.29x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     2652: contig of 2652 bp in length
            *    2653     2752: gap of 100 bp
            *    2753   111362: contig of 108610 bp in length
            *  111363   111462: gap of 100 bp
            *  111463   124950: contig of 13488 bp in length
            *  124951   125050: gap of 100 bp
            *  125051   128865: contig of 3815 bp in length
            *  128866   128965: gap of 100 bp
            *  128966   178608: contig of 49643 bp in length
            *  178609   178708: gap of 100 bp
            *  178709   217332: contig of 38624 bp in length
            *  217333   217432: gap of 100 bp
            *  217433   225328: contig of 7896 bp in length.
FEATURES             Location/Qualifiers
     source          1. .225328
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-178N6"
                     /clone_lib="DanioKey"
     misc_feature    1. .2652
                     /note="assembly_fragment:00004
                     fragment_chain:1"
     misc_feature    2753. .111362
                     /note="assembly_fragment:01763
                     fragment_chain:1"
     misc_feature    111463. .124950
                     /note="assembly_fragment:00219
                     fragment_chain:1"
     misc_feature    125051. .128865
                     /note="assembly_fragment:00045
                     fragment_chain:2"
     misc_feature    128966. .178608
                     /note="assembly_fragment:01061
                     fragment_chain:2"
     misc_feature    178709. .217332
                     /note="assembly_fragment:00426
                     fragment_chain:2"
     misc_feature    217433. .225328
                     /note="assembly_fragment:00093.0"
ORIGIN

  Query Match             5.1%;  Score 55.6;  DB 2;  Length 225328;
  Best Local Similarity   40.6%;  Pred. No. 0.45;
  Matches  205;  Conservative    0;  Mismatches  300;  Indels    0;  Gaps    0;

Qy         23 TAACTTTCACTCTCTATGTTCTTATAAGCAAATGTTACAATGAGATATGAGTTTTAAAGC 82
              ||  | |   | | ||| |   ||||   | || |    ||        |  | || |
Db     111984 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATA 111925

Qy         83 CAGATCTTCCTTATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGG 142
              | || ||  ||  | | | |       | | |  || |     | | | ||||   |||
Db     111924 TATATATTTTTTTTTTTTTTTTATTTATATTTTTATATTTTTTTTTAATATTTTTTTAA 111865

Qy        143 TTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAG 202
                |    | |   |    |   | | |    |||       | ||  |     |
Db     111864 AAAATTTNTTTATTTTTATNATTAATATTAATATATTTATATATATATATATAAATTTTT 111805

Qy        203 ACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTA 262
               ||| |||| ||| ||||      | || | || |||| |    | | | | | | |
Db     111804 TTATTATTTTTTTATTATATAAAAATAATATATTAAAAAAAAAAAAAAAAAAAAAATTT 111745

Qy        263 TTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCAT 322
              ||  ||| || ||  |||   |   |      ||| || ||||   |  | | |||  | |
Db     111744 TTTTTTTTTTNTTNNTTTNNTNTTATTTTTTATNAATAAATTTTTAAATAAAAAATTTTT 111685

Qy        323 ATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCA 382
               | ||       || ||||  ||| | ||||  |          |
Db     111684 TTTTTTTTATTTTAAAAAAAAAAAAAAAAAAAAAAAAAANNNANNNAANNNNNNNNNNNNNN 111625

Qy        383 GTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATA 442
                    |    |     |||  | |      | |   ||   |    ||||
Db     111624 NNNNNNNNNANNNNTTNNNNANAAANAAAAATATTTNTTTATTAAATAAAAAAATAAATT 111565

Qy        443 GAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTA 502
                 |   |   ||   | |    ||  |   | | |||| ||| | | | |   | |
Db     111564 TTTTATTATAATTATAAAAATAAATAATAATATTAAAAAAAAATATTATATAAATAAAAA 111505

Qy        503 TTAAAACATTTACTTTCCAAATATT 527
                | ||| |  |  |||  | ||| |
Db     111504 ATTAAATAAATTATTTAAATATAAT 111480


RESULT 40
BX936449
LOCUS       BX936449              232621 bp    DNA     linear   HTG 17-FEB-2004
DEFINITION  Danio rerio clone CH211-221N14, *** SEQUENCING IN PROGRESS ***, 11
            unordered pieces.
ACCESSION   BX936449
VERSION     BX936449.3  GI:42592546
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 232621)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-FEB-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Feb 17, 2004 this sequence version replaced gi:42538933.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zC221N14
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 229064 bases at least Q40
            Consensus quality: 229907 bases at least Q30
            Consensus quality: 230610 bases at least Q20
```

```
            Insert size: 231621; sum-of-contigs
            Insert size: 215663; 3.1% error; agarose-fp
            Quality coverage: 7.74x in Q20 bases; sum-of-contigs Quality
            coverage: 8.79x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     6275: contig of 6275 bp in length
            *     6276     6375: gap of 100 bp
            *     6376    65163: contig of 58788 bp in length
            *    65164    65263: gap of 100 bp
            *    65264   138959: contig of 73696 bp in length
            *   138960   139059: gap of 100 bp
            *   139060   141155: contig of 2096 bp in length
            *   141156   141255: gap of 100 bp
            *   141256   163189: contig of 21934 bp in length
            *   163190   163289: gap of 100 bp
            *   163290   201923: contig of 38634 bp in length
            *   201924   202023: gap of 100 bp
            *   202024   208476: contig of 6453 bp in length
            *   208477   208576: gap of 100 bp
            *   208577   215288: contig of 6712 bp in length
            *   215289   215388: gap of 100 bp
            *   215389   219000: contig of 3612 bp in length
            *   219001   219100: gap of 100 bp
            *   219101   223490: contig of 4390 bp in length
            *   223491   223590: gap of 100 bp
            *   223591   232621: contig of 9031 bp in length.
FEATURES             Location/Qualifiers
     source          1. .232621
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-221N14"
                     /clone_lib="CHORI-211"
     misc_feature    1. .6275
                     /note="assembly_fragment:00281
                     fragment_chain:1"
     misc_feature    6376. .65163
                     /note="assembly_fragment:01195
                     fragment_chain:1"
     misc_feature    65264. .138959
                     /note="assembly_fragment:02005
                     fragment_chain:1"
     misc_feature    139060. .141155
                     /note="assembly_fragment:00080"
     misc_feature    141256. .163189
                     /note="assembly_fragment:00430"
     misc_feature    163290. .201923
                     /note="assembly_fragment:00706
                     fragment_chain:2"
     misc_feature    202024. .208476
                     /note="assembly_fragment:00229
                     fragment_chain:2"
     misc_feature    208577. .215288
                     /note="assembly_fragment:00184
                     fragment_chain:2"
     misc_feature    215389. .219000
                     /note="assembly_fragment:00093
                     fragment_chain:2"
     misc_feature    219101. .223490
                     /note="assembly_fragment:00140
                     fragment_chain:2"
     misc_feature    223591. .232621
                     /note="assembly_fragment:00349
                     fragment_chain:2
                     clone_end:T7
                     vector_side:right"
ORIGIN

  Query Match             5.1%;  Score 55.6;  DB 2;  Length 232621;
  Best Local Similarity   50.8%;  Pred. No. 0.45;
  Matches  133;  Conservative    0;  Mismatches  129;  Indels    0;  Gaps    0;

Qy        117 TCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATA 176
              | || ||  || | |||     |||    |   | || ||  ||| || | |||| ||| |
Db      54622 TTTTTAACAGTTTTATACCCTTTTAACTACAATATTTTACTCAATAAACATCAATTATTG 54681

Qy        177 GTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTT 236
              | |  |   ||||  |    ||  || | || || | || |  |  |   ||   || |
Db      54682 GCTAAGATTAAGAGATGCACTACAAAATTTTCTTATTTTGAGGTTTTGTTTTTTATTCTA 54741

Qy        237 TAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATA 296
              | || |   ||   | | |     | |  |||| ||||    || || |||||  || ||
Db      54742 AATAATATTCTAAGTAAAAAATATTGTCAATTTTTTTTAAAGTTTTTAAAAAAGTGTCTA 54801

Qy        297 AAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAG 356
              |||   |   |   ||     ||  |||||  |||| ||   ||| | |  ||
Db      54802 CTTAAAACAAGCAAACCAATTCAGATTATTTAGCTTGTTGTAAGGGAAAAAAATTATTTA 54861

Qy        357 TGCATATTACATTACTTAATCT 378
               |  ||||   | | ||| | |
Db      54862 AGTTTATTCATTCATTTATTTT 54883


RESULT 41
AF527403
LOCUS       AF527403                532 bp    DNA     linear   INV 07-AUG-2002
DEFINITION  Cydia pomonella haplotype A16 12S ribosomal RNA and tRNA-Met genes,
            partial sequence; mitochondrial genes for mitochondrial products.
ACCESSION   AF527403
VERSION     AF527403.1  GI:22128529
KEYWORDS    .
SOURCE      mitochondrion Cydia pomonella (codling moth)
  ORGANISM  Cydia pomonella
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
            Tortricoidea; Tortricidae; Olethreutinae; Cydia.
REFERENCE   1  (bases 1 to 532)
  AUTHORS   Thaler,R. and Dallinger,R.
  TITLE     Phylogeny and pesticide resistance of Cydia pomonella
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 532)
  AUTHORS   Thaler,R., Dallinger,R. and Zelger,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JUL-2002) Inst. of Zoology & Limnology, University of
            Innsbruck, Technikerstr. 25, Innsbruck A-6020, Austria
FEATURES             Location/Qualifiers
     source          1. .532
                     /organism="Cydia pomonella"
                     /organelle="mitochondrion"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:82600"
                     /haplotype="A16"
     rRNA            <1. .94
                     /product="12S ribosomal RNA"
     misc_feature    95. .514
                     /note="A-T rich region"
     tRNA            515. .>532
                     /product="tRNA-Met"
ORIGIN

  Query Match             5.1%;  Score 55.2;  DB 3;  Length 532;
  Best Local Similarity   49.2%;  Pred. No. 1.5;
  Matches  176;  Conservative    0;  Mismatches  178;  Indels    4;  Gaps    1;

Qy        156 ATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTT 215
              || |||| ||||||   || | || | || |  || |  | ||  | | ||| |||| ||
Db        133 ATTATTATAAATCATAAAAAATTTATTTATATAATTTTTTTTTGTATAGATTTTTTTTT 192
```

```
Qy        216 AATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTT 275
Db        193 TTTTTTTTATATTAAATATTTAATAATAATAATAAATATTAAATAATTTCTTTTTTTTTT 252

Qy        276 GGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTT 335
Db        253 TATTTATAATATTCATATTAAAAATTACATTTGCTATTTAAAATTTTATTAATTAATATT 312

Qy        336 GAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAA 395
Db        313 CAATATTAAATTTTAAATATTCATATTCATAATTACATAAAAATTTATATAAATAATAAT 372

Qy        396 AGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAA 455
Db        373 ATAAAATTTAATTTTAAATATTTATTTATAATTTATTATTATTATT----ATTATTAAAT 428

Qy        456 ATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTT 513
Db        429 ATTTAATATTAAAATAATATTAATAAATATTTAATAATAATATTTTATTAAACCATTT 486


RESULT 42
PFMAL8P1_07
WPCOMMENT
  Sequence split into 14 fragments  LOCUS PFMAL8P1 Accession AL844507
     Fragment Name      Begin     End
     PFMAL8P1_00            1    110000
     PFMAL8P1_01       100001    210000
     PFMAL8P1_02       200001    310000
     PFMAL8P1_03       300001    410000
     PFMAL8P1_04       400001    510000
     PFMAL8P1_05       500001    610000
     PFMAL8P1_06       600001    710000
     PFMAL8P1_07       700001    810000
     PFMAL8P1_08       800001    910000
     PFMAL8P1_09       900001   1010000
     PFMAL8P1_10      1000001   1110000
     PFMAL8P1_11      1100001   1210000
     PFMAL8P1_12      1200001   1310000
     PFMAL8P1_13      1300001   1325595
Continuation (8 of 14) of PFMAL8P1 from base 700001 (AL844507 Plasmodium falciparum 3D7

  Query Match             5.0%;  Score 55;  DB 2;  Length 110000;
  Best Local Similarity   47.2%;  Pred. No. 0.66;
  Matches  230;  Conservative    0;  Mismatches  255;  Indels    2;  Gaps    2;

Qy         52 AAATGTTACAATGAGATATGAGTTTTAAAGCCAGATCTTCCTTATCTCTCTGCCCCATCT 111
Db     103802 ATATGTAAATATGATGTATAAATTTTAATATCATATCAT-CTGTTATATATAAAAAATAT 103860

Qy        112 CTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGT 171
Db     103861 ATAATTTTTATAGTACTATTTTATATATTATATATATATTTTATTATAATAATATTGTAT 103920

Qy        172 TAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGAT 231
Db     103921 GTATTATATAATCAGTTATAAAATGTAGAAAAAATTTTCTTTTTTTTTTTTTTTTTTTT 103980

Qy        232 TTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAG 291
Db     103981 TTTTGTTATATATATTTGAATGAAAAAAATATAAATATTATATTATATATATTATTATCT 104040

Qy        292 GTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAA 351
Db     104041 ATATAAATAAATATAGTGATATTAATTAAATATATATATATATTATAGTATTATCATA 104100

Qy        352 ATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTA 411
Db     104101 TGTTTTAATAAATAAATACATTATTATTATATATAAAAATAAAAAAAAAAAGTAATTTTT 104160

Qy        412 TACTACCTGCTTTCTTATCGCACT-GCAAATAGAAATTACCACAAATTTTATTTTGGAAA 470
Db     104161 TAATAATTTTTTTTTATACATTATTAATATGAATATATAAAAATATATATAATATATAATA 104220

Qy        471 TAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTG 530
Db     104221 TATCATTATAGAATAATAATAAAAAAATACTTATTTATTTTTTATTTTTTTTTAATACTT 104280

Qy        531 TCATTCA 537
Db     104281 TGTTTTA 104287


RESULT 43
PFMAL8P1_08
WPCOMMENT
  Sequence split into 14 fragments  LOCUS PFMAL8P1 Accession AL844507
     Fragment Name      Begin     End
     PFMAL8P1_00            1    110000
     PFMAL8P1_01       100001    210000
     PFMAL8P1_02       200001    310000
     PFMAL8P1_03       300001    410000
     PFMAL8P1_04       400001    510000
     PFMAL8P1_05       500001    610000
     PFMAL8P1_06       600001    710000
     PFMAL8P1_07       700001    810000
     PFMAL8P1_08       800001    910000
     PFMAL8P1_09       900001   1010000
     PFMAL8P1_10      1000001   1110000
     PFMAL8P1_11      1100001   1210000
     PFMAL8P1_12      1200001   1310000
     PFMAL8P1_13      1300001   1325595
Continuation (9 of 14) of PFMAL8P1 from base 800001 (AL844507 Plasmodium falciparum 3D7 c

  Query Match             5.0%;  Score 55;  DB 2;  Length 110000;
  Best Local Similarity   47.2%;  Pred. No. 0.66;
  Matches  230;  Conservative    0;  Mismatches  255;  Indels    2;  Gaps    2;

Qy         52 AAATGTTACAATGAGATATGAGTTTTAAAGCCAGATCTTCCTTATCTCTCTGCCCCATCT 111
Db       3802 ATATGTAAATATGATGTATAAATTTTAATATCATATCAT-CTGTTATATATAAAAAATAT 3860

Qy        112 CTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGT 171
Db       3861 ATAATTTTTATAGTACTATTTTATATATTATATATATATTTTATTATAATAATATTGTAT 3920

Qy        172 TAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGAT 231
Db       3921 GTATTATATAATCAGTTATAAAATGTAGAAAAAATTTTCTTTTTTTTTTTTTTTTTTTT 3980

Qy        232 TTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAG 291
Db       3981 TTTTGTTATATATATTTGAATGAAAAAAATATAAATATTATATTATATATATTATTATCT 4040

Qy        292 GTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAA 351
Db       4041 ATATAAATAAATATAGTGATATTAATTAAATATATATATATATTATAGTATTATCATA 4100

Qy        352 ATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTA 411
Db       4101 TGTTTTAATAAATAAATACATTATTATTATATATAAAAATAAAAAAAAAAAGTAATTTTT 4160

Qy        412 TACTACCTGCTTTCTTATCGCACT-GCAAATAGAAATTACCACAAATTTTATTTTGGAAA 470
Db       4161 TAATAATTTTTTTTATACATTATTAATATGAATATATAAAAATATATATAATATATAATA 4220

Qy        471 TAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTG 530
Db       4221 TATCATTATAGAATAATAATAAAAAAATACTTATTTATTTTTTATTTTTTTTTAATACTT 4280

Qy        531 TCATTCA 537
Db       4281 TGTTTTA 4287


RESULT 44
```

```
AC117076
LOCUS       AC117076               146570 bp    DNA     linear   INV 21-FEB-2004
DEFINITION  Dictyostelium discoideum chromosome 2 map 3323568-3470138 strain
            AX4, complete sequence.
ACCESSION   AC117076 AC115610
VERSION     AC117076.3  GI:42733926
KEYWORDS    HTG.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 146570)
  AUTHORS   Gloeckner,G., Eichinger,L., Szafranski,K., Pachebat,J., Dear,P.,
            Lehmann,R., Baumgart,C., Parra,G., April,J.F., Guigo,R., Kumpf,K.,
            Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and
            Noegel,A.A.
  TITLE     Sequence and analysis of chromosome 2 of Dictyostelium discoideum
  JOURNAL   Nature 418 (6893), 79-85 (2002)
  MEDLINE   22092622
   PUBMED   12097910
  REMARK    The Dictyostelium Genome Sequencing Consortium
REFERENCE   2  (bases 1 to 146570)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-APR-2002) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   3  (bases 1 to 146570)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   4  (bases 1 to 146570)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   5  (bases 1 to 146570)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-FEB-2004) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
COMMENT     On or before Feb 21, 2004 this sequence version replaced
            gi:28828883, gi:28829496.
            CDS predictions from GeneID do not necessarily reflect true genes.
            Further Information is available from IMB Jena, Department of
            Genome Analysis
            (http://genome.imb-jena.de/dictyostelium/)
            and the Univerity Colonge, Institute for Biochemistry I
            (http://www.uni-koeln.de/dictyostelium/project.shtml
            Funding
            Agency : Deutsche Forschungsgemeinschaft (DFG).
FEATURES             Location/Qualifiers
     source          1. .146570
                     /organism="Dictyostelium discoideum"
                     /mol_type="genomic DNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /chromosome="2"
                     /map="3323568-3470138"
     CDS             complement(join(<63. .86,436. .603,692. .704))
                     /note="GeneID exon scores (in order of location ranges):
                     1.46, 2.66, -0.45 - GSCJ_ID dd_03180"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAS38882.1"
                     /db_xref="GI:42733988"
                     /translation="MTILGLISSISNVKSISKSNNLSSLSNSLSSLQSMNSIERRGGI
                     SISNPGFLGGIGRNPGESMVCDSG"
     CDS             complement(join(1507. .1763,1872. .1884))
                     /note="GeneID exon scores (in order of location ranges):
                     19.50, 3.22 - GSCJ_ID dd_01276"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAS38835.1"
                     /db_xref="GI:42733927"
                     /translation="MTILGSISSIGNVKSITKSNNLSSLSNSSSSTQSMNSIQCGGGC
                     GNGGLLGGVGGLVGGVLVGTGVIIGSVLHGVGSILTGGSNNCGCN"
     CDS             join(2611. .2623,2691. .2947)
                     /note="GeneID exon scores (in order of location ranges):
                     3.18, 23.32 - GSCJ_ID dd_01275"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAS38836.1"
                     /db_xref="GI:42733928"
                     /translation="MTILGSISSIGNVKLISKSNNLSSLSNSSSSLQSMNSIQCGGGC
                     GNGGLLGGVGGLVGGVLVGTGVIVGSVLHGVGSILTGGSNNCGCN"
     CDS             complement(join(3484. .3740,3824. .3836))
                     /note="GeneID exon scores (in order of location ranges):
                     25.27, 0.79 - GSCJ_ID dd_01274"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAS38837.1"
                     /db_xref="GI:42733929"
                     /translation="MTILGSISSIGNVKSISKSNNLSSLSNSSSSLQSMNSIQCGGGC
                     GNGGLLGAVGGLVGGVLTGTGVIVGSVLHGVGSILTGGPNNCGCN"
     CDS             join(4219. .4228,4659. .4915)
                     /note="GeneID exon scores (in order of location ranges):
                     -7.18, 24.38 - GSCJ_ID dd_03039"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAS38838.1"
                     /db_xref="GI:42733930"
                     /translation="MVKCSISSIGNVKSISKSNNLSSLSNSSSSLQSMNSIQCGGGCG
                     NGGLLGGVGGLVGGVLVGTGVIVGSVLHGVGSILTGGSNNCGCN"
     CDS             complement(join(5405. .5655,5746. .5758))
                     /note="GeneID exon scores (in order of location ranges):
                     25.79, -3.80 - GSCJ_ID dd_03235"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAS38839.1"
                     /db_xref="GI:42733931"
                     /translation="MTILASISSIGNVKSISKSNNVSLSNSSSSLQSMNSIQCGGCGN
                     GGLLGAVGGLVGGVLTGTGVIVGSVLHGVGSILTGGSNNCGCN"
     CDS             join(6487. .6499,6572. .6825)
                     /note="GeneID exon scores (in order of location ranges):
                     3.30, 26.87 - GSCJ_ID dd_03141"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAS38840.1"
                     /db_xref="GI:42733932"
                     /translation="MTILASISSIGNVKSISKSNNLSSLSNSSSSLQSMNSIQCGGCG
                     NGGLLGAVGGLVGGVLTGTGVIVGSVLHGVGSILTGGSNNCGCN"
     CDS             join(7061. .7150,7227. .8279,8516. .9178)
                     /note="GeneID exon scores (in order of location ranges):
                     -0.28, 67.69, 40.02 - GSCJ_ID dd_03341"
                     /codon_start=1
                     /product="similar to Dictyostelium discoideum (Slime
                     mold). MkpA protein"
                     /protein_id="AAS38841.1"
                     /db_xref="GI:42733933"
                     /translation="MDTTKNNDEFDISYFLTCSIFGFILWILTEQILKYYNKTNKNNK
                     YNLPKGPSFLKWFINYLFNFYDLKLSNNKEEDNNNNNNKSNNSLSQEELIEDTSENTV
                     LKWFNQLNSDNYSVSFFGRPMIFTRDTTISKYILSSNNIDNYTKPPDSSGVLIRLAQN
                     SILMSEGDQWRYHRSIINQPFSSKNVKLMIPTIITTINKLINHLNNNNNNNNNNNNNN
                     NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTIIIDIHSYCTKLTFDII
                     GKLSIGYDFNSIESSDNDNDNNDDDDISKQFDFILNEMIRPIRRFSSYLPLYNDIKLF
                     KFLNELESIIKGAINSRSLITDNNNNKTYKKNFLLDNLLDDNVKEKDIIGNINTFLLA
                     GHETSANLLTFIFYLLSTHNNVQNDLYNHLIENQKKKINKDNKFTEEDEDYQSIEFLD
                     WVIYETLRLFPPAPMIGRTSKNDDILKSGNNNNNNNNNISIPSETLILISVYAIHRDP
                     KLWKDPNIFNPYRWKNIENINNRSDFIPFSSGGRVCVGQKFSIVEARIIISKLILNFE
                     LSFNNLKSKPFKIYQRATLTPKYPVFLNFKKRENK"
     CDS             complement(9415. .11361)
                     /note="GeneID exon scores (in order of location ranges):
                     133.26 - GSCJ_ID dd_03256"
                     /codon_start=1
                     /product="similar to Dictyostelium discoideum (Slime
                     mold). CIGB protein (Fragment)"
```

```
                     /protein_id="AAS38842.1"
                     /db_xref="GI:42733934"
                     /translation="MMEPINYKNISFNDIKLDSNCNSSSSSSSSSIGSNRSGSGGGGA
                     ISRKSFDCSTSNGISEFGGDDGGCGGGCGGDMGNNINISINKRHNSVNSISSNKNINC
                     LYHDLNDLITYCNDCERMICVQCIVSNDHCGHKFTEISKPFASSLLEKFQKQQFPNLL
                     GYRESDKHIISALDYCCNDVERQHIESVKQVKQHFQKLIESIQSIEKDIIEKLQSTVE
                     WNKELQNSIKQQFERDSKKIDSIISKHHSKINGYQIDQIYDFETKQHIEIIKDSDQCT
                     PLLKRRLEILNNNSSSKIKFKNDLLTFNMEIFDEIKEKVLNFTKIENVEYKDLKSTKQ
                     HLMLNKKIQPYNIDYYLCHGEPIPLGETSVALDSKYSSKSICIQSNKSIDRLLLCNGF
                     KLSLSEFKIPDSIKTLRIGDIKYSWDANSIPSSVTELHLCDGFNHILSSGMIPNTVTK
                     LYLYDIKQPLTMGSIPSSVNELHMCDGPSLQLAPFVLPNSITSLSIGEIKYSILPEFI
                     PSSVKKIRLYSSFNHPIIPQNISESVETLVLFNIKLNLENKVIPKSIKELHLCDGFEQ
                     QIQPFSIENGGKLEKLVLCDIKEKLLKNSIPNSVKKLIFLGYNHQLENGILNDSLELI
                     YFSKSFTHPIDNNLIPNNIEIIKK"
     CDS             complement(join(12693. .12990,13133. .13227))
                     /note="GeneID exon scores (in order of location ranges):
                     18.88, 1.20 - GSCJ_ID dd_02598"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAO52031.2"
                     /db_xref="GI:42733935"
                     /translation="MSLNNFSPEEKQNIVRSIFAEEEIVYNVFFERGDDSFKNLKLSE
                     IKAILSILKSSRSMAYPLSGKKQVLVDLLTSIGFVGPINTSPQPQPPQPPPSPTPPGL
                     FQEELQPHSFLELDLGITVELDNLLGEN"
     CDS             join(13881. .13926,14100. .14230,14342. .14589,14703. .15129)
                     /note="GeneID exon scores (in order of location ranges):
                     6.31, 2.00, 17.53, 44.25 - GSCJ_ID dd_02599"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAO52033.2"
                     /db_xref="GI:42733936"
                     /translation="MTLVEYSETLKKSLEEFCKQDKKVVTPELNSIIENISKTGITCY
                     PWNILKELFYFKLNEILDIFEKEYIATSSTTTTSTTTTTTNTSSSSTNSPQSPQSPQS
                     PQLNVNIKDEEMQKIKNQVNLSGLIKMKKEFLDSFKESKLSPFTVQRLCELIINYKMY
                     TSFSKYLCAVEKMLNVSTLPHLTPEEVIEFNKNQNSGSRLNSSTNTTTTTTTTTTTTT
                     PTEKGNNDQPNLDSFAFSSSFTTSFPSDTTSSTSTSLSTDQEMKDLDVEQQKDDTDNT
                     ANENESK"
     CDS             complement(15517. .15972)
                     /note="GeneID exon scores (in order of location ranges):
                     30.20 - GSCJ_ID dd_02600"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAO52034.2"
                     /db_xref="GI:42733937"
                     /translation="MDPGRSHIETIIESSTHKFILFDEPTEETIPYFKDLMKKNSCIN
                     IVRCCEINYDASLFEGVKIHELCFKDGNVPPKDIIERWLEILKQAFIENGKQKTTVGI
                     HCIAGLGRTPLLVCIALIEDGMKPLQAVEFVRSKRKNAINSPQIKFLKE"
     CDS             complement(join(17629. .18615,18760. .18992,19147. .19282))
                     /note="GeneID exon scores (in order of location ranges):
                     132.77, 35.41, 19.03 - GSCJ_ID dd_02602"
                     /codon_start=1
                     /product="similar to seryl-trna synthetase, cytoplasmic
                     (EC 6.1.1.11) [Schizosaccharomyces pombe]"
                     /protein_id="AAS38843.1"
                     /db_xref="GI:42733938"

  Query Match             5.0%;  Score 55;  DB 3;  Length 146570;
  Best Local Similarity   45.6%;  Pred. No. 0.63;
  Matches  193;  Conservative    0;  Mismatches  230;  Indels    0;  Gaps    0;

Qy        179 TTTGTAGAAGATCTCATCTTAAAGACATTGTTTGTTAATATACTCCCTTGATTTTTTTA 238
              ||  |  |  ||||| | | |   | ||     || || |||| ||  | |   ||| |
Db     111033 TTGTTTAACAATCTCTTTTAATGTATATATACATGATATTATAATCTTTGGTGCTTTATT 111092

Qy        239 AAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAA 298
              |     ||||      | |    ||||    | | | | |  | ||||||   | |||
Db     111093 ATTACTATTACTATTTGATAATATTTCACAAATATATTGCTAATAAAAAAAAAAAAAAAA 111152

Qy        299 GAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTG 358
               |||    |   | | ||| |  | | |  |    |     ||||  ||| | ||||| | |
Db     111153 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATTATTT 111212

Qy        359 CATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACC 418
              |||||   || |   | ||   || |||   || || | ||   | |  || ||||
Db     111213 ATTATTAATATAATATTTGTTTTTGTAAAAAAGAATAAATAAATAAATAAATAAACTAAA 111272

Qy        419 TGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCA 478
                  |||    |    |  |  |||| | || |||  || |   ||  |||
Db     111273 CTTACAATTACAATATATGGACCAATAATTTCAACTAATAATAATAAACAATCAATTGAT 111332

Qy        479 GAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGTCATTCAG 538
                ||  | ||    |       ||| ||||| ||     | |   ||    | |
Db     111333 TTTAATTTTATGGAATCACGTGAATTGAAACAGTTGTCAACGATTGATGTAATTAAATCA 111392

Qy        539 GAGTATGGAAGTATCGATGGCTTCTTTAAAATGAAGCAGGAGGGTCTGGCAGAGAGTATC 598
              || ||      ||  | ||       ||||    |   | || | | |  |  ||| |
Db     111393 GAATAGAAGGATAATGGTGTTGATGGTAAATGTACAAAAGATGCTTTATCTAAGAATGCA 111452

Qy        599 TAT 601
              |||
Db     111453 TAT 111455


RESULT 45
AL607043/c
LOCUS       AL607043              169841 bp    DNA     linear   PRI 22-MAY-2002
DEFINITION  Human DNA sequence from clone RP11-309P22 on chromosome 10,
            complete sequence.
ACCESSION   AL607043
VERSION     AL607043.12  GI:21211881
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 169841)
  AUTHORS   Heath,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAY-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On May 25, 2002 this sequence version replaced gi:18476790.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 10, constructed by the Sanger Centre Chromosome 10
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr10
            RP11-309P22 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6.
FEATURES              Location/Qualifiers
     source           1. .169841
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:9606"
                      /chromosome="10"
                      /clone="RP11-309P22"
                      /clone_lib="RPCI-11.2"
```

ORIGIN

Query Match 5.0%; Score 55; DB 9; Length 169841;
Best Local Similarity 45.8%; Pred. No. 0.62;
Matches 302; Conservative 0; Mismatches 350; Indels 7; Gaps 3;

```
Qy         37 TATGTTCTTATAAGCAAATGTTACAATGAGATATGAGTTTTAAAGCCAGATCTTCCTTAT 96
Db      25398 TATATAAATATACATTTATATTATATATATTTATATTATATAAATATACATTTATATTAT 25339

Qy         97 CTCTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATATTGA 156
Db      25338 ATATCTTTATATTATATAAATATACATTTATATTATATATCTTTATATTATATAAATATA 25279

Qy        157 TCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTA 216
Db      25278 CATTTATATTATATATCTTTATATTATATAAATATACATTTATATTATATATCTTTATAT 25219

Qy        217 ATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTG 276
Db      25218 TATATAAATATACATTTATATTATATATCTTTATATTATATAAATATACATTTATATTAT 25159

Qy        277 GTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTG 336
Db      25158 ATATCTTTATATTATATAAATATACATTTATATTATATATCTTTATATTATATAAATATA 25099

Qy        337 AAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAA 396
Db      25098 CATTT---ATATTATATATCTTTATATTATATAAATATACATTTATATTATATATATTTA 25042

Qy        397 GTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAA 456
Db      25041 TATTATATATATTTATATTATATATAATATAAATATATTTATATTATATATATTTATATT 24982

Qy        457 TTTTATTTTGGAAATAATCTCAGAAAACATAATTTT---TTATGTACTATTAAAACATTT 513
Db      24981 ATATATAATATAAATATATTTATATTATATATATTTATATTATATATAATATATATATTT 24922

Qy        514 ACTTTCCAAATATTCTGTCATTCAGGAGTATGGA-AGTATCGATGGCTTCTTTAAAATGA 572
Db      24921 ATATTATATATAATATCGATATTTATATTATATATAATATCGATATTATATATAATATAT 24862

Qy        573 AGCAGGAGGGTCTGGCAGAGAGTATCTATGAAATAAGTTCCTCTGACCTTCACGCTTAAT 632
Db      24861 ATATTATATATAATATATATATTATATATAATATATATATCTATATTATATATAATATAT 24802

Qy        633 TTTCTGAATGGAGTGGAGCAAATTACTTCAAGCTTCACTTAACTTGCATATGAAATGAA 691
Db      24801 ATTTATATTTTATATTTTATAATATATAAATATATATTTTATATTTTATAATATATAAA 24743
```

RESULT 46
AX345933/c

```
LOCUS       AX345933               6118 bp    DNA     linear   PAT 01-FEB-2002
DEFINITION  Sequence 1004 from Patent WO0200928.
ACCESSION   AX345933
VERSION     AX345933.1  GI:18493819
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Olek,A., Piepenbrock,C. and Berlin,K.
  TITLE     Diagnosis of diseases associated with the immune system
  JOURNAL   Patent: WO 0200928-A 1004 03-JAN-2002;
            Epigenomics AG (DE)
FEATURES             Location/Qualifiers
     source          1. .6118
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="chemically treated genomic DNA (Homo sapiens)"
ORIGIN
```

Query Match 5.0%; Score 54.8; DB 6; Length 6118;
Best Local Similarity 48.4%; Pred. No. 1.2;
Matches 210; Conservative 0; Mismatches 222; Indels 2; Gaps 2;

```
Qy         94 TATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATAT 153
Db       2830 TAAACCTATAACCTTAATCAAATTATTTAAATTCCCTTTATCTCCTTTATAAACTAAAAA 2771

Qy        154 TGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTG 213
Db       2770 TAATAATAAC-ACCCTCCAAATTATATTAAAAATTACCTAAATTAATATACACAAAATCC 2712

Qy        214 TTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTT 273
Db       2711 TTATAATTACACCTTTTTATTTTATAAACACTATATAAATACTTCAATATACTATATATA 2652

Qy        274 TTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAG 333
Db       2651 TATAAAAATCTATACCCAAAATATATTTAAAAAATAAATAATAAATCAAAAAATATTTTA 2592

Qy        334 TTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAA 393
Db       2591 TTATTACTTCATACAAATAACATTTTAAATTCTTTTAAATTATTTTCTCTTTAATTCAAT 2532

Qy        394 AAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCAC 453
Db       2531 AACAATATTATAAACAAATACTAAATCCC-AAATATCCTAAAATAAATAAAAATTTTAAA 2473

Qy        454 AAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTT 513
Db       2472 CAATTAATTTTCATAAATAATATCAAAAAATAAAATATCATAAACATTTCTATAACTCCA 2413

Qy        514 ACTTTCCAAATATT 527
Db       2412 ATTACACTAATATT 2399
```

RESULT 47
AC116100/c

```
LOCUS       AC116100               36148 bp    DNA     linear   INV 12-MAR-2003
DEFINITION  Dictyostelium discoideum chromosome 2 map 6163571-6199719 strain
            AX4, complete sequence.
ACCESSION   AC116100
VERSION     AC116100.2  GI:28828263
KEYWORDS    HTG.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 36148)
  AUTHORS   Gloeckner,G., Eichinger,L., Szafranski,K., Pachebat,J., Dear,P.,
            Lehmann,R., Baumgart,C., Parra,G., April,J.F., Guigo,R., Kumpf,K.,
            Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and
            Noegel,A.A.
  TITLE     Sequence and analysis of chromosome 2 of Dictyostelium discoideum
  JOURNAL   Nature 418 (6893), 79-85 (2002)
  MEDLINE   22092622
   PUBMED   12097910
  REMARK    The Dictyostelium Genome Sequencing Consortium
REFERENCE   2  (bases 1 to 36148)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-MAR-2002) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   3  (bases 1 to 36148)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   4  (bases 1 to 36148)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-2003) Genome Analysis, Institute of Molecular
```

```
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
COMMENT     On Mar 4, 2003 this sequence version replaced gi:19703102.
            CDS predictions from GeneID do not necessarily reflect true genes.
            Further Information is available from IMB Jena, Department of
            Genome Analysis
            (http://genome.imb-jena.de/dictyostelium/)
            and the Univerity Colonge, Institute for Biochemistry I
            (http://www.uni-koeln.de/dictyostelium/project.shtml
            Funding
            Agency : Deutsche Forschungsgemeinschaft (DFG).
FEATURES              Location/Qualifiers
     source           1. .36148
                      /organism="Dictyostelium discoideum"
                      /mol_type="genomic DNA"
                      /strain="AX4"
                      /db_xref="taxon:44689"
                      /chromosome="2"
                      /map="6163571-6199719"
     CDS              join(<1647. .3930,4045. .4129)
                      /note="GeneID exon scores (in order of location ranges):
                      183.85, 8.31 - GSCJ_ID dd_01612"
                      /codon_start=3
                      /product="hypothetical protein"
                      /protein_id="AAL93567.2"
                      /db_xref="GI:28828279"
                      /translation="FLNTDTTITFEWNLNDNNYYTKWIIIRPNDAKLEMPFNCVKGET
                      SLKCVYTNSDSTVNQLYGKNVTACINSLGGNCDTPLSPLHYPTPILKSPIPEIPVDGG
                      EVVITGSFLKFVESPNYYEIVGTENKITFIDDLTKQEFNPTNVTLRFPSGCGLNKSLL
                      LPNDQKIPFNYKKSNITSSSLISDNVVELIGTNFCKDQSISFDGKLIPSSDVKILSST
                      KLQFNIVQEYSSTGTPITIGDSEVKLVFPPIVNSYDPFCGLDGIVVIHGKRLKSNNPN
                      SEITVKVENTNCNVLSGASNTTYLQCHIESGYDKNLIVRIDNIASQPIMFYYEQPIIY
                      KTTIVENTVNIDGFCLSKINDIQVDGTTKNNKTKPLSTSDSKVTFNLDPQYYGETFIT
                      ISNEKYTATTNVTAKLFAKVSENPKAGDKGILLDIYYNDPQGTLKCNDKSKLTGSEVK
                      DYTFEINPICGTINLTLTDGTNEFTFQATSMPGSIDKCELLSNGETRCNGSYLVPDEF
                      TSPVGTVKFAGKNVKTTFLNSTIFTFTTLDDMYAGDLYLDSCVISKKTEYKLEPVILS
                      YDANGFSKTGGVASFTGKYFVLNSNHMKIQNCGNNQTSQCTVTSHTIIDCPISIDGPR
                      DRTCYLTKDDNSSILIGDQKDGSFIISFGEPIPIASTSLGIEGGSLIIFGKNFYNTTD
                      YPLLVSVGNVTCNNPILVDESNIHCTLAPNSNTNLFGTIFDNLQDVNVTLAGRSGVGK
                      IFSYSRNLIATTTNDNSFNSIKQCKKQQHYVRPHLQMTEKRSQSFANY"
     CDS              complement(join(4424. .4729,5222. .5770,5887. .5912,
                      6188. .6206))
                      /note="GeneID exon scores (in order of location ranges):
                      38.39, 79.78, 3.11, -3.71 - GSCJ_ID dd_01611"
                      /codon_start=1
                      /product="similar to Gallus gallus (Chicken). 60S
                      ribosomal protein L7A"
                      /protein_id="AAO50940.1"
                      /db_xref="GI:28828264"
                      /translation="MVHKCCFVTLYQVRMSQKQTKPAAATKAAPAKTAVATTKSKKVV
                      KKGEKKIKTRTLFNALYTKNVKNFGTGFGVQPKRDLTHFTHWPRYIKLQRQRRVLLKR
                      LKVPPTINQFTRVFDKNTAVHLFKLLDKYRPEEASVKKARLLKIAEARAATPKGQAAP
                      KAEKPVQHLRFGIDSVTKLIEKKKAKLVVIAHDVDPVELVLYLPTLCRRMDVPYCIVK
                      SKSRLGELVHMRNASCVALTGVNSADSNELALLVESAKQMFNNNSEHRKTWGGNTLSG
                      PARAILAKRQKAEAKESLAKSKI"
     CDS              complement(join(6256. .6418,6544. .6963,8555. .8589))
                      /note="GeneID exon scores (in order of location ranges):
                      20.13, 38.68, 2.95 - GSCJ_ID dd_01610"
                      /codon_start=1
                      /product="similar to Hypothetical ORF; Pth2p
                      [Saccharomyces cerevisiae]"
                      /protein_id="AAO50941.1"
                      /db_xref="GI:28828265"
                      /translation="MSTDLYIEGILIGVLLGYFQAKCFKRSTITKNKEEVEQIKEIEN
                      NNQKEKEKEEKEGDDDDDDDESESIDGDYDENDDYEIEYDEEIGDCKQVLVIRTDLG
                      MTKGKIIAQCCHASLGARQKALKDSKKVLRQWENFGCAKITLKANSKEELVDLRNKAK
                      ELGVVSYLVLDSGHTQIPSGSATVLAVGPAPKEAVDAVTKHLKLY"
     CDS              join(10513. .11358,11462. .11929,12200. .12211)
                      /note="GeneID exon scores (in order of location ranges):
                      71.61, 38.24, -2.42 - GSCJ_ID dd_00360"
                      /codon_start=1
                      /product="hypothetical protein"
                      /protein_id="AAL93570.2"
                      /db_xref="GI:28828266"
                      /translation="MGNCCSSSKRSNGYDSTDDRILEESLDIKGGHNNNNHGGGHHHH
                      HHHHHHIKRNSNGGLYNNEMKSISSENNNNNNNNNNNNNNNNNNNNKRNSLKSSSQ
                      HKLLVDEEKEDLISASVHQKDPYDLSTSSTLTNHSSTNTLVSSGVQTTSSSSTSSIDN
                      LLDATTTTTTTTSTTNNISSQSKNTGSHSYPHTHTIQMSTEKVGPQFKTPITSIPHHM
                      RMEGFQRVKHNSTSSLYIKSTLSTPDNDEILRCMANALLYHIERGSQFPQKTVEIFSE
                      EKYPITKNKIDLKSNPTVETIYKFIRDIFKAEKLDSECAIMCLAYIERIITYTGITLS
                      SINWRRIVLSALILASKVWEDQSVWNVDFLPVFDNLTAADLNCLERQFLAMLQYNVSL
                      NASIYAKYYFELRNFSKLDSNQFPLKPLDKSGVRRLEDHSKASEYRNQK"
     CDS              join(12615. .12639,12966. .13201)
                      /note="GeneID exon scores (in order of location ranges):
                      -2.47, 10.64 - GSCJ_ID dd_00359"
                      /pseudo
                      /codon_start=1
     CDS              complement(join(13370. .13554,13830. .13901,14628. .14643))
                      /note="GeneID exon scores (in order of location ranges):
                      8.81, 5.58, -8.59 - GSCJ_ID dd_00358"
                      /codon_start=1
                      /product="hypothetical protein"
                      /protein_id="AAL93571.2"
                      /db_xref="GI:28828267"
                      /translation="MELNLPLLYFLYVPHSNANSNDNWKPSALDCSTLCCPQGYHCEI
                      KNNERSCVIDQSSNYNNYKCKDVQCPIGNECEYDSFKDEICCTKKL"
     CDS              join(15231. .15626,15715. .16065,16162. .16797)
                      /note="GeneID exon scores (in order of location ranges):
                      25.22, 18.35, 37.16 - GSCJ_ID dd_00529"
                      /codon_start=1
                      /product="hypothetical protein"
                      /protein_id="AAO50942.1"
                      /db_xref="GI:28828268"
                      /translation="MQDYNSDLVIRSCFKLDNVKIQEYQNQENLKLNENNSEFEKIQS
                      KDIKDHIINNLDWRFLHGKWRCVDSFSEYHDDNFNTFGGSLGEFIMALHVFTDNTNQI
                      IDSDSKIIYELMFKFLKLNPEKYLIYHSNFNTFEKLKKTIFGIYPDMDEKYFDIYHIK
                      KQHQSQILELVVRDEFIGCEFIKQIILGNKKSLGISPKSISPILVKNIIQSFFKIIWN
                      EKVKISNALFEIDFHNDMVAILNININPPFICGGPEIYPTFTQELFNDSHMNYMVIQG
                      EHQVITGIRKSLSQYFHDEFHCDDHHKVEHRNKIEISFKDYLNNLNLVANLYKEVSLN
                      LYKTLPIFNVTFNILIPDSPPKINKQVIPIDNNNNKNISNNTNSSNNTNSNNNNSTP
                      NNLQVKKSKVQSFFKGIFIFLGIGVTFAVIIFGLLSYKKKKEKYRYHTKLQQEKVFSL
                      EEDIENEFDL"
     CDS              complement(join(17257. .17777,17920. .18049))
                      /note="GeneID exon scores (in order of location ranges):
                      49.00, 9.03 - GSCJ_ID dd_00530"
                      /codon_start=1
                      /product="similar to Arabidopsis thaliana (Mouse-ear
                      cress). Apospory-associated protein C (Hypothetical 35.4
                      kDa protein)"
                      /protein_id="AAO50943.1"
                      /db_xref="GI:28828269"
                      /translation="MYTLEQLQSLWNKGKYIQVLKGNNDLSKLVLKTSTSSCEIYIHG
                      AHITSFKTNSVEHLFMSEKSGFQDMKAIRGGVPLIWPQFGNGKIQTHGFARNCDWEIH
                      KALVDEKTQSVQIEFLLRNNEYSEKVWNCNITSGKETINNFHTIYRVTLHPDFLDLEF
                      DVKNTSTSPSSNDNILEFQLAFHTYYQISNIENVYVLGLDGREYIDKMNGHKITKE"
     CDS              join(18919. .19049,19137. .21735)
                      /note="GeneID exon scores (in order of location ranges):
                      -2.24, 179.96 - GSCJ_ID dd_00531"
                      /codon_start=1
                      /product="hypothetical protein"
                      /protein_id="AAL93574.2"
                      /db_xref="GI:28828270"
                      /translation="MGTFKKLFLIFFLIIINILYCYCGCEFSSIGENKKYNFTPAFPG
                      EYISTSDGTNTVFLNLCTTNQISKCPTTSACIKDSSKNTDIGQQVGPAIFLPKGILLT
                      YKSQAYPECVGLGGVVSTNFTIECNSTDNSFTLDKIQQFSICSYSIKFLTKLACVCPG
                      DCSSHGICNTNTIKCECNSYASGSSCNELKMIISTISSTTIYGGDVMISGDFSNLPLE
                      IKNNLKIKINWDDCKNVININNNTIQCTADKGYNGNGIEVEFSSGPLSTVIPVNFKYI
                      EIPCPFNCSYPNGVCNKILGTCTCNHSWINGTGCELIKIKTHSIEPTTINGGITNIIG
                      DFTMIINNGTTTFLPNNLIIKIGDDDDDDSSLCQDIKILNDTNLQCTVLPGKQGKKNL
                      TISSGLSMDFNYQLFEYFNKTCLFNCSYPHGDCDLFTSYCSCDSQTNGTGCENSKLVI
                      ESIDPTDEPGGLTNVVGFFGIPTSNFSIKIGDLLCNNLLIINQSLIQCVVPPGKGFKN
                      VFISDRDLSFNGELMFRYFQPIVTNAPKQCTNCGALNQGYCASQGCICHPPWMGVDCQ
                      SQYIIIPQPPKNSTSPSVEIPTFSKPNSTNGGGGSTTSTDIDENENIEKNVFKSLIAI
                      IKLSELDFQSNEIKSFTFSEWIYKELTPSKYQYNTTIKVPTTNDGGGGVSPSSSSSSL
                      NESETNISVTLEWFESLTNITFANENLIMNPSSIKYTIEISKYKFSSKLNQLQLVMMA
                      ELSSNKTKDLCSNNEFGETSSGDNSNYIKIQVDNHSLYGRFIKRAIIDSVPRAIGNVQ
                      LDETMRPLKNASSSQSYIGIVVPYFEKSIIIDPDFSVLIDSSKLESNSNSNSICLIQD
```

```
                QLISGLTKSQIAGIIIGSIAFLCSITAMVIYSIHKSNKDKSLFSKITLRNLNK"
     CDS        complement(21934. .22398)
                /note="GeneID exon scores (in order of location ranges):
                21.03 - GSCJ_ID dd_00532"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AAO50944.1"
                /db_xref="GI:28828271"
                /translation="MDNLKEKPLSYNINNNNLNNNNNNNNNNNNNNNINNNINNNNF
                KYVSFSESLEDIGYQNQYIIDKDEDYYIQEQRYIRQNNQSDDEEDFNNNNYEDSIKIS
                LIQKSFIKNKKNKIIITTIVVLLMIAVSLGLILAWQGVFDQIHNNNNNSKDE"
     CDS        complement(22774. .23184)
                /note="GeneID exon scores (in order of location ranges):
                28.48 - GSCJ_ID dd_03186"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AAO50945.1"
                /db_xref="GI:28828272"
                /translation="MVSIDFNSSIVQNNQNNFSYPATGNDTISTTISLEKPSDNINYA
                NEELPFRDPSVNINNNNNSRNNYRNNNNSNVYRNNNTTTTTTTTTSSNSRHSVSNQG
                VNGGLSRRTKFRIISTIISLVICIVVLVIYFGKN"

  Query Match             5.0%;  Score 54.8;  DB 3;  Length 36148;
  Best Local Similarity   50.8%;  Pred. No. 0.87;
  Matches  131;  Conservative    0;  Mismatches  127;  Indels    0;  Gaps    0;

Qy        132 TATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATC 191
              | || |||||||||||||| | |    || | |||| | |  |||    |    || ||
Db       1526 TTTGGGTTTGGTTGAGAAAAAATAGGAATAATAAATAAATATATAAAAATAATAAAAATA 1467

Qy        192 TCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAG 251
                ||  |    ||  | |  | | ||  |     | |  || || || |  | | |
Db       1466 ATTAATTTCATTTTTTCATGTAATTAAATTTTATAGGGTTAATTATATAATATATGATAT 1407

Qy        252 ACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAG 311
              |     | || ||  ||| |||||  ||| ||   || ||     ||  |||     ||
Db       1406 ATTATAATCTTTTTTTTTTTTTTTTTTTTTTTTGAAGAGAGTCTAAATAATATTAACAT 1347

Qy        312 AGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTAC 371
              | ||| |  | | |  |   |    ||||| |||   ||||| | |  | ||   || |
Db       1346 AAAAAAAAAAAAAAAAAAAAAAAACAAAAATAAAAATAAAAATAAATAAAAATATAATAAA 1287

Qy        372 TTAATCTTGCAGTCAAAG 389
                |||  | ||   ||||
Db       1286 AAAATTATCCATCTAAAG 1269


RESULT 48
AF402141
LOCUS       AF402141               49704 bp    DNA     circular PLN 04-MAR-2002
DEFINITION  Schizophyllum commune mitochondrial DNA, complete genome.
ACCESSION   AF402141 AF275272
VERSION     AF402141.1  GI:15077911
KEYWORDS    .
SOURCE      mitochondrion Schizophyllum commune
  ORGANISM  Schizophyllum commune
            Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
            Agaricales; Schizophyllaceae; Schizophyllum.
REFERENCE   1  (bases 11104 to 15465)
  AUTHORS   Bullerwell,C.E., Burger,G. and Lang,B.F.
  TITLE     A novel motif for identifying rps3 homologs in fungal mitochondrial
            genomes
  JOURNAL   Trends Biochem. Sci. 25 (8), 363-365 (2000)
  MEDLINE   20377911
   PUBMED   10916154
REFERENCE   2  (bases 1 to 49704)
  AUTHORS   Forget,L., Ustinova,J., Wang,Z., Huss,V.A. and Franz Lang,B.
  TITLE     Hyaloraphidium curvatum: a linear mitochondrial genome, tRNA
            editing, and an evolutionary link to lower fungi
  JOURNAL   Mol. Biol. Evol. 19 (3), 310-319 (2002)
  MEDLINE   21851207
   PUBMED   11861890
REFERENCE   3  (bases 11104 to 15465)
  AUTHORS   Lang,F.B.F. and Butterwell,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUN-2000) Departement de Biochimie, Universite de
            Montreal, 2900 Edouard-Montpetit, Montreal, Quebec H3T 1J4, Canada
REFERENCE   4  (bases 1 to 49704)
  AUTHORS   Lang,F.B.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JUL-2001) Departement de Biochimie, Universite de
            Montreal, 2900 Edouard-Montpetit, Montreal, Quebec H3T 1J4, Canada
  REMARK    Sequence update by submitter
COMMENT     On Aug 2, 2001 this sequence version replaced gi:9966501.
FEATURES             Location/Qualifiers
     source          1. .49704
                     /organism="Schizophyllum commune"
                     /organelle="mitochondrion"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:5334"
     gene            90. .3665
                     /gene="rnl"
     rRNA            90. .3665
                     /gene="rnl"
                     /product="large subunit ribosomal RNA"
     gene            3981. .4054
                     /gene="trnW(cca)"
     tRNA            3981. .4054
                     /gene="trnW(cca)"
                     /product="tRNA-Trp"
                     /anticodon=(pos:4016. .4018,aa:Trp)
     gene            4203. .4276
                     /gene="trnD(guc)"
     tRNA            4203. .4276
                     /gene="trnD(guc)"
                     /product="tRNA-Asp"
                     /anticodon=(pos:4237. .4239,aa:Asp)
     gene            4688. .4759
                     /gene="trnG(ucc)"
     tRNA            4688. .4759
                     /gene="trnG(ucc)"
                     /product="tRNA-Gly"
                     /note="copy number 1"
                     /anticodon=(pos:4721. .4723,aa:Gly)
     gene            4965. .5035
                     /gene="trnR(ucu)"
     tRNA            4965. .5035
                     /gene="trnR(ucu)"
                     /product="tRNA-Arg"
                     /anticodon=(pos:4997. .4999,aa:Arg)
     gene            5127. .5198
                     /gene="trnG(ucc)"
     tRNA            5127. .5198
                     /gene="trnG(ucc)"
                     /product="tRNA-Gly"
                     /note="copy number 2"
                     /anticodon=(pos:5160. .5162,aa:Gly)
     gene            6090. .7952
                     /gene="rns"
     rRNA            6090. .7952
                     /gene="rns"
                     /product="small subunit ribosomal RNA"
     gene            8226. .8297
                     /gene="trnV(uac)"
     tRNA            8226. .8297
                     /gene="trnV(uac)"
                     /product="tRNA-Val"
                     /anticodon=(pos:8259. .8261,aa:Val)
     gene            8463. .8538
                     /gene="trnL(uag)"
     tRNA            8463. .8538
                     /gene="trnL(uag)"
                     /product="tRNA-Leu"
                     /anticodon=(pos:8497. .8499,aa:Leu)
     gene            8956. .9027
                     /gene="trnR(ucg)"
```

```
tRNA            8956. .9027
                /gene="trnR(ucg)"
                /product="tRNA-Arg"
                /anticodon=(pos:8989. .8991,aa:Arg)
gene            9374. .10162
                /gene="atp6"
CDS             9374. .10162
                /gene="atp6"
                /EC_number="3.6.1.34"
                /codon_start=1
                /transl_table=4
                /product="ATP synthase F0 subunit 6"
                /protein_id="AAK83398.1"
                /db_xref="GI:15077912"
                /translation="MTYNMLFSNDYVILSPLNQFEVTNLIGLNIPFMNIVNFSFTNLS
                LYTCLVLVVIIGLHFYGLNDTFLIPNNWSITLESFYASLSNMVREQISKTSEVYIPFI
                YSLFFFILTANLISNVPYSFAVTASAVVSLGLSMTIFIGVTILALSIHKIKFFSFFVP
                GGTPIALVPLLVLIEFVSYFARSLSLGIRLFANLFAGHVLLAILSSFLYKLFSSGILV
                AILTLIPFALMIGIIGLELAVSFIQAFVFTILTCSYLKDAILLH"
gene            11104. .15465
                /gene="rps3"
CDS             11104. .15465
                /gene="rps3"
                /note="hybrid protein"
                /codon_start=1
                /transl_table=4
                /product="ribosomal protein S3"
                /protein_id="AAG10295.1"
                /db_xref="GI:9966502"
                /translation="MNDIKIQHSLGERQNLYKTHHIDVNYSNWPEINLDKRVLPLFTT
                PLLLKKRLERKVNKYKFEVIDSKRKNNLNYKIVNFSQILNVLVTKVITFGYEIDYNLK
                YNENKVFLVKQIKEMNESANLYSTVNNIFDRSFANNLKGKESNLNLTFDNNSLDNSSH
                YNSVLNTVKNSSELPFINSELLTLLNKNDNNNLDEMNNSLTSFNHRINNLLSHIITKA
                YGIDNKLGIQLNTNKKAPIKNTFNSKDQNKFYYNFDLYINYNNKKSKNNRNKGNTIII
                RNFDEILNNILPYYNNNKVLNDINNFNNMAQESYNFNIKLPKINSGKESNLINKFANL
                QEFNKIFALLNSNLPGIDRNLIFKEVVKVQPNLINLVYYDAISDSIKYLNGQDINFKI
                NNNRTNNSIKTNDKINDKVLNKLLLKILENDKIEIWVKSMIIKSIFNKKGLDLNIFLQ
                NRGKGLIKNYNNFNNIFKKVKSTKAINKSNKIISKIKNKNKINNSNNNLVKINEFVNK
                AQLSSHVSNFFNNVFNDKLINNGELNIISLLKFNPSNSIKSTVNSNYISNVMSIIRGK
                IISSAITLNYLEKNKKREIIFLINMNNWFNKFAMDIKNTNISNIKELYYLAYMSAFNF
                VIIKNFLHFICNWENSSYNYKKNSINFDSNNLNNTLNNLTKSKFDINSQMKEFVNLIL
                YLKADKIFFNKFITKINNRQSSLEKSFNLFFNINNEPMPKFNFSSTNYSFPSMEKSAI
                KELNIIDGVDTIFKLFSRKFDIVAKNNKNNFEEFNNSIKEIVESDNLVIENNEETLDN
                IKINNNSNIDFSLYNPLLFIKPNNYFNFSFLRNMNNFTSKDNLKNLFDLRTNASEYSI
                IDSLNNNKVKEINNKDNNFLIKSLSFKNQFTSLNNRKALVNDNQNKLGQGWWNSWIFD
                KNYYSNLNKDTIISKLSLNGININNFSKTNKLVQSPSNDFYNLFIKKNLNNNNYENLF
                SLIKDKIDSNRFANNYFDFKNFSDLINTNPVNMVYNVFRKGRNIEKKISHIFNLINLL
                QSNIFNFNNMKEALTANNVLSNKEINQLQKLFVVNNNSSLKLIDNGWKYSIDNNYNYI
                RIGFNNQNFDKEFNLLLIKEILALFTNSFFDNQYNSIIENPLFPNTALQINSEDNNYAI
                LYDYLKEMSYYNMNQKGVIFSFNKFISFFFNTRNTNKNNKLIKYLYDLLAQSFRPMFC
                LISKPIFVITPDKVTFKLFYYLFIPYLLKFKKRYNLNLLNNFKNSGLLSNKAVRELKR
                EGQDIIRKRKNNNRKKRFFLNLRKRKREFRKKILLVSRNSLSSLFRKRFQVLCHILSK
                LFKRPVELKLTRLHYHYTDSNILANFLAIMINKRKLPLILIKLFSNTVFKRYNNIHVK
                RSMIPSFISGLNIRVAGRLLTSKISPRKTVKKRRVGISATGKVNYLDFARYTNKNKRG
                AFSITVESGQNLFNFNK"
gene            complement(15991. .16063)
                /gene="trnH(gug)"
tRNA            complement(15991. .16063)
                /gene="trnH(gug)"
                /product="tRNA-His"
                /anticodon=(pos:16028. .16030,aa:His)
gene            complement(16251. .16324)
                /gene="trnM(cau)e"
tRNA            complement(16251. .16324)
                /gene="trnM(cau)e"
                /product="tRNA-Met"
                /anticodon=(pos:16287. .16289,aa:Met)
gene            complement(16569. .18152)
                /gene="cox1"
CDS             complement(16569. .18152)
                /gene="cox1"
                /EC_number="1.9.3.1"
                /codon_start=1
                /transl_table=4
                /product="cytochrome c oxidase subunit 1"
                /protein_id="AAK83399.1"
                /db_xref="GI:15077913"
                /translation="MRWLMSTNAKEIGTLYLIYSLFAGMIGTAFSVLIRLELSAPGVQ
                FLQGDHQLFNVIITAHAFLMIFFMVMPGLIGGFGNYFLPIHCGAIDMAFPRLNNISFW
                LLVPSLLLLLLSALVGNGAGTGWTVYPPLSGIQSHSGPSVDLAIFSLHLSGVSSLLGA
                INFITTALNMRTNGMGLHKLPLFVWAIFVTAILLLLSLPVLAGGITMLLTDRNFNTSF
                FDPAGGGDPILYQHLFWFFGHPEVYIIIIPGFGVISHVVSTFSGKTIFGYLGMVYAMF
                SIGILGFLVWSHHMFAVGMDVDSRAYFTAATMVIAVPTGIKIFSWLATLFGGSLRFNT
                PLVFAMGFLALFTIGGLTGVMLSNASLDVAFHDTYYVVAHFHYVLSMGAVFALFAGFY
                FWMPKIVGKTYSEFLSNVHFWTLFVGVNLTFFPQHFLGMSGMPRRIPDYPDAFTGWNL
                VSSIGSIISVVSVIIFGYIIYDMFAYQGLVSNNPWEVPSYFVDMELFENDEQGAGSLE
                WSFDTPIAEHAYNMQPVQS"
gene            complement(18412. .18493)
                /gene="trnS(gcu)"
tRNA            complement(18412. .18493)
                /gene="trnS(gcu)"
                /product="tRNA-Ser"
                /anticodon=(pos:18459. .18461,aa:Ser)
gene            complement(18609. .18680)
                /gene="trnE(uuc)"
tRNA            complement(18609. .18680)
                /gene="trnE(uuc)"
                /product="tRNA-Glu"
                /anticodon=(pos:18645. .18647,aa:Glu)
gene            complement(18809. .18881)
                /gene="trnM(cau)f"
tRNA            complement(18809. .18881)
                /gene="trnM(cau)f"
```

Query Match 5.0%; Score 54.8; DB 8; Length 49704;
Best Local Similarity 47.8%; Pred. No. 0.83;
Matches 219; Conservative 0; Mismatches 237; Indels 2; Gaps 2;

```
Qy        52 AAATGTTACAATGAGATATGAGTTTTAAAGCCAGATCTTCCTTATCTCTCTGCCCCATCT 111
             |||   || ||| | | |  ||  | ||| | | ||  |  |||  | |||    ||
Db     45999 AAAATCTAAAATTAAAAAACAGGATAAAATCTAAATAATATTTAATTTTCTTTATCACAC 46058

Qy       112 CTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGT 171
              || || || | |    |  ||| | ||    | | | | ||| |  |  | ||| ||
Db     46059 ATATTTTTTTACGCTAATTTTATAAATTCTAATTAAATAAATTTAAAAAAATAAAACAAA 46118

Qy       172 TAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGAT 231
             ||| | |||| ||  | ||   |  |||      |   |||    | | ||||     | ||
Db     46119 TAAAATTTTTTTACCAAATTAAAATTTATTTTTAAAATTTAACTTAAATACATTAATTAT 46178

Qy       232 TTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAG 291
             ||| |  ||| || ||| |   |    |   |||  | | || | |||  | |
Db     46179 TTTGTATAAA-ACATTATAATTAATAAAAAATTTTTAGGGTTATAGTTATTCCCTCTCTT 46237

Qy       292 GTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAA 351
              ||| | |||    ||  | ||  | ||||      |||      |||      |||
Db     46238 TTATCTAAAAAAAATTTATAATAATCTTTATATTAAGACCAAAAATTTATTTTTATTAAA 46297

Qy       352 ATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAA-GGT 410
             |   |  |    | |||| || ||  |      | |  |||    |  |  ||||   |
Db     46298 CTTTTTTAAACCAATATTATTTTATTATTGGTCTACATATAATCTTTATAATTAAACTAT 46357

Qy       411 ATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAA 470
              ||||      |   ||  | |  ||||  ||||||  | |||||| ||||  |
Db     46358 TGTCTACAATAAATTAAATAATATTATAAATTTAAATTATTAAAAATTTTCTTTTTTATT 46417

Qy       471 TAATCTCAGAAAACATAATTTTTTATGTACTATTAAAA 508
             |    || | | | |||  | ||     |  |||  |
Db     46418 TTTAAAAAGGATAGAGAATAATATAATATGTTATAATA 46455
```

RESULT 49
PFMAL7P1_09/c
WPCOMMENT
Sequence split into 14 fragments LOCUS PFMAL7P1 Accession AL844506
Fragment Name Begin End

```
     PFMAL7P1_00            1    110000
     PFMAL7P1_01       100001    210000
     PFMAL7P1_02       200001    310000
     PFMAL7P1_03       300001    410000
     PFMAL7P1_04       400001    510000
     PFMAL7P1_05       500001    610000
     PFMAL7P1_06       600001    710000
     PFMAL7P1_07       700001    810000
     PFMAL7P1_08       800001    910000
     PFMAL7P1_09       900001   1010000
     PFMAL7P1_10      1000001   1110000
     PFMAL7P1_11      1100001   1210000
     PFMAL7P1_12      1200001   1310000
     PFMAL7P1_13      1300001   1351552
Continuation (10 of 14) of PFMAL7P1 from base 900001 (AL844506 Plasmodium falciparum 3D7

  Query Match             5.0%;  Score 54.8;  DB 2;  Length 110000;
  Best Local Similarity   46.4%;  Pred. No. 0.73;
  Matches  216;  Conservative    0;  Mismatches  247;  Indels    3;  Gaps    1;

Qy        177 GTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTT 236
              || ||   ||  | |||| ||||    ' |   ||   | ||| | |     | | | |
Db      99855 GTATTTATAAAAGTTTCATGTTAACTCTGTGAATTACATTATAAAGTATGGTAAATATAA 99796

Qy        237 TAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATA 296
                |   |||         |  |  | |    ||  ||| |||  |  |  | | |  ||||
Db      99795 ATGTACTACTTTTTATATTTAATTTTTAAGTTCTTTTATGGAATAATATATATATATATA 99736

Qy        297 AAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAG 356
              | | ||  ||  | | | ||  || ||| | ||    |  |  |||  ||| | ||||   |
Db      99735 TATATATATATATATATAATAATAATACAATATATATATTTAAAAAAAAAAAAAAAAAGG 99676

Qy        357 TGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGG---TATA 413
                |||  || | | | | | | |    ||  |  ||||   |  |    | | |   ||||
Db      99675 ATTCTATGACGTCATTAATTATATAGAGTTCTTAATTAAACCTTACAAAGATATGAATTATA 99616

Qy        414 CTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAA 473
              ||      |   ||||   |     | ||| | ||||| | |||   ||  |    | | |
Db      99615 TTATAATGTACTTTATAAAATATTACATAAATATTACAAAAAAAAATAAATATATATTTA 99556

Qy        474 TCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGTCA 533
              | |   | |   ||  | ||||  |      ||| | |||||  |   ||||||  | |
Db      99555 TATATAATATTTTATATATTTAATTTTATATAACATATTTATATATAAAATATAATATAT 99496

Qy        534 TTCAGGAGTATGGAAGTATCGATGGCTTCTTTAAAATGAAGCAGGAGGGTCTGGCAGAGA 593
              |       | |  || | |  ||     |  ||| ||  ||  | |      |   | | |
Db      99495 ATAGTTCTTTTATAATTTTATATAATATTATTATAAGTAAAAAATATATATTTTTATATA 99436

Qy        594 GTATCTATGAAATAAGTTCCTCTGACCTTCACGCTTAATTTTCTGA 639
              |   |||  ||||   |  | | ||    |   ||| | || | |
Db      99435 TTTCATATATAATATAATATTTTTACAAATATCATTATTATTATAA 99390
```

```
RESULT 50
CR376844
LOCUS       CR376844              210050 bp    DNA     linear   HTG 12-AUG-2004
DEFINITION  Danio rerio clone DKEYP-51B7, *** SEQUENCING IN PROGRESS ***, 8
            unordered pieces.
ACCESSION   CR376844
VERSION     CR376844.5  GI:51174212
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 210050)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-AUG-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: cloneregquest@sanger.ac.uk
COMMENT     On Aug 12, 2004 this sequence version replaced gi:51100619.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zKp51B7
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 205725 bases at least Q40
            Consensus quality: 206694 bases at least Q30
            Consensus quality: 207460 bases at least Q20
            Insert size: 209350; sum-of-contigs
            Insert size: 206069; 1.8% error; agarose-fp
            Quality coverage: 7.51x in Q20 bases; sum-of-contigs Quality
            coverage: 7.77x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    53231: contig of 53231 bp in length
            *    53232    53331: gap of 100 bp
            *    53332    88409: contig of 35078 bp in length
            *    88410    88509: gap of 100 bp
            *    88510   124254: contig of 35745 bp in length
            *   124255   124354: gap of 100 bp
            *   124355   134156: contig of 9802 bp in length
            *   134157   134256: gap of 100 bp
            *   134257   140548: contig of 6292 bp in length
            *   140549   140648: gap of 100 bp
            *   140649   188003: contig of 47355 bp in length
            *   188004   188103: gap of 100 bp
            *   188104   191278: contig of 3175 bp in length
            *   191279   191378: gap of 100 bp
            *   191379   210050: contig of 18672 bp in length.
FEATURES             Location/Qualifiers
     source          1. .210050
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEYP-51B7"
                     /clone_lib="DanioKeypilot"
     misc_feature    1. .53231
                     /note="assembly_fragment:02162
                     fragment_chain:1"
     misc_feature    53332. .88409
                     /note="assembly_fragment:00501
                     fragment_chain:1"
     misc_feature    88510. .124254
                     /note="assembly_fragment:00994
                     fragment_chain:1"
     misc_feature    124355. .134156
                     /note="assembly_fragment:00153
                     fragment_chain:2"
     misc_feature    134257. .140548
                     /note="assembly_fragment:00063
                     fragment_chain:2"
     misc_feature    140649. .188003
                     /note="assembly_fragment:01545
                     fragment_chain:3"
     misc_feature    188104. .191278
                     /note="assembly_fragment:00035
                     fragment_chain:3"
     misc_feature    191379. .210050
                     /note="assembly_fragment:00255.0"
ORIGIN
```

```
  Query Match             5.0%;  Score 54.8;  DB 2;  Length 210050;
  Best Local Similarity   31.2%;  Pred. No. 0.65;
  Matches  194;  Conservative    0;  Mismatches  427;  Indels    0;  Gaps    0;

Qy        146 AGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACA 205
              | | ||||  ||   || | |  | |  |||  | | || |    |   || |  |
Db     140121 ATATATATATATATATATATAAAAATATATATATATATATATTTTTNTATATATANATAT 140180

Qy        206 TTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTC 265
              | | |  |||       |   ||  | | | |  |        |   | | |     ||
Db     140181 ATATATATTTATATATATATTTTTTTATATATNNANNNNNANATNNANAAANNNAAATAT 140240

Qy        266 ATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATA 325
              | || || |     |      ||||||   | |          |||   | |||    |  |
Db     140241 ANTTTTTATATAATAAAAAAAAAAAAAAAAANANNNNNNNNAATTNNAAAAAAATNNAATTT 140300

Qy        326 TTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTC 385
              ||      |   | | |    |||   |          | ||   ||       |
Db     140301 TTTTTANANTANTANTATATATAAAANATATAATTNNNNNNANTAAAAAAAAAAAAAAAA 140360

Qy        386 AAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAA 445
              |||   |  ||       |||         ||  |  ||| || |    |     ||| | ||
Db     140361 AAAAAAANNAAANNTTTNTAATATATATTTATTTTTTTTTTTTTNNNAAAAAAAAAAAAA 140420

Qy        446 ATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTA 505
              |  |  | || |||| ||||       |  |     || | || |||   | |  | ||
Db     140421 AAAAAAAAAANTTTTTTTTTNATNNNNANNTTNNNNAAAAAAAATTTAATTTTTTTTTTT 140480

Qy        506 AAACATTTACTTTCCAAATATTCTGTCATTCAGGAGTATGGAAGTATCGATGGCTTCTTT 565
                                                  |  |  | |       ||  |
Db     140481 NNNNNNNNNAAANNNNNNNNNNNNNNNNNNNTTTNTTTTATATTTTTTAATNTTTNTA 140540

Qy        566 AAAATGAAGCAGGAGGGTCTGGCAGAGAGTATCTATGAAATAAGTTCCTCTGACCTTCAC 625
              ||||  ||
Db     140541 AAAAAAAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 140600

Qy        626 GCTTAATTTTCTGAATGGAGTGGAGCAAATTACTTCAAGCTTCACTTAACTTGCATATGA 685
                                                        |  |  ||||
Db     140601 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNATATATATATAT 140660

Qy        686 AATGAACCGTACAAAAATACAAGAGTGTCAGGAGAAAGTTATGCTCTGGTAAATATTTTG 745
              |   |    || | | ||| |    | |    | | |  |||       || ||||  |
Db     140661 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 140720

Qy        746 CAAAACAGATAAAAGATAATA 766
              | | | ||| |     |||
Db     140721 ATATATATATATATATATATA 140741


RESULT 51
CR450749/c
LOCUS       CR450749              246589 bp    DNA     linear   HTG 24-MAY-2004
DEFINITION  Danio rerio clone DKEY-2J6, *** SEQUENCING IN PROGRESS ***, 11
            unordered pieces.
ACCESSION   CR450749
VERSION     CR450749.1  GI:47604407
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 246589)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAY-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     ---------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            ---------- Project Information
            Center project name: zK2J6
            ---------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 243112 bases at least Q40
            Consensus quality: 243477 bases at least Q30
            Consensus quality: 243996 bases at least Q20
            Insert size: 245589; sum-of-contigs
            Insert size: 241608; 2.3% error; agarose-fp
            Quality coverage: 7.87x in Q20 bases; sum-of-contigs Quality
            coverage: 8.07x in Q20 bases; agarose-fp
            ----------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    28228: contig of 28228 bp in length
            *    28229    28328: gap of 100 bp
            *    28329    39447: contig of 11119 bp in length
            *    39448    39547: gap of 100 bp
            *    39548    48874: contig of 9327 bp in length
            *    48875    48974: gap of 100 bp
            *    48975    97551: contig of 48577 bp in length
            *    97552    97651: gap of 100 bp
            *    97652   184544: contig of 86893 bp in length
            *   184545   184644: gap of 100 bp
            *   184645   194757: contig of 10113 bp in length
            *   194758   194857: gap of 100 bp
            *   194858   209127: contig of 14270 bp in length
            *   209128   209227: gap of 100 bp
            *   209228   224943: contig of 15716 bp in length
            *   224944   225043: gap of 100 bp
            *   225044   230389: contig of 5346 bp in length
            *   230390   230489: gap of 100 bp
            *   230490   236916: contig of 6427 bp in length
            *   236917   237016: gap of 100 bp
            *   237017   246589: contig of 9573 bp in length.
FEATURES             Location/Qualifiers
     source          1. .246589
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-2J6"
                     /clone_lib="DanioKey"
     misc_feature    1. .28228
                     /note="assembly_fragment:01036
                     fragment_chain:1"
     misc_feature    28329. .39447
                     /note="assembly_fragment:00514
                     fragment_chain:1"
     misc_feature    39548. .48874
                     /note="assembly_fragment:00184
                     fragment_chain:1"
     misc_feature    48975. .97551
                     /note="assembly_fragment:01406
                     fragment_chain:1"
     misc_feature    97652. .184544
                     /note="assembly_fragment:02042
                     fragment_chain:1"
     misc_feature    184645. .194757
                     /note="assembly_fragment:00397
                     fragment_chain:1"
     misc_feature    194858. .209127
                     /note="assembly_fragment:00656
                     fragment_chain:1"
     misc_feature    209228. .224943
                     /note="assembly_fragment:00820
                     fragment_chain:1"
```

```
     misc_feature    225044. .230389
                     /note="assembly_fragment:00033
                     fragment_chain:1"
     misc_feature    230490. .236916
                     /note="assembly_fragment:00101
                     fragment_chain:1"
     misc_feature    237017. .246589
                     /note="assembly_fragment:00288.0"
ORIGIN

  Query Match             5.0%;  Score 54.8;  DB 2;  Length 246589;
  Best Local Similarity   32.4%;  Pred. No. 0.63;
  Matches  182;  Conservative    0;  Mismatches  378;  Indels    2;  Gaps    1;

Qy        146 AGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACA 205
              | | ||||  ||   || | ||   |  |||  | | || |       || |  |
Db     185221 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 185162

Qy        206 TTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTC 265
              || |||| ||  | |  |   ||  |||||||        ||  | |  || |  ||||
Db     185161 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTATTTTATATTTATATATATTT 185102

Qy        266 ATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATA 325
              ||    | |   | | |  ||||||   ||||| ||||      | |  || |  | | |
Db     185101 ATNATNTNTNTNTNTNTAAAAAAAAAAAATAAAAAAATTTTAAAAAAATAAAAAAAAAAA 185042

Qy        326 TTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTC 385
                | |    ||  || || | ||||  |     |||||  ||   || || |  |
Db     185041 AAAAAAAAAAAAATTTTTAAAAAAAAAAAAAAAAATTATTATATATTTTTATATATAAA-- 184984

Qy        386 AAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAA 445
              |||    ||||   |          | |  |   |  ||| ||||   |     | | | ||
Db     184983 AAAATATAAAAANAANNNNNNTNTTTTTTTTTTTTTTTTTTTATNTAAAAAAATAAAAAA 184924

Qy        446 ATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTA 505
              |  |  |  |       | |     | | |          |  ||||  | |      |
Db     184923 AAAAAAATTATNNNNNNTNTNNNTTTTTTTTTTTTTTTTTTTTTTTTNTNTNNNNAAANN 184864

Qy        506 AAACATTTACTTTCCAAATATTCTGTCATTCAGGAGTATGGAAGTATCGATGGCTTCTTT 565

Db     184863 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 184804

Qy        566 AAAATGAAGCAGGAGGGTCTGGCAGAGAGTATCTATGAAATAAGTTCCTCTGACCTTCAC 625
                          |                                  |   |
Db     184803 NNNNNNNNNNNNANNNNNNNNNNNNNNNNNANANNNNANNNANNNNNNNNNNNA 184744

Qy        626 GCTTAATTTTCTGAATGGAGTGGAGCAAATTACTTCAAGCTTCACTTAACTTGCATATGA 685
                  |   | |                |  ||     | ||       | || |
Db     184743 NNNNNNNNANTTNNTTTAAANNNNTNNNNNNTNTNNAANNNNNANTTNNNNANAAAATTA 184684

Qy        686 AATGAACCGTACAAAAATACAA 707
              |||      ||  || |   ||
Db     184683 AATTNTAANTANNAATAANNAA 184662


RESULT 52
AL356287/c
LOCUS       AL356287              182118 bp    DNA     linear   PRI 04-OCT-2001
DEFINITION  Human DNA sequence from clone RP11-57O2 on chromosome 13, complete
            sequence.
ACCESSION   AL356287
VERSION     AL356287.16  GI:15795407
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 182118)
  AUTHORS   Griffiths,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-OCT-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 27, 2001 this sequence version replaced gi:15723066.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 13, constructed by the Sanger Centre Chromosome 13
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr13
            RP11-57O2 is from the library RPCI-11.1 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            This sequence is the entire insert of clone RP11-57O2 The true left
            end of clone RP11-72P19 is at 122681 in this sequence. The true
            right end of clone RP11-124N19 is at 6499 in this sequence.
FEATURES             Location/Qualifiers
     source          1. .182118
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="13"
                     /clone="RP11-57O2"
                     /clone_lib="RPCI-11.1"
     misc_feature    24162. .24876
                     /note="Single clone region. Assembly confirmed by
                     restriction digest."
     misc_feature    24877
                     /note="Tandem repeat. Forced join. Gap size estimated to
                     be approximately 200bp by restriction digest."
     misc_feature    103758. .103772
                     /note="Sequence from uni-directional dGTP reads only."
ORIGIN

  Query Match             5.0%;  Score 54.6;  DB 9;  Length 182118;
  Best Local Similarity   45.9%;  Pred. No. 0.73;
  Matches  186;  Conservative    0;  Mismatches  219;  Indels    0;  Gaps    0;

Qy        148 AAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATT 207
              | ||||        ||| |      |  ||| || | || || |  | | | ||| | |  |
Db      25013 ATATATAATAATATACTATATTATATATATTTATATATAATACATTGTATTATATATTTA 24954

Qy        208 GTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCAT 267
                | |  ||||| | |    ||   | ||| | | |   ||   | ||| |   ||  ||
Db      24953 TATATAATAATACATTATATTATATATATTTATATATAATATATTTATATATATTAATATAT 24894

Qy        268 TTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATT 327
              |    | |   | | |  ||||  |  ||  |  | ||  | |       | || |||||||
Db      24893 TATATATATTTATATATAAAATATATAATATAATATATAATATATTTTATATAATATATT 24834

Qy        328 AGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAA 387
              |   | |  || ||   ||| | |||    |       |       |    | |  | |  |
Db      24833 ATATATTATAATATATTACATATAATATATTATATATTATAATATATTACATATAATATA 24774

Qy        388 AGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAAT 447
                || |   | |       |  |||| ||  |  | | |||| | |    | ||| ||
Db      24773 TTATATATTATAATATATTACATATAATATATTATATATTATAGTATATTACATATAATA 24714

Qy        448 TACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAA 507
```

```
Db      24713 TATTATATATTATAGTATATTACATATAATATATTATATATTATAGTATATTATACATTA 24654

Qy        508 ACATTTACTTTCCAAATATTCTGTCATTCAGGAGTATGGAAGTAT 552
Db      24653 GAATATATTATATATATAATATATATTAGAACATATTATATATAT 24609


RESULT 53
CQ410653
LOCUS       CQ410653                461 bp    DNA     linear   PAT 23-JAN-2004
DEFINITION  Sequence 17724 from Patent WO0170979.
ACCESSION   CQ410653
VERSION     CQ410653.1  GI:41318434
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Lee,J. and Lillie,J.
  TITLE     Genes, compositions, kits, and method for identification,
            assessment, prevention, and therapy of ovarian cancer
  JOURNAL   Patent: WO 0170979-A 17724 27-SEP-2001;
            Millennium Pharmaceuticals, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .461
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

  Query Match             5.0%;  Score 54.4;  DB 6;  Length 461;
  Best Local Similarity   47.1%;  Pred. No. 2.2;
  Matches  154;  Conservative    0;  Mismatches  173;  Indels    0;  Gaps    0;

Qy        199 AAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACA 258
Db        110 AAAATAAAAATTTTTTTTAAAATAATAAAAAATTTTAAAAAAATTAAAAAAAAAAAAAA 169

Qy        259 GCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGA 318
Db        170 AATATAAAAATAATTTATTTTTTTTNAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 229

Qy        319 TCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCT 378
Db        230 AATTATTTTTTTTNNNAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATTTAAAAA 289

Qy        379 TGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCA 438
Db        290 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATTTTTTAAAAAAAAA 349

Qy        439 AATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGT 498
Db        350 AAAAAAAATTTTTTTTTTTTATTTTTTAAATAAAAAAAAAAAAATTNNNTTTTTTTTT 409

Qy        499 ACTATTAAAACATTTACTTTCCAAATA 525
Db        410 TTTTTTTTTTTTTTTTTTTAAAAAAA 436


RESULT 54
CEY53C12D
LOCUS       CEY53C12D               1496 bp    DNA     linear   INV 12-OCT-2004
DEFINITION  Caenorhabditis elegans YAC Y53C12D, complete sequence.
ACCESSION   AL033123
VERSION     AL033123.1  GI:3820757
KEYWORDS    HTG.
SOURCE      Caenorhabditis elegans
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1
  AUTHORS   none.
  TITLE     Genome sequence of the nematode C. elegans: a platform for
            investigating biology. The C. elegans Sequencing Consortium
  JOURNAL   Science 282 (5396), 2012-2018 (1998)
  MEDLINE   99069613
   PUBMED   9851916
  REMARK    The C.elegans Sequencing Consortium.
REFERENCE   2  (bases 1 to 1496)
  AUTHORS   Lennard,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-NOV-1998) Nematode Sequencing Project, Sanger
            Institute, Hinxton, Cambridge CB10 1SA, England and Department of
            Genetics, Washington University, St. Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rw@nematode.wustl.edu E-mail: worm@sanger.ac.uk
COMMENT     Coding sequences below are predicted from computer analysis, using
            predictions from Genefinder (P. Green, U. Washington), and other
            available information.
            Current sequence finishing criteria for the C. elegans genome
            sequencing consortium are that all bases are either sequenced
            unambiguously on both strands, or on a single strand with both a
            dye primer and dye terminator reaction, from distinct subclones.
            Exceptions are indicated by an explicit note.
            IMPORTANT:  This sequence is NOT necessarily the entire insert of
            the specified clone.  It may be shorter because we only sequence
            overlapping sections once, or longer because we arrange for a small
            overlap between neighbouring submissions.
            For a graphical representation of this sequence and its analysis
            see:- http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?
            name=Y53C12D
            For a graphical representation of this sequence and its analysis
            see:- http:
            IMPORTANT: This sequence is not the entire insert of clone Y53C12D.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we arrange for a small overlap between
            neighbouring submissions.
            The start of this sequence (1. .117) overlaps with the end of
            sequence Z99277.
            The end of this sequence (1390. .1496) overlaps with the start of
            sequence Z99278.
FEATURES             Location/Qualifiers
     source          1. .1496
                     /organism="Caenorhabditis elegans"
                     /mol_type="genomic DNA"
                     /strain="Bristol N2"
                     /db_xref="taxon:6239"
                     /chromosome="II"
                     /clone="Y53C12D"
ORIGIN

  Query Match             5.0%;  Score 54.4;  DB 3;  Length 1496;
  Best Local Similarity   45.4%;  Pred. No. 1.8;
  Matches  238;  Conservative    0;  Mismatches  281;  Indels    5;  Gaps    1;

Qy         33 TCTCTATGTTCTTATAAGCAAATGTTACAATGAGATATGAGTTTTAAAGCCAGATCTTCC 92
Db        917 TATTTATTTAATTAATATTTATTTATTTATTTAATTAATAATTATTTATGTAAATAATAT 976

Qy         93 TTATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATA 152
Db        977 TTATTTATTTAATTAATATTTATTTAATTATTTAATTAATATTTATTTATTTAATTAATA 1036

Qy        153 TTGA-----TCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATT 207
Db       1037 TTTATTTATTTATTTAATTAATATTTATTTATTTAATTAATATTTATTTATGTAAATAAT 1096

Qy        208 GTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCAT 267
Db       1097 ATTTATTTATTTAATTAATATTTATTTATTTAATTAATATTTATTTATTTAATTAA 1156

Qy        268 TTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATT 327
Db       1157 TATTTATTTTTCTAATTAATATTTATTTATTTAATTAATAATATATAATAATAAATATTA 1216
```

```
Qy        328 AGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAA 387
              |   | |  | || ||       |||| | | .|||||  | | | |||      | |
Db       1217 ATTAAATAAATAAATATTATTTAAATAAATAAATATTAATTAAATAAATAAATAAATATT 1276

Qy        388 AGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAAT 447
              |  |||| |   ||      |  || |       | |     |          ||||| | ||
Db       1277 AATTAAATAAATAAATATTAATTAAATATATAAATATTAATTGAATAAATAAATAAATAT 1336

Qy        448 TACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAA 507
              ||    ||    ||  |   || |||    | |||  |  |||  |||  ||  |  ||
Db       1337 TAATTAAATAAATAAATATTAATTAAATAAATAAATAAATATTAATTAAATAAAAACAAT 1396

Qy        508 ACATTTACTTTCCAAATATTCTGTCATTCAGGAGTATGGAAGTA 551
              |||  | |    |||||||   | | | |   |  || ||
Db       1397 TAATTAAATAAATAAATATTAATTAAATAAATAAATTTTAATTA 1440


RESULT 55
AY047363/c
LOCUS       AY047363               5701 bp    DNA     linear   INV 20-NOV-2002
DEFINITION  Dictyostelium discoideum cGMP-stimulated cGMP phosphodiesterase
            (pdeD) gene, complete cds.
ACCESSION   AY047363
VERSION     AY047363.1  GI:18031838
KEYWORDS    .
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 5701)
  AUTHORS   Meima,M.E., Biondi,R.M. and Schaap,P.
  TITLE     Identification of a Novel Type of cGMP Phosphodiesterase That Is
            Defective in the Chemotactic stmF Mutants
  JOURNAL   Mol. Biol. Cell 13 (11), 3870-3877 (2002)
  MEDLINE   22317259
   PUBMED   12429831
REFERENCE   2  (bases 1 to 5701)
  AUTHORS   Meima,M.E., Biondi,R.M. and Schaap,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JUL-2001) School of Life Sciences, University of
            Dundee, Dow Street, Dundee, Angus DD1 5EH, UK
FEATURES             Location/Qualifiers
     source          1. .5701
                     /organism="Dictyostelium discoideum"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:44689"
     gene            <1565. .>4387
                     /gene="pdeD"
     mRNA            join(<1565. .2672,2769. .3129,3253. .>4387)
                     /gene="pdeD"
                     /product="cGMP-stimulated cGMP phosphodiesterase"
     CDS             join(1565. .2672,2769. .3129,3253. .4387)
                     /gene="pdeD"
                     /codon_start=1
                     /product="cGMP-stimulated cGMP phosphodiesterase"
                     /protein_id="AAL06059.1"
                     /db_xref="GI:18031839"
                     /translation="MMFSKRKSIRFNPEGDYTELPRGGYVVPSKLGGIQFGVPPETIK
                     DSMALKIDVPTIYVFPEELWDRKTGINAAEAEFPAYFNYFILKRKVSFVCTKEQERI
                     RIVFQETLLGPPEFNTGIIINSSSSTTDTSKTSPIKKQTSSSSPPLSPQQQQPPPPLV
                     KQPSQQQLEELATMDTCHYHHHHHHQEVNDNDNNNNTTTNNNNIEILEQQQQQQQQQQ
                     QQQDEDSTDVDEEFQKEFSSTFPRSEIPNLEKECKYLRTFNSVDELFDFILFDDNGIA
                     KLSDDVEIHFQEDQSLFKVLQFEETGKGNKVQTLVATIPSKILFPDMINSISTINDNK
                     IFDPPTFGITIIGSSHGFDPKKSTTGFVLWINKRGIMVDPPLNSSSFLQSQGVPTRMV
                     DHIILTHCHADHDSGTFQKLLEEYQITVVTTPTILGSFLRKYGALSNLSTDLLRRLFI
                     FRPVMIGEPMIISGAEFRFFYTIHTIPTISFEVFYGGKSIFYSGDTCYDPNRIKDMNK
                     RGIMKTSRMKFFLRPWNHTVVLHEAGVPPIHTPVSVLRALPDEVKNRLYLVHISEHTL
                     PAGSGLKIAKEGVAHTLSLDVMKSSHSEAVDILKLVESVDIFRSIPLTQACEILQTAT
                     KRKYSQGSVIIARDTEPDAFYVVASGVVCVNIGELKKNLIVGDYFGEMSLVMGGLRSA
                     NVQAVTDVEVLSFNKEDFLSITRNSTESIQFITRLWEMRNEKSWETMSLNSVFSRCTN
                     SQKTAIQSILVRELIKKDETLWCKGEEALFGCLVAEGSFVFKEDDSLESFSQGSFLGD
                     INAMTTQPPSIHKTTVVAKEESVIYKVLSQDLIKFFSNNPGIQLAFLDTIFVDALRDQ
                     VQQLVNSKLTY"
ORIGIN
```

```
  Query Match             5.0%;  Score 54.4;  DB 3;  Length 5701;
  Best Local Similarity   47.2%;  Pred. No. 1.4;
  Matches  166;  Conservative    0;  Mismatches  186;  Indels    0;  Gaps    0;

Qy        172 TAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGAT 231
              | ||  |||| ||  |     |  | |||      ||| || |  | || |  |   ||| |
Db       5339 TTATTATTTTTTATTATTATTATTATTATTATTATTATTATTATTTTTTAATTTGTT 5280

Qy        232 TTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAG 291
              | ||||  |       ||  | | | | |  |||| ||   ||| |  ||   |   || |
Db       5279 TATTTTTTATCTTTTTTTTATATAAATATGTATTTATAGATTTATTTTTATATGGTAATAA 5220

Qy        292 GTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAA 351
               |||||          ||| |  || || ||  | ||    |     | ||  ||| | |||
Db       5219 ATATAATCCTTCCTCATCATACGAAAATAATTAAATAAAAAAGAAATAAAAAAAAAAAAA 5160

Qy        352 ATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTA 411
              |  | | || |  |  |||      | |  || |||||  ||||  |       | |   ||
Db       5159 AAAAATACAAAAAAGTTTAAACTTTTTCCCAATCAAATATAAATAAAATCTATCATTATA 5100

Qy        412 TACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAAT 471
              |  |   | ||      |     |      | | | || |  | | | | | | |    |
Db       5099 AAAAAAAGGTTTAAAACTATTCCATACCAAATACATAAATAGATAATATGTATTTATATA 5040

Qy        472 AATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAA 523
              || |    |   |  |||||||| |  | ||   |  |||| |||  |||
Db       5039 TATATTTTTATTTTTTTTTTTTATTTTTTTTTTTTATTTTTATTTTAAAAA 4988


RESULT 56
HS127L4/c
LOCUS       HS127L4                47812 bp    DNA     linear   PRI 04-MAR-2003
DEFINITION  Human DNA sequence from clone RP1-127L4 on chromosome 22, complete
            sequence.
ACCESSION   Z83839
VERSION     Z83839.1  GI:2769655
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 47812)
  AUTHORS   Burton,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jan 13, 1998 this sequence version replaced gi:2511647.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
```

```
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 22, constructed by the Sanger Centre Chromosome 22
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr22
            RP1-127L4 is from the library RPCI-1 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pCYPAC2.
FEATURES             Location/Qualifiers
     source          1. .47812
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="RZPD:RPCIP704L04127"
                     /db_xref="taxon:9606"
                     /chromosome="22"
                     /clone="RP1-127L4"
                     /clone_lib="RPCI-1"
ORIGIN

  Query Match             5.0%;  Score 54.4;  DB 9;  Length 47812;
  Best Local Similarity   46.5%;  Pred. No. 0.99;
  Matches  213;  Conservative    0;  Mismatches  241;  Indels    4;  Gaps    1;

Qy         33 TCTCTATGTTCTTATAAGCAAATGTTACAATGAGATATGAGTTTTAAAGCCAGATCTTCC 92
              | | |||  |   ||  |   ||   || | | | |||| |  | || |    | || ||
Db      37044 TTTATATATAAATAATAAATAAATATATATTTATATATCAAATATATAAAAATATATTTG 36985

Qy         93 TTATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATA 152
               ||| |   |     || | || |   | |  | |   |||| |  ||   | | | ||
Db      36984 ATATATAAATAATAAATATATAATATATAATATATTATATATTATATTATATTATATATT 36925

Qy        153 TTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTT 212
               | ||   |  |    |  |||||  | |      | || | || ||| | | ||| | ||
Db      36924 ATTATATATTAATTATATATAATATATATAACATATATATTATATTATATATATTATATT 36865

Qy        213 GTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTT 272
              |  |    |   | | |  | ||| | |  ||| | |    |  ||||   |  |
Db      36864 ATATAATATATAATATTAATAATATAATATATATTATATATTATTATATATTATATAATA 36805

Qy        273 TTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCA 332
              | |  | ||||  | ||   |||| |  |    | |    | | || ||||  ||
Db      36804 TATTATATGTTATATAATAATATATAATATAATAATATATAATATATTATATTATATAAT 36745

Qy        333 GTTGAAAATTAAACACAAAATGAG----TGCATATTACATTACTTAATCTTGCAGTCAAA 388
               |  | ||   | | || || |      |  |||||| ||||  | || ||  | |  |
Db      36744 AATATATATGTTATATAATATAATATAATATATATTATATTATATTATATTATATTATAT 36685

Qy        389 GGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATT 448
              || | |  |    |   |||| |   |  | |  |||       | |||  |  |
Db      36684 ATTATATAATATATAATATATTATATTGTATTATATTATATATATGATAATATATTATAT 36625

Qy        449 ACCACAAATTTTATTTTGGAAATAATCTCAGAAAACAT 486
                 ||  | ||| ||  | || |||| | | |  ||
Db      36624 TATCTAATATATATCTTATATATTATCTAATATATTAT 36587


RESULT 57
AC023279/c
LOCUS       AC023279               117585 bp    DNA     linear   PLN 28-JUN-2000
DEFINITION  Genomic sequence for Arabidopsis thaliana BAC F12K21 from
            chromosome I, complete sequence.
ACCESSION   AC023279
VERSION     AC023279.2  GI:7159339
KEYWORDS    HTG.
SOURCE      Arabidopsis thaliana (thale cress)
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 117585)
  AUTHORS   Shinn,P., Brooks,S., Buehler,E., Chao,Q., Johnson-Hopson,C.,
            Khan,S., Kim,C., Altafi,H., Bei,Q., Chin,C., Chiou,J., Choi,E.,
            Conn,L., Conway,A., Gonzales,A., Hansen,N., Howing,B., Koo,T.,
            Lam,B., Lee,J., Lenz,C., Li,J., Liu,A., Liu,K., Liu,S.,
            Mukharsky,N., Nguyen,M., Palm,C., Pham,P., Sakano,H., Schwartz,J.,
            Southwick,A., Thaveri,A., Toriumi,M., Vaysberg,M., Yu,G.,
            Federspiel,N.A., Theologis,A. and Ecker,J.R.
  TITLE     Genomic sequence for Arabidopsis thaliana BAC F12K21 from
            chromosome I
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 117585)
  AUTHORS   Ecker,J.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-FEB-2000) Arabidopsis thaliana Genome Center,
            Department of Biology, University of Pennsylvania, 38th Street and
            Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
REFERENCE   3  (bases 1 to 117585)
  AUTHORS   Ecker,J.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAR-2000) Arabidopsis thaliana Genome Center,
            Department of Biology, University of Pennsylvania, 38th Street and
            Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
REFERENCE   4  (bases 1 to 117585)
  AUTHORS   Cheuk,R., Shinn,P., Brooks,S., Buehler,E., Chao,Q.,
            Johnson-Hopson,C., Khan,S., Kim,C., Altafi,H., Bei,B., Chin,C.,
            Chiou,J., Choi,E., Conn,L., Conway,A., Gonzalez,A., Hansen,N.,
            Howing,B., Koo,T., Lam,B., Lee,J., Lenz,C., Li,J., Liu,A., Liu,J.,
            Liu,S., Mukharsky,N., Nguyen,M., Palm,C., Pham,P., Sakano,H.,
            Schwartz,J., Southwick,A., Thaveri,A., Toriumi,M., Vaysberg,M.,
            Yu,G., Davis,R., Federspiel,N., Theologis,A. and Ecker,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUN-2000) Arabidopsis thaliana Genome Center,
            Department of Biology, University of Pennsylvania, 38th and
            Hamilton Walk, Philadelphia, PA 19104-6018, USA
COMMENT     On Mar 6, 2000 this sequence version replaced gi:6960378.
FEATURES             Location/Qualifiers
     source          1. .117585
                     /organism="Arabidopsis thaliana"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:3702"
                     /chromosome="1"
                     /clone="F12K21"
     CDS             complement(join(84. .362,587. .704,804. .901))
                     /note="hypothetical protein"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="F12K21.1"
                     /protein_id="AAF79277.1"
                     /db_xref="GI:8778268"
                     /translation="MDSSHNDSSSDEEGIDSNNRRHHSNHQVQRLEAFFHECPHPDDS
                     QRRQLGNELNLKHKQIKFWFQNRRTQARIHNEKADNIALRVENMKIRCVNEAMEKALE
                     TVLCPPCGGPHGKEEQLCNLQKLRTKNVILKTEVCVYANTYPKNYIFFLLININFMSS
                     YVTK"
     CDS             join(5485. .5726,6116. .6206)
                     /note="hypothetical protein"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="F12K21.2"
                     /protein_id="AAF79252.1"
                     /db_xref="GI:8778243"
                     /translation="MGNEAALRSSMVGLAVVMVVVWLWTQSLKKTVVTYAVGISLIGG
                     ILLPDWDFFDRSFSRWGYPVTAEERSAALARKSHPSRFRVYPMRMVLYGTVYGYAVYR
                     WWMFVSNY"
     CDS             complement(join(6477. .6773,6844. .6910,7004. .7084,
                     7169. .7305,7431. .7638,7848. .7981,8092. .8272,8309. .8595))
                     /note="hypothetical protein"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="F12K21.3"
                     /protein_id="AAF79276.1"
                     /db_xref="GI:8778267"
                     /translation="MSPSDESTNGSRCSSCISSDNDGNDFGKSDEFTCKQCKNSDSKS
                     KSKSVNDSDKLRRIIVASVKGFTIGTGLKGGLAIFSIVARFARRRRSSPQSSSSDLKS
                     NRSRFRKTGEFSNSEAIAMGIKETLRYGLFLGTFAGTFVSVDEAIAALAGDKRHNCKM
                     EGPVCWIGGWTIYASHRAKYATYEFGCLYIDACCGSRIALRSAYILKQESLPSSYKSF
```

LNKQGGKDLSILQGVKDIATAQPFTNLRAIEKYYKSVGVDIKLDPTMKVPCTIIHGNE
SCVKHGVTFFLQAYKRALPVYVPVYLIPALIVHRQDLLKKQYSILGKGLLGTARSSLF
LATYCSSAWAWTCLLFRTFETCNIPLVAIATFPTGLALAIEKKSRRIEISLYCLARAI
ESFPTCMTEAGYIRPPKSLRRADVVVFSVSTAIIMHCYAQERDVFRSKYLNVLDWVFG
VPPPPPAPSCETT"

CDS join(9844. .10717,10900. .11771,11895. .12087,12178. .12812)
/note="similar to mutator-like transposase
gb|AAD23710.1|AC006918_9"
/codon_start=1
/evidence=not_experimental
/product="F12K21.4"
/protein_id="AAF79253.1"
/db_xref="GI:8778244"
/translation="MTYEELSQKVKDKLGLRGRGITVKMAYQFPEWMAIDEGNGSPPQ
YISEDSDVTMFIQMRRHIEEVNLCVTVVQHTIPSSPVAVTDKNFNSEDSDDSDYDSEE
ELHQFALDSAVLGPTLEPTAATTETEQHRRTRRELFPTGGISIREPEECIRLKSPDCL
PTDKGKGIMDPIGKKGKAGPSRATVTDTSTDSDDDLLIVPFTLPPNPPAITIPENADI
EGADSSRAAVGTPHGNHVSPIGNAVAETDLQRQNILFWGRAQEALNTILSDFSDDPIL
FGRDAPPVFNDGKGEGVDSAFFDVKYEGDKLFVGRVFKSKSDCKIKIAIHAINRKFHF
RTARSTPKFMVLKCISKTCPWRVYASKVDTSDSFQVRQANQRHTCTIDQRRRYHRLAT
TQVIGELMQSRFLGIKRGPNAAVIRKFLLDDYHVSISYWKAWRAREVAMEKSLGSMAG
SYALIPAYAGLLQQANPGSLCFTEYDDDPTGPRRFKYQFIAFAASIKGYAFMRKVIVV
DGTSMKGRYGGCLISACCQDGNFQIFPLAFGIVNSENDSAYEWFFQRLSIIAPDNPDL
MFISDRHESIYTGLSKVYTQANHAACTVHLWRNIRHLYKPKSLCRLMSEAAQAFHVTD
FNRIFLKIQKLNPGCAAYLVDLGFSEWTRVHSKGRRFNIMDSNICESWNNVIREAREY
PLICMLEYIRTTLMDWFATRRAQAEDCPTTLAPRVQERVEENYQSAMSMSVKPICNFE
FQVQERTGECFIVKLDESTCSCLEFQGLGIPCAHAIAAAARIGVPTDSLAANGYFNEL
VKLSYEEKIYPIHSVGGEVAPEIAAGTTGELQPPFVRRPPGRPRKIRILSRGEFKAST
Y"

CDS complement(join(14609. .14766,14852. .15317,15431. .15515,
15599. .15993,16011. .16109,16199. .16318,16394. .16555,
16621. .17352,17444. .18328))
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/product="F12K21.5"
/protein_id="AAF79275.1"
/db_xref="GI:8778266"
/translation="MGDVELPKRLFRLGSEPEGRKRINNYFTLRWIEIIKAALDQNHL
KQLEGSQFGFLLQMGTHQFSVMFLHYILSRQLVTDKKYEMWWVFAGKPIRFAIDDFAL
VTGLNCSKSGFNKKDARTRGKMKAVQNPTRFYKALFGKERKPTSQWIVGKLSMGKKYK
DPLTRFRLSLLLLVDGILCPTCSRTIIDPDHVEMVKDVESFLKYPWGRTSFLATIRSA
KQRTAKELAQDTTAIQGFPHALVLVTVACCPAIIYTADSVFNILDPTIPVEKLVQYVL
GLSLNINVHDARTIDQSGMDRVTSLLNEELLAKEIDLSFSDEEDDEEVSHLLSLIMEE
HPFEHNSWSGGVDANDVEHPSEDEFMADDCVNQAKGDGDDEMDVDGSGVAAEEREECA
AGEGVDGLDGEANKDGPVTDKDGVSFADGEGLQQEVPRAGEMENNDSGSAGHGLREAV
IKEVADEFEKRVLKIQQSNNDELKSYFEAELVKQKKELVAILTESVREAVSGLNRQAG
IVGECSKCVADAHSKQVKHLNLLVETSSIALEQIRADQPGNKHGVSQQVEPNIGLSEG
SDNLPAACGINVGVSVPERSNPNTRMEVREEDSIQKGGKKKSCSQMDAQSNRPLKRAK
DMEHVASTKLENKEAVHEPKTVRVPIYVNVTLYVVAPLNYVHYVSPTEKLDDEVDIVE
PNFSLGLTQEINDNRSRLPPRGKKKTFPSKVSVLRSSNRIRDRPIQSNSIENMTSIPV
QPHQTASRGAGKSHVSSSAKVEKQTPNQSPAFIGPFKPFSVPSPSSVKRFLKQMERGV
SYVLADGVVLENTEFLNIFEGQTLLDESGPDLLVRFIQLRLDKNKPLRFDFLPSSFLS
ELRRQYNKFVGAADKSAFYFSTLARVPFMVRPKWFAEIDVIYCPMQVERQHWVGLVID
LKKWGVTVLDCNRDVLKDEQVEKYVEHITIMLPYLIRKHGVNGYMLTQRLEPMTVSRP
KIQCSCSSLGLVGIASVVLLELHSVNAMEYCGKLDDGKMRDAAKQYAIETFNALSPTE
RGA"

CDS join(21389. .21877,22131. .22883,23181. .23486,24079. .24250,
24525. .24826,25245. .25490)
/note="unknown protein; similar to EST gb|AI996433.1"
/codon_start=1
/evidence=not_experimental
/product="F12K21.6"
/protein_id="AAF79254.1"
/db_xref="GI:8778245"
/translation="MKLVRSQMHSAVSWAPNIDRILEESHNTPFTHRISNAIISDPEK
LRIEYFNGSSDPKGHLKSFIISVARAKFRPEERDAGLCHLFVEHLKGPALDWFSRLEG
NSVDSFHELSTLFLKQYSVLIDPDTSDADLWSLSQQPNEPLRDFLANQRRGGSLCSEE
STVADKPQGEAARGRGRGHGRESYTWTKDQPAGNEQEYCELHKSYGHHTSRCRSLGAR
LAAKFLAGEISGGLTIEDLEAKKGKTEQVNTVANPEQAAPAANPEGPKRGRGNREADD
DEPEAARGRIFTILGDSAFCQDTAASIKAYQRKADANRNWARPFNGPNDEVIFHESDT
NGLDRPHNDPLVITLTIGDFNVEQVLVDTGSTLDIVFLTTLREMKIDMTQIMPTPRPV
LGFSGETTMTLGTIKLPVRAKGLKIEEARESTTLTPDPVILICLDDAKPERCVEIGGD
LGEELTAELTAFLKENVNAFAWSPEDLPGVSVDVVSHELIIDSTFKPIKQKRRKLGRE
RAEAVKAELVANQFNGEYTARDENMEAYLIHVQKLAKNFNEFELTRIPRGENTSADAL
VEFIEKPNCKHFSKRCEKCQRHAPTIHQPAELLSSIASPYPFIRWSMDIIGPMHLSKQ
KKLVLVLTDYFSKWIEAESYASIKDAQVENFVWKHILCRHGIPYEIVTDNNEMARHYN
SNVRQRRFHEGDRVLQKVFQNTAEPNTGKLGTNWEGPYLISKIVRPEVYELADLSGKA
VPRSWNAMHLRKYYY"

CDS complement(join(26094. .26531,26640. .26807,27008. .27705,
27782. .28745))
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/product="F12K21.7"
/protein_id="AAF79274.1"
/db_xref="GI:8778265"
/translation="MSSTQSPSIPSASFVDSSDSNHLDDLPPIYKRRSVWTSSEEDTV
SSSDAPEQATPFTAREETNADITRELDLPDDPELPLVRRSFASTADEAGTSNWQDVPE
PFMPTVKIEDFLCFGPNETEDILRLNEQKAFEKAEKKKRKRNKKVIMPDSPGSTLCTE
RSLSDLRARFSLGAVTLRVSSPDERADNPPANFYTLYEGFFYGCFLWLPIPRLVLECV
TSYQIALSQITMRSLRHLLGILIRGYESEKEITLAHLRNFLEIRRVPKSKVDRYYISS
AKGNKIIDGFSSKDEPYTDHFFFVAIEDAVHEDLLGTVLTRWGILERTLKFIEPIPDD
FLSAFHALSARKCDWLKHFSRERVERALRLLHGVSCPTSLESFDHRTQFFVDMQSTKL
TLWEVYAKKKEDKERRLAEEKRLVDAGFISPRAAPETTRDGNVIPDAAAPVDAGLISP
RAAPEASRDGNVIPDAAAPVDAGLISLRDAPEASRDGNVIPDAAAPVDAAPSEAQEAE
PSVAASEAVVALPVSDTAAGKRVRVDDESSKKKKTKKKTSGSEAEKAVASMNRMVHSY
DSAMRSNMEVAGKLAEAESRIQAAEREKNEALSEAAAAKLEREERGLRDSEVARAIQT
TRREVSEAFTAKIKTAKHKMSLLDEVNDRFMYLSQARANAQLIEAFEGGGVLEREKEQ
VDEWLKDFADAEANLNRFIAELKDDLKAPALESAPLSPGGHRSVESVADKAGVTDQAG
SLLPAKDNRQSEDLD"

CDS complement(join(30359. .30784,31644. .32202,32697. .33063,
33146. .33281))
/note="similar to monosaccharide transport protein
pir|S25015"
/codon_start=1
/evidence=not_experimental
/product="F12K21.8"
/protein_id="AAF79273.1"
/db_xref="GI:8778264"
/translation="MAGGGLALDVSSAGNIDAKITAAVVMSCIVAASCGLIFGYDIGI
SGGVTTMKPFLEKFFPSVLKKASEAKTNVYCVYDSQLLTAFTSSLYVAGLVASLVASR
LTAAYGRRTTMILGGFTFLFGALINGLAANIAMLISGRILLGFGVGFTNQVAIYSSNF

Query Match 5.0%; Score 54.4; DB 8; Length 117585;
Best Local Similarity 45.8%; Pred. No. 0.86;
Matches 187; Conservative 0; Mismatches 221; Indels 0; Gaps 0;

Qy 173 AATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATT 232
Db 58666 ATTAATTATAAATAAATGATTTTTTTTTTCACTTTTATTTGTTCTTATATCTAATTTGTT 58607

Qy 233 TTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGG 292
Db 58606 TTTCTTTTATATATAAAATACATATGGTTAAAAAGTATTATATATTTTCTGCATATGATG 58547

Qy 293 TATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAA 352
Db 58546 TAATTCAAAAAATTTTAAAATAAAAATATATTATTCAATCGATAAAGAAAACAAATATGT 58487

Qy 353 TGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTAT 412
Db 58486 ATAAACGGTCCCACAATGTTTAAAAAACTAGCCAATGGTTGAGAGTCATACTATAAAAAA 58427

Qy 413 ACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATA 472
Db 58426 AGACCAAAATGATTTGAATATGAACAAGCAAAAAGCTTGATTTATCAGGCATCCAAAATT 58367

Qy 473 ATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGTC 532
Db 58366 ATAAGTGTTATCAAGGTTTACTATGGTTCTTTATTTTAAATAAATAGAAGTTTTAAATAT 58307

Qy 533 ATTCAGGAGTATGGAAGTATCGATGGCTTCTTTAAAATGAAGCAGGAG 580
Db 58306 TTTCAGAATATTATAAATATTTAAAATTTATTAAAAGTTAAATATTAG 58259

```
RESULT 58
AC117176
LOCUS       AC117176               182871 bp    DNA     linear   INV 12-MAR-2003
DEFINITION  Dictyostelium discoideum chromosome 2 map 5018074-5200947 strain
            AX4, complete sequence.
ACCESSION   AC117176 AC115586 AC117178 AC115596 AC116923 AC115587 AC117071
            AC115614
VERSION     AC117176.2  GI:28829570
KEYWORDS    HTG.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 182871)
  AUTHORS   Gloeckner,G., Eichinger,L., Szafranski,K., Pachebat,J., Dear,P.,
            Lehmann,R., Baumgart,C., Parra,G., April,J.F., Guigo,R., Kumpf,K.,
            Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and
            Noegel,A.A.
  TITLE     Sequence and analysis of chromosome 2 of Dictyostelium discoideum
  JOURNAL   Nature 418 (6893), 79-85 (2002)
  MEDLINE   22092622
   PUBMED   12097910
  REMARK    The Dictyostelium Genome Sequencing Consortium
REFERENCE   2  (bases 1 to 182871)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-APR-2002) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   3  (bases 1 to 182871)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   4  (bases 1 to 182871)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
COMMENT     On or before Mar 4, 2003 this sequence version replaced
            gi:19569950, gi:20069691, gi:19570012, gi:19909381, gi:19569956,
            gi:20066238, gi:19570143, gi:20069689.
            CDS predictions from GeneID do not necessarily reflect true genes.
            Further Information is available from IMB Jena, Department of
            Genome Analysis
            (http://genome.imb-jena.de/dictyostelium/)
            and the Univerity Colonge, Institute for Biochemistry I
            (http://www.uni-koeln.de/dictyostelium/project.shtml
            Funding
            Agency : Deutsche Forschungsgemeinschaft (DFG).
FEATURES              Location/Qualifiers
     source           1. .182871
                      /organism="Dictyostelium discoideum"
                      /mol_type="genomic DNA"
                      /strain="AX4"
                      /db_xref="taxon:44689"
                      /chromosome="2"
                      /map="5018074-5200947"
     CDS              201. .4191
                      /note="GeneID exon scores (in order of location ranges):
                      188.00 - GSCJ_ID dd_03459"
                      /pseudo
                      /codon_start=1
     CDS              join(4599. .4898,5000. .5833)
                      /note="GeneID exon scores (in order of location ranges):
                      10.51, 75.05 - GSCJ_ID dd_00370"
                      /codon_start=1
                      /product="hypothetical protein"
                      /protein_id="AAO52088.1"
                      /db_xref="GI:28829571"
                      /translation="MFGSKSNFILLIITTLICISLTHCKSFTPSLLSSSSSSFSSSSS
                      SKNKYLYSISSVNPNIESGQIVLVIDPLTNEVIKNKTVYIPYRIIDLLDVDESSNSLL
                      FFCLINDAKNLLLSLNLDTFEFKEIGGTYGGWDYTYARQPYIFVNDNVFLPAKLSFYN
                      ATGLTVLHWDFTDKNSKGGSFKVIEIPIDDFDVESGIPSMGGYDYVNDLLYVTYQSLN
                      DLSENNSTSIVCFNPFSSDTNSTHKIHTGITDLIPGYIHLLFTDQNEGGSLYALSKSF
                      NTQSSIEVCKINFQTNNCDQILSKEMGAYYNSFDYRPYFIGGPQNSDLILLEFTEKNA
                      KPFINFQSVNLLTWNESSPTLLSNYWTSTFGPTYLTFTFFAYI"
     CDS              complement(6175. .6489)
                      /note="GeneID exon scores (in order of location ranges):
                      36.92 - GSCJ_ID dd_00368"
                      /codon_start=1
                      /product="hypothetical protein"
                      /protein_id="AAO52089.1"
                      /db_xref="GI:28829572"
                      /translation="MGHHHHHHGHHGHHHHDTITYTTGVYGAPVVPMVQPMVQPMVQP
                      MPIYGQPPMMGAPVMVPPPMMGAPVMVPQPMVYPTTTVVYETGHHHHHGHHHGHHHGH
                      FF"
     CDS              complement(join(7112. .7299,7429. .7809,7910. .8072))
                      /note="GeneID exon scores (in order of location ranges):
                      6.62, 14.60, -0.44 - GSCJ_ID dd_00365"
                      /codon_start=1
                      /product="hypothetical protein"
                      /protein_id="AAO52090.1"
                      /db_xref="GI:28829573"
                      /translation="MISSKTTQIKPFYILCIISLIVPITFITTYAITEVRGDDDQIIP
                      YISDTIDFAPESCIGTFGLSIIAFLMIIVVFIKHLIIKNNIIEKEYFDDKKEKKELRF
                      NRFSTSMGFISGLSMHGVSSFQFHNANLVHVIFAALFFLCGFIYLLTQTILDIKNSSD
                      IPLKVLIIKKILICISCLFIGYIITQFFIDNLAAIFEIMSALSIFAYMVTFFYEFSNL
                      TLNIEFSKSIKKINGNGGYHNIESN"
     CDS              9857. .10129
                      /note="GeneID exon scores (in order of location ranges):
                      5.94 - GSCJ_ID dd_00385"
                      /codon_start=1
                      /product="hypothetical protein"
                      /protein_id="AAO52091.1"
                      /db_xref="GI:28829574"
                      /translation="MQLFFFFLGNPLGDSIDATNLLLDYKGGCGTRIIQWPNGYKYSF
                      NHSNPVIDNISIESSSFISLGHNFCNSPESIQVYLDGVQIDKSKIK"
     CDS              join(10249. .10614,10675. .10957,11041. .11405,11600. .12025)
                      /note="GeneID exon scores (in order of location ranges):
                      20.49, 7.68, 14.12, 22.22 - GSCJ_ID dd_00382"
                      /codon_start=1
                      /product="hypothetical protein"
                      /protein_id="AAO52092.1"
                      /db_xref="GI:28829575"
                      /translation="MPLKINSVPRLKGGSITIIGERLSSQVNNSTIIVKIGQYDCKNV
                      ISSKNEILCNLESVSNVEKVKLVDLPVNISINGIFNENNLLFSFDTPIISDFILPQGE
                      VKLVGHCFGYSQITQINIDNPILPIKKSKLYIIVNGTKSNTIEIDSSFFVKSVPSSPS
                      VNGQTVNFTLFNINTNNFNATPIMTFQDNSSIKSTDYKNSNEYSTHIYSFEIPKSCGP
                      FDQLCKIKFNGKDEENNISISYYPPLVLNSTTISNSSIGGIITIFGNEFYNEIDLITV
                      GKSKCLNSTFINSTSISCFVEPSNFIINQAKQQQQQQYVNITINGKSGGNNLNIYFDN
                      SFIEKKWAVVYWKYIIKEKIKKFKNRKEQKCIDGMREILNNAEKNFPDIKLDNVITYK
                      KYKMPQEKSVEGELKEQAPPLSEYDQFPSYYLHSPKDSQPSNQIETLPKDSPPPNQNE
                      KLTEDSLPPSNQIETLPEDSSPPPNQTEK"
     CDS              complement(join(12236. .12342,12410. .12826,13073. .13304,
                      13350. .16187))
                      /note="GeneID exon scores (in order of location ranges):
                      8.99, 49.32, 13.83, 125.04 - GSCJ_ID dd_00379"
                      /codon_start=1
                      /product="similar to Plasmodium falciparum. Hypothetical
                      220.3 kDa protein"
                      /protein_id="AAO52093.1"
                      /db_xref="GI:28829576"
                      /translation="MINNETDEFKKENDSINNLDTNKEDIDMESLESKSSRELIKVAK
                      TLKLKANGRKSEIIERIKKFVNDSRNIKQELTKEKLINKFNSSIYQLPIKDIEKHEQL
                      FWRVFKNKMLFKTIFGNLNCKYHSYDDLIYIDWILNNGYLNLLKDKIKRNEPIIGNYY
                      YDGRCPQSEWYLNIFINIQKEYKSNKKETNAPTNNNNNNNNNNKNNNNNNNNNNNNNN
                      NNNNNNNKEFFELLFRNYGKYFISFKKVLKELIEVGNFQGFQVLLDNYFNYCKFGPKN
                      GKIKEQSLVNILNLINIHSHLKMFKYFWETYTDKTNCIIKGLLTNLLINQSNGNSHRS
                      NPPKRYKFLNYLFLQSSVVTKNLFDDLIKINIKSTTNITKFTINDLINYCYSISYINN
                      NNNNNNNNNNDNDDLNLIVPLINEIKSKLSNSQLKLQIKSIIPKLINDEDSNNNNYQ
                      QTTANIKKLIDFCEEKSPIRRTIIQQIEIRIVFNLDFDKDIFKAHLPSSPYPREYFLI
                      ISLEYGNYELFKFIVNHYGDSKVLPPIFKNFKSVSNHGEEYPYILFSNCHSKKQRIQF
                      INNAINDHNNNNNNNNNNNNNNNNNNNNNNNNNNNIIINNKNKEWSNLSVFTIFHLL
                      ISYDDLELLEIAIKGYGGEDSIVNNHCSNIIKVLEKESMMALHQGYLSSFSYFHRHIK
                      SIKMLDYCFHHLNSFFKRVQDMVSWYSLGRIDLLERFEQLLGITNNNNSNSSDNNNNN
                      NNNSNSNIDLLLPFSKNILEENNSNFERSNLIKTAAYIINKPIGIYKYTEDAITHYLE
```

```
                INRNPRLDIDDIITILSKTNNRMEFNDHLNDNYMNSNKIKLLDWMFNFRYEDLISGRF
                TISKGYYIGLFISNKLDNNNNNSIKSIFSNDKINNKWIFENNYRIPRYPSPYYQVILE
                LLDWVKKSGDINLLCKLLNEFYFPIFDGTDFEGVIKNGWLRIIQHLFYDQQYRRVLTK
                KLSNNNSIYTQKELSLLLEHAISWDQFYIVEFFINQMGFKISEAQIVKAIEIRERAKQ
                VIDLLQDDQRIKEERDKAKTNQNKYVGIGNDNSDFGYGGGSYGGGYDSDSYGSNQRDS
                YGGNQRDSYGGNQRYSYGGNQRETTRRDSFNGRDEGYGNNNNNNNNNNNNNNNSYDSD
                PYSNTRAEYENYSNRAASGERSRPRAASGSGPSPATPNFQSSQQQQPTLID"
     CDS        complement(join(16716. .16960,17218. .17290))
                /note="GeneID exon scores (in order of location ranges):
                7.26, 1.45 - GSCJ_ID dd_00372"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AAO52094.1"
                /db_xref="GI:28829577"
                /translation="MINNETDYNNNEEDIDMESLESKSVSIESAKETLFEFLKGVIKN
                GWLRIIQHLFYDQQYRRVLTKKLSSNNSIYTQTELSLLLRHAISWDQFYIAEFFVDQM
                GLK"
     CDS        complement(17816. .18070)
                /note="GeneID exon scores (in order of location ranges):
                26.78 - GSCJ_ID dd_00371"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AAO52095.1"
                /db_xref="GI:28829578"
                /translation="MPPHGHHHHGHHGHHEHVTYTTTPIVQPMVQPMPVYGQPPMMVQ
                PQVVMGAPVVYPPTNVVIETSHHHGHHGHHGHHGHHGHFF"
     CDS        complement(18665. .18934)
                /note="GeneID exon scores (in order of location ranges):
                27.69 - GSCJ_ID dd_00369"
                /codon_start=1
                /product="similar to F22H10.2.p [Caenorhabditis elegans]"
                /protein_id="AAO52096.1"
                /db_xref="GI:28829579"
                /translation="MPPHGHHHHHGHHGHHEHITITPVYAPPQVVQPIYAPPPVVQP
                VYVQPPVVYPQATIVLETGHHHGHHHHHGHHGHHDHHHHGHHGHH"
     CDS        complement(19591. .19872)
                /note="GeneID exon scores (in order of location ranges):
                34.42 - GSCJ_ID dd_00366"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AAO52097.1"
                /db_xref="GI:28829580"
                /translation="MGHHHHHHGHYEEITYTTTPVYGAQVVQPISPVVGIQPIYGQPP
                IVQPIYGQPPVVQPIYGQPPVVYPQTTTIVYETDNHHHGHHHHHHGHFF"
     CDS        complement(join(20614. .22325,22463. .23814,23918. .24106,
                24226. .24464))
                /note="GeneID exon scores (in order of location ranges):
                176.52, 128.57, 13.58, 19.46 - GSCJ_ID dd_00363"
                /codon_start=1
                /product="similar to Dictyostelium discoideum (Slime
                mold). MkpA protein"
                /protein_id="AAO52098.1"
                /db_xref="GI:28829581"
                /translation="MYQNALNHTNSGYGGMKAEEKTNYLNQSVITGYIKGISSYIPSS
                IKSSIKKQIMNSPVTFEEKDTIVGTYFDECFIQGEKHKILINCYNNGFQVWDLNHTDG
                VKELLSCRDGLTKFCKVLTNPLEPDNESSSFYGKRPLLAVVSGEDNPKVTKNMVRIFS
                LQTTELISMYKFKSPIYNVLSNQQIILVVLKERIVGFNANDMSKIKGLSLPCYPSVTS

  Query Match             5.0%;  Score 54.4;  DB 3;  Length 182871;
  Best Local Similarity   46.5%;  Pred. No. 0.79;
  Matches  175;  Conservative    0;  Mismatches  201;  Indels    0;  Gaps    0;

Qy       146 AGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACA 205
            || ||  ||  |  ||  |  |   ||| || || |  | ||| | |  | |       |
Db    173106 AGGAAGTTTTTTATTTTTATTTTTATTACTATTTCTAAAAAAGTTATTTTTTATTTTTTA 173165

Qy       206 TTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTC 265
            || ||| | | ||       || ||  || |  ||    |||       ||    | ||
Db    173166 TTTTTTGGATTATTACTATTTTTAATAATTATTTAATTTTTTATTTTATTAATATTTTTT 173225

Qy       266 ATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATA 325
            ||| |||||  ||| ||          | ||| ||  |  |   || |||| | ||| |
Db    173226 TTTTTTTTTTTTTTTTTTTTTTTTTCAAAAAAAAACAGACTGTTGATAAAGTTAATAAA 173285

Qy       326 TTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTC 385
            ||             |||| |   | | | ||   |||| |   ||    || | || | |
Db    173286 TTTTTAGAAATTTTATTAGAAGAAGATTCAGATTATATAAATATAGAATATATACCAATG 173345

Qy       386 AAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAA 445
            | | || |  | ||    |  ||| | ||    ||   ||   || |   || | |
Db    173346 ATGGAAAATATTTTAAATAGAGTTTATTCAACAAATTTATCTACAACATTAATAAAACCA 173405

Qy       446 ATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTA 505
              | | | ||  |||| |||   | | |  | || |||  ||    ||  ||  |   |
Db    173406 TCTCCAAGAATCTTTAATTTCATATTGAATTTAGGAAAATTACAAGTTGCTGATATTGAA 173465

Qy       506 AAACATTTACTTTCCA 521
            || |||  |   | ||
Db    173466 AATCATAAAAAGTTCA 173481


RESULT 59
CR762392
LOCUS       CR762392              205475 bp    DNA     linear   HTG 24-SEP-2004
DEFINITION  Danio rerio clone DKEY-265A11, *** SEQUENCING IN PROGRESS ***, 6
            unordered pieces.
ACCESSION   CR762392
VERSION     CR762392.2  GI:51988245
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 205475)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-SEP-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 11, 2004 this sequence version replaced gi:51987771.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK265A11
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 203834 bases at least Q40
            Consensus quality: 204157 bases at least Q30
            Consensus quality: 204331 bases at least Q20
            Insert size: 204975; sum-of-contigs
            Insert size: 155101; 5.9% error; agarose-fp
            Quality coverage: 8.84x in Q20 bases; sum-of-contigs Quality
            coverage: 11.71x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    26435: contig of 26435 bp in length
            *    26436    26535: gap of 100 bp
            *    26536    97866: contig of 71331 bp in length
            *    97867    97966: gap of 100 bp
            *    97967   101958: contig of 3992 bp in length
            *   101959   102058: gap of 100 bp
            *   102059   117477: contig of 15419 bp in length
            *   117478   117577: gap of 100 bp
            *   117578   153352: contig of 35775 bp in length
```

```
          *   153353   153452: gap of 100 bp
          *   153453   205475: contig of 52023 bp in length.
FEATURES             Location/Qualifiers
     source          1. .205475
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-265A11"
                     /clone_lib="DanioKey"
     misc_feature    1. .26435
                     /note="assembly_fragment:00267
                     fragment_chain:1"
     misc_feature    26536. .97866
                     /note="assembly_fragment:01739
                     fragment_chain:1"
     misc_feature    97967. .101958
                     /note="assembly_fragment:00011"
     misc_feature    102059. .117477
                     /note="assembly_fragment:00048.0"
     misc_feature    117578. .153352
                     /note="assembly_fragment:00584.0"
     misc_feature    153453. .205475
                     /note="assembly_fragment:01027.0"
ORIGIN

  Query Match             5.0%;  Score 54.4;  DB 2;  Length 205475;
  Best Local Similarity   32.3%;  Pred. No. 0.78;
  Matches  211;  Conservative    0;  Mismatches  443;  Indels    0;  Gaps    0;

Qy        146 AGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACA 205
              | | |||| ||   || | ||       ||| | | || |    | | || | ||
Db      97433 ATATATATATATATATATATATATAAATATATATATNTATATATTTTTATATAAATATAT 97492

Qy        206 TTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTC 265
              | | |    |     |   |  |||| | ||         | | |   | |  | ||
Db      97493 ATATATANAAATATATATAAATATATTTATATATTTATANAAAAAAATTAANATTTTTTA 97552

Qy        266 ATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATA 325
              ||   |||     |  |  | ||||     ||| ||    |   |   |   || |||||
Db      97553 ATNNTTTTANATATAATAAANAAAAANANAAAAAAANAANAAAAAANTATNTATTATATA 97612

Qy        326 TTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTC 385
                  |    | | |||||     | |       | |    |||     |
Db      97613 ANNTANAAAATATATTTAAAATATTANTANAAAAAAAAANNNATTTNAANAATNNNNNTAA 97672

Qy        386 AAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAA 445
              | |||    |   |||   |||  |  |   ||  |             ||
Db      97673 TTTNNAANAAAANNNANNAAAAAAAATAAAAANTTTATTAAAAATATNNATNTAANNNNN 97732

Qy        446 ATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTA 505
                |  | ||| ||| | |   ||| ||       |  |    ||||  |||  |  ||
Db      97733 NAAAAAAAAAAATTTTTNTAAAAAAAAAANANNNANNANNNTATTTNNTATAAAAAATNN 97792

Qy        506 AAACATTTACTTTCCAAATATTCTGTCATTCAGGAGTATGGAAGTATCGATGGCTTCTTT 565
              ||   | | ||    |   |  |   |  |  |  |  ||             || |
Db      97793 AANNNTNNATTTATAATTATTNNTAAAAAAAAATATAAAAAAATNNNANNNNNNTTTTNA 97852

Qy        566 AAAATGAAGCAGGAGGGTCTGGCAGAGAGTATCTATGAAATAAGTTCCTCTGACCTTCAC 625
                |   |  | |
Db      97853 NNATAAANNAATTANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 97912

Qy        626 GCTTAATTTTCTGAATGGAGTGGAGCAAATTACTTCAAGCTTCACTTAACTTGCATATGA 685
                                                                  ||||
Db      97913 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNATATAT 97972

Qy        686 AATGAACCGTACAAAAATACAAGAGTGTCAGGAGAAAGTTATGCTCTGGTAAATATTTTG 745
              |   |    || | | ||| |    | |    | | |  |||         | |||
Db      97973 ATATATATATATATATATATATATATATATATATATATATATATAAAATATAAAAAA 98032

Qy        746 CAAAACAGATAAAAGATAATACTAGAGCTCTGTCCTCAAAGAGTTAAGCAGCTA 799
              |||| | | |||| | |    || |  | | |    | | |  ||   |  ||
Db      98033 AAAAAAAAAAAAAAAAAATATATATATATATATATATATATATATATATATA 98086
```

```
RESULT 60
AC116956
LOCUS       AC116956              266544 bp    DNA     linear   INV 12-MAR-2003
DEFINITION  Dictyostelium discoideum chromosome 2 map 1418423-1684967 strain
            AX4, complete sequence.
ACCESSION   AC116956 AC116980 AC115678
VERSION     AC116956.2  GI:28828496
KEYWORDS    HTG.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 266544)
  AUTHORS   Gloeckner,G., Eichinger,L., Szafranski,K., Pachebat,J., Dear,P.,
            Lehmann,R., Baumgart,C., Parra,G., April,J.F., Guigo,R., Kumpf,K.,
            Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and
            Noegel,A.A.
  TITLE     Sequence and analysis of chromosome 2 of Dictyostelium discoideum
  JOURNAL   Nature 418 (6893), 79-85 (2002)
  MEDLINE   22092622
   PUBMED   12097910
  REMARK    The Dictyostelium Genome Sequencing Consortium
REFERENCE   2  (bases 1 to 266544)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-APR-2002) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   3  (bases 1 to 266544)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   4  (bases 1 to 266544)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
COMMENT     On or before Mar 4, 2003 this sequence version replaced
            gi:20042935, gi:19682974, gi:19919988.
            CDS predictions from GeneID do not necessarily reflect true genes.
            Further Information is available from IMB Jena, Department of
            Genome Analysis
            (http://genome.imb-jena.de/dictyostelium/)
            and the Univerity Colonge, Institute for Biochemistry I
            (http://www.uni-koeln.de/dictyostelium/project.shtml
            Funding
            Agency : Deutsche Forschungsgemeinschaft (DFG).
FEATURES             Location/Qualifiers
     source          1. .266544
                     /organism="Dictyostelium discoideum"
                     /mol_type="genomic DNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /chromosome="2"
                     /map="1418423-1684967"
     CDS             join(<842. .1240,1322. .1916)
                     /note="GeneID exon scores (in order of location ranges):
                     29.76, 65.85 - GSCJ_ID dd_02843"
                     /codon_start=2
                     /product="similar to Dictyostelium discoideum (Slime
                     mold). Hypothetical 98.7 kDa protein (Fragment)"
                     /protein_id="AAO51167.1"
                     /db_xref="GI:28828559"
                     /translation="VAISNFNINNNNKNNNDNKEIRFPTDIIIDDSTEINFNNHEDNN
                     NNNNNNNNNNAYNEDSGNSVQLGDIECVVCLDFLDSYLPTLVKIISEYGVIESCSKI
                     CGLLNKTDEIDICTGLCDLVGVDTFWKIFVTNDINPFYACELITACSTPKNPAADFIS
                     LGVSPAIGQSGSKFLIDLTFQVVNTTGAGQFAFIVFYTQTQSKFINYQLFEGFSPGNY
                     SVKYDFETTNNSTFINGQYPVTVYMCNGECGTMYSTLLNQTTSYFTISNPTPTPTPTP
                     SNSTTPTPTPTNSTPTPTSTSTPTSTPTSTPTPTPTSSSSTKTHSSHYKNKINK"
     CDS             complement(join(2119. .2662,2752. .2943,3024. .3055))
                     /note="GeneID exon scores (in order of location ranges):
                     32.18, 11.63, -1.19 - GSCJ_ID dd_02844"
                     /codon_start=1
```

/product="similar to Mus musculus (Mouse). 16 days neonate thymus cDNA, RIKEN full-length enriched library, clone:A130067E11 product:cyclin C, full insert sequence"
/protein_id="AAO51105.1"
/db_xref="GI:28828497"
/translation="MAANFWESSHCTEWLLDRNKIEESNPKDKQYLTPIELKKLRTHY CFVIQNLGNALKLRQRAISTAIVYFKRFYLKNSFVDCEPRLVAVTCLYLSSKVEECIT QAKKCAAKMKEIDHSFNYLMNDILECEFFVLEELDFCLIIYHPYKSLPFYLQSSGLDP ASIEIIWGIVNDSYRTDVCLLYPPFVVGLGCILLGSYLLKKDIKQWLSELNVEMKDIW EVSKDLIDYYEFEKQQSLQNQSELDLLYNKLPIRNKK"
CDS complement(join(3639. .4158,4352. .4422))
/note="GeneID exon scores (in order of location ranges): 63.23, 9.00 - GSCJ_ID dd_02845"
/codon_start=1
/product="similar to Dictyostelium discoideum (Slime mold). Vacuolar ATP synthase proteolipid subunit (EC 3.6.1.34)"
/protein_id="AAO51106.1"
/db_xref="GI:28828498"
/translation="MFQLLSFLLSGEATAVERIITDACPVYAPFFGAMGVTAALVFTV MGAAYGTAKASVGISNMGVMKPDLVIKAFIPVIFAGVIAIYGLIICVILVGGIKPNAN YTLMKSFTDLGAGLTVGLCGLAAGMAIGIVGDSGVRAFGQQPKLYVIMMLILIFSEAL GLYGLIIGILLSSVSDTYCPGQALVPLNSGNVIGKN"
CDS join(5902. .6004,6157. .6199,6351. .6486,6853. .6962, 7066. .7249)
/note="GeneID exon scores (in order of location ranges): 6.21, -0.71, 4.93, 17.98, 20.83 - GSCJ_ID dd_02846"
/codon_start=1
/product="similar to Mus musculus (Mouse). Similar to ATPase, H+ transporting, lysosomal (Vacuolar proton pump) 21kD"
/protein_id="AAO51107.1"
/db_xref="GI:28828499"
/translation="MPDINPLTPIPDNNYYDQCTYFLVTISPSTWAALGIGLSLALSV VGSAWGIWVTASSLMGAAVKEPRIRSKNIISIIFCEAVAIYGIILAIILNGKIDKFLN IWDPASDYMAGYMMFGAGITVGLCNVFSGVCVGIAGSGCALGDAQNPSLFVKMLIIEI FAGALGLYAVIVGILMTTNALGVTQQPALKP"
CDS complement(7573. .8283)
/note="GeneID exon scores (in order of location ranges): 66.20 - GSCJ_ID dd_02847"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAO51108.1"
/db_xref="GI:28828500"
/translation="MSTTIYFSYPTYSNEREFTTYDDDSEEEYYDQSYFQDYSFQNKP QPSSSNINKSKSVKVTGYGNGSKNAGQAIPRTDKDYKAAIEYKKEKKLAIKQIKENKL STSPTTITTTTTTNTATSYENVNIESPNDENNKTKKLTKLEVKSIIEKETDLKKKERL PSSIEVYNRILHDEQLKDKIQDFIITYEDRFDGLCDSELLKLDEAIPFHRIRQIRYLN QIVWSRSERFYDFSNLLD"
CDS complement(join(9480. .10271,10379. .11125,11253. .11858, 11963. .12121))
/note="GeneID exon scores (in order of location ranges): 50.34, 56.92, 38.65, -2.17 - GSCJ_ID dd_03351"
/codon_start=1
/product="similar to Plasmodium falciparum (isolate 3D7). Hypothetical protein"
/protein_id="AAO51109.1"
/db_xref="GI:28828501"
/translation="MEDTALMGYFHFIRYLHAATTPKLRFAFEIFSLIVISLFVSTLF VMHNQFVGKSDCLKLPNNTRFDILSIEIKGETSIFSRQHIYKKSTTTETTPTTTTTT TTTVNNDIDQQSYNKYLQVLKGAVDKHEPCIYDENYDNDNDNDIDKDSVTTATTTTAA AAAAAAAATTTTTTTSLITESTYFKTAIEFIIDLIRQMYHSLYNFISIFFNSQFDEYD QQYKKDRLEREHLKRLKEELKKHIKDIDRNHVNFEKLQHHQQQQLQQPQQQQQQQQQQ QQQQQQQQQQQQQQQQEREEKKEKDKQHNNNHHNNGIGSGVLNDLNDIVQYFIDSHNP RYEFSYQNAFLMLTTEQRNKFNITTLHIDINANDQCLGNGFHKAIIGMLGYDILIINN LVYGFPLIPNNNINNNNNNNNGRGGYLRIAGNSQLINLYQYSSPKLSLNNESILYIGY RFFETNFIFFLTLIVFGSSLKLCESSFAHYVDRDPRGVLKCMLHNLLIFALFIFIQIG YISIFSIILSTEYLVFFISSLITSYFSTWGLRSKESIRYYPIIFLSILYGLILYVVFF PSGFHPIAFYACYSLIEYLLIICLFNFEIPAVLNNRILPNNQNNNNQNNNNQNNNNNN QNNNQNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNQ NNNQNNNNNQNNNNNNTNINNHNNNNINNQNNINNTQNENQNNNNQNDGNNQNIYNVP INQQTIVNNSSQRRLAASAILNNSNEM"
CDS complement(join(13033. .14475,14560. .15048))
/note="GeneID exon scores (in order of location ranges): 138.82, 42.31 - GSCJ_ID dd_00640"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAO51110.1"
/db_xref="GI:28828502"
/translation="MNIINYFKNGNNNNGHSSGVGSSGHSYYSNNKGTSFVPSPVLYM TPMAPIDYGCPDDMLDDGRSDTAPSSQNTSPMYQHHNFHNNHNNTHNHNNNNNNNNNN HEQPNCALEFLEHHLSFETKYGVNVSQGNRKYQEDRHKVKMGLENNQYLSLFGVFDGH GGDRASNFVKKKIVNCVNKYVKENKAGYSSKNLNSSSSPTGSTSSASSSGSSSSSASS SGGGGGGGGPLNGSSSSIGMPHASTISPTPSSGNSSTSSGASSSGVSSANSPIGVGFD SPGWMKFNHLYNQNNFQQEIQNRSEFLQAALYNTFQFLDNRYCKKYRQKGDGGTTCLV ALLSNPPNAQPLLVVANAGDSRGVLCRNGKAYALSYDHKPGNPKEKQRITSSGGKIEW DFNERIWRVSGILSVSRGIGDIPLKKWVICDPEFVVFPLKGPIRKSNSSRYTSPSSSS SSSSSSSSSSSSSPSSPKAFRSSPQINRFFFSPSFKNKTISNSTGSLPPPSTNNSPQ QPPPILHQNNISSPVSSISEAQPFNSIDNQNNNNNNNNNSKLISTSPPNHKMYYCSNH QYNYGEVDQFFVLATDGIWDVFENQELVEFINAIIEESYHSKRLDWDPNEISKRVVQE AYRKGSGDNATVLIIKLYW"
CDS complement(join(17600. .18734,18858. .19218,19315. .20422))
/note="GeneID exon scores (in order of location ranges): 69.94, 30.45, 124.05 - GSCJ_ID dd_00639"
/codon_start=1
/product="similar to Dictyostelium discoideum (Slime mold). cGMP-stimulated cGMP phosphodiesterase (cGMP-stimulated cGMP-specific phosphodiesterase)"
/protein_id="AAO51111.1"
/db_xref="GI:28828503"
/translation="MMFSKRKSIRFNPEGDYTELPRGGYVVPSKLGGIQFGVPPETIK DSMALKIDVPTIYVFPEELWDRKTGINAAEAEFPAYFNYFILKRKVSFVCTKEQEQRI RIVFQETLLGPPEFNTGIIINSSSSTTDTSKTSPIKKQTSSSSPPLSPQQQQPPPPLV KQPSQQQLEELATMDTCHYHHHHHHQEVNDNDNNNNTTTNNNNIEILEQQQQQQQQQQ QQQDEDSTDVDEEFQKEFSSTFPRSEIPNLEKECKYLRTFNSVDELFDFILFDDNGIA KLSDDVEIHFQEDQSLFKVLQFEETGKGNKVQTLVATIPSKILFPDMINSISTINDNK IFDPPTFGITIIGSSHGFDPKKSTTGFVLWINKRGIMVDPPLNSSSFLQSQGVPTRMV DHIILTHCHADHDSGTFQKLLEEYQITVVTTPTILGSFLRKYGALSNLSTDLLRRLFI FRPVMIGEEPMIISGAEFRFFYTIHTIPTISFEVFYGGKSIFYSGDTCYDPNRIKDMNK RGIMKTSRMKFFLRPWNHTVVLHEAGVPPIHTPVSVLRALPDEVKNRLYLVHISEHTL PAGSGLKIAKEGVAHTLSLDVMKSSHSEAVDILKLVESVDIFRSIPLTQACEILQTAT KRKYSQGSVIIARDTEPDAFYVVASGVVCVNIGELKKNLIVGDYFGEMSLVMGGLRSA NVQAVTDVEVLSFNKEDFLSITRNSTESIQFITRLWEMRNEKSWETMSLNSVFSRCTN SQKTAIQSILVRELIKKDETLWCKGEEALFGCLVAEGSFVFKEDDSLESFSQGSFLGD INAMTTQPPSIHKTTVVAKEESVIYKVLSQDLIKFFSNNPGIQLAFLDTIFVDALRDQ VQQLVNSKLTY"
CDS complement(21566. .21865)
/note="GeneID exon scores (in order of location ranges): 24.15 - GSCJ_ID dd_00638"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAO51112.1"
/db_xref="GI:28828504"
/translation="MTQLGYQQIPPGMHNNNNNNPNNNSNSSNNNNNSNFIDFLRQSG ITMNDDVNSSGMMTPSSLLNDSSLEHRVASTNKELKDLEDKLQKSSPLNTQSYYR"
CDS complement(join(21982. .25164,25267. .25396,25496. .25821, 25923. .26082,26307. .26464,26602. .26691))
/note="GeneID exon scores (in order of location ranges): 340.18, 15.33, 31.36, 13.30, 15.31, 10.95 - GSCJ_ID dd_00637"

Query Match 5.0%; Score 54.4; DB 3; Length 266544;
Best Local Similarity 47.2%; Pred. No. 0.75;
Matches 166; Conservative 0; Mismatches 186; Indels 0; Gaps 0;

Qy 172 TAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGAT 231
Db 16648 TTATTATTTTTTATTATTATTATTATTATTATTATTATTATTATTATTTTTTAATTTGTT 16707

Qy 232 TTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAG 291
Db 16708 TATTTTTTATCTTTTTTTATATAAATATGTATTTATAGATTTATTTTTATATGGTAATAA 16767

Qy 292 GTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAA 351
Db 16768 ATATAATCCTTCCTCATCATACGAAAATAATTAAATAAAAAAGAAATAAAAAAAAAAAAA 16827

```
Qy        352 ATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTA 411
              |  | | || |  |  |||     | |  || |||||  |||| |       | |   ||
Db      16828 AAAAATACAAAAAAGTTTAAACTTTTTCCCAATCAAATATAAATAAAATCTATCATTATA 16887

Qy        412 TACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAAT 471
              |  |   | ||      |      | | | || | | | | | | ||   |
Db      16888 AAAAAAAGGTTTAAAACTATTCCATACCAAATACATAAATAGATAATATGTATTTATATA 16947

Qy        472 AATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAA 523
              || |    |    |  |||||||| |  | ||   |  |||| |||  |||
Db      16948 TATATTTTATTTTTTTTTTTTTATTTTTTTTTTTATTTTATTTTAAAAA 16999


RESULT 61
AC005140
LOCUS       AC005140               310779 bp    DNA     linear   HTG 12-AUG-2000
DEFINITION  Plasmodium falciparum chromosome 12 clone 3D7, *** SEQUENCING IN
            PROGRESS ***, 4 unordered pieces.
ACCESSION   AC005140
VERSION     AC005140.8  GI:9797735
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 310779)
  AUTHORS   Hyman,R.W., Fung,E.L., Qin,F., Rowley,D., Mao,J., Tamaki,T.,
            Kurdi,O.B., Conway,A.B. and Davis,R.W.
  TITLE     Plasmodium falciparum 3D7 chromosome 12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 310779)
  AUTHORS   Hyman,R.W., Qin,F., Fung,E.L., Conway,A.B. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-JUN-1998) Stanford DNA Sequencing and Technology
            Center, Stanford University, 855 California Avenue, Palo Alto, CA
            94304, USA
COMMENT     On Aug 12, 2000 this sequence version replaced gi:8810455.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   187308: contig of 187308 bp in length
            *  187309   187508: gap of unknown length
            *  187509   257820: contig of 70312 bp in length
            *  257821   258020: gap of unknown length
            *  258021   307255: contig of 49235 bp in length
            *  307256   307455: gap of unknown length
            *  307456   310779: contig of 3324 bp in length.
FEATURES             Location/Qualifiers
     source          1. .310779
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:5833"
                     /chromosome="12"
                     /clone="PFYACB8-420"
                     /clone="3D7"
ORIGIN

  Query Match             5.0%;  Score 54.4;  DB 2;  Length 310779;
  Best Local Similarity   51.7%;  Pred. No. 0.73;
  Matches  124;  Conservative    0;  Mismatches  116;  Indels    0;  Gaps    0;

Qy        131 ATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGAT 190
              |||||||         | | ||| | ||  |    |||    || |    | | | |
Db     292569 ATATGAGGGATAAATAAATATAATAATGCTGCATAATTTTATATTTAAAATTTTGGATGG 292628

Qy        191 CTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACA 250
              |  | ||  | | ||     |   ||| |||||   ||  |||||||||| | |  |   |
Db     292629 ATATTATTTTATATATAAAATAAATAAAATACTTTTTTCCTTTTTTTAATATATATATAA 292688

Qy        251 GACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCA 310
              | |       | ||  ||| |||||  ||| |||| || |  ||||     ||    | |
Db     292689 TATAATATATTTTATTTTTTTTTTTTTTTTTCATAATATATATATGTTTATTTTATAA 292748

Qy        311 GAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTA 370
              ||||| |  | | |  |  |||| |||| ||  | ||||  | |  |||| | | ||
Db     292749 AAGAAAAAAAAAAAAAAAAAAAAGTTCAAAAAGAAGAAAAAAAAAACTATATATAAAAGTA 292808


RESULT 62
PFMAL13_05
WPCOMMENT
  Sequence split into 28 fragments  LOCUS PFMAL13 Accession AL844509
     Fragment Name      Begin     End
     PFMAL13_00             1    110000
     PFMAL13_01        100001    210000
     PFMAL13_02        200001    310000
     PFMAL13_03        300001    410000
     PFMAL13_04        400001    510000
     PFMAL13_05        500001    610000
     PFMAL13_06        600001    710000
     PFMAL13_07        700001    810000
     PFMAL13_08        800001    910000
     PFMAL13_09        900001   1010000
     PFMAL13_10       1000001   1110000
     PFMAL13_11       1100001   1210000
     PFMAL13_12       1200001   1310000
     PFMAL13_13       1300001   1410000
     PFMAL13_14       1400001   1510000
     PFMAL13_15       1500001   1610000
     PFMAL13_16       1600001   1710000
     PFMAL13_17       1700001   1810000
     PFMAL13_18       1800001   1910000
     PFMAL13_19       1900001   2010000
     PFMAL13_20       2000001   2110000
     PFMAL13_21       2100001   2210000
     PFMAL13_22       2200001   2310000
     PFMAL13_23       2300001   2410000
     PFMAL13_24       2400001   2510000
     PFMAL13_25       2500001   2610000
     PFMAL13_26       2600001   2710000
     PFMAL13_27       2700001   2732359
Continuation (6 of 28) of PFMAL13 from base 500001 (AL844509 Plasmodium falciparum 3D7 ch

  Query Match             5.0%;  Score 54.2;  DB 2;  Length 110000;
  Best Local Similarity   45.0%;  Pred. No. 0.94;
  Matches  247;  Conservative    0;  Mismatches  298;  Indels    4;  Gaps    1;

Qy        198 TAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATAC 257
              |||| |      |||  ||  |||  |   ||  |||  ||  | |       |    ||||
Db      76912 TAAATATGAATATTTATTTTTATTTTTGATTCTTTTAATTTATAATTAAAAATCGCATAT 76971

Qy        258 AGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAG 317
              |  |  | |      |     | | ||  | | | | |   | | ||    || | | |
Db      76972 AAGTGATTAACCTCTAAGTTTCTATTATATAGAAGAAAGTAAGACTGATAACAAAAAGAA 77031

Qy        318 ATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATC 377
              | | | | |    | |||||  |||||| | ||  ||||||||||     ||  || |
Db      77032 AAAAAAGTTAAAAAGATAGAAAAATAAACATATAAAAAGTGCATATTCATATATCTACTA 77091

Qy        378 TTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTAC----CTGCTTTCTTATCGCA 433
              | ||| | |    | |  |    | |     ||  ||       | |   |||   |
Db      77092 AAGAAGTAACAACATATATATTCTAATTATTCATTATCACAACATAACATATATATATGA 77151

Qy        434 CTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTT 493
              |  |    | || ||  | ||  | ||| |    |   || |    | | |||||   |
Db      77152 GAGATATATGGAAATAAAAAAATGTGTATATATATATATATATATATATATAATACAT 77211

Qy        494 TATGTACTATTAAAACATTTACTTTCCAAATATTCTGTCATTCAGGAGTATGGAAGTATC 553
              |     |  | ||     | | ||  || ||  || ||||    ||||  ||
Db      77212 TTATATACAGAAGAAATACATATATACATGTAATCAATCTTTCACATATATGTAATATGT 77271
```

```
Qy        554 GATGGCTTCTTTAAAATGAAGCAGGAGGGTCTGGCAGAGAGTATCTATGAAATAAGTTCC 613
              ||    | ||||||||  |     |||       | | |   |  |||| ||||||
Db      77272 AATATTCTGTTTAAAATTTATACAAAGGAAAAAAAAAAAAAAAAAAAAAAAATTAGTTCC 77331

Qy        614 TCTGACCTTCACGCTTAATTTTCTGAATGGAGTGGAGCAAATTACTTCAAGCTTCACTTA 673
              |||       |    ||  ||| | |||  | |  |   |  |||||  |    | ||  |
Db      77332 TCTTTTACAAATATATACATTTGTAAATACATTAAATTTATATACTTTATAAATGACAAA 77391

Qy        674 ACTTGCATATGAAATGAACCGTACAAAAATACAAGAGTGTCAGGAGAAAGTTATGCTCTG 733
              |     | |  |||  ||   || |||||   |      | || |\| |  |||   ||
Db      77392 AAAAAAAAAAAAAAAAAAAAATAAAAAAAAGTATCTCCATAAGAAAATATATATTTCCTT 77451

Qy        734 GTAAATATT 742
              |   || |
Db      77452 ATTTTTAAT 77460


RESULT 63
PFMAL1P2_1
WPCOMMENT
  Sequence split into 4 fragments  LOCUS PFMAL1P2 Accession AL031745
     Fragment Name      Begin     End
     PFMAL1P2_0             1    110000
     PFMAL1P2_1        100001    210000
     PFMAL1P2_2        200001    310000
     PFMAL1P2_3        300001    384550
Continuation (2 of 4) of PFMAL1P2 from base 100001 (AL031745 Plasmodium falciparum DNA f

  Query Match             5.0%;  Score 54.2;  DB 3;  Length 110000;
  Best Local Similarity   47.4%;  Pred. No. 0.94;
  Matches  193;  Conservative    0;  Mismatches  213;  Indels    1;  Gaps    1;

Qy         89 TTCCTTATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGA 148
              ||  || | | | |   | | |  |  ||  | | | |  | |   ||  ||
Db      51694 TTTTTTTTATTTTTTTTTATTTTTTTTATTTTATTTTATTTTATTTTTATTTTATTTTTAT 51753

Qy        149 AATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTG 208
              |||  |    |||| |  ||  |  |||  |||||| ||  |    |  ||||     ||
Db      51754 AATTATATATATTATATGTCTATATATATATTTGTAAAACTTTAGAATTTAATATTTTTA 51813

Qy        209 TTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATT 268
              ||  |||| |  | |   || ||||||  | ||     | ||| | ||| |    || | |
Db      51814 TTAGGTTATTTAATTTTATTTATTTTTAAAGAACTTTTAACATATATATATAATTTTAGT 51873

Qy        269 TGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTA 328
              || ||   | ||    |  ||| ||  |||||||| | |     ||    | | |
Db      51874 ATACGTTTGTACATACACTTTACATATGAAT-AATGCATTTATAATTTAATACATTTTAA 51932

Qy        329 GCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAA 388
              ||  |  |   |     | | | |||| ||||  |  | || | |  | ||  | |
Db      51933 TTTTATTATATTTCTTATTTTATTATATTTCTTATTTTATTATATTTTATTTTATTTTCT 51992

Qy        389 GGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATT 448
              |||||  ||  ||   ||  || ||| | | | |     |||  ||  |   |||  |   |
Db      51993 CTTAAAATGTTAATATATTGGCATATTCCTTCCGGATTTAATGCTCATTAAAGATTCAAC 52052

Qy        449 ACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTA 495
                 ||| | ||| ||    | ||    | |||| | ||   | ||
Db      52053 TAGTTAAAATATATATTTACCAAAAAAAAAAAAAAAAAAAAAATATA 52099


RESULT 64
AC119571/c
LOCUS       AC119571               163703 bp    DNA     linear   HTG 29-MAY-2004
DEFINITION  Pan troglodytes clone rp43-48d22, WORKING DRAFT SEQUENCE, 2 ordered
            pieces.
ACCESSION   AC119571
VERSION     AC119571.15  GI:47825091
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 163703)
  AUTHORS   Deschamps,S., Wang,K., Qu,C., Zhu,H., Shaikh,T. and Roe,B.A.
  TITLE     Pan troglodytes BAC Clone rp43-48d22
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 163703)
  AUTHORS   Wang,K., Qu,C., Zhu,H., Shaikh,T. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-APR-2002) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE   3  (bases 1 to 163703)
  AUTHORS   Deschamps,S., Keqing,K., Qu,C., Zhu,H., Shaikh,T. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAY-2004) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On May 29, 2004 this sequence version replaced gi:34576530.
            ---------- Genome Center
            Center: Department Of Chemistry And Biochemistry
            The University Of Oklahoma
            Center code:UOKNOR
            ----------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1    51208: contig of 51208 bp in length
            *   51209    51308: gap of unknown length
            *   51309   163703: contig of 112395 bp in length.
FEATURES            Location/Qualifiers
     source         1. .163703
                    /organism="Pan troglodytes"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:9598"
                    /clone="rp43-48d22"
                    /clone_lib="RPCI - 43 Male Chimpanzee BAC Library"
ORIGIN

  Query Match             5.0%;  Score 54.2;  DB 2;  Length 163703;
  Best Local Similarity   48.3%;  Pred. No. 0.88;
  Matches  181;  Conservative    0;  Mismatches  193;  Indels    1;  Gaps    1;

Qy        149 AATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTG 208
              |||||  ||  |||| | || | |  ||| || |||| |  || | || | | | | | |
Db      47853 AATATATATAATTATATATAATTATGTATATTATTTTATGTGATATATAATT 47794

Qy        209 TTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATT 268
              | | | | ||||       ||  | || | | |    | | | |   | |  || | ||
Db      47793 ATATATAATTATATAATTATGTATATTATTATATGTGTAATATAATAATATATAATTATA 47734

Qy        269 TGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATC-ATATATT 327
              |||  |    |  |  | || |   |  |   |||  ||| | |  |  ||  |||   |
Db      47733 TGTAATATAATAATATATAATATACAATATATAATATATTGATATTATTATATATAATAT 47674

Qy        328 AGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAA 387
              |    ||  | |||| | |  | |  |    | || | ||   |||    |    |
Db      47673 ATTGTATTATATATTATATAATACTATAATACATAATATATATTATATATAATATATC 47614

Qy        388 AGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAAT 447
              |    | |   ||   || |  ||   || |    |   |  | | | | | | |  |
Db      47613 ATAATATATTACATATTATATATAATATATCATAATATATTACATATTATATAAAATATA 47554

Qy        448 TACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAA 507
              ||  | | || |||| |   | || || | | |  | ||||| | | || |  ||| |||
Db      47553 TATAATATATATTATATATAATATTATATAAAATTATATAATATATAATATTATATAAAA 47494

Qy        508 ACATTTACTTTCCAA 522
                 || || | |  ||
```

```
Db      47493 TTATATAATATATAA 47479


RESULT 65
CR376765/c
LOCUS       CR376765              191040 bp    DNA     linear   HTG 06-OCT-2004
DEFINITION  Danio rerio clone DKEY-208L2, WORKING DRAFT SEQUENCE, 13 unordered
            pieces.
ACCESSION   CR376765
VERSION     CR376765.4  GI:53850263
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 191040)
  AUTHORS   McLaren,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 6, 2004 this sequence version replaced gi:45772264.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK208L2
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 183560 bases at least Q40
            Consensus quality: 184917 bases at least Q30
            Consensus quality: 186193 bases at least Q20
            Insert size: 189840; sum-of-contigs
            Insert size: 184621; 1.4% error; agarose-fp
            Quality coverage: 7.23x in Q20 bases; sum-of-contigs Quality
            coverage: 8.24x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 13 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    27377: contig of 27377 bp in length
            *   27378    27477: gap of 100 bp
            *   27478    35427: contig of 7950 bp in length
            *   35428    35527: gap of 100 bp
            *   35528    50047: contig of 14520 bp in length
            *   50048    50147: gap of 100 bp
            *   50148    90203: contig of 40056 bp in length
            *   90204    90303: gap of 100 bp
            *   90304   107309: contig of 17006 bp in length
            *  107310   107409: gap of 100 bp
            *  107410   124728: contig of 17319 bp in length
            *  124729   124828: gap of 100 bp
            *  124829   140974: contig of 16146 bp in length
            *  140975   141074: gap of 100 bp
            *  141075   147663: contig of 6589 bp in length
            *  147664   147763: gap of 100 bp
            *  147764   154007: contig of 6244 bp in length
            *  154008   154107: gap of 100 bp
            *  154108   156153: contig of 2046 bp in length
            *  156154   156253: gap of 100 bp
            *  156254   158750: contig of 2497 bp in length
            *  158751   158850: gap of 100 bp
            *  158851   161831: contig of 2981 bp in length
            *  161832   161931: gap of 100 bp
            *  161932   191040: contig of 29109 bp in length.
FEATURES             Location/Qualifiers
     source          1. .191040
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-208L2"
                     /clone_lib="DanioKey"
     misc_feature    1. .27377
                     /note="assembly_fragment:01631
                     fragment_chain:1"
     misc_feature    27478. .35427
                     /note="assembly_fragment:00282
                     fragment_chain:1"
     misc_feature    35528. .50047
                     /note="assembly_fragment:00785
                     fragment_chain:1"
     misc_feature    50148. .90203
                     /note="assembly_fragment:02065
                     fragment_chain:1"
     misc_feature    90304. .107309
                     /note="assembly_fragment:00393
                     fragment_chain:1"
     misc_feature    107410. .124728
                     /note="assembly_fragment:00988
                     fragment_chain:1"
     misc_feature    124829. .140974
                     /note="assembly_fragment:00586
                     fragment_chain:1"
     misc_feature    141075. .147663
                     /note="assembly_fragment:00182
                     fragment_chain:1"
     misc_feature    147764. .154007
                     /note="assembly_fragment:00114
                     fragment_chain:1"
     misc_feature    154108. .156153
                     /note="assembly_fragment:00095"
     misc_feature    156254. .158750
                     /note="assembly_fragment:00100"
     misc_feature    158851. .161831
                     /note="assembly_fragment:00106"
     misc_feature    161932. .191040
                     /note="assembly_fragment:01201.0"
ORIGIN

  Query Match             5.0%;  Score 54.2;  DB 2;  Length 191040;
  Best Local Similarity   45.7%;  Pred. No. 0.86;
  Matches  161;  Conservative    0;  Mismatches  191;  Indels    0;  Gaps    0;

Qy        205 ATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATT 264
              ||| |||| ||  | |  |   ||  |||||||  | |   || | | | | |    |
Db     124438 ATTTTTTTTTTTTTTTTNTNTTTTTTTTTTTTTATNAAATAATAAAAAAAAAAAAANTTA 124379

Qy        265 CATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATAT 324
              || | |||||  ||| |           | |       |   ||      ||   | | |
Db     124378 TATNTTTTTTTTTTTTTTTNTTNTTTTTTTTTTTTTTTTTTTNTTTTATTAATTTTTAATT 124319

Qy        325 ATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGT 384
              |||   | || ||||  ||| |  ||||   |  |||||  |||  ||||| ||     |
Db     124318 ATTTTTTATTTTAAAAAAAAANAATAAATATATATATTTTATTTTTTAATTTTATTNT 124259

Qy        385 CAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGA 444
                |  || ||  |  |  || |  | || |       | | || |      |  | | |
Db     124258 TTATTTTATAATATATATTTATAAATTTAATTATATATATTTTTTTTTTATTTTATTTAA 124199

Qy        445 AATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATT 504
              |||      |||| | |  ||    ||  || |       | |||  || | | |   | ||
Db     124198 TATTTATNTAAATATAAAATTTTTAAATATTTATATTTATATATATTATAAAATTNNAAT 124139

Qy        505 AAAACATTTACTTTCCAAATATTCTGTCATTCAGGAGTATGGAAGTATCGAT 556
              | |   |  | |  | ||||     ||  |   |||  |  |||  ||
Db     124138 ATATATATATATATATATATATATATATATATATATATATATATATATATAT 124087
```

```
RESULT 66
AC026698
LOCUS       AC026698              192731 bp    DNA     linear   PRI 30-JAN-2002
DEFINITION  Homo sapiens chromosome 5 clone CTC-457H17, complete sequence.
ACCESSION   AC026698
VERSION     AC026698.6  GI:18425273
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 192731)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 192731)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-MAR-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 192731)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JAN-2002) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Jan 30, 2002 this sequence version replaced gi:15290376.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 99.9% of Sequence;
            Estimated Total Number of Errors is 0.2.
FEATURES             Location/Qualifiers
     source          1. .192731
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTC-457H17"
ORIGIN

  Query Match             5.0%;  Score 54.2;  DB 9;  Length 192731;
  Best Local Similarity   45.9%;  Pred. No. 0.86;
  Matches  185;  Conservative    0;  Mismatches  218;  Indels    0;  Gaps    0;

Qy        154 TGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTG 213
              | || ||      ||  ||||    |  | || || || | || |  || | ||| | |
Db     168904 TTATAATATATTATATGTTATATATAATATATAATATATTATATATAATATATTATATAT 168963

Qy        214 TTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTT 273
              |  ||||| |   |    | || ||| | |   || | |          |||  |||   |
Db     168964 TATATATAATATATAATATATTATAATATATAATATATAATATATTATATATATTATATA 169023

Qy        274 TTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAG 333
              |     | |    | |   |||| | |     ||  |    || || | ||||     |
Db     169024 TATTGATATATTGATATATTATATATATCAATATATATGATAATATGAGATATATTATAT 169083

Qy        334 TTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAA 393
              | ||| ||  |  |||| ||  |   | | || |    |  || ||  | ||| |  | |
Db     169084 ATCAATATATATTACAATATATGATTACAATATACAGATATATATTATAATCATATATTA 169143

Qy        394 AAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCAC 453
              ||   |    |     ||  | |   | | ||    || |  |      || | ||
Db     169144 TAATATACAGATATATATTATAATCATATATTATAATATACAGATATATATTATAATCAT 169203

Qy        454 AAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTT 513
              | ||| || | |    |||         || ||||  || | || ||||  || |   | |
Db     169204 ATATTATAATATACTGATATATATTATAATCATATATTATAATATACTGATATATATTAT 169263

Qy        514 ACTTTCCAAATATTCTGTCATTCAGGAGTATGGAAGTATCGAT 556
              | |     ||||| |||  ||  |  |  ||  |   ||  ||
Db     169264 AATCATATAATATACTGATATATATTATAATATACTGATATAT 169306
```

```
RESULT 67
PFMAL13P2_3/c
WPCOMMENT
  Sequence split into 4 fragments  LOCUS PFMAL13P2 Accession AL049185
     Fragment Name      Begin     End
     PFMAL13P2_0            1    110000
     PFMAL13P2_1       100001    210000
     PFMAL13P2_2       200001    310000
     PFMAL13P2_3       300001    377835
Continuation (4 of 4) of PFMAL13P2 from base 300001 (AL049185 Plasmodium falciparum 3D7 c

  Query Match             5.0%;  Score 54;  DB 2;  Length 77835;
  Best Local Similarity   47.6%;  Pred. No. 1.1;
  Matches  159;  Conservative    0;  Mismatches  175;  Indels    0;  Gaps    0;

Qy        201 AGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGC 260
              | ||| |  |||   || |||  |  |  |        | || |   |||  | | |
Db     65820 ATACAATTCTTTTAAAAAATATACATTATAACAAAAAAGAAAAAAATACCAAAAAATTAA 65761

Qy        261 TATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATC 320
              |  |     | ||| |  | | |  | | ||  |||| | ||||  |   | |   | ||
Db     65760 TTATTGCATTTTTCTTTTATATAAATATAATATATATATAAATATAAATATATGGACATA 65701

Qy        321 ATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTG 380
              | | ||   | |  ||| |  || | | ||  | || ||  ||   ||| |  ||| |
Db     65700 TAAAAATATAAAATAAAAATTATAAAAAATAAATAATGTATTTTTTTTTTATTATATCCTT 65641

Qy        381 CAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAA 440
              | | |  |   |  |  | || |  ||| ||  |||||   | ||   |   |||
Db     65640 TTTTTATATATTGTATATTTATATATATATATCATATGTGCTTGTATTTCTTTCAAAAAA 65581

Qy        441 TAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTAC 500
              |    ||         ||||| ||    ||  || |    ||| ||| |||||||| |
Db     65580 AAATGTTTTTTTTTTTTTTTAATTATCTAAATATATATATAAATATATTTTTTTATTTTT 65521

Qy        501 TATTAAAACATTTACTTTCCAAATATTCTGTCAT 534
              |||||  | |   | |   || |     ||||
Db     65520 ATATAAAATAATACATATAATAAAAAAAAATCAT 65487
```

```
RESULT 68
PFMAL7P1_00
WPCOMMENT
  Sequence split into 14 fragments  LOCUS PFMAL7P1 Accession AL844506
     Fragment Name      Begin     End
     PFMAL7P1_00            1    110000
     PFMAL7P1_01       100001    210000
     PFMAL7P1_02       200001    310000
     PFMAL7P1_03       300001    410000
     PFMAL7P1_04       400001    510000
     PFMAL7P1_05       500001    610000
     PFMAL7P1_06       600001    710000
     PFMAL7P1_07       700001    810000
     PFMAL7P1_08       800001    910000
     PFMAL7P1_09       900001   1010000
     PFMAL7P1_10      1000001   1110000
     PFMAL7P1_11      1100001   1210000
     PFMAL7P1_12      1200001   1310000
     PFMAL7P1_13      1300001   1351552
LOCUS       PFMAL7P1              1351552 bp    DNA     linear   HTG 03-APR-2004
DEFINITION  Plasmodium falciparum 3D7 chromosome 7, *** SEQUENCING IN PROGRESS
            ***, 12 ordered pieces.
ACCESSION   AL844506
VERSION     AL844506.1  GI:23498713
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Plasmodium falciparum 3D7
  ORGANISM  Plasmodium falciparum 3D7
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 1351552)
  AUTHORS   Seeger,K., Murphy,L., Harris,D., Berriman,M., Pain,A., Hall,N.,
            Quail,M. and Barrell,B.
```

```
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1351552)
  AUTHORS   Seeger,K., Murphy,L., Harris,D., Berriman,M., Pain,A., Hall,N.,
            Quail,M. and Barrell,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-SEP-2002) P.falciparum Genome Sequencing Consortium,
            The Welcome Trust Sanger Institute, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA, UK
COMMENT     This chromsome is unfinished and still contains gaps. The contigs
            have been orderd as accuratly as possible using mapping data but
            there will be errors. There may also be some sequencing errors.
            Contigs are seperated with 100 Ns.
            For more information about this sequence or the Malaria Project,
            see http://www.sanger.ac.uk/Projects/P_falciparum.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 12 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1    14518: contig of 14518 bp in length
            *   14519    14618: gap of 100 bp
            *   14619    40653: contig of 26035 bp in length
            *   40654    40753: gap of 100 bp
            *   40754    41735: contig of 982 bp in length
            *   41736    41835: gap of 100 bp
            *   41836    62916: contig of 21081 bp in length
            *   62917    63016: gap of 100 bp
            *   63017    68625: contig of 5609 bp in length
            *   68626    68725: gap of 100 bp
            *   68726   252889: contig of 184164 bp in length
            *  252890   252989: gap of 100 bp
            *  252990   545136: contig of 292147 bp in length
            *  545137   545236: gap of 100 bp
            *  545237   714553: contig of 169317 bp in length
            *  714554   714653: gap of 100 bp
            *  714654   987810: contig of 273157 bp in length
            *  987811   987910: gap of 100 bp
            *  987911   992126: contig of 4216 bp in length
            *  992127   992226: gap of 100 bp
            *  992227  1204744: contig of 212518 bp in length
            * 1204745  1204844: gap of 100 bp
            * 1204845  1351552: contig of 146708 bp in length.
FEATURES             Location/Qualifiers
     source          1. .1351552
                     /organism="Plasmodium falciparum 3D7"
                     /mol_type="genomic DNA"
                     /isolate="3D7"
                     /db_xref="taxon:36329"
                     /chromosome="7"
     gene            complement(join(15188. .16408,17145. .26585))
                     /gene="MAL7P1.1"
                     /note="synonym: VAR"
     CDS             complement(join(15188. .16408,17145. .26585))
                     /gene="MAL7P1.1"
                     /note="Similar to Plasmodium falciparum variant-specific
                     surface protein var-3 SWALL:Q26032 (EMBL:L40609) (3006 aa)
                     fasta scores: E(): 8e-125, 38.49% id in 3673 aa"
                     /codon_start=1
                     /product="erythrocyte membrane protein 1 (PfEMP1)"
                     /protein_id="CAD50783.1"
                     /db_xref="GI:23498714"
                     /db_xref="GOA:Q8I098"
                     /db_xref="UniProt/TrEMBL:Q8I098"
                     /translation="MGNASSSEGEAKTPSLTESHNSARNILEGYAESIKEQASKDAKI
                     HGHHLKGDLAKAVFRHPFSAYRPNYGNPCELDYRFHTNVWHRNAEDRNPCLFSRAKRF
                     SNEGEAECNGGIITGNKGECGACAPYRRRHICDYNLHHINENNIRNTHDLLGNLLVMA
                     RSEGESIVKSHEYTGYGIYKSGICTSLARSFADIGDIIRGKDLYRRDSRTDKLEENLR
                     KIFANIYKELKNGKKWAEAKEYYQDDGTGNYYKLREAWWALNRKDVWKALTCSAPRDA
                     QYFIKSSVRDQTFSNDYCGHGEHEVLTNLDYVPQFLRWFEEWAEEFCRIKKIKLGKVK
                     EACRDDSKKLYCSHNGYDCTKTIRNKDILSDNPKCTGCSVKCKVYELWLRNQRNEFEK
                     QKKKYYKEIQTYTSKDAKTDSNINNEYYKEFYDKLKNEGYETLNKFIKLLNEGRYCKE
                     KISGERNIDFTMTGDKDAFYRSDYCQICPECGVQCSGTTCTPKKVIHPNCKDKETYEP
                     GDAKTTDITVLYSGDEEGDIAQKLQDFCNDKNKENDENYEKWQCYYKSSEINKCQMTP
                     SSHKVPKHGYIMSFYAFFDLWVKNLLIDSINWKNDLTNCINNTNVTDCKNDCNTNCKC
                     FENWAKTKENEWKKVKTIYKNENGNTNNYYKKLNNHFQGYFFHVMKELNKEEKWYKLM
                     EDLKEKIDSSNLKNGTKDSEGAIKVLFDHLKDIAERCIDNNSKDSCPPSVDTKTNPCA
                     KPPGSKPTKSVKQLAEHMQQKAQKLLGTRGGESKLKGDATRGTYNLGGQGNTLNGDIC
                     KITKNHTNDSRPNGEPCTGKDKVKNGFRLKIGTPWTNIVQKKKKKSYKDFYLPPRRQH
                     MCTSNLENLSTSSKGLSNGSFASHSLLGDVLLAAKFEAQKIILVYKNKNNINIRKRIT
                     DPNDQATVCRAIRYSFADLGDIIRGKDMWNINSDAKDLQDRLEKIFKTINEKLPNEIQ
                     KRYTNRENKHLDLRSDWWEANRHQVWRAMKCATKGISNNNCNGIPIEDYIPQRLRWMT
                     EWAEWYCKKQSQEYEKLEEKCGMCTGKGQGDGKDCTQKDKECSPCKKACDAYKKEIEK
                     WEKQWKTVSAIYQILYAKARIVASNGGPGYYNTEVQKKDRSVYDFLYELHLQNGGKKG
                     PPPATHPYKSVNTRDKRDATDDTTPTVYSTAAGYVHQEAHIGDCKEQHVFCDNNGNKE
                     KYAFKNPPNVYVEACKCMTREAPPPPTTPSTPNPCAETGGVHTIKTVTDVAKILQGEA
                     NETMLKNSSNGNDKDESKLKGKAEEGDYSRGGTPSDFNNNLCGITQKHSNAHNDSQQP
                     CYGKDQKRFNVGTEWSFKDNHRKRTHPEAYMPPRREHICTSNLEYLIHKRKKPIIEGD
                     PNKIIHSLLGDVLLAAKYEAENIKKLYEENNNRKDQEGICRAMKYSFADIGDIIRGKD
                     MWIENNDAKRLQTNLKEIFTKIKEKTGGTTYNEDNDPYLKLRADWWEANRAKVWKAMK
                     CKTNGVDITCDSDHTPLDDYIPQRLRWMTEWAEWYCKAQSQEYKKLEEKCSQCKSKGK
                     GGNECYRETKECNDCKQACEEYKRKIKTWADQWKVISNKYEDLYKKAQNPTNAVLKDN
                     KDEKDKNVIDFLTQLQKANNGEKTGVHTVYSTAAGYIHQEARTRECQEQREFCDKKNG
                     IDNTSYAFKDPPHGYATACDCINRSQTEEPKKKEENVESACKIVEEVLSKPRDKTTGG
                     IDHCNPKYYPRKENYPGWNCTPGQFKSGHAGACMPPRRIKLCVINLQYLNEKKSPEEL
                     RKAFIQCAAIETYWLWQKYKKDKNGGVAQAKLNSGTIPDDFKRQMFYTFGDYRDLCLD
                     TDISSKADTSTGVGKVKINIDSVFQKIDITNVEQRKPWWGKNAEAIWDGMLCALSYNT
                     TNKNMDYNAHTKLNPTYGYNAIKSELEDFVNRPQFLRWFTEWSDEFCTERSIKIKELE
                     TKCNDCTVSESGTSDATGNKTCDDKDKCDECKRACTTYKTWLKNWKTQYKTQSKKYFD
                     DKRKELYKSIDDVASSTQAYQYLHAQLKKLCGNADCKCMDGESKETTGQPDNSHDSHM
                     PASLDDEPEEVNGKCNCKVKHRPQPPLALPPPAPSGPPAEDQIEHDNRGRSERGDQGP
                     LPARPPPPPQAAQPPQPKPKRTGEGLGRNLPPADRNTNLSDSEEEDDEDDDEVQEEEE
                     TPPSEAEEGEGHVETEEETKPVKEKTEGAGATEVTKQGSAPTATTPTVEDICATVAKA
                     LKGDKSLNAACALKYGKNNSRLGWKCIPTSGDKTDTSENGAPRRARSAHGGKSDSEKG
                     SICVPPRRRRLYIKKIVDWAESQSKTVTSVNGDGNGSQEVVSVNGASESGGSGSGTES
                     QASDVSQGNGASTSPQVALLHAFVKSAAIETFFAWHKYKVDKEIEEKEKQAAQNHLVQ
                     RKTSENPQKKLEGGEIPEDFKRQMFYTLGDYRDILVGDKTMIEALEKSGDTKIEDISE
                     KIPKILDGENNKAAGGGPKQPNSGKTPQEWWKENAKHIWHGMICALTYNTDSNGKDKK
                     IQQVKATDNTDLFQKLKKDNDYETVSFGASGTGAKSNDDTKLKNFVVRPTYFRWLEEW
                     GEEFCRKQKHKLYIIKKDCRDNKFCSGDGLRCDEKVPDKKDIFKHFDCPSCARHCRSY
                     RKWIERKKTEYEKQESAYSKQKSNYVNGSNGDGGNNNDKEFYTKLETCTKATNFLESL
                     KGQCIGNNNGGTDIKFSNTNITFGSAEDCKPCSEFKVNCENGSCGSAKQKDCPNNTIT
                     SQNIKGLTDQVDMRVSDNTESGFEGDLGICQGAGIFKGIRKDEWKCGDFCGIDICTLE
                     KTNNGKESDKKYIIMKEFVKRWLEYFFEDYNRIQKKLKTCKENGKGSTCIRSCVDEWI
                     KLKKDEWQKINSNYLDQNTKENPEGNNLSSFLEDGPFKNEVDKAIKPCGNLTDFKKSK
                     KCNGTSRSGNSEESTKYDGVICLLDNLKNIIKTCQNVPSGKPDTPCQKSPAPVGDDDD
                     PLEEENPVTQPNICPQTSVEEKKKEEEEKCDEKEEEEEKEEEKDKGDEEVKEEEKDKG
                     DEEEEAEEEEEEEEETDSHIYEDYSDSDAEEDDEDEAVTESLSPSESQPKRLLREFPS
                     PELKNAMLFSTILWMVGIGFAAFTYFFLKKKPKSPVDLLRVLDIHKGDYGTPTPKSKN
                     RYIPYVSDTYKGKTYLYVEGDTDEEKYMFLSDTTDITSSESEYEELDINDIYVPGSPK
                     YKTLIEVVLEPSKSDGNTPGKGDGNTLGDDMVPTTNTFTDEEWNELKQDFVSQYIQSR
                     LPMDVPQYDVSTESPMNIGGNVLDDGMDEKPFITSIHDRDLNSGEEISYNIHMSTNTN
                     NDIPKYVSNNVYSGIDLINDTLSDNKHIDIYDEVLKRKENELFGTNYKKNTSNNSVAK
                     NTNSDPIMNQLDLLHKWLDRHRDICENWGKKEDILNKLNEQWNKDNDGGDIPNDNKKL
                     NTDVSIQIDMDETKGKKEFSNMDTILDDMEDDIYYDVNDENPSVDDIPMDHNKVDVPK
                     KVHVEMKILNNTSNGSLEQQFPISDVWNI"
     gene            join(29398. .29451,29607. .30668)
                     /gene="MAL7P1.2"
     CDS             join(29398. .29451,29607. .30668)
                     /gene="MAL7P1.2"
                     /note="Similar to Plasmodium falciparum RifiN, mal1p4.04
                     rifiN SWALL:Q9NFB3 (EMBL:AL031747) (372 aa) fasta scores:
                     E(): 6.3e-63, 53.66% id in 382 aa"
                     /codon_start=1
                     /product="RIFIN"
                     /protein_id="CAD50784.1"
                     /db_xref="GI:23498715"
                     /db_xref="UniProt/TrEMBL:Q8IC55"
                     /translation="MKIHYTNILLFPLKLNILVNTHKKPYITHHTQKIPTTRSLSECE
                     LFSPQNYDNDPEMKRVMQQFHDRTTQRFQEYDENLKEKRQKCKDRCDKEIQKIILKDK
                     LEKQMAQQFSTLHTDIQSDAIPTCVCEKSLADKVEKGCLRCGSVFGGGIAPSVGLLGG
                     IGEAAISAWKVAALKAAARYAASKGAAQGLAAGIQAGKGVAIKSLEKLGVKYFWTGMS
                     SEILKMNHYKEVANLTDVIYTAKLKVCDELTFDDFVNMCEQFDIKIGVYTEEVKNALL
```

```
                PKYAVPNALNEIVSNAETTAKEVFEAESTRIAAEITEQQTAVINATYSSWQIAITASV
                IAIVVIVLIMMIIYLVLRYRRKKKMKKKLQYIKLLEE"
     gene       complement(join(32811. .33821,33981. .34049))
                /gene="1838.m00006"
                /note="synonyms: PF07_0001, PFG0015c"
     CDS        complement(join(32811. .33821,33981. .34049))
                /gene="1838.m00006"
                /note="Similar to Plasmodium falciparum Pfc0030c, mal3p8.5
                protein pfc0030c, mal3p8.5 SWALL:O97326 (EMBL:AL034560)
                (359 aa) fasta scores: E(): 5.4e-85, 73.25% id in 359 aa,
                and to Plasmodium falciparum rifin, Mal1p4.06 mal1p4.06
                SWALL:Q9NFB2 (EMBL:AL031747) (359 aa) fasta scores: E():
                1.1e-75, 66.01% id in 359 aa"
                /codon_start=1
                /product="RIFIN"
                /protein_id="CAD50785.1"
                /db_xref="GI:23498716"
                /db_xref="GOA:Q8IC54"
                /db_xref="UniProt/TrEMBL:Q8IC54"
                /translation="MKLHYTKILLFFFPLNILLTSYHAHNKNKPSITPNHTRTTTSRL
                LSEYDTESSIYDSDEEIDSVKEIFERQTSRRFEEYEERMITQRQKYKEQRDKNIQKII
                HKDKMEKNLAEKIEKGCLRCGCLAGVAGSIGLFGAVAINIWKPGAITAAIEKALELK
                SAYISATAEAARILAGKEAVIEGLKKMGISTLEGKDLGTYFATTPYEKVASIYVAVNE
                QHFEKCAFGSLKKEFIPVGDGSRDSLFCQSVWKQTSAVSQPRQYISPEEVIEQTVQNI
                VNKAEGPANTAAEVARESATNAIEVQQTNVINAIFMSKQTAIIASVVAILIIVLIMLI
```

```
  Query Match             5.0%;  Score 54;  DB 2;  Length 110000;
  Best Local Similarity   46.3%;  Pred. No. 1;
  Matches  177;  Conservative    0;  Mismatches  205;  Indels    0;  Gaps    0;

Qy        143 TTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAG 202
              || | |   ||  |  |||  | ||    ||  |  ||  |   | | | |   ||
Db      49898 TTAATATTAATAAAATATTTTATATTTTCTATAATATTAATTATATGTATTAGAATAGGA 49957

Qy        203 ACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTA 262
              | ||      | ||  |||    | ||||| || ||  | |  |   |    || |  ||
Db      49958 AAATAAAAAGGATATCATATAAACATGATTATTATAGTACAATTATAATTATTATATATA 50017

Qy        263 TTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCAT 322
              |   ||||| ||  ||| |  || |    |  | | |   || | |   |  |
Db      50018 ATATATTGTTATTATTTTATAATAATACATATTGTATGATTAGGTTAATTCATTATTATT 50077

Qy        323 ATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCA 382
              ||||||    ||    |  || | | | | || || | | ||  ||  ||  | |  |
Db      50078 ATATTATAATGTGTTCATATATAGATATATTGTGTTCAAAATATTTTTGTTGTTACTATA 50137

Qy        383 GTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATA 442
                | |   |  |  |  |  || |  ||||        ||||||||     | | |  |
Db      50138 TAAACAAAAATTATATATATATATATATATATATTACATTTTCTTATATGTGTCCTATAA 50197

Qy        443 GAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTA 502
              ||  || |||| ||||| ||||   || |     | ||||  |  ||    |||     |
Db      50198 TAATATAACACATATTTTGTTTTATTAAAAGGGGGAAAAAAGAAGATAAAATATAATTAA 50257

Qy        503 TTAAAACATTTACTTTCCAAAT 524
                ||    |||  || || ||
Db      50258 CATAATATCTTATATTACACAT 50279
```

```
RESULT 69
AB063522
LOCUS       AB063522               349174 bp    DNA     linear   BCT 20-FEB-2004
DEFINITION  Wigglesworthia glossinidia endosymbiont of Glossina brevipalpis
            DNA, complete genome, section 2/2.
ACCESSION   AB063522 BA000021
VERSION     AB063522.2  GI:25166251
KEYWORDS    .
SOURCE      Wigglesworthia glossinidia endosymbiont of Glossina brevipalpis
  ORGANISM  Wigglesworthia glossinidia endosymbiont of Glossina brevipalpis
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Wigglesworthia.
REFERENCE   1
  AUTHORS   Akman,L., Yamashita,A., Watanabe,H., Oshima,K., Shiba,T.,
            Hattori,M. and Aksoy,S.
  TITLE     Genome sequence of the endocellular obligate symbiont of tsetse
            flies, Wigglesworthia glossinidia
  JOURNAL   Nat. Genet. 32 (3), 402-407 (2002)
  MEDLINE   22297718
   PUBMED   12219091
REFERENCE   2  (bases 1 to 349174)
  AUTHORS   Yamashita,A., Oshima,K., Furuya,K., Yoshino,C., Shiba,T. and
            Hattori,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JUN-2001) Masahira Hattori, Kitasato University,
            School of Science; Kitasato 1-15-1, Sagamihara, Kanagawa 228-8555,
            Japan (E-mail:hattori@genome.sci.kitasato-u.ac.jp,
            Tel:81-42-778-9926, Fax:81-42-778-9926)
COMMENT     On Nov 21, 2002 this sequence version replaced gi:19071886.
FEATURES             Location/Qualifiers
     source          1. .349174
                     /organism="Wigglesworthia glossinidia endosymbiont of
                     Glossina brevipalpis"
                     /mol_type="genomic DNA"
                     /specific_host="Glossina brevipalpis"
                     /db_xref="taxon:36870"
                     /note="equivalent name: Wigglesworthia brevipalpis"
     gene            138. .404
                     /gene="rpsT"
     CDS             138. .404
                     /gene="rpsT"
                     /note="Protein synthesis"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAC24442.1"
                     /db_xref="GI:25166252"
                     /translation="MANIKSSKKRSIQSEKKRKYNSSKKSMIRSFIKKVKNSILKKEK
                     DTINNSFIEMQSILDRYAKKNLIHKNKASRLKSRLSKKIKFIFK"
     gene            complement(424. .1614)
                     /gene="nhaA"
     CDS             complement(424. .1614)
                     /gene="nhaA"
                     /note="Transport and binding proteins"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAC24443.1"
                     /db_xref="GI:25166253"
                     /translation="MKKFFICKSKFFKNPTFSGLLLILFCFLAIFISNTNFWKTYNYI
                     INYPLIKINTHDKNFSLTNIVNDILMTFFFLEIGIEIKHEMLVGSLKNKSRAILPGIA
                     AIGGMIFPALIYNFITKEDSSISSGWAITVATDIAFAVGVLKILGHSIPHSLLIFLLS
                     LAIFDDIGAILIIAFFYSNHIDQYMILLSTLVILTILSINYLRVTCIYIYIIFGILLW
                     ESIFLSGIHSTISGVILGILMPHSSYLSSSKYEKSMAFLKKSLSFLNKYFILPIFAFF
                     NSGINFSNFENLSSSLLPFGIFFGLVLGKPIGVFLFSYLSVKFKLSKLPVGISFKEIA
                     GISFLCGIGFTMSIFISNLAFQNINEKIIYIAKFSILISSIVSSVIGFLFLYFLYKKI
                     KLKN"
     gene            complement(1675. .2799)
                     /gene="dnaJ"
     CDS             complement(1675. .2799)
                     /gene="dnaJ"
                     /note="Protein fate"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAC24444.1"
                     /db_xref="GI:25166254"
                     /translation="MAKSDYYDILGISKNASDREIKTAYKRLAVKFHPDRNPGNLEAE
                     SKFKEIKEAYEVLLDPKKRAAYNQYGHTAFDQGSSNTEFTSNADFSDIFGDVFGDIFG
                     NNRRKRSNRGSDLQYNINLSLEEAVKGISKEITIPKLERCNICSGKGMQPGTSAQNCS
                     TCNGQGQIQMRQGFFSVQQTCPSCRGKCKIIRFPCRNCSGQGRVERSKNISIRIPAGV
                     DNGDKIRLSGEGEAGKLGGSSGDLYVKISVKEHPIFKREEHNLYCEVPINFAMAALGG
                     DIEVPALDGRVKLRIPAETQTGRLFRMRGKGVKSIRSGHQGDLLCRVVVETPVRLSEK
                     QKSLLKELGISFVGRTGMYNSPRSKSFFDGVKKFFDGLTN"
     gene            complement(2874. .4811)
                     /gene="dnaK"
     CDS             complement(2874. .4811)
                     /gene="dnaK"
                     /note="Protein fate"
                     /codon_start=1
```

```
                     /transl_table=11
                     /protein_id="BAC24445.1"
                     /db_xref="GI:25166255"
                     /translation="MEILMGNIIGIDLGTTNSCVAIIENGKVKVIENSEGDRTTPSII
                     AYTEENEILVGQPAKRQSVTNPKNTFFAIKRLIGRKFTDHEVQRDVNIMPYKIVSSEN
                     GDVWLNVKNQKVAPPQISAEILKKMKKTAEDYIGKSITEAVITVPAYFNDTQRQATKD
                     AGKIAGLDVKRIINEPTAAALAYGLDKKTGNRIIAVYDLGGGTFDISIIEIDDVDGEK
                     TFEVLSTNGDTHLGGEDFDSRLINYLVNEFKKEQGIDLRNDPLAMQRLKESSEKAKIE
                     LSSVHQTDVNLPYITADSSGPKHMNIKVTRAKLESLVEELIYKTLEPVKTSLKDAKLK
                     IIDIKDVILVGGQTRMPLVQKKVSDFFGKEPRKDVNPDEAVAIGAAVQGGVLAGDVKD
                     VLLLDVTPLSLGIETMGGVMTTLISKNTTIPTKHSQIFSTAEDNQSAVTIHVLQGERK
                     RSIDNKSLGQFNLDGIAPAMRGMPQIEVTFDIDADGILHVSAKDKNSGREQKITIKAS
                     SGLSENEIDKMLKESEANAELDIKFEELVKTKNQADHLLHSTRKQIKEAKNLPLEKKT
                     EIEKCINELELSIKGEDKKDIEIKIQSLIQISSCLVDFSKNKENLNKEDIIKTNKKEN
                     NNKTNDDVVDAEFEEIKDKKN"
gene                 5858. .6391
                     /gene="fabA"
CDS                  5858. .6391
                     /gene="fabA"
                     /note="Fatty acid and phospholipid metabolism"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAC24446.1"
                     /db_xref="GI:25166256"
                     /translation="MVKKHKFYTKEDLLCSSRGELFGKYGPQLPAPNMLMIDRLVKVT
                     ENGGNYNKGFIKAELDINPNMWFFSCHFIGDPVMPGCLGLDAMWQLVGFYLGWLGGKG
                     KGRALGVREVKFSGQILPTSKIVVYYIHFRRIINRKLFMGMADGEVFCDGKIIYTAND
                     LKVGLFQDITSFKKDFK"
gene                 complement(7084. .7707)
                     /gene="yhgI"
CDS                  complement(7084. .7707)
                     /gene="yhgI"
                     /note="Unknown function"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAC24447.1"
                     /db_xref="GI:25166257"
                     /translation="MFNKYIIINFIYFKLKKIIFYKNMIIISKSAKHYLSKLLSKKKK
                     GTNIKLSIENKDKENLTCQIKYFDKKEENFNYTKFNFIDFKVYVKKNIFIYIKKIKID
                     IESKSLKDEIIIKIFKKRENLKNRISNFIKYKINSKLEYHGGFVELIDIKENMFVILK
                     FFGGCNGCSMAKVTLKEGIEKEIKKNFPNIKGVIDITDHIHSESSFY"
gene                 complement(7730. .7803)
                     /gene="pro-TGG"
tRNA                 complement(7730. .7803)
                     /gene="pro-TGG"
gene                 complement(7876. .7950)
                     /gene="his-GTG"
tRNA                 complement(7876. .7950)
                     /gene="his-GTG"
gene                 complement(7960. .8034)
                     /gene="arg-CCG"
tRNA                 complement(7960. .8034)
                     /gene="arg-CCG"
gene                 complement(8250. .9398)
                     /gene="wecE"
CDS                  complement(8250. .9398)
                     /gene="wecE"
                     /note="Cell envelope"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAC24448.1"
                     /db_xref="GI:25166258"
                     /translation="MKKKNKIIPFNLPVYSGKELKYIDFAIRNKTFSGEGYFSNLCTL
                     WIENKLNIKKVFLTPSCTSSLEMAALLLKINKDDEIIMPSYTFVSTANPFVLRGAKIV
                     FVDIRSDNLNINENLIESAITKKTKAIVVMHYAGISCEMDFIKNISNKYKLWLIEDSA
                     QSMMSKYKNLYLGSLGHIGCFSFHETKNYTSGGEGGAILINDESLIKRASIIKEKGTN
                     RSLFIKGKTNKYVWKDIGSSYTISDIQAAYLWSQLKISEKIQKKREMLWKNYFLSLEN
                     IKLFRIPFFNKNCKHNANTFYLIFNDKKTCNLFIEYMKNNKITTLRHYVPLHISPAGK
                     KFGKFHGNDINTIKTSNNLVRLPLFFNLKNKDQKYIISNIYNFIDKYS"
gene                 complement(9418. .10677)
                     /gene="rho"
CDS                  complement(9418. .10677)
                     /gene="rho"
                     /note="Transcription"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAC24449.1"
                     /db_xref="GI:25166259"
                     /translation="MNLTELKNKLISDLITLGENMGLENLARMRKQDIIFSTLKQHAK
                     IGEDIFGDGVLEILQDGFGFLRSSDSSYLAGPDDIYVSPSQIRRFNLRTGDTISGKIR
                     PPKEGERYFALLKVDEVNYDKPESSRNKILFENLTPLHANSRLKMERGNGSTEDLTTR
                     VLDLAAPIGRGQRGLIVAPPKAGKTMLLQNIAQSISYNHPDCVLIVLLIDERPEEVTE
                     MQRLVKGEVIASTFDEPSSRHVQVSEMVIEKAKRLVEHKKDVIILLDSITRLARAYNT
                     VVPASGKVLTGGVDANALHRPKRFFGAARNMEEGGSLTIIATALIDTGSKMDEVIYEE
                     FKGTGNMELHLSRKIAEKRVFPAIDYNRSGTRKEELLTSSEELQKMWILRRIIHPMNE
                     IDGMEFLMNKLSMTKTNNEFFDMMKRS"
gene                 complement(10812. .11150)
                     /gene="trxA"
CDS                  complement(10812. .11150)
                     /gene="trxA"
                     /note="Energy metabolism"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAC24450.1"
                     /db_xref="GI:25166260"
                     /translation="MNNKEIISVNDEDFEINVLKNKKPVLVDFWAQWCMPCKSMIPIL
                     KKISEKYVNKLEVFKINIEKNSIYQEKYLVKSIPSFFLFFKEKIIDKKIGIVSQKQLE
                     IFIDENLNRL"
gene                 11604. .12788
                     /gene="pgk"
CDS                  11604. .12788
                     /gene="pgk"
                     /note="Energy metabolism"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAC24451.1"
                     /db_xref="GI:25166261"
                     /translation="MKIIKVTDLNLKDKKVLIRSDLNVPIKNGKIMCNMRIIKSLETI
                     NFVIKNEAKCVIVASHLGNPIEEKYDYNFSLTHIVNHMKKIIKYPIRLIQNYLNGFEA
                     KEKEIIVLENVRFNPGERKNDKNLSKKYANLCDILVMDSFGSSHRSESSTTGIIKFAP
                     ISCIGLLFLKEIKYLKKALFNSKRPIVTIFGGSKISTKLGVIEKLSNISENVLVGGGI
                     ANTILFSKGHNIGKSLHDKNNISKIKKFLYKKNIIIPKDFIVTDNINKFSSYKEKSIK
                     EIKNEDYIVDIGKKSCECFIKIIKKAKTIFWNGPLGLIEIKQFRNGTKEIANSVIHSS
                     AKSIIGGGETVLAMRMFNFIDRFSYISTGGGAFLSFIEKQELPVLSELKKFKKYTNFK
                     QN"

  Query Match             5.0%;  Score 54;  DB 1;  Length 349174;
  Best Local Similarity   49.5%;  Pred. No. 0.85;
  Matches  220;  Conservative    0;  Mismatches  220;  Indels    4;  Gaps    3;

Qy       132 TATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATC 191
            ||  | |||| || | |  ||      |   ||  |||   |   |  ||   || |  ||
Db    300260 TAATACTTTGTTTAAAACTTAACTTTATATAATAATATTTACTTTATTAATTAAATAATA 300319

Qy       192 TCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTA-AAAGACCTTACA 250
            | || | | | |   | ||   ||||||     || ||||||  | ||| |  |  ||
Db    300320 TGATGTAATATATAAAATATTTCAAATATAAATATTTTATTTTTAAATAAATAAATCTCA 300379

Qy       251 GACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCA 310
            || |      |   | | || | ||| |  || | |  |||  ||||   | ||  || |
Db    300380 GAAACTATACATATAAATTATATTTTGAATGATAATTAAATTTATAGCTATATTGATAAA 300439

Qy       311 GAGAAAGATCATATATTAGCCAGTTGAAAATTAAAC-ACAAAATGAGTGCATATTACATT 369
               | |     | || |||    | || ||||| |||  ||||||          ||||  |
Db    300440 ACTATATTAAAGATTTTATTAAATTAAAAATAAAAATACAAAACATTAATTTATTTTAAA 300499

Qy       370 ACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTAT 429
               | ||| |   |     |  |||| | | ||  | |   | | | |     | ||| |
Db    300500 TGTAAATATAAAATCTTTACATAAATAATTAATATTATCATTTGCAATTAAATAACTTTT 300559

Qy       430 CGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAAT 489
               | |  |||| ||||||      || | |  ||  |||  ||   |  ||  |||||
Db    300560 AAAATTATTAATAAAAATTATATTTTATATAACATT--AAAATATTAAATTAATTATAAT 300617

Qy       490 TTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGTCATTCAGGAGTATGGAAG 549
               |   | | |  ||| |||         |||  |||||||  | | ||  |     |   |
```

```
Db     300618 TAAATTTATTTTATAAAATATAAATTTTTTAAAATATTTTTTTATAAAAATAAAAAAATA 300677

Qy        550 TATCGATGGCTTCTTTAAAATGAA 573
              ||  || | || || ||| | ||
Db     300678 AATAAATTGATTATTAAAATTTAA 300701


RESULT 70
AL683870
LOCUS       AL683870               162377 bp    DNA     linear   PRI 12-AUG-2002
DEFINITION  Human DNA sequence from clone RP11-261P4 on chromosome X, complete
            sequence.
ACCESSION   AL683870
VERSION     AL683870.15  GI:22213668
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 162377)
  AUTHORS   Whitehead,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 13, 2002 this sequence version replaced gi:22003159.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome X, constructed by the Sanger Centre Chromosome X Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/ChrX
            RP11-261P4 is from the library RPCI-11.1 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6.
FEATURES             Location/Qualifiers
     source          1. .162377
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /clone="RP11-261P4"
                     /clone_lib="RPCI-11.1"
ORIGIN

  Query Match             4.9%;  Score 53.8;  DB 9;  Length 162377;
  Best Local Similarity   46.5%;  Pred. No. 1.1;
  Matches  246;  Conservative    0;  Mismatches  277;  Indels    6;  Gaps    2;

Qy         51 CAAATGTTACAATGAGATATGAGTTTTAAAGCCAGATCTTCCTTATCTCTCTGCCCCATC 110
              || || ||  | | | |||| | ||  | |      |   |   ||| | | ||    ||
Db      51891 CATATATTTAATTAATATATAATTTACATATAATCACACTTTATATTATTATGTAATATG 51950

Qy        111 TCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAG 170
              | |   | |  | |   |    | |   | || | | |||||    ||||   ||
Db      51951 TATTAATATGCAGTTACATATATTAATCATATTTAATACATATTACATATTATCCATATA 52010

Qy        171 TTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGA 230
              || |||  ||| || |  || | |   || |   ||  |||      ||||| ||  || |
Db      52011 TTCATATATTTCTATATTATATAAATATATAATTATATTTTATAATATATAATCATTTTA 52070

Qy        231 TTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAA 290
              |  || || |     ||| | | |     ||   |  || ||  |   |  | | ||
Db      52071 TAATTATATATATAATTATATATAATTTATATTTATATAATTGTTTATAATTATATATAA 52130

Qy        291 GGTATAAAGAAATGCATTCAGAGAAAGATCA-TATATTAGCCAGTTGAAAATTAAACACA 349
              |  |  |  | | |  | | || | ||||      | || | || |     |
Db      52131 TTATATATCATTTATAATTAATTATATATAATTATAAATTATATTTATATATAATTATAA 52190

Qy        350 AAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGG 409
              | ||  ||   |||    |||  ||     |  | | |||  || ||| |  |   ||
Db      52191 ATATATGTAATTATATATTTATATATAAATAAAATTAAATATATAAAATATATAAAATTA 52250

Qy        410 TATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAA 469
              |||| |   |  |  ||||      | | | | |     |||||   | |
Db      52251 TATATTTTATATATAATTATAAATAATTATAAATATAAATTATATAATTTATATATAATT 52310

Qy        470 ATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTT-----ACTTTCCAAAT 524
              ||||||    | |||    | || |||   |||||  |||| ||     || |  || ||
Db      52311 ATAATCGATTATAACTACCTATTATATAATCTATTTTAACAGTTATATAACCTATCATAT 52370

Qy        525 ATTCTGTCATTCAGGAGTATGGAAGTATCGATGGCTTCTTTAAAATGAA 573
              | ||| | ||    |  |||  |  || |||||  | | | | ||| ||
Db      52371 AATCTATTATAACAGTATATATTATAATAGATGGTATATATTACATTAA 52419


RESULT 71
AE014842/c
LOCUS       AE014842               253151 bp    DNA     linear   INV 07-OCT-2002
DEFINITION  Plasmodium falciparum 3D7 chromosome 11 section 7 of 8 of the
            complete sequence.
ACCESSION   AE014842 AE014186
VERSION     AE014842.1  GI:23496321
KEYWORDS    .
SOURCE      Plasmodium falciparum 3D7
  ORGANISM  Plasmodium falciparum 3D7
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 253151)
  AUTHORS   Gardner,M.J., Hall,N., Fung,E., White,O., Berriman,M., Hyman,R.W.,
            Carlton,J.M., Pain,A., Nelson,K.E., Bowman,S., Paulsen,I.T.,
            James,K., Eisen,J.A., Rutherford,K., Salzberg,S.L., Craig,A.,
            Kyes,S., Chan,M.-S., Nene,V., Shallom,S.J., Suh,B., Peterson,J.,
            Angiuoli,S., Pertea,M., Allen,J., Selengut,J., Haft,D.,
            Mather,M.W., Vaidya,A.B., Martin,D.M.A., Fairlamb,A.H.,
            Fraunholz,M.J., Roos,D.S., Ralph,S.A., McFadden,G.I.,
            Cummings,L.M., Subramanian,G.M., Mungall,C., Venter,J.C.,
            Carucci,D.J., Hoffman,S.L., Newbold,C., Davis,R.W., Fraser,C.M. and
            Barrell,B.
  TITLE     Genome sequence of the human malaria parasite Plasmodium falciparum
  JOURNAL   Nature 419 (6906), 498-511 (2002)
   PUBMED   12368864
REFERENCE   2  (bases 1 to 253151)
  AUTHORS   Gardner,M.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-2002) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .253151
                     /organism="Plasmodium falciparum 3D7"
                     /mol_type="genomic DNA"
                     /isolate="3D7"
                     /db_xref="taxon:36329"
                     /chromosome="11"
```

```
repeat_region   complement(73. .101)
                /note="AT_rich"
                /rpt_type=tandem
repeat_region   173. .279
                /rpt_type=tandem
                /rpt_unit="(taaaaa)n"
repeat_region   complement(302. .328)
                /note="AT_rich"
                /rpt_type=tandem
repeat_region   complement(393. .466)
                /rpt_type=tandem
                /rpt_unit="(ta)n"
repeat_region   393. .442
                /rpt_type=tandem
repeat_region   440. .466
                /rpt_type=tandem
repeat_region   545. .578
                /rpt_type=tandem
repeat_region   complement(545. .578)
                /rpt_type=tandem
                /rpt_unit="(ta)n"
repeat_region   complement(580. .673)
                /note="AT_rich"
                /rpt_type=tandem
repeat_region   762. .795
                /rpt_type=tandem
repeat_region   complement(762. .795)
                /rpt_type=tandem
                /rpt_unit="(ta)n"
repeat_region   complement(869. .955)
                /note="AT_rich"
                /rpt_type=tandem
repeat_region   complement(956. .975)
                /rpt_type=tandem
                /rpt_unit="(a)n"
repeat_region   complement(983. .1040)
                /note="AT_rich"
                /rpt_type=tandem
repeat_region   1094. .1121
                /rpt_type=tandem
                /rpt_unit="(a)n"
repeat_region   complement(1149. .1221)
                /note="AT_rich"
                /rpt_type=tandem
repeat_region   1239. .1281
                /rpt_type=tandem
repeat_region   complement(1243. .1319)
                /rpt_type=tandem
                /rpt_unit="(ta)n"
repeat_region   complement(1323. .1379)
                /note="AT_rich"
                /rpt_type=tandem
repeat_region   complement(1571. .1605)
                /note="AT_rich"
                /rpt_type=tandem
repeat_region   complement(1621. .1699)
                /note="AT_rich"
                /rpt_type=tandem
repeat_region   complement(1820. .1872)
                /note="AT_rich"
                /rpt_type=tandem
repeat_region   complement(1921. .2010)
                /note="AT_rich"
                /rpt_type=tandem
repeat_region   2105. .2135
                /rpt_type=tandem
                /rpt_unit="(a)n"
repeat_region   complement(2135. .2176)
                /rpt_type=tandem
                /rpt_unit="(ta)n"
repeat_region   complement(2226. .2248)
                /rpt_type=tandem
                /rpt_unit="(taaaa)n"
repeat_region   complement(2289. .2362)
                /note="AT_rich"
                /rpt_type=tandem
repeat_region   2372. .2396
                /rpt_type=tandem
repeat_region   complement(2372. .2396)
                /rpt_type=tandem
                /rpt_unit="(a)n"
repeat_region   complement(2402. .2424)
                /note="AT_rich"
                /rpt_type=tandem
repeat_region   complement(2438. .2620)
                /note="AT_rich"
                /rpt_type=tandem
repeat_region   2636. .2671
                /rpt_type=tandem
repeat_region   complement(2636. .2671)
                /rpt_type=tandem
                /rpt_unit="(a)n"
repeat_region   2680. .2837
                /rpt_type=tandem
                /rpt_unit="(tataa)n"
repeat_region   complement(2763. .2865)
                /rpt_type=tandem
                /rpt_unit="(ta)n"
repeat_region   complement(2932. .2954)
                /note="AT_rich"
                /rpt_type=tandem
repeat_region   2970. .3008
                /rpt_type=tandem
                /rpt_unit="(a)n"
repeat_region   complement(3028. .3145)
                /note="AT_rich"
                /rpt_type=tandem
repeat_region   complement(3189. .3238)
                /rpt_type=tandem
                /rpt_unit="(tataa)n"
repeat_region   complement(3250. .3403)
                /rpt_type=tandem
                /rpt_unit="(ta)n"
repeat_region   3425. .3520
                /rpt_type=tandem
repeat_region   complement(3425. .3520)
                /rpt_type=tandem
                /rpt_unit="(a)n"
repeat_region   3556. .3580
                /rpt_type=tandem
                /rpt_unit="(a)n"
repeat_region   complement(3584. .3641)
                /note="AT_rich"
                /rpt_type=tandem
repeat_region   complement(3649. .3809)
                /rpt_type=tandem
                /rpt_unit="(ta)n"
repeat_region   complement(3842. .3884)
                /rpt_type=tandem
                /rpt_unit="(a)n"
repeat_region   3845. .3884
                /rpt_type=tandem
repeat_region   3897. .3944
                /rpt_type=tandem
                /rpt_unit="(tataa)n"
repeat_region   complement(3946. .3999)
                /note="AT_rich"
                /rpt_type=tandem
repeat_region   4000. .4024
                /rpt_type=tandem
repeat_region   complement(4000. .4024)
                /rpt_type=tandem
                /rpt_unit="(a)n"
repeat_region   complement(4026. .4104)
                /note="AT_rich"
                /rpt_type=tandem
```

```
     repeat_region   4163. .4226
                     /rpt_type=tandem
                     /rpt_unit="(taaaa)n"
     repeat_region   4232. .4267
                     /rpt_type=tandem
     repeat_region   complement(4232. .4267)
                     /rpt_type=tandem
                     /rpt_unit="(ta)n"
     repeat_region   complement(4339. .4359)
                     /rpt_type=tandem
                     /rpt_unit="(ta)n"
     repeat_region   4380. .4407
                     /rpt_type=tandem
     repeat_region   complement(4380. .4407)
                     /rpt_type=tandem
                     /rpt_unit="(a)n"
     repeat_region   complement(4408. .4575)
                     /note="AT_rich"
                     /rpt_type=tandem

  Query Match             4.9%;  Score 53.8;  DB 3;  Length 253151;
  Best Local Similarity   46.8%;  Pred. No. 0.98;
  Matches  240;  Conservative    0;  Mismatches  267;  Indels    6;  Gaps    2;

Qy         23 TAACTTTCACTCTCTATGTTCTTATAAGCAAATGTTACAATGAGATATGAGTTTTAAAGC 82
              ||  | ||| |    || |  | | ||   |   |  || ||  |||| | |||| |
Db       2968 TATGTATCAATAAGAATATAATAAAAAAATATATTATCATTGGAATATAAATTTTTACAA 2909

Qy         83 CAGATCTTCCTTATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGG 142
                | |  |  | |    | ||     ||   | ||  |  |      |||| |   |
Db       2908 TGGTTTATGATAAAAATTATGATTTCTCAGAAATTAATCGAAGTGTACATATAAACATAC 2849

Qy        143 TTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAG 202
                |  |||  || || ||   |   | |||||    |  ||  | || | || |||
Db       2848 ATATTAAAGGTTTATAAT--ATATATATTTAATTAAATATTATTATATATTATATTATGC 2791

Qy        203 ACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTA 262
              | |||  | | | ||||||      |  |||||||| ||| |||  |   | ||     |
Db       2790 ATATTAATATATAAATATATGTATATATTTTTTTTTGAAATACCACATTTATATTATTATT 2731

Qy        263 TTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCAT 322
              |  ||  |||||   |  | | || | |   |||  | | |||  |  |   ||| | |
Db       2730 ATATTTATTTTTTTTATATTCCACAGATATTATTTATATATGTGTATATTAAAACAACGA 2671

Qy        323 ATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCA 382
              | |  |   |    ||||  ||| | ||||  |    ||||  | |      |   |  |
Db       2670 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATTATATACCTTATACATAATATATA 2611

Qy        383 GTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGC----A 438
              |  |   |  |  |   | | | | |     | | |||| | | | |    |
Db       2610 TATATATACATGATATTTATTAATAATTCTTACAAGAAATGTTTTATTGAAATACTATAA 2551

Qy        439 AATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGT 498
              || || |    | |  | |    || | |||||    |  ||| ||| || | ||| |
Db       2550 TCTATAATTATTTAAAGTTATATCATTTGTAATAAAACAATTAAATATATTTATATATAT 2491

Qy        499 ACTATTAAAACATTTACTTTCCAAATATTCTGT 531
              |  || |||    ||  |||  |||| |  |||
Db       2490 ATAATAAAATACCTTTATTTAAAAATTTAATGT 2458


RESULT 72
AC116977
LOCUS       AC116977              302156 bp    DNA     linear   INV 25-FEB-2004
DEFINITION  Dictyostelium discoideum chromosome 2 map 5515173-5817331 strain
            AX4, complete sequence.
ACCESSION   AC116977
VERSION     AC116977.2  GI:28828573
KEYWORDS    HTG.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 302156)
  AUTHORS   Gloeckner,G., Eichinger,L., Szafranski,K., Pachebat,J., Dear,P.,
            Lehmann,R., Baumgart,C., Parra,G., April,J.F., Guigo,R., Kumpf,K.,
            Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and
            Noegel,A.A.
  TITLE     Sequence and analysis of chromosome 2 of Dictyostelium discoideum
  JOURNAL   Nature 418 (6893), 79-85 (2002)
  MEDLINE   22092622
   PUBMED   12097910
  REMARK    The Dictyostelium Genome Sequencing Consortium
REFERENCE   2  (bases 1 to 302156)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-APR-2002) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   3  (bases 1 to 302156)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   4  (bases 1 to 302156)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   5  (bases 1 to 302156)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
COMMENT     On Mar 4, 2003 this sequence version replaced gi:20042932.
            CDS predictions from GeneID do not necessarily reflect true genes.
            Further Information is available from IMB Jena, Department of
            Genome Analysis
            (http://genome.imb-jena.de/dictyostelium/)
            and the Univerity Colonge, Institute for Biochemistry I
            (http://www.uni-koeln.de/dictyostelium/project.shtml
            Funding
            Agency : Deutsche Forschungsgemeinschaft (DFG).
FEATURES             Location/Qualifiers
     source          1. .302156
                     /organism="Dictyostelium discoideum"
                     /mol_type="genomic DNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /chromosome="2"
                     /map="5515173-5817331"
     CDS             join(<1385. .3437,3519. .4001)
                     /note="GeneID exon scores (in order of location ranges):
                     178.05, 30.36 - GSCJ_ID dd_01590"
                     /codon_start=2
                     /product="similar to Dictyostelium discoideum (Slime
                     mold). Putative transcription factor"
                     /protein_id="AAO51288.1"
                     /db_xref="GI:28828685"
                     /translation="NNNNNFDPITSLTNNNNNKIISPLNSNRQQPNSNTNNNSNNKNN
                     NKSNLLDIILDNNNLENNIDNNNNNNNNNNNNNNNNNNNSKSIINLIDNNNNNNKNN
                     NNNNNNNNILKIPIDYTPVDKNTESIFKGFSPFSFKQQQPIDTSDQQQKETKKLQLNI
                     NSNNKNNNNIINNSNNNNNINNIESNNNNLNSASTSPIPTPTSLTAVSVPSSSPLPSP
                     TSPNTILQSPTTTTTTTTTTSNSSVSPQPTQINKTDKYLLPMSAPRSCEFCGSSQTPT
                     WRRGPSGKGSLCNACGIKWRLKGKDGIFKPSQKQQNRQKPIMSAQKQPKQQQNQPQQQ
                     QPQQPQQPQQPQQQNQQQNQQQSIQSQPQQRQNNHLILNQQTINQHLQMLSKQHQQQA
                     SQSMFQQQQLHNNQHSNDSNYNSYNNNNNNNNNNYSQHEQLPPISSYNSMNNNNNSNN
                     SNNNNNNNNYSQIPQYGSDDNVSVPNQSPDSYKPSKKRKPEFIHTDTSGNPIEFEKGY
                     YCKYCKKTWPQSSFRNSQQFGAHCSNCSRKPKGELEELISIKKKLSKKKSKEGNYYQE
                     EDGAVPYWDMNRYNNNNNNNGINNNNNNNNNNNNNNNSNNNGSYNNNNSNNNNNNRI
                     EPIVNNNINNFNSINNNNSNNNNNNNSTSQNAPLLQRLIHVVENQLLEPNELSSIRED
                     IDIMKNDLILRKKKRIAQLNDLKQKANNDVQILKANIEKGSNQRSNESKSFITDLKQE
                     IDKRLNQYTHIYIPPDHHLVPSSTNYLSIDSLQNSSSSNNYGTIHNSNNNNNNNNNNN
                     KRSLVNLSPELIHSPLKKPMVGGYRDEQGMKHFSIDGDNLSLPSIS"
     CDS             complement(4147. .4578)
                     /note="GeneID exon scores (in order of location ranges):
                     21.45 - GSCJ_ID dd_01589"
                     /codon_start=1
```

/product="hypothetical protein"
/protein_id="AAO51177.1"
/db_xref="GI:28828574"
/translation="MDSNSTISLMTDSKAELVAYYSGRVVNKSSKDIIVIAVPIIAGT
IFFTIICLGILFRKRISSIIQKNKSNKEKYNINDIKIPQSPSTIITTNSKNNNSCNYL
NSSDISMVLKNSNDNNDNNIENAFEIIEIPHLDIELIKFKK"
CDS complement(4816. .9204)
/note="GeneID exon scores (in order of location ranges):
317.53 - GSCJ_ID dd_01588"
/codon_start=1
/product="similar to Dictyostelium discoideum (Slime
mold). Hypothetical 12.5 kDa protein (Fragment)"
/protein_id="AAO51178.1"
/db_xref="GI:28828575"
/translation="MLSLKMINNYNCERLNKSSSEEVLFWKVFGNVYLLKLIFHHIHT
TEWVVYDAYKKINVQNRWKFKDVYSLDWMILKKQYPLLINKIKFGDFILFDTISIKVL
FSAIPKLKKQPPPPPPLTTTPSSNSTTKTLTLKAVKEKEKEEKRKEIEKEIENEKEIE
TLKYIIKLLLEKRRLEFEICDLTQLAVKTNCLDTLSLLLCEQYSTIVYPSTLEWAIEN
CDCSIVKQLLTYNNKQLLTEDIKKKAFFLASINTHYQNEMINLILDESLYKLPPPTPT
PTPPPPPPTSSSPQHQKQIDQLSKNQNKKIKPDSNSDYHNYGNDDNENDNDDDDDDDY
IFGDDYEEDNRDILDSYYDDYFGNHETLSEIERERVTEYPYIKFEKFLTITSIDTMSK
LMDLKVVKPIARQHYQMDLFKFEFSLNQLLVYLRIYLSLFKKDYYFYSKLYETLEIKI
NNEIINNEKLDQSAKLKELRRLVVIHSRQETFYKQYILHYEESIDEFHRDFPYDRIDS
SNLDSDCKKFLALLRYMGEFVAHKMPNRHKYFLEDLKKMKSSTSNDRWCPLLFKAIEC
VPGQLENFNITHKSIIDWIKNTNSIVINTAVCTSFTMVSYSLIELKHYAITEAFTNLD
MEDNKKSLSYFCKKNWFVKFYLKEMEFCNVERLREIEDFLTKNCLENPSSPININVDS
YGSFDFKIDQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQKQQQKQPNQNIVDDQFDSDV
FEKKIVALIEYLNKTNFKFIQHGMHAFLKKVIEHLPREISSTLVVNLNPLKFYLNSWL
TASIITLNLKMVEFLINNLEAIEFEEQDFTRIQSVERNLFNSSLNLYHNHDEILVFDF
VMSIFKTIVNKWRCTVNNDHQKEGGSGSGGSSSGTIKSLPESCIVILNYFFEMIIQTKY
ITIEQVRMAHQPLCLNSIPIKHSLFHAFYYLKIRSAKFIKYILSRPGIETFFDVTMDS
RYNSTREYFTERCFDIRDHLKYEMFTFDYVLGGESSSSGGASSTAASFDEILQNLFDQ
LSNDICFKENIVHNFYQLSKNADLSFCLDIRRVDLFFKHLDIIQSHLKDTTSEDYVLK
SNDPFPIQEYSYNSPMTTILPTSQLFSGGREFPKPYLPFCERNTIQYDHCIFKPLTII
NSFDYSEKVLDLIFMVITDLELIQKFLEYNFHVIQFEDTLFSKSLHWSRFDISTFLLA
NFDIQFFSFPRFALDTHYSFYQEAVFILNNHFDKVRYPNKFKYDLNVFNDTTFPIAFL
SVLIDFLSKRKNSNHYEKFMNIETRKDALIKLLRIQRDTYGASNVDDDMDADDINFET
SFQKEKQCRAILIQPETQSLLKSIYDDFKIELDKKREKEKLTGEAEYLDSDDNLFLSA
FIVGDIKLCDLLLCHYYHPIQFKITKSAITKAIKDQKIHIIKYYYNGHHLNLLNNNIK
NDKTLSDHLLKELLNHNDYKFDLNWL"
CDS complement(join(9977. .10749,10855. .11323,11395. .11661,
11921. .12127))
/note="GeneID exon scores (in order of location ranges):
79.65, 53.78, 36.02, 15.96 - GSCJ_ID dd_01587"
/codon_start=1
/product="similar to Homo sapiens (Human).
Phenylalanyl-tRNA synthetase beta-subunit (Fragment)"
/protein_id="AAO51179.1"
/db_xref="GI:28828576"
/translation="MKKNETGVEDLTVSDDVIYKIDVSANRYDLLCLEGIARALNVYN
HKASIPKYQIVPPKNSHEKLYISKEVESVRPVIVAGILRDITFTQESYDSFIDLQEKL
HANICKKRSLVSIGTHDLDTLSGPFYYKALAPKDIKFVPLSQTKEYNAEELFKFYDES
SSHLKKFLPIIKDSPVYPVIYDSKNVVCSLPPIINGEHSKIKLSTKNVFIEVTANDRT
KANIVLNTMLTMFSEYCKQPFTMEQVEVIDADGKSTGLYPQIQEKQINAQVDYINKSA
GINITPNDMVTLLKRMSLQSKLSDDEKSIIVDVPVTRSDIMHACDIMEDVAIGYGYDN
LKKEIPNCNTIGRVQPINKLSELLANEIALAGFTEIMTFVLCQNRDNFTALNKADDGS
SVKISNAVSEEFTEVRTNLVSTLLKSVSANKAAPLPLKMFEISDVSIKGSLGNKDLSD
PNSNNSDVGAYNKRMLGAIYCNQSAKIEVIHGLLDRIMLVLDIKLDATRSSNKGYYLE
LSNDKLFLPGTGINVIVNGKRVGHMGIVHPLVLKNYSCSFPCTILELELTIDTMAHNV
LLREN"
CDS 13521. .13895
/note="GeneID exon scores (in order of location ranges):
1.38 - GSCJ_ID dd_01586"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAO51180.1"
/db_xref="GI:28828577"
/translation="MNQQELIEEINKIESCSLVFERFFGYLPNDIKERKQYILKFNKT
LKDEKRKLKCKKTFKGEKKNKKQTFLKHHQSDDHSENKVYKSKKLEKKIQQLNKKKQL
IDTKINFLKEIKESNKINKSKI"
CDS complement(join(13981. .17874,17997. .18239))
/note="GeneID exon scores (in order of location ranges):
419.33, 16.26 - GSCJ_ID dd_01585"

/codon_start=1
/product="similar to putative DEAD/DEAH box helicase
[Schizosaccharomyces pombe]"
/protein_id="AAO51181.1"
/db_xref="GI:28828578"
/translation="MIELNIDSFLEELNNLENITSTNTTPVNKSNSSIKNNSNNTNVE
SFDLPNWYVNPTGLSAQPEITPIAPQSSIISPTIPNNCIQLFNFSDKEEILSDIRREY
LDFSIESHKLNDDDEQKRQSYLIWPRNIDTDSWDFKLLAERTLSIPCSTVELVIDKIQ
YSISITNYHEVFLNTEKTSSNSTSLDRPWSNSISKDNFHRGNTSNLPFLPGGIKSPNN
NSTNSTNNNNQTPIDWVQFWHSKSLLVKPPGLDNGILDFNENINKQDNKQQQQQQQQQ
QQQQQQQQQQQEKEKEEEEEVNLRYTLKDLEREIEEERKREEAEEAEVYEDDDQEDDL
DIEEEELDELDDIEKTKEEIDDIIKSPPQPTSSSLSTSSSSSSTTLKDKNKIEIEIEE
ETKRKEEESKKEKQWAFLEQKEITSPLSDLITNPAIEYPFDLDSFQKQAIVHMEQGES
VFITAHTSAGKTVIAEYAIAMAAKNMTRAIYTSPIKALSNQKFRDFKNTFNDVGLITG
DVSISPSSSCLVLTTEILRSMLYKGADLIRDIEWVIFDEVHYLNDLERGVVWEEVIIM
LPPYVKMVFLSATVSNPLEFAQWIGRTKQLPIYVIGTTKRPIPLEHYIHTPSNELFKI
VDSNRNFLPSGYNSAYNSLYKTDNNKGGGGGGRGGHNQQQQRGGSGNNISGWSKLIN
TLKDKQQLPVIIFSFSKNKCQEYAGSLGQSVNLTQGNEKSQIRVFIEQSLGRLCEDDK
SLPQILQMKELLERGIGVHHGGLLPIVKELVEILFSKSLVKVLFATETFAMGVNMPAK
TVVYSSTRKHDGITFRDLIPGEYTQMSGRAGRRGLDKVGTVIITYWKDIPEQATIESM
ILGTPSRLNSQFRLTYNMILNLLRVPDFKVEDMIKRSFSEFSSQKELPGIEKQIEKLQ
QQQKQLAQVDCILGEPDIDHYYKLFSQAKKYNQQIQNIILNLPNDNHLSIGRVVVLST
SSNIEKEVNDDDDDDDNNDDDEIFENEKYTIGLILECKSTLIKQYTNTKVDRVFKIF
SLKLKIENDIVLCVGNEIITSSGGKEIMKVCNEKLQLKQSDLKSILAGAESGGESEIL
INSLEQQLLRLIEEYPLPLGPKSIDPIKQLKLKDVDFVSTYDHLQSIEKLIPESKCHK
CPRLHDHYEQTEKRYQLQYAIRDAKYTASDENLKLMPQFNIRLDILHELGYIDDENTV
TLKGRVSREINTCEDLVITELIFENAFINLEPSEVVSVLSCLIFQEKDAVQPSLTPRL
EEAKQNLIKTAEKTYKVESDKGLDVVPDDKLETTLKFGLMQVVYEWARGTPFNDICTL
TNVLEGSIVRAITRIGETCQEVRNAARVIGDTKLLQKMEEAMRLIKRDIVFTSSLYVN
"
CDS complement(18767. .19162)
/note="GeneID exon scores (in order of location ranges):
41.07 - GSCJ_ID dd_02896"
/codon_start=1
/product="similar to Involved in global regulation of
transcription; Snf5p [Saccharomyces cerevisiae]"
/protein_id="AAO51182.1"
/db_xref="GI:28828579"
/translation="MEIINYLPPAACENCPQHKYCTRCYFQQLQFMSALTIHQKEISQ
PKKRKNTGHDEPNSNFYESDIESDDDDEQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ
QQQQQQESESRTVTGIQIEFDDEIEFYH"
CDS complement(join(20381. .22148,22235. .23160,23330. .23685,
24790. .24793))
/note="GeneID exon scores (in order of location ranges):

Query Match 4.9%; Score 53.8; DB 3; Length 302156;
Best Local Similarity 47.0%; Pred. No. 0.95;
Matches 166; Conservative 0; Mismatches 187; Indels 0; Gaps 0;

```
Qy      183 TAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAG 242
            || ||    |||||| ||||   | ||| |  | ||  |     |||  |  |  ||
Db   133298 TACAATTATTCATCTGAAAGCAGTAGTTATTATTATTAATAATATTGTATAATCAAACGT 133357

Qy      243 ACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAA 302
            |  | | | | | | |   |  |                    |||||  ||   |  ||
Db   133358 AATTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACCAACCAACAAAAATTTAATGATCAA 133417

Qy      303 TGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATA 362
            || ||  |  | ||||  | | |  |     | ||||  ||| | ||||| |   |
Db   133418 TGAATCAATTGTAAGAAAAAAAAAAAAAAACCATCAAAAAAAAAAAAAAAATTAAATTAAG 133477

Qy      363 TTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCT 422
             || | |  |  |   || | | ||| |||  ||  |||  ||| |||| |  |  |
Db   133478 ATATAATTTTCCAATCTGTAATAAAAAGTACCAAAACAATTTAATGGTAAAAAAAATAAA 133537

Qy      423 TTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAA 482
                |    |||||  |||  |  | | | |  |  |||||  ||     |
Db   133538 AATAAAAATAAAATAAAATAAAAAAAAAATAATATTATAAAAATAAAAATAGACTTTATTA 133597

Qy      483 ACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGTCATT 535
            | ||| |||||||     | ||  ||   ||  |||   | |||| |   |||
Db   133598 AAATAGTTTTTTAACCTTTTTTGGAAATATTTGTTTTTTATTATTTTTAGATT 133650
```

```
RESULT 73
CR394534/c
LOCUS       CR394534              156975 bp    DNA     linear   HTG 03-MAY-2004
DEFINITION  Danio rerio clone CH211-276A17, *** SEQUENCING IN PROGRESS ***, 7
            unordered pieces.
ACCESSION   CR394534
VERSION     CR394534.2  GI:46981158
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 156975)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-MAY-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On May 3, 2004 this sequence version replaced gi:46935091.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zC276A17
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 154476 bases at least Q40
            Consensus quality: 154894 bases at least Q30
            Consensus quality: 155360 bases at least Q20
            Insert size: 156375; sum-of-contigs
            Insert size: 163877; 6.0% error; agarose-fp
            Quality coverage: 9.23x in Q20 bases; sum-of-contigs Quality
            coverage: 8.88x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    80940: contig of 80940 bp in length
            *    80941    81040: gap of 100 bp
            *    81041    92335: contig of 11295 bp in length
            *    92336    92435: gap of 100 bp
            *    92436   125282: contig of 32847 bp in length
            *   125283   125382: gap of 100 bp
            *   125383   132339: contig of 6957 bp in length
            *   132340   132439: gap of 100 bp
            *   132440   147983: contig of 15544 bp in length
            *   147984   148083: gap of 100 bp
            *   148084   154842: contig of 6759 bp in length
            *   154843   154942: gap of 100 bp
            *   154943   156975: contig of 2033 bp in length.
FEATURES             Location/Qualifiers
     source          1. .156975
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-276A17"
                     /clone_lib="CHORI-211"
     misc_feature    1. .80940
                     /note="assembly_fragment:01281
                     fragment_chain:1
                     clone_end:SP6
                     vector_side:left"
     misc_feature    81041. .92335
                     /note="assembly_fragment:00209
                     fragment_chain:1"
     misc_feature    92436. .125282
                     /note="assembly_fragment:00647
                     fragment_chain:1"
     misc_feature    125383. .132339
                     /note="assembly_fragment:00110
                     fragment_chain:1"
     misc_feature    132440. .147983
                     /note="assembly_fragment:00384
                     fragment_chain:1"
     misc_feature    148084. .154842
                     /note="assembly_fragment:00043
                     fragment_chain:1"
     misc_feature    154943. .156975
                     /note="assembly_fragment:00007
                     fragment_chain:1
                     clone_end:T7
                     vector_side:right"
ORIGIN

  Query Match             4.9%;  Score 53.6;  DB 2;  Length 156975;
  Best Local Similarity   34.9%;  Pred. No. 1.2;
  Matches  185;  Conservative    0;  Mismatches  343;  Indels    2;  Gaps    1;

Qy         27 TTTCACTCTCTATGTTCTTATAAGCAAATGTTACAATGAGATATGAGTTTTAAAGCCAGA 86
              |||   | | | | || || |       | ||    |    | |  |||| |||   | |
Db     148428 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAAAAAAAAAA 148369

Qy         87 TCTTCCTTATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGA 146
                ||  |||| | | |      | | ||  | |  || |   |   |  |  |  |  | |
Db     148368 AATTATTTATATATTTATTTTTTTTTTAAATATATAAATAAATATATATATATATATATA 148309

Qy        147 GAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACAT 206
               | ||||  ||   || | ||   |   |  |||   | || |   |    ||| | |
Db     148308 TATATATATATATATATATATATATTTTATTTAAAAAAAAAAAAAAAAAAAAAAAAAAA 148249

Qy        207 TGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCA 266
              |  | | |   | || |   ||  ||||  ||||  |    | |     |    | ||
Db     148248 TTNTNTTTNTNTNTATTTATTTNTTTTTAAAAAAAAAAAAAAAAANNANNANTNTTTTTTN 148189

Qy        267 TTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATAT 326
              ||| ||||   |||    ||||||  |  |   |      | |   |    |||
Db     148188 TTTTTTTTNNTTTTNAAAAAAAAAAAAAAANNANNNNANANNNNNNAAANNAANNNNTAT 148129

Qy        327 TAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCA 386
              |          ||  | ||   | ||| |    ||||     ||  |
Db     148128 TTTTTNNNNTAANATNAAATTATAATAAANTAATAT--ATATATATNNNNNNNNNNNNN 148071

Qy        387 AAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAA 446

Db     148070 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 148011

Qy        447 TTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAA 506
                                        | | | |    || ||||| || || | ||
Db     148010 NNNNNNNNNNNNNNNNNNNNNNNNNTTTAAAATTTTTTATTTTTTAATTTATTTTTTT 147951

Qy        507 AACATTTACTTTCCAAATATTCTGTCATTCAGGAGTATGGAAGTATCGAT 556
               | |||  |||     | || | |   | |    |||  |  |||  ||
Db     147950 TATTTTTTTTTTTTTTTTTTTTTTTTATTATATTTATTAATATATATAT 147901


RESULT 74
AC098870/c
LOCUS       AC098870              163597 bp    DNA     linear   PRI 01-MAR-2002
DEFINITION  Homo sapiens BAC clone RP11-651C19 from 4, complete sequence.
ACCESSION   AC098870 AC040969
VERSION     AC098870.3  GI:17149797
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

REFERENCE 1 (bases 1 to 163597)
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
PUBMED 9847074
REFERENCE 2 (bases 1 to 163597)
AUTHORS Harris,A. and Haakenson,W.
TITLE The sequence of Homo sapiens BAC clone RP11-651C19
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 163597)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (04-NOV-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 163597)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (29-NOV-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 5 (bases 1 to 163597)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (03-JAN-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 6 (bases 1 to 163597)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (01-MAR-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On Nov 29, 2001 this sequence version replaced gi:16905363.
---------- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
---------- Summary Statistics
Center project name: H_NH0651C19
Drafting Center: WIBR
----------.

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at http://www.chori.org
VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-745K4. Actual start of this clone is at base position 1 of RP11-651C19; actual end is at base position 163597 of RP11-651C19.

Data from AC097108 and AC016295 was used to finish this clone, AC098870. Polymorphisms have been identified between AC097108 and AC098870.

The sequence of AC040969 has been incorporated into AC098870.

FEATURES Location/Qualifiers
source 1. .163597
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone="RP11-651C19"
/clone_lib="RPCI-11"
repeat_region 283. .363
/rpt_family="MIR"
repeat_region 1048. .1092
/rpt_family="MIR"
misc_feature 1236. .1724
/note="match to EST W04269 (NID:g1276168) za59g12.r1"
repeat_region 1763. .1826
/rpt_family="(TG)n"
misc_feature 1773. .1818
/note="match to EST W04269 (NID:g1276168) za59g12.r1"
misc_feature 1805. .2088
/note="match to EST BF894421 (NID:g12285880)"
misc_feature 2084. .2284
/note="match to EST BG398820 (NID:g13292280)"
misc_feature 2100. .2236
/note="similar to Rattus norvegicus EST BF524337 (NID:g11632304)"
misc_feature 2101. .2284
/note="similar to Mus musculus EST BI144071 (NID:g14604072)"
misc_feature 2124. .2284
/note="similar to Mus musculus EST AI747969 (NID:g5126233) ul04f02.y1"
misc_feature 2204. .2284
/note="similar to Bos taurus EST BE684791 (NID:g10072415)"
misc_feature 3055. .3424
/note="match to EST N70321 (NID:g1226901) za59g12.s1"
repeat_region 3514. .3538
/rpt_family="AT_rich"
misc_feature 4132. .4134
/note="match to EST BF894421 (NID:g12285880)"
misc_feature 4778. .5295
/note="match to EST AI767709 (NID:g5234218) wh38h02.x1"
repeat_region 5482. .5506
/rpt_family="AT_rich"
repeat_region 5931. .6226
/rpt_family="Alu"
repeat_region 6267. .6818
/rpt_family="ERVL"
repeat_region 8196. .8311
/rpt_family="MIR"
repeat_region 8580. .8646
/rpt_family="L2"
repeat_region 9091. .9214
/rpt_family="MER1_type"
repeat_region 9586. .9880
/rpt_family="Alu"
repeat_region 9990. .10281
/rpt_family="Alu"
repeat_region 10575. .10997
/rpt_family="MaLR"
repeat_region 11086. .11235
/rpt_family="MER1_type"
misc_feature 11306. .11844

```
                     /note="match to EST BF507877 (NID:g11591175)"
     repeat_region   12352. .12571
                     /rpt_family="MIR"
     repeat_region   12625. .12916
                     /rpt_family="Alu"
     repeat_region   13475. .13513
                     /rpt_family="AT_rich"
     repeat_region   13762. .14036
                     /rpt_family="Alu"
     repeat_region   14139. .14177
                     /rpt_family="(T)n"
     repeat_region   14148. .14474
                     /rpt_family="Alu"
     repeat_region   14328. .14357
                     /rpt_family="(T)n"
     repeat_region   14524. .14552
                     /rpt_family="(CAAAA)n"
     repeat_region   15633. .15799
                     /rpt_family="MIR"
     repeat_region   16169. .16198
                     /rpt_family="AT_rich"
     repeat_region   16599. .16772
                     /rpt_family="MIR"
     repeat_region   17117. .17335
                     /rpt_family="MIR"
     misc_feature    17259. .17261
                     /note="match to EST BE155404 (NID:g8618125)"
     repeat_region   17575. .17651
                     /rpt_family="MIR"
     repeat_region   17670. .17715
                     /rpt_family="MIR"
     repeat_region   17787. .17831
                     /rpt_family="AT_rich"
     misc_feature    18541. .18923
                     /note="similar to Homo sapiens EST BE155532
                     (NID:g8618253)"
     misc_feature    18545. .19165
                     /note="match to EST BE155365 (NID:g8618086)"
     misc_feature    18556. .19207
                     /note="match to EST BE155498 (NID:g8618219)"
     repeat_region   18598. .18702
                     /rpt_family="L2"
     repeat_region   18828. .18859
                     /rpt_family="(CA)n"
     misc_feature    18845. .19186

  Query Match             4.9%;  Score 53.6;  DB 9;  Length 163597;
  Best Local Similarity   48.4%;  Pred. No. 1.1;
  Matches  178;  Conservative    0;  Mismatches  189;  Indels    1;  Gaps    1;

Qy        159 ATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAAT 218
              |||| | || |  |||||  |  |   || || | || |  || | | | | |     |
Db      49454 ATTATATATAATATAATATATAAGATAAATATATTATATATAATATAATATATAAAACAA 49395

Qy        219 ATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGT 278
              |||     |  ||   | |||| | |   || | | |||      |  || |   | |  |
Db      49394 ATATATTATATATAAATATAATATATTATATATAAATAATATAGTATATATAAATATTTT 49335

Qy        279 TTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGA-AAGATCATATATTAGCCAGTTGA 337
              | |     | |  || | | | ||  ||  | | |  | || | |   ||   | |
Db      49334 ATATAACTTATATATAAATATATATTTATATATAAATTATATAAAAATATATTTATATAT 49275

Qy        338 AAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAG 397
              |||||| | | |||   | |     || ||||  |||   |  | | | |  ||||
Db      49274 AAATTATATAAAAATATATTTATATATAAATTATATAAAAATATATTTATATATAAATTA 49215

Qy        398 TCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAAT 457
              | |     |   |   ||     | | | |       |  | ||| ||||||    ||||
Db      49214 TATAAAATATATTTATATATAAATTATATAAAATATATTTATATATAAATTATATAAAAT 49155

Qy        458 TTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTT 517
              |  || |  | | | | |   |||| ||| || | |||  | |||  ||| || || ||
Db      49154 ATATTTATATATAAATTATATAAAAATATATTTATATATAAATTATATAAAAATATATTT 49095

Qy        518 TCCAAATA 525
              | | |||
Db      49094 TATATATA 49087


RESULT 75
AC008474/c
LOCUS       AC008474              186115 bp    DNA     linear   PRI 13-JUL-2002
DEFINITION  Homo sapiens chromosome 19 clone CTC-379B2, complete sequence.
ACCESSION   AC008474
VERSION     AC008474.9  GI:21743729
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 186115)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 186115)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 186115)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JUL-2000) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
REFERENCE   4  (bases 1 to 186115)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-OCT-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
REFERENCE   5  (bases 1 to 186115)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUL-2002) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Jul 13, 2002 this sequence version replaced gi:16197759.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 99.8% of Sequence;
            Estimated Total Number of Errors is 0.4.
FEATURES             Location/Qualifiers
     source          1. .186115
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /clone="CTC-379B2"
ORIGIN

  Query Match             4.9%;  Score 53.6;  DB 9;  Length 186115;
  Best Local Similarity   49.5%;  Pred. No. 1.1;
  Matches  218;  Conservative    0;  Mismatches  219;  Indels    3;  Gaps    3;

Qy         90 TCCTTATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAA 149
              | | ||| | | |       | | || || |  | | | |     |  ||  |
Db      61171 TACATATTTATTTATTTATTTTGTATTTATATATATATTTATATATAAATTTATATTTTT 61112

Qy        150 ATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGT 209
              ||||  || | |  | ||   || |||  ||| || |        || |  ||     |
Db      61111 ATATGTATAAATTTATATATATTTATATATTTATACATATAAATATATATAAATTTATAA 61052

Qy        210 TTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTT 269
              ||| |  |  ||     || || |  || | | |  ||| | |  || |  |||  |||
Db      61051 TTTATATAATTATATAATTTATATAATT-ATATAAATTATATAATTATATATATAAATTA 60993
```

```
Qy        270 GTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATA-TTA 328
              | ||  | | ||  |   | |    || || ||  |   | | | |  | |||||| |||
Db      60992 TATATTTATATATTAATTTATGTATAAATGATATATAGAAATATATAATTTATATATTTA 60933

Qy        329 GCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACT-TAATCTTGCAGTCAA 387
              | ||  | | || | | || || | |  | |||| || | | ||    |  | | |
Db      60932 TATATTTATATAATATATATAATATAAATATAAATTATATAAATATATAAATATATTTAT 60873

Qy        388 AGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAAT 447
              | | |||  | |   ||||  |||| ||  |  | | ||||   | |  ||||| | |
Db      60872 ATATTAAATATTATTTTAAATATATAATATATTATATATTATATAATTTTAAATATATAA 60813

Qy        448 TACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAA 507
              ||    | || |||  |   |||||||  |   | |   ||| | |  || ||  | ||||
Db      60812 TATATTATATATTAAATATAAAATAATATATTATATATTAAATATAAATATATAAATAAA 60753

Qy        508 ACATTTACTTTCCAAATATT 527
                |||  |   | |||||
Db      60752 TATATTATATATTATATATT 60733


RESULT 76
CR631129
LOCUS       CR631129               186979 bp    DNA     linear   HTG 08-OCT-2004
DEFINITION  Danio rerio clone DKEY-218L8, WORKING DRAFT SEQUENCE, 11 unordered
            pieces.
ACCESSION   CR631129
VERSION     CR631129.2  GI:54019781
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 186979)
  AUTHORS   McLaren,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail.enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 9, 2004 this sequence version replaced gi:51127408.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK218L8
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 182414 bases at least Q40
            Consensus quality: 182991 bases at least Q30
            Consensus quality: 183522 bases at least Q20
            Insert size: 185979; sum-of-contigs
            Insert size: 190733; 1.4% error; agarose-fp
            Quality coverage: 8.50x in Q20 bases; sum-of-contigs Quality
            coverage: 8.33x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2684: contig of 2684 bp in length
            *     2685     2784: gap of 100 bp
            *     2785    13549: contig of 10765 bp in length
            *    13550    13649: gap of 100 bp
            *    13650    27876: contig of 14227 bp in length
            *    27877    27976: gap of 100 bp
            *    27977    53494: contig of 25518 bp in length
            *    53495    53594: gap of 100 bp
            *    53595    58611: contig of 5017 bp in length
            *    58612    58711: gap of 100 bp
            *    58712    72392: contig of 13681 bp in length
            *    72393    72492: gap of 100 bp
            *    72493    86458: contig of 13966 bp in length
            *    86459    86558: gap of 100 bp
            *    86559    91162: contig of 4604 bp in length
            *    91163    91262: gap of 100 bp
            *    91263   133955: contig of 42693 bp in length
            *   133956   134055: gap of 100 bp
            *   134056   140947: contig of 6892 bp in length
            *   140948   141047: gap of 100 bp
            *   141048   186979: contig of 45932 bp in length.
FEATURES             Location/Qualifiers
     source          1. .186979
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-218L8"
                     /clone_lib="DanioKey"
     misc_feature    1. .2684
                     /note="assembly_fragment:02252
                     fragment_chain:1"
     misc_feature    2785. .13549
                     /note="assembly_fragment:02374
                     fragment_chain:1"
     misc_feature    13650. .27876
                     /note="assembly_fragment:00001
                     fragment_chain:1"
     misc_feature    27977. .53494
                     /note="assembly_fragment:00553
                     fragment_chain:1"
     misc_feature    53595. .58611
                     /note="assembly_fragment:02327
                     fragment_chain:1"
     misc_feature    58712. .72392
                     /note="assembly_fragment:00159
                     fragment_chain:1"
     misc_feature    72493. .86458
                     /note="assembly_fragment:02505
                     fragment_chain:1"
     misc_feature    86559. .91162
                     /note="assembly_fragment:02282
                     fragment_chain:1"
     misc_feature    91263. .133955
                     /note="assembly_fragment:00836
                     fragment_chain:1"
     misc_feature    134056. .140947
                     /note="assembly_fragment:00327.0"
     misc_feature    141048. .186979
                     /note="assembly_fragment:01386"
ORIGIN

  Query Match             4.9%;  Score 53.6;  DB 2;  Length 186979;
  Best Local Similarity   47.4%;  Pred. No. 1.1;
  Matches  182;  Conservative    0;  Mismatches  200;  Indels    2;  Gaps    1;

Qy        146 AGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACA 205
              | ||| |   |    || || | | |||   ||| |     |    | ||   |
Db     186303 AAAAAAAAAAAAAATATATAATTTATTTATATATTTATTTATTTTTTATTTATTTTTTATT 186362

Qy        206 TTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTC 265
              || |||| |  ||||| |   |  || | | || |      || | | ||| |  |||
Db     186363 TTTTTTTATATATATATTAAATAAATATATATATATATATATATATATATATATATATAT 186422

Qy        266 ATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATA 325
              || | | | |   |  | | | |  |||| | | ||  ||  | | ||| || |||
Db     186423 ATATATATATATATATATATATATATATATATATATATATATAAAAAAAAAAATAATAAT 186482

Qy        326 TTAGCCAGTTGAAAATTAAACACAAAATGAGTGCA--TATTACATTACTTAATCTTGCAG 383
              ||    | |  | |||  ||    ||   |    |  |||||  ||  || || ||  |
```

```
Db      186483 TTTTTAAATATATAATANAAATATAATAAATAAAAATTATTATTTTTATTTATTTTATAN 186542

Qy         384 TCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAG 443

Db      186543 NAANAAAANAAANTAATAAAAAAAAAAAAAATATTAAATAAAATATTATATTAAAAATAA 186602

Qy         444 AAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTAT 503

Db      186603 ATTTTATTTTAATTTTTTAAAATTTATTAAAAAAAAAAAAAAAAAATATTATAAAAATTAA 186662

Qy         504 TAAAACATTTACTTTCCAAATATT 527

Db      186663 TATAATTAAAATTATAAAAAAATT 186686


RESULT 77
AC111539/c
LOCUS       AC111539               234560 bp    DNA     linear   HTG 13-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-158A12, WORKING DRAFT SEQUENCE, 3
            unordered pieces.
ACCESSION   AC111539
VERSION     AC111539.5  GI:30579258
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 234560)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 234560)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-FEB-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 234560)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 13, 2003 this sequence version replaced gi:25007559.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GMNF
                Center clone name: CH230-158A12
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 221222 bases at least Q40
                Consensus quality: 224241 bases at least Q30
                Consensus quality: 226292 bases at least Q20
                Estimated insert size: 231347; sum-of-contigs estimation
                Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *         1   231343: contig of 231343 bp in length
            *    231344   231443: gap of unknown length
            *    231444   232446: contig of 1003 bp in length
            *    232447   232546: gap of unknown length
            *    232547   234560: contig of 2014 bp in length.
FEATURES             Location/Qualifiers
     source          1. .234560
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-158A12"
     misc_feature    1. .1537
                     /note="wgs_contig"
     misc_feature    3731. .7474
                     /note="wgs_contig"
```

```
     misc_feature    225356. .228724
                     /note="wgs_contig"
     misc_feature    230080. .231343
                     /note="wgs_contig"
ORIGIN

  Query Match             4.9%;  Score 53.6;  DB 2;  Length 234560;
  Best Local Similarity   64.0%;  Pred. No. 1.1;
  Matches  114;  Conservative    0;  Mismatches   59;  Indels    5;  Gaps    2;

Qy        887 TTACCACCAGTGCCTGAGGCTCCAGAGGAGCCATTGCCTTCTCAAGGTCACATAGGTGGT 946
              | | |  ||| ||| | | |||||||||  ||  || || ||||||||||||| ||||||
Db        476 TCATCTTCAGGGCCAGGGTCTCCAGAGGGTCC--TGTCTGCTCAAGGTCACATGGGTGGT 419

Qy        947 ---GGGTGAGTTAGGACCAAATCTAGAATTCCTGACTCCAGTAACTTCTGAAGTCATTTT 1003
                  |  ||   ||||||   || |   | |||| ||||||  |||||||||||| | |||
Db        418 AACAGCAGACAAAGGACCTTGTCCAAGCTACCTGTCTCCAGAGACTTCTGAAGTAACTTT 359

Qy       1004 GTTTTTTATTTTTATGGTTTTATTATAAGAATACTTGCTAAGCACACTTACCCCCTGC 1061
              | |||||||| || || || | || |   |  |  |  |   ||  | | |  |  ||
Db        358 GGTTTTTATTCTTCTGCTTCTGTTGTGTTACAAGATATTTGCCAAGCATGCTTCTAGC 301


RESULT 78
AC125563/c
LOCUS       AC125563               275631 bp    DNA     linear   HTG 09-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-9B12, WORKING DRAFT SEQUENCE, 4
            unordered pieces.
ACCESSION   AC125563
VERSION     AC125563.4  GI:24817906
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 275631)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 275631)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 275631)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 9, 2002 this sequence version replaced gi:23096552.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GDBK
                Center clone name: CH230-9B12
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329
                Consensus quality: 228152 bases at least Q40
                Consensus quality: 230310 bases at least Q30
                Consensus quality: 231872 bases at least Q20
                Estimated insert size: 235029; sum-of-contigs estimation
                Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   192878: contig of 192878 bp in length
            *  192879   192978: gap of unknown length
            *  192979   270295: contig of 77317 bp in length
            *  270296   270395: gap of unknown length
            *  270396   272405: contig of 2010 bp in length
            *  272406   272505: gap of unknown length
            *  272506   275631: contig of 3126 bp in length.
```

```
FEATURES             Location/Qualifiers
     source          1. .275631
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-9B12"
     misc_feature    30410. .30884
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH339524"
     misc_feature    complement(56153. .56817)
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH339528"
     misc_feature    57750. .59243
                     /note="wgs_end_extension
                     clone_end:Sp6"
ORIGIN

  Query Match             4.9%;  Score 53.6;  DB 2;  Length 275631;
  Best Local Similarity   64.0%;  Pred. No. 1.1;
  Matches  114;  Conservative    0;  Mismatches   59;  Indels    5;  Gaps    2;

Qy        887 TTACCACCAGTGCCTGAGGCTCCAGAGGAGCCATTGCCTTCTCAAGGTCACATAGGTGGT 946
              | | |  ||| ||| | | |||||||||  ||  || || ||||||||||||| ||||||
Db     141355 TCATCTTCAGGGCCAGGGTCTCCAGAGGGTCC--TGTCTGCTCAAGGTCACATGGGTGGT 141298

Qy        947 ---GGGTGAGTTAGGACCAAATCTAGAATTCCTGACTCCAGTAACTTCTGAAGTCATTTT 1003
                 |  ||   ||||||   || |   | |||| ||||||  |||||||||| | |||
Db     141297 AACAGCAGACAAAGGACCTTGTCCAAGCTACCTGTCTCCAGAGACTTCTGAAGTAACTTT 141238

Qy       1004 GTTTTTTATTTTTATGGTTTTATTATAAGAATACTTGCTAAGCACACTTACCCCCTGC 1061
              | |||||||| || || || | || |   |  |  |  |   ||  | | |  |  ||
Db     141237 GGTTTTTATTCTTCTGCTTCTGTTGTGTTACAAGATATTTGCCAAGCATGCTTCTAGC 141180


RESULT 79
AC117080/c
LOCUS       AC117080               85916 bp    DNA     linear   INV 12-MAR-2003
DEFINITION  Dictyostelium discoideum chromosome 2 map complement(821514-735598)
            strain AX4, complete sequence.
ACCESSION   AC117080 AC114260
VERSION     AC117080.2  GI:28850277
KEYWORDS    HTG.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 85916)
  AUTHORS   Gloeckner,G., Eichinger,L., Szafranski,K., Pachebat,J., Dear,P.,
            Lehmann,R., Baumgart,C., Parra,G., April,J.F., Guigo,R., Kumpf,K.,
            Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and
            Noegel,A.A.
  TITLE     Sequence and analysis of chromosome 2 of Dictyostelium discoideum
  JOURNAL   Nature 418 (6893), 79-85 (2002)
  MEDLINE   22092622
   PUBMED   12097910
  REMARK    The Dictyostelium Genome Sequencing Consortium
REFERENCE   2  (bases 1 to 85916)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-APR-2002) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   3  (bases 1 to 85916)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   4  (bases 1 to 85916)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
COMMENT     On or before Mar 5, 2003 this sequence version replaced
            gi:19424360, gi:20066270.
            CDS predictions from GeneID do not necessarily reflect true genes.
            Further Information is available from IMB Jena, Department of
            Genome Analysis
            (http://genome.imb-jena.de/dictyostelium/)
            and the Univerity Colonge, Institute for Biochemistry I
            (http://www.uni-koeln.de/dictyostelium/project.shtml
            Funding
            Agency : Deutsche Forschungsgemeinschaft (DFG).
FEATURES             Location/Qualifiers
     source          1. .85916
                     /organism="Dictyostelium discoideum"
                     /mol_type="genomic DNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /chromosome="2"
                     /map="complement(821514-735598)"
     CDS             join(9. .188,279. .1337)
                     /note="GeneID exon scores (in order of location ranges):
                     9.57, 89.77 - GSCJ_ID dd_00785"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAO53082.1"
                     /db_xref="GI:28850278"
                     /translation="MEDNNNSSSNENKQNELSTPSIKKERHLVIMQHGLHGTSLDFKT
                     IRNHFLKQKHLDNCIFISANSNSHFLATHDGIDKIGERLFNEVKELYEQYDHPEKISM
                     IGHSLGGLITRYAIGLLYDDGFFKKCKPDQFISLSSPHCGSRRPSTTIFNKVAHIFVD
                     NFLSVTGKQLILHDTEIPDNIKTFPSTSSPPPNEKLKSSKTIVNSSVKNETDLSLPLA
                     EAKEPSIYKEVGNNEKLMIIEKKEENEIITNDQEVPMPLLVRLSEGIFFNGLNSFRKR
                     TLYSNIYNDVQVNFCTSDISAKNPYTLGKLMKFSEKYRHIIEEESILDIDPNLLEQQS
                     PPPDKKPFDEKDLDEYFTHDTHHHFLKRILKNLNQLHFVRYHMYFKNMLSHTNIIVKR
                     EWINSEGPEIIEHLVSHFEG"
     CDS             join(2619. .2689,2806. .2917,3021. .3235,3389. .3938)
                     /note="GeneID exon scores (in order of location ranges):
                     4.00, 11.19, 30.82, 62.55 - GSCJ_ID dd_00787"
                     /codon_start=1
                     /product="similar to Mus musculus (Mouse). 26S proteasome
                     regulatory subunit S5A (Rpn10) (Multiubiquitin chain
                     binding protein)"
                     /protein_id="AAO53083.1"
                     /db_xref="GI:28850279"
                     /translation="MYTASKTQSNPESAVSIMSMAGKKPEVLVTLTQELSKILSGAQE
                     LKINGKIDFSTTMQIAQLALRHRQNNHQHPRIIAFVGSPLKETKEELIQLAKRLKKNG
                     VAVDIINFGEVTENSDKLEAFINDVNNNDESHLLTVPPGPHILSDIILQSPIVESGSG
                     QFGSEFINADTDPDLAMALKLSLEEEKQRQERERKAREEANGGSTNSGTTTTTAPTES
                     NMDVNFEDDPELAEALALSMATDKMEVQSSTTNTDSQPPQQQQQPPTDDTSSEAFKDQ
                     DFLNSTLNSLPGVDPNRIKNALENLSKKDEDKDKDNEKK"
     CDS             complement(join(4186. .5395,5501. .5689,5854. .6303,
                     6445. .7055))
                     /note="GeneID exon scores (in order of location ranges):
                     94.07, 12.83, 33.44, 41.83 - GSCJ_ID dd_00788"
                     /codon_start=1
                     /product="similar to Mus musculus (Mouse). 12 days embryo
                     male wolffian duct includes surrounding region cDNA, RIKEN
                     full-length enriched library, clone:6720456H20
                     product:similar to  KAIA4653 PROTEIN (Fragment)"
                     /protein_id="AAM45299.2"
                     /db_xref="GI:28850280"
                     /translation="MNNSPTLSNTRNRKNNNNNSNSNSNSNNNNNNNNNNNNSNNNN
                     NNNNNVNRNVNNRNNNNNNIINVNNNVSSTTMNMAKDFMNWNDGIIEREDTQLQQQKD
                     RCFKILKKSTIACIVLLFISCTIIYSMTQYPSVSGGDAGELIVVAHQFGVAHPPGYPL
                     FTFLGYIFSHIIPSNTLSVAWKISFMSSMIGSIASIFIYLTVYLWVNNHWCGLLSAYM
                     FTFSPLIWMYQIQGEVFSMNNMFVAMLMFLGVWYTRVRIFENERYNTAFWTSERIAYL
                     SAFSCAIGLTNQHTLVLIVIPFAFWLMFISGRDQLWNIKILSNLVFYTLIGLSPYLLL
                     FITPKLNRVKYSWGNTSTLRGFIKHFLREEYGTLQLYGGDGGVISLFTKISIYIENLI
                     IQFGYIGLALSLIGLLNLLLGYNIRTFKWKSFGTMIIFSFLFYITFFFNLCNLPIDKP
                     LYRGVFLRFFMQPNVIISITMGLGIKSIFGFLNRIQRDGGTAATTTTTISKIQKYLLP
                     IIIILLVGNQIGFNYNLQDQSDNYSFYDYGHSVLDGLPRNTLLLVGGDLVTNVPMYLH
                     LCEKVRPDIDILSMEIMSWEWFKITQSPLFQRVKFPGNVYHPYIPDGYSLKDFLDANI
                     NDRPIYIGGDFKSGDSSFQNDYYTISKGLTSQIIPKKDSHKFNTFKIIKTTFSTFPSF
                     HIPNNTEKYPNDSWEHFMMEEMAVSLERAAETLLKEYLSQQNTESFKALELSVEILEK
```

```
                ALTYVNDKCWSLKHLGICFDHLRYRVVQNVNNNNQQAQNKASNYSKQLLYFWKKYINQ
                CTHEENTDQDWETIKKVIQLI"
     CDS        7833. .8195
                /note="GeneID exon scores (in order of location ranges):
                30.55 - GSCJ_ID dd_00789"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AAO53084.1"
                /db_xref="GI:28850281"
                /translation="MINPLNKPVLGTRLMNSVFPPNIKLQVIRNSGKTFENTNKIMFK
                TDCSVGKTDIKTYMKALYDVNVDKVNTINVQGRIKSKTKGLQKRSKLNSKYKTPDYKK
                AIITVDPSLRAQLSRGKN"
     CDS        9000. .9674
                /note="GeneID exon scores (in order of location ranges):
                54.39 - GSCJ_ID dd_02815"
                /codon_start=1
                /product="similar to Homo sapiens (Human). Small GTP
                binding protein RAB32"
                /protein_id="AAM45300.2"
                /db_xref="GI:28850282"
                /translation="MYSNKNDKDKDKDQNNENNKNNDDEAISLKILVVGKLACGKTSI
                IQRYCHNNFQPKYKPTIGVDFQQKVLEIMGQKVLLQLWDIAGQERFGHMTRVYFQNAH
                GAVIVFDATRSGTFLGAKAWKDDIDYCFNNENLPTILLANKCDLLTPPYTFPEDINTF
                CEQNRFNKYFYTSAKEDTGINEALEELVKIILESYQTQEQSTGFKLSDQSQSTETTPT
                QSKTCC"
     CDS        complement(10728. .12734)
                /note="GeneID exon scores (in order of location ranges):
                120.50 - GSCJ_ID dd_03061"
                /codon_start=1
                /product="similar to Plasmodium falciparum. Hypothetical
                protein"
                /protein_id="AAO53085.1"
                /db_xref="GI:28850283"
                /translation="MNNGKILVLNSLPNYLQKRIIQLLCEHIDFNNKQLVVNCNFPKI
                LEYRENESLIKRLIINISLVSHQWFQIISNYLNVSVDFNFNSINNSYYSIIKATNVKI
                LKIHYNHQNNYFYKFIKGNEFKLTINEINEKIKQLPNLKQLLIRNINISKVINDIDGT
                IDTSNNSSYSNINIEGNFKIETNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
                NNNNGDGDGDGDGNSDTTTNSITNINNFNNNSDNYHYNNRIINNDNLNFLKDIEFKGN
                IELLEVEINKSNSMERLMELKSKISKVFSLNVVNTLSKELLFETFKNWLPSIRQLLIV
                AHESPMSVFPWYSLSNNGYFKNISYFDIYGIELRLSNISTIIKSSPNIKLLSFSICFD
                KLLYYIVKLNETQQKQQQQQQQQQLPFLQDKMIDDDNIIIEKCTCNCLLFYKSNINYK
                KSLNCTSGGIDIDDEKQQSYNLNFVDHQEEEKEIQDDLNREINKFSNDWKVLCELLKN
                SSIECLHLHQDCMDGLEIMQFNQLNRSSTTTKPNSIIYPDCFIQELGGSLISPINTLK
                KISIMGLNTPQLFKYIVDHNKTIQNYIYTESILDLGYYNHQDYTSIFESIINENQQIT
                ELKYSKAVFSLENEINMIDYEDYDADDDADELPYKDQVILHYKL"
     CDS        complement(13223. .14611)
                /note="GeneID exon scores (in order of location ranges):
                36.02 - GSCJ_ID dd_02813"
                /codon_start=1
                /product="similar to Plasmodium falciparum. Hypothetical
                protein"
                /protein_id="AAO53086.1"
                /db_xref="GI:28850284"
                /translation="MKVLSNLILKNIFLYLLNKNKFKQVLEYSLISKACFSIIQNLLT
                FDNDCIMVKMTFNSIKSYIENSKLSINRPTQILNQLMKRKLLQFKEIEQLVIFYDNNN
                NKNNNNENNDSNNDKELKRSEYNYLINNFKNLKRIYSESSKFAYFPSYQYYNDLVDYN
                KICKIYIKKLYINWFYFGPISSSQEHQLYSNSPYEESIYFKFSFEFLSHYNPKKLIIS
                TPNHQIQNNFYNLDNILNCKSIKSIYFKNIETPNSLIKKVFSETNIKSFKIILDNNNN
                NQNNDNDNDNNGFILNRIENNNFKLKKLVIKNFNFYNNNNSKKEEGDINEQNKFLNEI
                CLKNYNLKYIGLEFFVIKESNQLNQLFNLKFLNTISCSNDSFIAVIQHCNENKNIKTL
                KYHNKVLVQLIPEDFKIFDSFFKSNKTLEDFRIFYYYPHLERSFLKVKKILSDSKNKT
                IKNFEIKEVTKK"
     CDS        complement(join(14996. .16054,16142. .17131))
                /note="GeneID exon scores (in order of location ranges):
                89.73, 89.24 - GSCJ_ID dd_03199"
                /codon_start=1
                /product="similar to Dictyostelium discoideum (Slime
                mold). Myb2 protein"
                /protein_id="AAM45304.2"
                /db_xref="GI:28850285"
                /translation="MTAIFPNSNIPFYYWASTNAQPQQPQQSLQQQQQQQQQQQQQQQ
                QQQQQQQHHHQNIPPTPQPIFSPQLINKNFGCYGTSNQIPMIPQALFPQQTFNYSGL
                PLCPSPPNYNNYHTITNSPPHQIHSPQLTIDQHSPPTCVSGELSPTPLSSSTGFSNNN
                NNNNNNNNNNNSNSNNNINNNNNNNIKRSYSAMMEENNYPNNNNNNNNNNNNNNNNN
                NNNNNNNNNNNNNIINNNINSSGSISSYSNNSCIEDDYSSDKHCSDKESSNDCDDDAD
                SYFEEYEDDQEAIRVSITPFVNLICKTPGITIPIIKIFSECSKEEITVKQMLLKIKNL
                CLEFNLDNSIHNTLIHSLVKIDGDIKMVKYLIDNGISEQELEFTKDILLTSNSQEIIK
                KKRERKRESISRGIRSPPNKWTKEESQNLIKLVTENGDKQWKKIATKLGGGKTGAQCA
                QHWKRVLSPEIKKGSWDEAEEELLFQLVDKHGQSWKNVAIEIKTRTDIQCRYQYFKAI
                MSRQTEWNQLEDDILTKKIKLMTQNNEKISFQQVSKHLARAKTTKIPRTALECKSRWS
                QLNSTNVNNNNNNNNNSITTSSSNTNQQQQSTMVTPTSSPLSSPIPMSTPIINNSMPQ
                QQVQQIQQQQISIQPYPQSFIQETQGFYRGNDQMQYHIQNQTIPQYHHQHNNNPLLML
                "
     CDS        complement(join(18129. .18142,19312. .19449,22272. .22294,
                22670. .22874,23418. .23622,23954. .24313,24384. .24563,
                24644. .27214))
                /note="GeneID exon scores (in order of location ranges):
                0.17, 3.82, -0.39, 11.33, 8.29, 30.56, 7.20, 265.13 -
                GSCJ_ID dd_00357"
                /pseudo
                /codon_start=1
     CDS        join(28837. .28849,28942. .29126)
                /note="GeneID exon scores (in order of location ranges):
                -2.77, 3.32 - GSCJ_ID dd_01206"

  Query Match             4.9%;  Score 53.4;  DB 3;  Length 85916;
  Best Local Similarity   46.2%;  Pred. No. 1.4;
  Matches  177;  Conservative    0;  Mismatches  206;  Indels    0;  Gaps    0;

Qy        134 TGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTC 193
              ||  | |  |||||   | ||  |  ||  | || | || |||    | |       ||||
Db      17717 TGGTTATTTTTGAGTTTTTTTTTTTTTTTTATATAAATTTATATAAATTTTTTTTTTCTC 17658

Qy        194 ATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGAC 253
              || ||||  | |||   || ||  | |     ||  |||||||   |   |||||  ||
Db      17657 ATATTAATAAAATTTAATTTTTTCTTTTAATTTTTTTTTTTTTTTTTCATCACTTACTTAC 17598

Qy        254 ATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAG 313
              |    | | |  ||  ||       |     |||||     | |   |   |   | ||
Db      17597 ACTAGGTCACTTTTTATTTAACAAAATTAATTAAAAAAAATAACAATCAAAAAAAAAAAA 17538

Qy        314 AAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTT 373
              ||| |||| | |  |    |  | |  |  ||| | ||||       ||  ||   ||  |
Db      17537 AAAAATCAAAAAAAAAAAAAAATAATTAAAAAAAAAAAAAAATCAATCACTTTTTTTTTTG 17478

Qy        374 AATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCA 433
              | |  |   |   || |  ||  |  ||||  |  | ||  |  |  |         |
Db      17477 TACATAACCAATACCAATACATTGTATAAATAAAATTCAAATAAATAATAATAATAATAA 17418

Qy        434 CTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTT 493
                || || |||| |  | |||   ||   |  || ||||     |     | ||||||
Db      17417 TAATAATTATAAATAATTATAAAAAATAAAATAAAATTAATTAATAATTTTTTTATTTTT 17358

Qy        494 TATGTACTATTAAAACATTTACT 516
              | | || | ||  || || | ||
Db      17357 TTTTTATTTTTTCAAAATCTCCT 17335


RESULT 80
PFMAL8P1_00
WPCOMMENT
  Sequence split into 14 fragments  LOCUS PFMAL8P1 Accession AL844507
     Fragment Name      Begin      End
     PFMAL8P1_00            1   110000
     PFMAL8P1_01       100001   210000
     PFMAL8P1_02       200001   310000
     PFMAL8P1_03       300001   410000
     PFMAL8P1_04       400001   510000
     PFMAL8P1_05       500001   610000
     PFMAL8P1_06       600001   710000
     PFMAL8P1_07       700001   810000
     PFMAL8P1_08       800001   910000
     PFMAL8P1_09       900001  1010000
     PFMAL8P1_10      1000001  1110000
     PFMAL8P1_11      1100001  1210000
     PFMAL8P1_12      1200001  1310000
```

```
    PFMAL8P1_13       1300001   1325595
LOCUS       PFMAL8P1             1325595 bp    DNA     linear   HTG 29-JAN-2003
DEFINITION  Plasmodium falciparum 3D7 chromosome 8, *** SEQUENCING IN PROGRESS
            ***, 25 ordered pieces.
ACCESSION   AL844507
VERSION     AL844507.1  GI:23498988
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Plasmodium falciparum 3D7
  ORGANISM  Plasmodium falciparum 3D7
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 1325595)
  AUTHORS   Seeger,K., Murphy,L., Harris,D., Berriman,M., Pain,A., Hall,N.,
            Quail,M. and Barrell,B.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1325595)
  AUTHORS   Seeger,K., Murphy,L., Harris,D., Berriman,M., Pain,A., Hall,N.,
            Quail,M. and Barrell,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-SEP-2002) P.falciparum Genome Sequencing Consortium,
            The Welcome Trust Sanger Institute, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA, UK
COMMENT     This chromsome is unfinished and still contains gaps. The contigs
            have been orderd as accuratly as possible using mapping data but
            there will be errors. There may also be some sequencing errors.
            Contigs are seperated with 100 Ns.
            For more information about this sequence or the Malaria Project,
            see http://www.sanger.ac.uk/Projects/P_falciparum.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 25 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1    18663: contig of 18663 bp in length
            *   18664    18763: gap of 100 bp
            *   18764    28918: contig of 10155 bp in length
            *   28919    29018: gap of 100 bp
            *   29019    45729: contig of 16711 bp in length
            *   45730    45829: gap of 100 bp
            *   45830    47510: contig of 1681 bp in length
            *   47511    47610: gap of 100 bp
            *   47611    59948: contig of 12338 bp in length
            *   59949    60048: gap of 100 bp
            *   60049   319498: contig of 259450 bp in length
            *  319499   319598: gap of 100 bp
            *  319599   320200: contig of 602 bp in length
            *  320201   320300: gap of 100 bp
            *  320301   381318: contig of 61018 bp in length
            *  381319   381418: gap of 100 bp
            *  381419   506172: contig of 124754 bp in length
            *  506173   506272: gap of 100 bp
            *  506273   516293: contig of 10021 bp in length
            *  516294   516393: gap of 100 bp
            *  516394   708270: contig of 191877 bp in length
            *  708271   708370: gap of 100 bp
            *  708371   777620: contig of 69250 bp in length
            *  777621   777720: gap of 100 bp
            *  777721   923460: contig of 145740 bp in length
            *  923461   923560: gap of 100 bp
            *  923561   934223: contig of 10663 bp in length
            *  934224   934323: gap of 100 bp
            *  934324   948566: contig of 14243 bp in length
            *  948567   948666: gap of 100 bp
            *  948667   958135: contig of 9469 bp in length
            *  958136   958235: gap of 100 bp
            *  958236   966722: contig of 8487 bp in length
            *  966723   966822: gap of 100 bp
            *  966823   977022: contig of 10200 bp in length
            *  977023   977122: gap of 100 bp
            *  977123   991432: contig of 14310 bp in length
            *  991433   991532: gap of 100 bp
            *  991533   997040: contig of 5508 bp in length
            *  997041   997140: gap of 100 bp
            *  997141  1007805: contig of 10665 bp in length
            * 1007806  1007905: gap of 100 bp
            * 1007906  1024535: contig of 16630 bp in length
            * 1024536  1024635: gap of 100 bp
            * 1024636  1196984: contig of 172349 bp in length
            * 1196985  1197084: gap of 100 bp
            * 1197085  1246228: contig of 49144 bp in length
            * 1246229  1246328: gap of 100 bp
            * 1246329  1325595: contig of 79267 bp in length.
FEATURES             Location/Qualifiers
     source          1. .1325595
                     /organism="Plasmodium falciparum 3D7"
                     /mol_type="genomic DNA"
                     /isolate="3D7"
                     /db_xref="taxon:36329"
                     /chromosome="8"
     gene            complement(join(18766. .21639,21751. .21977,22145. .23009))
                     /gene="MAL8P1.1"
     CDS             complement(join(<18766. .21639,21751. .21977,22145. .23009))
                     /gene="MAL8P1.1"
                     /note="Similar to Plasmodium falciparum hypothetical 192.3
                     kDa protein mal1p3.07 SWALL:Q9U0N0 (EMBL:AL031746) (1555
                     aa) fasta scores: E(): 3.5e-51, 32.25% id in 1423 aa, and
                     to Plasmodium vivax Pvstp1 protein pvstp1 SWALL:Q9N894
                     (EMBL:AL360354) (826 aa) fasta scores: E(): 8.3e-16,
                     25.64% id in 815 aa, and to Plasmodium falciparum
                     hypothetical 192.3 kDa protein mal1p3.07 SWALL:Q9U0N0
                     (EMBL:AL031746) (1555 aa) fasta scores: E(): 3.3e-51,
                     32.25% id in 1423 aa"
                     /codon_start=1
                     /product="hypothetical protein, conserved in P.
                     falciparum"
                     /protein_id="CAD51069.1"
                     /db_xref="GI:23498989"
                     /db_xref="UniProt/TrEMBL:Q8IBD7"
                     /translation="MAIISSIPNYRKGLNPSLYNLRFSLFRDRVNAYTISKIKKLKKS
                     RNYKRLCRYLNYKLDDIKDVFMLNNVLGIPRTARMGLWENVMENKLKDIIEENSDNKC
                     RRIYPYFQRIQRRRRKALENFCEDKEKKQEELKNVKFDMKKCLEFNEWVKESENKIFK
                     LFKNNITERDKRKKAYIISDKCKLGNLNTLFGQMDCSDKIDRHFPYIFLEGNPVYPSE
                     KDDDKNIKDPEEGKDEEEEEEDGEVTINIRSNGEKGFRFVNNTLVLEGIEGSGFDLDF
                     ETYDISPHSLNSGEIANLFEFNNESHLEHMNIESNRSPNMVDIDFSDVPLTAAHMTSS
                     ITAKNISSYLYVPLGVLFIIISFIWFFLNRTKHLNMLIKEKEKVKTENIVPSEAYVKL
                     DDLEVMSEEIKEKQKEKLKKKIHAQKKLEIIEDPYKIKLLERKKWLWKTIIEIHMLIL
                     EEQRKAEWEFNKGDFLNICIDEFSNQEKKPDTNIITSDLPFDEPEVIETIRMMNKQSD
                     IWYRWVERHAHMLEKWKEEEWFLRLKESWKDEEEEYMKRAYKELLISLRGDKYHMSQR
                     QKIIWRKWIAKHPYRIREFVIDKWFNQLFEELEKEGIISDEAIRNFLSQYEEAYDEEK
                     IQEFIKYKKKVLVVILWIKIYMSMLEEYVKEEDIESEKLFIDTSLEELKKKKEIEEEV
                     LEDLKKDIYESDEMKEIEKYKGEDWFKQMQQDWIRRDYKFVSSYNLESDDEEQFHEFV
                     KKPTVEIQKDLLKEQWKNIELKWIEEDNYDEYVLEDELLKETKYIKPKDKMKHRIIKE
                     DNIYGEQLKVFDEIEKRSIYLSSKRVLKWKTIIEIHLEVLHDCKKEEWDKNKRDFLNI
                     CIEELVKGQNEDKKNKENYNYITNESDNYDIVEYTKSFCNTWKDRHNYMLEKWNKENW
                     FVQLKNDWVEEQNEYMRKTYKELLISLKGDKYNMSQRQKIIWRKWIAKHPYKINEEII
                     NIWLNRLSDERDKNGIISDEAIKKFLSLEKDISDKDFLKYRRKKLEVILWVKIYMSVL
                     EEEKNEELEKSKESFLNTCIDELKKEEYSEENERMITIVDEMKKGFRICDKVEDYKKK
                     KWKNEEWYKDLKKEWIKEEDKYIDSIYLMKKDEIYEDDLLNKSIVELEKSVSLKHWKD
                     MYIKWIDEDNERDWLRIAIDKKNLKTDEINEDINRMIELNMLEDYSNDLYDKKKIKWK
                     TIIEIHMEVLNDCKNDEWEMNKGDFLDICLEEFGTMENKEYSHIINNMLMVGELEYKH
                     LNMDVIERQNDLWNKWIEGHRYMLEKWKKEKWFIVLKENWKNEEHKYMRKAYLELLMS
                     LR"
     gene            complement(join(25328. .26401,26530. .26736))
                     /gene="MAL8P1.2"
     CDS             complement(join(25328. .26401,26530. .26736))
                     /gene="MAL8P1.2"
                     /note="Similar to Plasmodium falciparum protein with DNAJ
                     domain pfb0085C SWALL:O96122 (EMBL:AE001370) (900 aa)
                     fasta scores: E(): 2.7e-09, 24.7% id in 340 aa"
                     /codon_start=1
                     /product="hypothetical protein with DNAJ domain"
                     /protein_id="CAD51070.1"
                     /db_xref="GI:23498990"
```

```
                     /db_xref="UniProt/TrEMBL:Q8IBD6"
                     /translation="MEKYIHTSKSNSVPSKGYVNRYIENTRNDCTLYKCAKKGQYRNR
                     KNILLKCGRFFRLPLCVVIYYIILMSQFIKTINCTSVLPSCNESEKNVSKNIKDICNE
                     GKDDNISNKWYEENKRSLNGEYSDNPLFDYKLFKYHDILYKGILNKELSEEFDLLFSD
                     NTNYPFVKNEKDKEIMKQIKLLSINEDENKKEHMKKIWTDFIKNELSNFLFLKKKISD
                     LFEELKSQYNIKDEYVNKIWNECLDLIETSGMSMNTNLNYAFYKWYNKTKVLDINEYI
                     FFICGIKLVWMKLFSSLEISCKDILMKCFEGKKNGKTLYTKCCSDKYKNFFKDINKKS
                     FSKKCENDENNNDSYSHINHLNTLLNELDHLISLNEKDDDYYTLRGYSSNKRVELTDI
                     FRLLCKMDSADMLTWFVHNFFGFSKKFRQYLYKM"
     gene            join(30003. .30071,30235. .30744)
                     /gene="2255.m00020"
                     /note="synonym: PF08_0001"
     CDS             join(30003. .30071,30235. .30744)
                     /gene="2255.m00020"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAD51071.1"
                     /db_xref="GI:23498991"
                     /db_xref="UniProt/TrEMBL:Q8IBD5"
                     /translation="MLSFNVIYITCVVVIFSLQYNNKVNLHNKVHNIYKTRYSLIIYF
                     KRLLIEYNETVQLEKNPDENSSTELDDNESTVSVEECEKSLKEFFDDCKRENYPGNVI
                     HLCDNVEEPEEIKKGGSKTNLLLEEYGDPVVYDKEEMKLISELLDSYLEDERHLCMIK
                     RILDYTKIYRNPLLLARYRIKKFLKRRCKKIK"
     gene            complement(join(31389. .35831,36001. .37707))
                     /gene="PF08_0002"
     CDS             complement(join(31389. .35831,36001. .37707))
                     /gene="PF08_0002"
                     /note="Similar to Plasmodium falciparum hypothetical 192.3
                     kDa protein mal1p3.07 SWALL:Q9U0N0 (EMBL:AL031746) (1555
                     aa) fasta scores: E(): 1.7e-19, 29.19% id in 1360 aa, and
                     to Plasmodium falciparum hypothetical 192.3 kDa protein
                     mal1p3.07 SWALL:Q9U0N0 (EMBL:AL031746) (1555 aa) fasta
                     scores: E(): 1.7e-19, 29.19% id in 1360 aa"
                     /codon_start=1
                     /product="hypothetical protein, conserved in P.
                     falciparum"
                     /protein_id="CAD51072.1"
                     /db_xref="GI:23498992"

  Query Match             4.9%;  Score 53.4;  DB 2;  Length 110000;
  Best Local Similarity   47.2%;  Pred. No. 1.3;
  Matches  162;  Conservative    0;  Mismatches  181;  Indels    0;  Gaps    0;

Qy       164 AAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACT 223
            ||  ||  |    | | || | ||   |  |   | ||       | |  | | |  | |
Db     86302 AATACACATTCGGGATATGAAAAAATACAAAAAATCAATCAGACATATAATAATTTAAAT 86361

Qy       224 CCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTT 283
              || ||   | ||||  |   |  | | | | |  || | | |  |  |    ||
Db     86362 ATATTCATAAATATAAATAATGATCTAAATAAAAATGTAATAAATACTCCTAATCTTCAA 86421

Qy       284 CAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTA 343
              ||| ||  |  | |||||  | | |     ||| ||   |||  |   |    |||   |
Db     86422 AAAACAAAATTAAGAGAAAAGGAATATACAAAAAATATCATAGAAAAAAAAAAAAAATAGA 86481

Qy       344 AACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACC 403
            ||| |  |   |    | |    | || | ||| ||||||  |||| | |||||   ||
Db     86482 AACCCTTATAAATGTAAAAACGAAATATTGAATATTGCAGGCAAAAGAAAAAAAAAAAAA 86541

Qy       404 TAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATT 463
              |||   | | ||  | |  || |||   | |  ||  |||  | |    |||    | |
Db     86542 AAAAAAAAAAATAATTTCAATCATATAATAATAAAATCAGACTTAAAAGAAAAGAAAAAT 86601

Qy       464 TTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAA 506
               |||  || |  ||||| | |   |  || |  |||||
Db     86602 CAAAAAAATATATATCAAAACAAACTCCCTGTTGAAAAATTAA 86644


RESULT 81
CR847833/c
LOCUS       CR847833              149238 bp    DNA     linear   HTG 05-OCT-2004
DEFINITION  Danio rerio clone DKEY-97I15, *** SEQUENCING IN PROGRESS ***, 13
            unordered pieces.
ACCESSION   CR847833
VERSION     CR847833.2  GI:53828268
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 149238)
  AUTHORS   McLay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 5, 2004 this sequence version replaced gi:53755884.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK97I15
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 145546 bases at least Q40
            Consensus quality: 146309 bases at least Q30
            Consensus quality: 146880 bases at least Q20
            Insert size: 148038; sum-of-contigs
            Insert size: 158903; 0.8% error; agarose-fp
            Quality coverage: 6.03x in Q20 bases; sum-of-contigs Quality
            coverage: 5.89x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 13 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    47030: contig of 47030 bp in length
            *   47031    47130: gap of 100 bp
            *   47131    51832: contig of 4702 bp in length
            *   51833    51932: gap of 100 bp
            *   51933    59948: contig of 8016 bp in length
            *   59949    60048: gap of 100 bp
            *   60049    64594: contig of 4546 bp in length
            *   64595    64694: gap of 100 bp
            *   64695    73602: contig of 8908 bp in length
            *   73603    73702: gap of 100 bp
            *   73703    76984: contig of 3282 bp in length
            *   76985    77084: gap of 100 bp
            *   77085    79394: contig of 2310 bp in length
            *   79395    79494: gap of 100 bp
            *   79495    81796: contig of 2302 bp in length
            *   81797    81896: gap of 100 bp
            *   81897    87366: contig of 5470 bp in length
            *   87367    87466: gap of 100 bp
            *   87467   104174: contig of 16708 bp in length
            *  104175   104274: gap of 100 bp
            *  104275   113737: contig of 9463 bp in length
            *  113738   113837: gap of 100 bp
            *  113838   140256: contig of 26419 bp in length
            *  140257   140356: gap of 100 bp
            *  140357   149238: contig of 8882 bp in length.
FEATURES             Location/Qualifiers
     source          1. .149238
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-97I15"
                     /clone_lib="DanioKey"
```

```
     misc_feature    1. .47030
                     /note="assembly_fragment:01218
                     fragment_chain:1"
     misc_feature    47131. .51832
                     /note="assembly_fragment:00172
                     fragment_chain:1"
     misc_feature    51933. .59948
                     /note="assembly_fragment:00270
                     fragment_chain:1"
     misc_feature    60049. .64594
                     /note="assembly_fragment:00214
                     fragment_chain:1"
     misc_feature    64695. .73602
                     /note="assembly_fragment:00459
                     fragment_chain:1"
     misc_feature    73703. .76984
                     /note="assembly_fragment:00100
                     fragment_chain:1"
     misc_feature    77085. .79394
                     /note="assembly_fragment:00079
                     fragment_chain:1"
     misc_feature    79495. .81796
                     /note="assembly_fragment:00062
                     fragment_chain:2"
     misc_feature    81897. .87366
                     /note="assembly_fragment:00131
                     fragment_chain:2"
     misc_feature    87467. .104174
                     /note="assembly_fragment:00687
                     fragment_chain:2"
     misc_feature    104275. .113737
                     /note="assembly_fragment:00362
                     fragment_chain:2"
     misc_feature    113838. .140256
                     /note="assembly_fragment:00869
                     fragment_chain:2"
     misc_feature    140357. .149238
                     /note="assembly_fragment:00559.0"
ORIGIN

  Query Match             4.9%;  Score 53.4;  DB 2;  Length 149238;
  Best Local Similarity   44.2%;  Pred. No. 1.3;
  Matches  180;  Conservative    0;  Mismatches  227;  Indels    0;  Gaps    0;

Qy        146 AGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACA 205
              | ||||||| | ||       |   ||  |  ||    | ||| |    |    ||| | |
Db      52477 AAAAATATTTTTTATATTTTTTTTTTTTTTTTTTAAAAAAAAAAAAAAAAATAAAAAAATA 52418

Qy        206 TTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTC 265
              |       |||||||     || |||  |||  || |    | | | | | |   |
Db      52417 ATAAAAAAATAATATAAAATTTTAATTAATTTTTAAAAAAAAAAAAAAAAAAAAAAAAA 52358

Qy        266 ATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATA 325
              .  || |||||  ||| || ||||||   ||| | |||   |   | |  ||    |
Db      52357 TATTTTTTTTTTTTTTTTTAAAAAATTAATATATAAAAAAAAAAAAAATAAANNNAANAN 52298

Qy        326 TTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTC 385
                      |||    || | || |  | |  | |  |  | | ||||  |  |
Db      52297 NNNNANTNNNNAAATANNAAAAAAATAATAATAAAAAAATTTAAATTAATTATAAATAA 52238

Qy        386 AAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAA 445
              |     ||||| | ||   ||   |  | ||  |     |   ||   | |  | ||
Db      52237 ATTATAAAAAAATAAAAAAAATAATTAATTATATATAATTAAATTATAATAAATATTATT 52178

Qy        446 ATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTA 505
              || |    |||| |      |  |||  |    | | || || || || ||| |   ||
Db      52177 ATAAAATAAAATATATAAATAAAAAATAAAAAAAATAAAATTATATTATATATTTAATAT 52118

Qy        506 AAACATTTACTTTCCAAATATTCTGTCATTCAGGAGTATGGAAGTAT 552
              | |   | | |  | ||      | |||  |   |  ||  ||  ||
Db      52117 TTAAAAAAATTATAAATATTAATATTAATTAATAATAATTTAATAAT 52071
```

```
RESULT 82
CR339045/c
LOCUS       CR339045              156978 bp    DNA     linear   HTG 04-MAR-2004
DEFINITION  Danio rerio clone CH211-248G22, *** SEQUENCING IN PROGRESS ***, 10
            unordered pieces.
ACCESSION   CR339045
VERSION     CR339045.2  GI:45126645
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 156978)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-MAR-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Mar 5, 2004 this sequence version replaced gi:44909358.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zC248G22
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 152900 bases at least Q40
            Consensus quality: 153614 bases at least Q30
            Consensus quality: 154373 bases at least Q20
            Insert size: 156078; sum-of-contigs
            Insert size: 161304; 4.6% error; agarose-fp
            Quality coverage: 9.83x in Q20 bases; sum-of-contigs Quality
            coverage: 9.66x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    14299: contig of 14299 bp in length
            *    14300    14399: gap of 100 bp
            *    14400    19363: contig of 4964 bp in length
            *    19364    19463: gap of 100 bp
            *    19464    25301: contig of 5838 bp in length
            *    25302    25401: gap of 100 bp
            *    25402    36604: contig of 11203 bp in length
            *    36605    36704: gap of 100 bp
            *    36705    40222: contig of 3518 bp in length
            *    40223    40322: gap of 100 bp
            *    40323    52179: contig of 11857 bp in length
            *    52180    52279: gap of 100 bp
            *    52280   123007: contig of 70728 bp in length
            *   123008   123107: gap of 100 bp
            *   123108   133219: contig of 10112 bp in length
            *   133220   133319: gap of 100 bp
            *   133320   154448: contig of 21129 bp in length
            *   154449   154548: gap of 100 bp
            *   154549   156978: contig of 2430 bp in length.
FEATURES             Location/Qualifiers
     source          1. .156978
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-248G22"
                     /clone_lib="CHORI-211"
     misc_feature    1. .14299
```

```
                     /note="assembly_fragment:00827
                     fragment_chain:1
                     clone_end:T7
                     vector_side:left"
     misc_feature    14400. .19363
                     /note="assembly_fragment:00122
                     fragment_chain:1"
     misc_feature    19464. .25301
                     /note="assembly_fragment:00192
                     fragment_chain:1"
     misc_feature    25402. .36604
                     /note="assembly_fragment:00632
                     fragment_chain:1"
     misc_feature    36705. .40222
                     /note="assembly_fragment:00077
                     fragment_chain:1"
     misc_feature    40323. .52179
                     /note="assembly_fragment:00290
                     fragment_chain:2"
     misc_feature    52280. .123007
                     /note="assembly_fragment:01459
                     fragment_chain:2"
     misc_feature    123108. .133219
                     /note="assembly_fragment:00450
                     fragment_chain:2"
     misc_feature    133320. .154448
                     /note="assembly_fragment:01084
                     fragment_chain:3"
     misc_feature    154549. .156978
                     /note="assembly_fragment:00033
                     fragment_chain:3"
ORIGIN

  Query Match             4.9%;  Score 53.4;  DB 2;  Length 156978;
  Best Local Similarity   42.4%;  Pred. No. 1.3;
  Matches  180;  Conservative    0;  Mismatches  243;  Indels    2;  Gaps    1;

Qy        113 TAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTT 172
              || ||  | ||   | |  | |   |||  |    | ||||  ||   || |||  | |
Db     123615 TAATTAATAAATATTTTTTTTTTTTTTTTATATNTATATATATATATATATAAAAAATA 123556

Qy        173 AATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATT 232
              |||  | | || || |     |    | | | ||   | | |  |||||      | . |
Db     123555 TATATATATATANAAAAANATAATANATATATATATATATATATATATATATATATAT 123496

Qy        233 TTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGG 292
              | |  |   |  ||| |   ||| |  |||  || | |   |   |   | | | |
Db     123495 ATATATATTTATTTTATATTTATATATATATATATATATAAATATATATATATATATATA 123436

Qy        293 TATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAAC--ACAA 350
              |||| | | ||  ||  | |  || |    | |  |   |    |||| |||   | |
Db     123435 TATATATATATATATATAAAANAATAANNNANANAAAATATAAAAAAAAAAAAANTAAAAN 123376

Qy        351 AATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGT 410
              ||| | |  | |  |  ||  |  || ||      |||   |||||   ||   |||
Db     123375 AATNAATTTAAAAAAAANTTNTTNNATTTTNTNNNAAAAAAAAAAAAAAAAAAAAAAAAN 123316

Qy        411 ATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAA 470
              | |  |    | | |      | ||| | |||  | | |    |   |   |||
Db     123315 ANAANAATATATATNTATATNATATATAAAAATAAAAAAATAAATTAATATAATATAAAA 123256

Qy        471 TAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTG 530
              |||  |     || | ||||||                    |   |      | || |
Db     123255 AAATNTTTTTTAATAAAATTTTNNNANNNNNNNNNNTNNNNNTNNNTNNNTTTTTTTTT 123196

Qy        531 TCATT 535
              | | |
Db     123195 TTAAT 123191


RESULT 83
BX908786/c
LOCUS       BX908786              194354 bp    DNA     linear   HTG 23-AUG-2004
DEFINITION  Danio rerio clone CH211-13O8, *** SEQUENCING IN PROGRESS ***, 9
            unordered pieces.
ACCESSION   BX908786
VERSION     BX908786.5  GI:51507706
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 194354)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-AUG-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 23, 2004 this sequence version replaced gi:48675589.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zC13O8
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 190992 bases at least Q40
            Consensus quality: 191356 bases at least Q30
            Consensus quality: 191641 bases at least Q20
            Insert size: 193554; sum-of-contigs
            Insert size: 170624; 10.7% error; agarose-fp
            Quality coverage: 9.30x in Q20 bases; sum-of-contigs Quality
            coverage: 10.61x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    25786: contig of 25786 bp in length
            *    25787    25886: gap of 100 bp
            *    25887    44730: contig of 18844 bp in length
            *    44731    44830: gap of 100 bp
            *    44831    51520: contig of 6690 bp in length
            *    51521    51620: gap of 100 bp
            *    51621    71285: contig of 19665 bp in length
            *    71286    71385: gap of 100 bp
            *    71386    82471: contig of 11086 bp in length
            *    82472    82571: gap of 100 bp
            *    82572   115645: contig of 33074 bp in length
            *   115646   115745: gap of 100 bp
            *   115746   159198: contig of 43453 bp in length
            *   159199   159298: gap of 100 bp
            *   159299   191741: contig of 32443 bp in length
            *   191742   191841: gap of 100 bp
            *   191842   194354: contig of 2513 bp in length.
FEATURES             Location/Qualifiers
     source          1. .194354
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-13O8"
                     /clone_lib="CHORI-211"
     misc_feature    1. .25786
                     /note="assembly_fragment:00910
                     fragment_chain:1
                     clone_end:SP6
                     vector_side:left"
     misc_feature    25887. .44730
```

```
                     /note="assembly_fragment:00617
                     fragment_chain:1"
     misc_feature    44831. .51520
                     /note="assembly_fragment:00058
                     fragment_chain:1"
     misc_feature    51621. .71285
                     /note="assembly_fragment:00331
                     fragment_chain:1"
     misc_feature    71386. .82471
                     /note="assembly_fragment:00156
                     fragment_chain:1"
     misc_feature    82572. .115645
                     /note="assembly_fragment:01310
                     fragment_chain:1"
     misc_feature    115746. .159198
                     /note="assembly_fragment:02401
                     fragment_chain:1"
     misc_feature    159299. .191741
                     /note="assembly_fragment:01801
                     fragment_chain:1"
     misc_feature    191842. .194354
                     /note="assembly_fragment:00023
                     fragment_chain:1
                     clone_end:T7
                     vector_side:right"
ORIGIN

  Query Match             4.9%;  Score 53.4;  DB 2;  Length 194354;
  Best Local Similarity   44.0%;  Pred. No. 1.2;
  Matches  264;  Conservative    0;  Mismatches  334;  Indels    2;  Gaps    1;

Qy         42 TCTTATAAGCAAATGTTACAATGAGATATGAGTTTTAAAGCCAGATCTTCCTTATCTCTC 101
              | ||| ||  || |  |  |   |  ||| |    || |      |   |  | || |
Db     116401 TTTTAAAATTAATTTATTTATATATTTATAAAAAATATATAATTAATATATTAATATAAT 116342

Qy        102 TGCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATT 161
              |    ||    |  | |  || |   | |||| |  ||  |  | |||  |  |  | |
Db     116341 TTATATATAATAATATATAAAAATTAATAATATAATATTAATATATAAAATTAAAATAAT 116282

Qy        162 ACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATA 221
              |  |     |||||| || | ||  | || |    ||| |   |  |  ||| | |||  |
Db     116281 AATATATTATTAATAATTATTTATTAAATATATAATTATAATAAATAATTTATAAATTAA 116222

Qy        222 CTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTG 281
                |  ||  ||| || |      || |   ||| |    ||  ||||  | ||  |
Db     116221 TAATATATATAATTTAAATAATTAATATATNAATATATAAATATATTTAATATTAATAAT 116162

Qy        282 TTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAAT 341
              || ||  |   |  || |||   ||  | | |  ||    || ||       |||||
Db     116161 TTAAATTA--TAAATAATAAAATAATAAAAAATATTATATATTAATAATATAAATAAAAT 116104

Qy        342 TAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAA 401
              | | | || ||   | | |  |   ||   ||       |   ||   |||| | | ||
Db     116103 AATATATAATATATATATAAAATTATATATAAAAAAAATAAAATAATTTTAAATATTTAA 116044

Qy        402 CCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTA 461
                || |  |||| ||      |||  |||   | |  || || | || |  |  |   |
Db     116043 ATTATATATATATTATTATATTTAATATAATAATTAAATTAAATATAAAAAATATAAATT 115984

Qy        462 TTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCA 521
              || |  | ||||| |   | |  |||||     ||    |||| ||   |  |  |   |
Db     115983 TTATAAATATAATTTATTATATATAATAATAAAAAATTATTTATTTAATATTAAATATAATA 115924

Qy        522 AATATTCTGTCATTCAGGAGTATGGAAGTATCGATGGCTTCTTTAAAATGAAGCAGGAGG 581
                  | ||  |  | ||    | ||   |      | | |||  || ||     |
Db     115923 TTATAAAATTAATATATTAATAATTTATTAAATAAAAAATTTATATTATAAATTTAAATA 115864

Qy        582 GTCTGGCAGAGAGTATCTATGAAATAAGTTCCTCTGACCTTCACGCTTAATTTTCTGAAT 641
              |  |   |      | ||| ||| || ||          || |   |||||  | |  ||
Db     115863 TTATAATAATATAATTTTATAAAANAAATTTAATAATAATTAATATTTAATAATATNTAT 115804

RESULT 84
```

```
AC006281
LOCUS       AC006281              199551 bp    DNA     linear   HTG 12-AUG-2000
DEFINITION  Plasmodium falciparum chromosome 12 clone 3D7, *** SEQUENCING IN
            PROGRESS ***, 2 unordered pieces.
ACCESSION   AC006281
VERSION     AC006281.8  GI:9797738
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Plasmodium falciparum (malaria parasite P. falciparum)
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 199551)
  AUTHORS   Hyman,R.W., Fung,E.L., Qin,F., Rowley,D., Mao,J., Tamaki,T.,
            Kurdi,O.B., Conway,A.B. and Davis,R.W.
  TITLE     Plasmodium falciparum 3D7 chromosome 12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 199551)
  AUTHORS   Hyman,R.W., Fung,E.L., Qin,F., Tamaki,T., Kurdi,O.B. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JAN-1999) Stanford DNA Sequencing and Technology
            Center, Stanford University, 855 California Avenue, Palo Alto, CA
            94304, USA
COMMENT     On Aug 12, 2000 this sequence version replaced gi:8810457.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    41526: contig of 41526 bp in length
            *   41527    41726: gap of unknown length
            *   41727   199551: contig of 157825 bp in length.
FEATURES             Location/Qualifiers
     source          1. .199551
                     /organism="Plasmodium falciparum"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:5833"
                     /chromosome="12"
                     /clone="PFYACB8-95"
                     /clone="3D7"
ORIGIN

  Query Match             4.9%;  Score 53.4;  DB 2;  Length 199551;
  Best Local Similarity   48.3%;  Pred. No. 1.2;
  Matches  215;  Conservative    0;  Mismatches  221;  Indels    9;  Gaps    2;

Qy        129 TCATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAG 188
              | ||||   ||   |  | | ||||  ||   || | ||   |  |||  |||  |   |
Db     144713 TAATATATATTATATATATATATATATATATATATATATATATTTATTATAT 144772

Qy        189 ATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTA 248
              || ||||   ||  |    || |  || || |  ||   |  | |||   |  ||   |
Db     144773 ATATCATGAAAACAATTAAGTGTATTTTATTTTTTCATATTTTATTTCCTACGGAAGAAA 144832

Qy        249 CAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGA-------- 300
               ||    | | || ||| ||| |||||  ||| || | | ||  || | | |
Db     144833 AAGTGTAAAAACTTTTCTTTTTTTTTTTTTTTTTTAAGACAAATTAAATATAATTATTAT 144892

Qy        301 AATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCA 360
              |||    | | |||||||  | | |  |  |    ||||  |||  || |||   |  |
Db     144893 AATATTGTTATAGAAAGAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAATCATAATATTTATA 144952

Qy        361 TATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTG 420
              | ||||| |  | |   |  ||  |||      ||    | | |||     | ||| |
Db     144953 GAATACATAAAATTAAAATAAAGAAAAACACTTGAATAAGAACAAAAAAGTAAATACTTT 145012

Qy        421 CTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGA 480
              ||||  |||      |  ||||||  ||| |  | |  | || |    ||||||  |  |
Db     145013 TTTTC-TATTTTTTTTTAAATAGTAATAATTATATGTATTTTAAATCAAATAAATTATGT 145071

Qy        481 AAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGTCATTCAGGA 540
              |    |||  ||  |||  |  ||  |  |  |  |     ||  ||  |    ||
```

```
Db     145072 ATTAGTAAGTTAATATTCATAATAATTATTCTAATATATTTTATTTTTTTATTTTTTATT 145131

Qy        541 GTATGGAAGTATCGATGGCTTCTTT 565
              |        |||   | | ||||||
Db     145132 TTTATTCCTTATTTTTTGTTTCTTT 145156


RESULT 85
CR790362/c
LOCUS       CR790362              223166 bp    DNA     linear   HTG 05-OCT-2004
DEFINITION  Danio rerio clone DKEY-219H14, *** SEQUENCING IN PROGRESS ***, 13
            unordered pieces.
ACCESSION   CR790362
VERSION     CR790362.3  GI:53828288
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 223166)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 5, 2004 this sequence version replaced gi:52626676.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK219H14
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 218162 bases at least Q40
            Consensus quality: 219035 bases at least Q30
            Consensus quality: 219848 bases at least Q20
            Insert size: 221966; sum-of-contigs
            Insert size: 231484; 3.1% error; agarose-fp
            Quality coverage: 9.07x in Q20 bases; sum-of-contigs Quality
            coverage: 9.03x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 13 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *      1    45629: contig of 45629 bp in length
            *  45630    45729: gap of 100 bp
            *  45730    47743: contig of 2014 bp in length
            *  47744    47843: gap of 100 bp
            *  47844    71492: contig of 23649 bp in length
            *  71493    71592: gap of 100 bp
            *  71593    80196: contig of 8604 bp in length
            *  80197    80296: gap of 100 bp
            *  80297    94270: contig of 13974 bp in length
            *  94271    94370: gap of 100 bp
            *  94371   105308: contig of 10938 bp in length
            * 105309   105408: gap of 100 bp
            * 105409   113062: contig of 7654 bp in length
            * 113063   113162: gap of 100 bp
            * 113163   125638: contig of 12476 bp in length
            * 125639   125738: gap of 100 bp
            * 125739   129703: contig of 3965 bp in length
            * 129704   129803: gap of 100 bp
            * 129804   143077: contig of 13274 bp in length
            * 143078   143177: gap of 100 bp
            * 143178   148174: contig of 4997 bp in length
            * 148175   148274: gap of 100 bp
            * 148275   180179: contig of 31905 bp in length
            * 180180   180279: gap of 100 bp
            * 180280   223166: contig of 42887 bp in length.
FEATURES             Location/Qualifiers
     source          1. .223166
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-219H14"
                     /clone_lib="DanioKey"
     misc_feature    1. .45629
                     /note="assembly_fragment:02641
                     fragment_chain:1"
     misc_feature    45730. .47743
                     /note="assembly_fragment:00088
                     fragment_chain:1"
     misc_feature    47844. .71492
                     /note="assembly_fragment:01125
                     fragment_chain:1"
     misc_feature    71593. .80196
                     /note="assembly_fragment:00289
                     fragment_chain:1"
     misc_feature    80297. .94270
                     /note="assembly_fragment:00715
                     fragment_chain:1"
     misc_feature    94371. .105308
                     /note="assembly_fragment:00568
                     fragment_chain:2"
     misc_feature    105409. .113062
                     /note="assembly_fragment:00191
                     fragment_chain:2"
     misc_feature    113163. .125638
                     /note="assembly_fragment:00919
                     fragment_chain:2"
     misc_feature    125739. .129703
                     /note="assembly_fragment:00071
                     fragment_chain:3"
     misc_feature    129804. .143077
                     /note="assembly_fragment:00426
                     fragment_chain:3"
     misc_feature    143178. .148174
                     /note="assembly_fragment:00118.0"
     misc_feature    148275. .180179
                     /note="assembly_fragment:01481"
     misc_feature    180280. .223166
                     /note="assembly_fragment:01968"
ORIGIN

  Query Match             4.9%;  Score 53.4;  DB 2;  Length 223166;
  Best Local Similarity   43.2%;  Pred. No. 1.2;
  Matches  162;  Conservative    0;  Mismatches  213;  Indels    0;  Gaps    0;

Qy        187 AGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCT 246
              | ||   ||  ||||   | |  | | |||    | |    | || || ||  ||
Db     180091 ATATTAAATTATAAAATTAATTATATATTATATAAATAATATNATNTTATTTTAATTTAA 180032

Qy        247 TACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCA 306
              || | |  |     |||| || | | | |    | ||  |  ||  || ||  | ||  |
Db     180031 TATAAAATTTATTATATTAATATATATATAATTTTATANTAAAATAAAATAATATATA 179972

Qy        307 TTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTAC 366
              |  |   | | |  ||||| ||   | ||   |  |||| | |  |      ||||
Db     179971 TATAATAATAAAAAATATAATATTAATTTATTATATAAAAATTATTTATAAAAATATNTA 179912

Qy        367 ATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCT 426
              | ||  |||| ||   | |    |   || | |  ||   | ||  | |  ||  |
Db     179911 AATATATAATTTTATTTTAATTTATTTNAATTTTATTAAATATTTTATAAATTTATTNTT 179852

Qy        427 TATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACAT 486
              |||   | |    | |    ||   |   | |||||||   ||| || | |  ||| ||
Db     179851 TATTTAAATATTTAAATTTTTTTTTATTTAAATTATTTTTTTAATTATTTTAATAAATAT 179792
```

```
Qy        487 AATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGTCATTCAGGAGTATGG 546
              ||||   |||      ||  |  | |   |||    |   |     | |     |
Db     179791 AATTAAATATAATAATTTTTATAAATATATTTANTTAATNTNNNNNANTANNNNNTNNNN 179732

Qy        547 AAGTATCGATGGCTT 561
                  ||    |
Db     179731 NNNNTAANATNNNNT 179717
```

RESULT 86
AE014845/c
LOCUS AE014845 250531 bp DNA linear INV 11-FEB-2003
DEFINITION Plasmodium falciparum 3D7 chromosome 12, section 2 of 9 of the complete sequence.
ACCESSION AE014845 AE014188
VERSION AE014845.1 GI:23496588
KEYWORDS .
SOURCE Plasmodium falciparum 3D7
ORGANISM Plasmodium falciparum 3D7
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 250531)
AUTHORS Gardner,M.J., Hall,N., Fung,E., White,O., Berriman,M., Hyman,R.W., Carlton,J.M., Pain,A., Nelson,K.E., Bowman,S., Paulsen,I.T., James,K., Eisen,J.A., Rutherford,K., Salzberg,S.L., Craig,A., Kyes,S., Chan,M.S., Nene,V., Shallom,S.J., Suh,B., Peterson,J., Angiuoli,S., Pertea,M., Allen,J., Selengut,J., Haft,D., Mather,M.W., Vaidya,A.B., Martin,D.M., Fairlamb,A.H., Fraunholz,M.J., Roos,D.S., Ralph,S.A., McFadden,G.I., Cummings,L.M., Subramanian,G.M., Mungall,C., Venter,J.C., Carucci,D.J., Hoffman,S.L., Newbold,C., Davis,R.W., Fraser,C.M. and Barrell,B.
TITLE Genome sequence of the human malaria parasite Plasmodium falciparum
JOURNAL Nature 419 (6906), 498-511 (2002)
MEDLINE 22255705
PUBMED 12368864
REFERENCE 2 (bases 1 to 250531)
AUTHORS Hyman,R.W., Fung,E., Conway,A., Kurdi,O., Mao,J., Miranda,M., Nakao,B., Rowley,D., Tamaki,T., Wang,F. and Davis,R.W.
TITLE Direct Submission
JOURNAL Submitted (13-SEP-2002) Stanford Genome Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA
REFERENCE 3 (bases 1 to 250531)
AUTHORS Hyman,R.W., Fung,E., Conway,A., Kurdi,O., Mao,J., Miranda,M., Nakao,B., Rowley,D., Tamaki,T., Wang,F. and Davis,R.W.
TITLE Direct Submission
JOURNAL Submitted (29-JAN-2003) Stanford Genome Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA
FEATURES Location/Qualifiers

```
     source          1. .250531
                     /organism="Plasmodium falciparum 3D7"
                     /mol_type="genomic DNA"
                     /isolate="3D7"
                     /db_xref="taxon:36329"
                     /chromosome="12"
     repeat_region   156. .206
                     /rpt_type=tandem
     repeat_region   284. .328
                     /rpt_type=tandem
     repeat_region   604. .629
                     /rpt_type=tandem
     repeat_region   766. .827
                     /rpt_type=tandem
     repeat_region   982. .1022
                     /rpt_type=tandem
     repeat_region   1150. .1201
                     /rpt_type=tandem
     repeat_region   1405. .1434
                     /rpt_type=tandem
     repeat_region   1467. .1511
                     /rpt_type=tandem
     gene            <1661. .>2629
                     /locus_tag="PFL0285w"
     mRNA            <1661. .>2629
                     /locus_tag="PFL0285w"
     CDS             1661. .2629
                     /locus_tag="PFL0285w"
                     /note="HMMPfam hit to PF00753, Metallo-beta-lactamase superfamily, score 5.4e-19"
                     /codon_start=1
                     /product="glyoxalase II family protein, putative"
                     /protein_id="AAN36146.1"
                     /db_xref="GI:23496589"
                     /translation="MRFLKTLFSLNCLASVGFYKKFACKNLFFENSFIRKDLHTYFID
                     KRKYFHLNKKDICTNTIIIPFYKDNYSYIFYDDKEEGIVVDPADYNIINDISKKENIK
                     IKHVLCTHKHSDHNNGNQYYYEKNINVYGIKEYDNKYINQDISNLTHFQINNFKINIF
                     LSNFHSKNQVSYLIENDNNKSKKNIFFTGDFLFISGIGKNFEQDNEDLYNSINKLKLL
                     DKQNIYIFCGHEYTLDNLKFALTVDSTNKNLLSFYDHVVNSNKNYPTVPTLLEHEYLY
                     NPFLRCDQNDVRKSIDLYAKKKNIKIQQESDYIVILRLMKDNFKAS"
     repeat_region   2728. .2760
                     /rpt_type=tandem
     repeat_region   2787. .2826
                     /rpt_type=tandem
     repeat_region   3829. .3868
                     /rpt_type=tandem
     repeat_region   4261. .4302
                     /rpt_type=tandem
     repeat_region   4794. .4857
                     /rpt_type=tandem
     gene            <4873. .>12237
                     /locus_tag="PFL0290w"
     mRNA            join(<4873. .10503,10697. .10856,11016. .11921,12080. .>12237)
                     /locus_tag="PFL0290w"
     CDS             join(4873. .10503,10697. .10856,11016. .11921,12080. .12237)
                     /locus_tag="PFL0290w"
                     /note="match to ESTs N97875.1, BI814869.1"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAN36147.1"
                     /db_xref="GI:23496590"
                     /translation="MDGAIELLPNVKNDHFLENKNLINEYNKLYMHRYYHINNHTTNN
                     DNIYNDIICNNNYNISNIHTCNSNISNHTVLGCTSQRKSFNKYKFLLNLSKTLLSKKY
                     LSINSSTRFEKKNVKNKDFKNKNSKNTIIKEKAKSYMLSLLNKKICNHLYEENKIKQS
                     NEEKVKGPTENYDPIYNLKINNKNEDQEINEQNKSVLNTNKDAHKYVNHVNGKYLNNY
                     NVSESMGDNTIILGVYNNTKDIKNYELKDKTKNESINFDMRNIQLCNADIEEGERVTD
                     RNNTNDTNSNYSKKNEDIEDHKNDNNNDDNEDDNNDDGNYDDGNYDDGNYDDGNNNDG
                     NNDDGNNNDGNNDDDNNDDGNNDDDNNDDDNNDDDNNDDDNNDDGNNDDDNND
                     DDDNNNNDNNNNDNNNNDNNNNNNNNNNNNNNNNNNDDDNNKKNNEYKTYYEKIKELKSN
                     VFIKEYKNSEKFVKKKILKYICKNEIIENVSNVFNYVKNIYFVIIKNEKMMFYFLYNT
                     KDVLNNIIYMENENKVSYNDSKFLLYFLLNEEEIYKKLFEKLNLDIEKERLISKLKKI
                     NLKFNELKYDFEKNLHKNNIHQIYFMKKRKIFINNMKGLQFNKNSFNTSNATTISGLK
                     MVSLKNVIPKLYNCKDYKKTNHFFFNMYENISKENNGVDNVNEAKKKNTSHQNISKKG
                     EEKYLPPEEQTKKRKKKTTQKNNKKNNMNNMNNMNNMNNMNIMNNMNNMNNMNNMNNM
                     NNMNNMNNMNNMKHDNFTFDGNNKPIANINNNNVNTNIINNNNNDKHNNNDDNNNDDN
                     NNDDNNNDDNNNNNNNNKEEVVKKPAIRGITSFTLFAREKRKELLDQKIYLGSSLTEQ
                     TSAVAKIWNNLSDEQKKEWAVKASKINEENFLLQKKKKKRKFTAFSIFAREKRKEYKE
                     KNIDMGLTLAQQNSHVSKLWKQLTAEEKNKYKVLSNTVNATIAKAYSKSESGGTTGER
                     GSGFKGRIKALDNMMSQHYANNNILNNDSNTKNDETVKKSNMDNININENINVSHGKN
                     VTNIENVANVQNMQNIKNIQNIQNMQNMQNIQNGHINNNHISYGANSNVATMHHTNNT
                     HMAHINYNYYDQTNIGNNNNNNYTMNMSLPANQNITDKPAMELIPTINVLHPYANNFN
                     GNVCNTSTNVINNNMNINNMNSMNNINNINNNHNSSSGSSAIAPPCDIMTDDHLNSAA
                     TTTGTISTSYNGMTTNNMNNKNSNQVMIMMVNKELNQINIKNNNSNVDMHQLKGDNNM
                     NNNNILMDNNLVNGDVINNQNINNNINFNTFQMNYKNVEHTHINNDPYQKIDVTQVDD
                     LILKKKLKEDEKLKVKEKKNILRIFDKQRKQLLKEEKKMDKIKINENQYNNTTLYYPT
                     THLMNNNLNNNHYDISTNQYINEPNTNHMNYYTNIDNKMVTTNVANNIMNFNNVNNVN
                     VTNNNMGNIHMATINNIDNVNNVNHINNMNNGNTFNYSNHYVPYNHVNHVVVNRKDKN
                     IQMNNNNNNNNNNNNMFVNNNNKNAPISNYNKYQHDNMNNNILKNDQGNNNMNNYNNYT
                     YNNIYFNGTVNHGKMPPVQNGNAIKHMDEKEININLNVPTKKLEKKNFNKIMNINEET
                     LYKNNNELNYIQSTLHYNQMSDNNINNKMIPAHHIHKDNMNNMNNMNNMNNMNNMNNM
                     NNMNNINNINNMNNMNNINNINNMNNMNNMNNINNMNNINKHFDSHVLKKRANDINPN
                     DANYFNEKLKIQNSLNDDFFDFTTTNNNCETMHSTTTQENINLNNTINNDLNMDINSN
                     LNLNYTTNELKKNNTYYYPYNKNNKNNDVNHIYDTHQSERNIYHKNNMNNNNHIINND
                     NNYSALVNTKNFIYYPEQNFSQTIGKENQHIYYDTDLDKPERTLQDTEIVRPQNKLFF
                     NNASHFLELVNNKIIKKKKNHKSSLNSVFNEETCTIWDLVMKNKKKKVKKKRDKKRER
                     DGCGNDNDKELNEMRNIEYDYNNFDLMMLNNDLNNYLKDTNRSENNEDDIFNSILNNI
                     DDNNIYNNSNKEVAILNNKNNVVEENRKNDNTDNLPEDDNNNCDDFLNLLNIGSSNFS
```

```
                NKLNINEHGDVIYDEKKRLGANNLINENNNNNDKLGLLMEKKNEINQNLNKSQIEKYN
                YAYKKSKLCKWTDEETDKFYQAIEMFGIDLMMVRALIPKFNDKQIRDKYKKEKKSNPL
                KIEQAIKKNKEINLDAYENEHGKIDHSTHYMYSDYSSTDEKNDNNNNAKAFTNVTGES
                EINILTIFDDKNENANIDDEYATNEEPSEFNVLTFF"
    repeat_region   5780. .6053
                    /rpt_type=tandem
    repeat_region   5780. .5960
                    /rpt_type=tandem
    repeat_region   5935. .6170
                    /rpt_type=tandem
    repeat_region   6056. .6170
                    /rpt_type=tandem
    repeat_region   6997. .7121
                    /rpt_type=tandem
    repeat_region   7209. .7319
                    /rpt_type=tandem
    repeat_region   7232. .7317
                    /rpt_type=tandem
    repeat_region   7986. .8061
                    /rpt_type=tandem
    repeat_region   9815. .9998
                    /rpt_type=tandem
    repeat_region   12315. .12401
                    /rpt_type=tandem
    repeat_region   13069. .13103
                    /rpt_type=tandem
    gene            complement(<13121. .>16369)
                    /locus_tag="PFL0295c"
    mRNA            complement(<13121. .>16369)
                    /locus_tag="PFL0295c"
    CDS             complement(13121. .16369)
                    /locus_tag="PFL0295c"
                    /note="Similar to Mus musculus 4632413c14riK protein
                    4632413c14riK TR:Q9D675 (EMBL:AK014571) (631 aa) fasta
                    scores: E(): 2.4e-30, 35.337% id in 416 aa;
                    match to EST BI814307.1"
                    /codon_start=1
                    /product="hypothetical protein, conserved"
                    /protein_id="AAN36148.1"
                    /db_xref="GI:23496591"
                    /translation="MLKNNFNPDISDDDYVSEKQIFQGSNKTILTVESIHVVKGIDKD
                    IKKQIEEGKVFLSDISPVKNAEIEEAFLKSLCKSSSTCYDNNNNNNKIFHHHFNDKKK
                    DYSNENSSEEININIEENYDRFNIDKSKYYNNDNLLSYNKKNKNNDKDDGIDDNVLIR
                    PPSDLQEHKDDEISSQYFKSYSFVKNLSSNIFEKYTFNDLPDLLEQETINYIKKQERL
                    LQKKRLNANLEALRLQYIPINNDKFVTIPTDIKENILALFACIIIKHIDSFEDEDIQN
                    HLDELCKNFLKQLADLLVIDPIAYQCVIKPFDQRKILREVTGDKTPYVLYEKPFVDNI
                    LEKIGSLQANIDAKFKNIVEELNNEMDFWQNDHIDDEQSDQDKEYLKNEMLLEHENLK
                    IEDNKLYIKNNISPTKTFPEREIDDHYKEEAINDCFTINNKNEEKIVKKINKKHDTKK
                    GNQINDSNGCTNIGECNNDPLNVKLVDKNKLDTSKVENKDNDENKDDNDENKDDNDEN
                    KDDNDDNKDNDDNMDDDNPNEVKECNKKKKKKNKKKKKKKKKIKDNPNGRKEKDKLDDL
                    TEEDKIKSNGCYEKVKKKNEEFGDFIKKDSNTSEQLNNSNNSLNNLENNKIRKNINEK
                    NVKHDMINDADILSKIKSHIVNKMDIYESESYYDESDMEKHLNHGVPIKAPEVKSVLL
                    RDLITTYIMTGNNNARIHLFFQKFAKCLKINELLILRIEENLAADLISALQASTNETA
                    KRRNIRRMKIAAAALGGGALIAFTAGLAAPGIIAGLTALGAGGSSLTAFLASASGLAF
                    IVSLFGAGGAGLTGYKYSRRIANIKTFEFIMLNGNISKSLSVCVCVSGEIKSDDDITN
                    PWLEVFPNCYCDLYCLKWENHLLKTLGSLIETMLSQEFAITASRIWLQYTIASTLSAA
                    LTWPLALIKYASNLDNVYLLIRERAQQAGRILADALSDRNTVGQRPVILIGYSVGARV
                    IFYCLKYLYSKKLYNIVSNAIFIGLPATTSTKVWEKIRMVVTNRVINVYSKNDWLLGF
                    LYRYMEWKINVAGLIAVKVPNVENYDASGIIHSHLDYKRKLKDIFHLINFDM"
    repeat_region   14724. .14767
                    /rpt_type=tandem
    repeat_region   14798. .14902
                    /rpt_type=tandem
    repeat_region   14838. .14901
                    /rpt_type=tandem
    repeat_region   16383. .16409

Query Match             4.9%;  Score 53.4;  DB 3;  Length 250531;
Best Local Similarity   48.3%;  Pred. No. 1.2;
Matches  215;  Conservative    0;  Mismatches  221;  Indels    9;  Gaps    2;

Qy        129 TCATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAG 188
              | ||||   ||   |  | | |||||  ||   || | ||   |  |||  ||| |   |

Db        829 TAATATATATTATATATATATATATATATATATATATATATATATATATATATTTATTATAT 770

Qy        189 ATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTA 248
              || ||||   ||  |    || |  || || |  ||   |  | |||   | ||    |

Db        769 ATATCATGAAAACAATTAAGTGTATTTTATTTTTTCATATTTTATTTCCTACGGAAGAAA 710

Qy        249 CAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGA-------- 300
               ||    | | || ||| ||| |||||  ||| || | | ||  || | | |

Db        709 AAGTGTAAAAACTTTTCTTTTTTTTTTTTTTTTTTAAGACAAATTAAATATAATTATTAT 650

Qy        301 AATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCA 360
              |||    | | |||||||  | | |  |    |    ||||  |||  || |||   | |

Db        649 AATATTGTTATAGAAAGAATAAAAAAAAAAAAAAAAAAAAAAAAAAAATCATAATATTTATA 590

Qy        361 TATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTG 420
              | ||||| |  | |   |  ||  |||       ||    | | |||     | ||| |

Db        589 GAATACATAAAATTAAAATAAAGAAAAACACTTGAATAAGAACAAAAAAGTAAATACTTT 530

Qy        421 CTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGA 480
              |||| |||      |  ||||||  ||| |  | |  | || |     ||||||  |  |

Db        529 TTTTC-TATTTTTTTTTAAATAGTAATAATTATATGTATTTTAAATCAAATAAATTATGT 471

Qy        481 AAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGTCATTCAGGA 540
              |    ||| ||  |||  |  || |  |   | |  |     || || |    ||

Db        470 ATTAGTAAGTTAATATTCATAATAATTATTCTAATATATTTTATTTTTTTATTTTTTATT 411

Qy        541 GTATGGAAGTATCGATGGCTTCTTT 565
              |      |||   | | ||||||

Db        410 TTTATTCCTTATTTTTTGTTTCTTT 386


RESULT 87
PFA929358/c
LOCUS       PFA929358            254050 bp    DNA     linear   INV 29-JAN-2003
DEFINITION  Plasmodium falciparum strain 3D7, chromosome 9; segment 4/5.
ACCESSION   AL929358 AL844508
VERSION     AL929358.1  GI:23505147
KEYWORDS    .
SOURCE      Plasmodium falciparum 3D7
  ORGANISM  Plasmodium falciparum 3D7
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 254050)
  AUTHORS   Hall,N., Pain,A., Berriman,M., Churcher,C., Harris,B., Harris,D.,
            Mungall,K., Bowman,S., Atkin,R., Baker,S., Barron,A., Brooks,K.,
            Buckee,C.O., Burrows,C., Cherevach,I., Chillingworth,C.,
            Chillingworth,T., Christodoulou,Z., Clark,L., Clark,R., Corton,C.,
            Cronin,A., Davies,R., Davis,P., Dear,P., Dearden,F., Doggett,J.,
            Feltwell,T., Goble,A., Goodhead,I., Gwilliam,R., Hamlin,N.,
            Hance,Z., Harper,D., Hauser,H., Hornsby,T., Holroyd,S.,
            Horrocks,P., Humphray,S., Jagels,K., James,K.D., Johnson,D.,
            Kerhornou,A., Knights,A., Konfortov,B., Kyes,S., Larke,N.,
            Lawson,D., Lennard,N., Line,A., Maddison,M., Mclean,J., Mooney,P.,
            Moule,S., Murphy,L., Oliver,K., Ormond,D., Price,C., Quail,M.A.,
            Rabbinowitsch,E., Rajandream,M.A., Rutter,S., Rutherford,K.M.,
            Sanders,M., Simmonds,M., Seeger,K., Sharp,S., Smith,R., Squares,R.,
            Squares,S., Stevens,K., Taylor,K., Tivey,A., Unwin,L.,
            Whitehead,S., Woodward,J., Sulston,J.E., Craig,A., Newbold,C. and
            Barrell,B.G.
  TITLE     Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13
  JOURNAL   Nature 419 (6906), 527-531 (2002)
  MEDLINE   22255708
   PUBMED   12368867
REFERENCE   2  (bases 1 to 254050)
  AUTHORS   Sanders,M., Hauser,H., Baker,S., Unwin,L., Mungall,K., Berriman,M.,
            Pain,A., Hall,N., Bowman,S., Churcher,C., Quail,M. and Barrell,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-SEP-2002) P.falciparum Genome Sequencing Consortium,
            The Welcome Trust Sanger Institute, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA, UK
COMMENT     For more information about this sequence or the Malaria Project,
            see http://www.sanger.ac.uk/Projects/P_falciparum.
FEATURES             Location/Qualifiers
     source          1. .254050
```

/organism="Plasmodium falciparum 3D7"
/mol_type="genomic DNA"
/isolate="3D7"
/db_xref="taxon:36329"
/chromosome="9"

gene 763. .2532
/gene="PFI1215w"

CDS 763. .2532
/gene="PFI1215w"
/note="Moderately similar to Mus musculus splicing factor 3a, subunit 3, 60kd TR:AAH09141 (EMBL:BC009141) (501 aa) fasta scores: E(): 1.7e-24, 30.756% id in 595 aa. No significant pfam hit; SMART hit to SM00355, zinc finger;ProfileScan hit to PS50171, Zn-finger (C2H2) in matrin and other RNA-binding proteins"
/codon_start=1
/product="splicesome-associated protein, putative"
/protein_id="CAD51929.1"
/db_xref="GI:23505148"
/db_xref="GOA:Q8I2R0"
/db_xref="UniProt/TrEMBL:Q8I2R0"
/translation="MSQFLIEQIRYLHEEIEVIEKAIAELIDEKVKNKKKNILYDYAI
NYLVEKIQEKSQLLLEYYNDEDKLKKEEMLFLSGKTGEDKNDIWKNYYERVKYIRDYH
KKTNIKKIEIKSSKLYKLEALKCSKYKQSFSPLEKKGKYVDLNKFYNDFINMKKIKEF
RISMFKKKEMASSHNRKRKKTGDNNNNNNNNNNVTTANINEFKEMDYVTYLHNFTRFI
YIPRYCKYRNEEYKSYVENLLNYMISFFEKINVLVDCKTTYKEYEDDFYNKFNNKEIK
GWEKYTYDLDFYCNINNKLYASEGTYQSYLKCKNYDNDLKKYLNKKYTVEQLEEIKKN
IEKEDMNLARCEYLIQLYANILNKIIQRTIQNIQRKQAYTIDEIQKKKYKEEEQEDEK
KRKKEYHLLSLMDHNNNELNNQNILLQDLENNNLSISMSDDDSSDMSNISNNEDQQKD
LDEENEDDNDDNKPIYNPLNLPLGFDNKPIPYWLYKLHGLSKEYKCEICGNYSYFGRA
TFEKHFYEWRHSFGMKCLKIPNTLHFKEITKIEDALNLYEKLKKETQMNIFKPDQEVE
CEDSKGNVMNIKAYDDLKRQGLL"

gene join(4370. .4451,4555. .4604,4733. .4964,5141. .5186, 5417. .5444,5735. .5773,5790. .5825)
/gene="PFI1220w"

CDS join(4370. .4451,4555. .4604,4733. .4964,5141. .5186, 5417. .5444,5735. .5773,5790. .5825)
/gene="PFI1220w"
/note="1 probable transmembrane helix predicted for PFI1220w by TMHMM2.0 at aa 74-96"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAD51930.1"
/db_xref="GI:23505149"
/db_xref="UniProt/TrEMBL:Q8I2Q9"
/translation="METFNHIDIKKYTREDLLYLRNTNNVLFKMFQKQVDEIACLSSK
EQKLCGTLTSINNIINENYTIYVNTKKEHIIYIYVYVYVFICLCICSFIYFYMYLILS
CQVLNTHRWLNRIYKSNNKLYMCVRLKIKKMKGSTQRHFCFLFMPNHFVIFSNYFKNY
MREFLINMCI"

gene join(6108. .6182,6392. .6547,6648. .6699,6851. .6903, 7009. .7062,7151. .7204,7343. .7424,7603. .7710,7795. .7853)
/gene="PFI1225w"

CDS join(6108. .6182,6392. .6547,6648. .6699,6851. .6903, 7009. .7062,7151. .7204,7343. .7424,7603. .7710,7795. .7853)
/gene="PFI1225w"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAD51931.1"
/db_xref="GI:23505150"
/db_xref="UniProt/TrEMBL:Q8I2Q8"
/translation="MNEKEKNGIFSVHKEKDNLSTHQKFDLLIDEYYCDLKEEKRERS
TEKNTKEGDFVFKSVDRNKADYAVRNRRKMLHHRLDENDEPPIFTDLIENEKKIGTVR
ITNPLPKIERRDELKDNYMKWLKTERLMRPNERRLEYEQMQEKYLESIKLTHDIKKDV
KLAKAIKSHYPRGIVGVDSIYNENTVLYKDQHEEIKKKYEIKEKRLKEKGEVLEKHNQ
KNGNFLTFQENN"

gene complement(8468. .9508)
/gene="PFI1230c"

CDS complement(8468. .9508)
/gene="PFI1230c"
/note="asparagine-rich
Signal anchor predicted for PFI1230c by SignalP 2.0 HMM (Signal peptide probabilty 0.001, signal anchor probability 0.995) with cleavage site probability 0.001 between residues 27 and 28
4 probable transmembrane helices predicted for PFI1230c by TMHMM2.0 at aa 12-34, 38-56, 63-82 and 97-116"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAD51932.1"
/db_xref="GI:23505151"
/db_xref="UniProt/TrEMBL:Q8I2Q7"
/translation="MEERKWSYIMVSVIFYLCYFFYFLTYSYEIIFLFQTKIDYIVFC
LLFITSLFSIYFCSKGLKYICTFLITIISLIEIYEYTNNIPKDLYDEEFLKLFLLVRL
FSAVNLVVFLYLHIKNLKNIKDEKLNNDKDSTIQSVAEPTNTTTENNHNQNNNHNNNN
IHNNNHNNNHNNHNNIHNNHNDPCNNHNDPCNNYYKNQDNFLKSNTVTYSHNNLFKEN
IPHVHQHTGIITSKTLPQENIYVLKHVSIPIFEKEIINKDTNISHMYEMQTNDSFNHY
QVNNLSSSNNINDNKITEPITTHNLNKNNNAYTNHIIQNNHVPNYTYEHPKGLQYYPG
NFHTHNMGNNIQ"

gene join(11425. .12434,12559. .12655,12774. .12908)
/gene="PFI1235w"

CDS join(11425. .12434,12559. .12655,12774. .12908)
/gene="PFI1235w"
/note="No pfam hit; ScanProsite hit to PS00430; Interpro hit to Ton-B dependant receptor protein"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAD51933.1"
/db_xref="GI:23505152"
/db_xref="UniProt/TrEMBL:Q8I2Q6"
/translation="MTDINNNKYSKLKVKKNKNKKNNNKNNDKNNTLLVSPNDKLKKE
KRKKKHKNLNQKVIHNNNNNNNIHKDEHINEQSNNHIYVNNRTKKKKKNIELIKDKWI
NKKKKKKTKKKTKKKTNNTTIYDNDNNINVDEEEEKKKKNVFPYISVLRENVVNNKIH
IHDTNEAYKYDYADLENNKQKYKKKKKIENSPEDIVNSSLFRYINEYMYTNNSEVVQN
KLNQTKNVFNIYHQGYKNQKNKWPHNPVSIIIKHLKKYFNKNNKIADLGCGEAEIART
LDGWYIKSFDLIQYNHYVTPCNITQLPLNNNSYDCFILSLSLMNTDWPKIIFESVRCL
KKGATLIIAEVVSRFTNYKAFIKFMNNVGFKLSNKINLDDFFYVFFFEKNQEVDISSS
INEKRIKKVSSLLTPCIYKRR"

gene complement(13154. .14893)
/gene="PFI1240c"

CDS complement(13154. .14893)
/gene="PFI1240c"
/EC_number="6.1.1.15"
/note="Similarities to prolyl-t-RNA synthases of bacterial origin; N-terminal signal peptide. Similar to Lactococcus lactis prolyl-trna synthetase pros or l12127 SWALL:Q9CDT4 (EMBL:AE006441) (616 aa) fasta scores: E(): 1.6e-09, 35.21% id in 213 aa, and to Escherichia coli prolyl-trna synthetase pros or drpa or b0194 SWALL:SYP_ECOLI (SWALL:P16659) (572 aa) fasta scores: E(): 7.3e-08, 31.91% id in 235 aa. Pfam match to PF00587, tRNA synthetase class II (G, H, P, S and T)
Signal peptide predicted for PFI1240c by SignalP 2.0 HMM (Signal peptide probabilty 0.720, signal anchor probability 0.018) with cleavage site probability 0.591 between residues 18 and 19"
/codon_start=1
/product="prolyl-t-RNA synthase, putative"
/protein_id="CAD51934.1"
/db_xref="GI:23505153"
/db_xref="GOA:Q8I2Q5"
/db_xref="UniProt/TrEMBL:Q8I2Q5"
/translation="MLIKIIILIGLFHHVILSLKLVNLKKKRSNVFFVNNKILKRNVP
KKKYRILSNKNQSYIGSNYIFNRIFNDNIKSNGDKTSELLQRANYIRDVNGIYNILPL
GFRVINKIIDFLNKHLEKLNSHAMSLSILQAKKLWNISDRSKLYSDEFLYVYKQKQSK
NDKSSEESLKKKIEDDGYILSPTCEESSLSLINQIYNENITIKCLPLLIHQYNYKFRN
EKRFEKSLFKSKEFLMKDGYSFHSNEKCLNETYEKYKECYKNIFEELKLSFNIIKKRK
KDKMNALESHEFQVLSRDGKYKEAAHIFKLGDYYSNKLDIKYLDKKNEKKNILMGSYG
IGIYRLLYFLIENFYDEEGIKLPQQVAPFSVYLIQTNQKSKYSATKISKILNMYKDSM
EKKGNISQQKNEQDDQGDDECGNSTFVSNIHNDDKVNKDDTLNNNNDTLNSNNTLNHD
NGVNINNTLTYSPLTSNDIEYILTLWLFNTFKNKNVDIYYDDTDLHLSRKLKHCDLIG
APNRIIINLSNLDKKVKVPTNFYNYIDNSDNILLNKLYLTFQNITIEYKHRFSQEKKI
MTIRELFKYFKFI"

gene complement(join(15848. .16581,16944. .17610))
/gene="PFI1245c"

CDS complement(join(15848. .16581,16944. .17610))
/gene="PFI1245c"

/note="Plasmodium falciparum protein phosphatase-beta
TR:O15920 (EMBL:U89025) (466 aa) fasta scores: E():
1.9e-180, 99.142% id in 466 aa. Pfam match to PF00149,
Ser/Thr protein phosphatase; SMART hit to SM00156, Protein
phosphatase 2A homologues, catalytic domain."
/codon_start=1
/product="Protein phosphatase-beta"
/protein_id="CAD51935.1"
/db_xref="GI:23505154"
/db_xref="GOA:Q8I2Q4"
/db_xref="UniProt/TrEMBL:Q8I2Q4"
/translation="MKNCELSKRDLIFNNVINTKNNLKRLNKNVKGDLTYAEKKPKDT
SLIIEEDKDEDSELDCGQSEKKYKISNGEEINNTFPNFNKEDNYNIYDNLSVKELSDE
EVSSENEYNKKIKKNVLNNPINNQNSIELYTKTSINDEQNYKNINEQNNNSSGKISYN

Query Match 4.9%; Score 53.4; DB 3; Length 254050;
Best Local Similarity 44.6%; Pred. No. 1.2;
Matches 251; Conservative 0; Mismatches 311; Indels 1; Gaps 1;

```
Qy      254 ATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAG 313
            | | |  ||   || |   ||   |  | | |  |||   | ||| ||||  ||  | |
Db   238099 AAATAAATAAATATATTAATTAAATAAGATAATTAAATAAAGAAATAAATAAATAAATAA 238040

Qy      314 AAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTT 373
            ||  |    ||    |    | ||| ||| |   ||  |    | ||| |  | | |
Db   238039 AATAAATTAATTAAATAAATAAAATAATAAAATAATTAAATAAATAAAATAAATTAATTA 237980

Qy      374 AATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCA 433
            |||     |   |||  |  || | ||   ||  ||||     |   |||   |
Db   237979 AATAAGATAAATTAAGTAATTAAATAAATAAAATAATATATATATAAATAATATATTAAA 237920

Qy      434 CTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTT 493
              |  |||| | |  |  |  ||   ||||    ||||||  | |  ||| | ||||  |
Db   237919 ATAATAATAAATAAGATAAATAAACATATTAATTAAATAAAATAAATAAATA-AATTAAT 237861

Qy      494 TATGTACTATTAAAACATTTACTTTCCAAATATTCTGTCATTCAGGAGTATGGAAGTATC 553
            ||  || |   |||| || ||  |     ||||||   | | | |  | |     |   ||
Db   237860 TAATTAATTAAAAAAAATATATATATATAATATTAATTAAATAATTAATTAAAATAAATT 237801

Qy      554 GATGGCTTCTTTAAAATGAAGCAGGAGGGTCTGGCAGAGAGTATCTATGAAATAAGTTCC 613
              |||  |  |||||  |  |  |      |  | | |  |  ||  ||| | |
Db   237800 AATGAATAAGTTAAATAAACATATTATTAAATAAAATAAATAAATTAATTAATTAAATAA 237741

Qy      614 TCTGACCTTCACGCTTAATTTTCTGAATGGAGTGGAGCAAATTACTTCAAGCTTCACTTA 673
            | |     |   ||| ||  | |||  | |  |  |||| |  | ||   | | |||
Db   237740 ATTAAATAAAATAAATAAATTAATTAATTAAATAAATTAAATAAAATAAATAATTAATTA 237681

Qy      674 ACTTGCATATGAAATGAACCGTACAAAAATACAAGAGTGTCAGGAGAAAGTTATGCTCTG 733
            |          || | |   || || |    |  | |       |  || | ||    |
Db   237680 ATAAAATAGATAATTAATTAATAAAATAGATAATTAATTAATAAAATAAATAATTAATTA 237621

Qy      734 GTAAATATTTTGCAAAACAGATAAAAGATAATACTAGAGCTCTGTCCTCAAAGAGTTAAG 793
            ||||     |   ||  |||||| | ||||    |   |   | ||  |  |||
Db   237620 ATAAAATAAATAATTAATTAATAAAATAAAATAAATAATTAATTAAATAAATTAAATAAT 237561

Qy      794 CAGCTAATCTAAGGAGGTAAACT 816
              | ||||  ||  |  ||||  |
Db   237560 TAATTAATTAAATAAAATAAATT 237538
```

RESULT 88
AF125965/c
LOCUS AF125965 26894 bp DNA linear INV 21-SEP-2004
DEFINITION Caenorhabditis elegans fosmid H43I07, complete sequence.
ACCESSION AF125965
VERSION AF125965.1 GI:4262641
KEYWORDS HTG.
SOURCE Caenorhabditis elegans
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 26894)
AUTHORS .
CONSRTM WormBase Consortium
TITLE Genome sequence of the nematode C. elegans: a platform for investigating biology. The C. elegans Sequencing Consortium
JOURNAL Science 282 (5396), 2012-2018 (1998)
MEDLINE 99069613
PUBMED 9851916
REFERENCE 2 (bases 1 to 26894)
AUTHORS Cordes,M. and Broy,M.
TITLE The sequence of C. elegans fosmid H43I07
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 26894)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (04-FEB-1999) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 4 (bases 1 to 26894)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2001) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA
REFERENCE 5 (bases 1 to 26894)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (07-DEC-2001) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA
REFERENCE 6 (bases 1 to 26894)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (23-MAY-2002) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA
REFERENCE 7 (bases 1 to 26894)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (19-NOV-2002) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA
REFERENCE 8 (bases 1 to 26894)
AUTHORS .
CONSRTM WormBase Consortium
TITLE Direct Submission
JOURNAL Submitted (21-SEP-2004) Department of Genetics, Washington University, Genome Sequencing Center, 4444 Forest Park Avenue, St. Louis, MO 63110, USA
COMMENT Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University
St. Louis , MO 63110, USA, and
Sanger Centre, Hinxton Hall
Cambridge CB10 IRQ, England
email: submissions@watson.wustl.edu and jes@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one m13 subclone.

For a graphical representation of this clone sequence and its analysis see:
http://www.wormbase.org/db/seq/sequence?name=H43I07;class=Sequence

NEIGHBORING CLONE INFORMATION

The 5' clone is Y45G5AR, 200 bp overlap; the 3' clone is E03D2, 200

```
          bp overlap. Actual start of this clone is at base position 1 of
          H43I07; actual end is at 8142 of E03D2.

          NOTES:

          Coding seqences below are the result of integration and manual
          review of the following data : computer analysis using the program
          Genefinder (P. Green and L. Hillier, personal communication), the
          large scale EST projects of Yuji Kohara
          (http://www.ddbj.nig.ac.jp/c-elegans/html/CE_INDEX.html) and The C.
          elegans ORFeome cloning project (http://worfdb.dfci.harvard.edu/),
          similarity to other proteins from BlastX analyses
          (http://blast.wustl.edu/), sequence conservation with C. briggsae
          using Jim Kent's WABA alignment program (Genome Research
          10:1115-1125, 2000), individual C. elegans GenBank submissions,
          and personal communications with C. elegans researchers. tRNAs
          are predicted using the program tRNAscan-SE (Lowe, T.M. and
          Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).
FEATURES             Location/Qualifiers
     source          1. .26894
                     /organism="Caenorhabditis elegans"
                     /mol_type="genomic DNA"
                     /strain="Bristol N2"
                     /db_xref="taxon:6239"
                     /chromosome="V"
                     /clone="H43I07"
     gene            3701. .7177
                     /gene="H43I07.3"
                     /locus_tag="H43I07.3"
     CDS             join(3701. .3795,3846. .4029,4096. .4288,5106. .5277,
                     5327. .5480,6310. .6498,7145. .7177)
                     /gene="H43I07.3"
                     /locus_tag="H43I07.3"
                     /standard_name="H43I07.3"
                     /note="contains similarity to Pfam domain PF00535
                     (Glycosyl transferase); coded for by the following C.
                     elegans cDNAs: yk1143d09.3, yk1535h07.3"
                     /codon_start=1
                     /product="Hypothetical protein H43I07.3"
                     /protein_id="AAL38962.1"
                     /db_xref="GI:17402836"
                     /db_xref="WormBase:H43I07.3"
                     /translation="MWDLISPEIFSIQNLCLLGFLVASIAVSSLFTLSYATPWVRRRR
                     AEYLEPEIVYRSSPEQKFPVLLQEDRDAPKRYTPSSCYLSVIIPAMNEVERIEIMLDD
                     CCDYLEARAEKDKDFTYEIIVVDDGSTDETADIVVQIGARRQNLRVLKMKANRGKGGA
                     VKMGVLHSSGKLILFADADGATKFADFENLEKEMLRTAGGEPLDESFPAVIVASRAHL
                     EAESMAVRSVPRTILMLGFHLLVYTFAARTIKDTQCGFKLFTRSIAARIFPVLHIERW
                     AFDVELIFLCERWTVPVKEVSVRWTEIEGGSKITPIWSWLQMGRDLVLIWFRYTIGIWT
                     DDQPR"
     gene            7576. .11644
                     /gene="H43I07.2"
                     /locus_tag="H43I07.2"
     CDS             join(7576. .7659,8379. .9058,10102. .10205,10258. .10374,
                     11529. .11644)
                     /gene="H43I07.2"
                     /locus_tag="H43I07.2"
                     /standard_name="H43I07.2"
                     /note="contains similarity to Pfam domains PF01000 (RNA
                     polymerase Rpb3/RpoA insert domain), PF01193 (RNA
                     polymerase Rpb3/Rpb11 dimerisation domain); coded for by
                     the following C. elegans cDNAs: yk187a1.5, yk258a7.3,
                     yk258a7.5, yk265h10.5, yk282d11.3, yk282d11.5, yk587h6.5,
                     yk645b9.3, yk645b9.5, yk1254g05.3"
                     /codon_start=1
                     /product="Hypothetical protein H43I07.2"
                     /protein_id="AAL38961.1"
                     /db_xref="GI:17402835"
                     /db_xref="WormBase:H43I07.2"
                     /translation="MVKKAAATSKQATYHSDEIEMKEEVVLNTYDECTDYVEYDDTKQ
                     FDINEYCDQIQVAMVNESDDGMTLEFDITRIEAPIANALRRVLIAEVPTMALEKIYLY
                     QNTSVIQDEVLCHRLGLLPLRVDPRGFQFPKEKVVGINEKGVDCDEEPEGDPAKNLIF
                     KINVSCTKNRNAVPTATDPKQLYHNSSVYSRAFEWVPIADQKTQFTEEAHPRMVSDDI
                     LVAKLRPGQEIEASCHAVKGIGRDHAKFSPVATASYRLLPTIRLNAEISGEAAERLKS
                     VFSEGVIAIEKKGAKRIAVVKDARKDTCSRNVFRHEDLSKVVQLGKNKQHFIFSVEST
                     GALKSSELVVEACKVMEIKCRSLRKQIEALIQ"
     gene            12875. .16451
                     /gene="H43I07.1"
                     /locus_tag="H43I07.1"
     CDS             join(12875. .12980,13557. .13795,14576. .14828,16249. .16451)
                     /gene="H43I07.1"
                     /locus_tag="H43I07.1"
                     /standard_name="H43I07.1"
                     /note="contains similarity to Pfam domain PF07047 (Optic
                     atrophy 3 protein (OPA3)); coded for by the following C.
                     elegans cDNAs: OSTF193H6_1, OSTR193H6_1"
                     /codon_start=1
                     /product="Hypothetical protein H43I07.1"
                     /protein_id="AAD14757.1"
                     /db_xref="GI:4262642"
                     /db_xref="WormBase:H43I07.1"
                     /translation="MIAARQVSKPIADAVLRYGKDHPVFRNKLLLPIGRGLVRFTSRL
                     RMKRLGLGEPISQAPISESAALEQASDFVQQLVLFSYSVGVFAGYYFYTKMTTPESLK
                     IEDYEEFKQQQERAIKELRLQVESLEQRLAAQQKRNFFSQLGFSKENGDESPPKPSQQ
                     APPPEKPSEEKKPATESLKQRFQRISSLPIDAAASLILDDEEYLVVRRGSSVGIELSV
                     NFAARGDRLRYRIPSPKLRESCAQMALKEAACHVIGAKDRSRFRYVNS"
ORIGIN

  Query Match             4.9%;  Score 53.2;  DB 3;  Length 26894;
  Best Local Similarity   47.9%;  Pred. No. 1.9;
  Matches  184;  Conservative    0;  Mismatches  198;  Indels    2;  Gaps    1;

Qy        132 TATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATC 191
              |||   |||      ||   |  | |  ||| | ||| |   |||  |||| |      |
Db      15659 TATTGATTTTTCATAGTTTTCCTCACAATTTCTAATTAAAAAATGTTTTTTTTTTTTTTG 15600

Qy        192 TCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAG 251
              | |||||  |||  |||||  || | |    |  |||, |  |  ||  |
Db      15599 GTTTTTTAAACTCATAATTTTGAAAAAAAATGTTTTGAAAAATTTAATTATTTATTTAAT 15540

Qy        252 ACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAG 311
              ||      ||  || ||||  | |  ||   ||||   | |  | | |  |||
Db      15539 TTTTATTAAATTTAGCTTTTTTTAAGGTGTTTTTTAAAATCAAAATTGTATTTAATTTTT 15480

Qy        312 AGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTAC 371
              ||||| ||  | | ||   | |||      |||||   ||  | ||   | |||  ||||
Db      15479 AGAAAAATATTTTTTTTTTCGGTTTTTTTTTAAATTTAATTTCAGAAAACATTTAATTAA 15420

Qy        372 TTAATCTTGCAGTCAAAGGTAAA--AAGTCAACCTAAAGGTATACTACCTGCTTTCTTAT 429
              | | |   |  ||  |||  ||        ||| |||| || | | ||  || |
Db      15419 ATTTTTATTTCATTGGAGAAAAATTAAAATTTAGAAAAAGTATTTTAACAGTTTAATTTT 15360

Qy        430 CGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAAT 489
                  ||||    |   |  | |||| ||||  | | |  |||     |
Db      15359 TTTTAAATGTTTTTAAATATTTAATTTTAATTTTTTAAAAATTTTTTTAAGAAAATAGTT 15300

Qy        490 TTTTTATGTACTATTAAAACATTT 513
              |||| |||||| |||   |||||
Db      15299 TTTTAATGTACTCTTACTTCATTT 15276


RESULT 89
HS101G11/c
LOCUS       HS101G11              150236 bp    DNA     linear   PRI 04-MAR-2003
DEFINITION  Human DNA sequence from clone RP1-101G11 on chromosome 22q12,
            complete sequence.
ACCESSION   AL021877
VERSION     AL021877.1  GI:3820976
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 150236)
  AUTHORS   Griffiths,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-2003) Wellcome Trust Sanger Institute, Hinxton,
```

```
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Nov 2, 1998 this sequence version replaced gi:3724203.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep ----------
            Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            ----------
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            This sequence was generated from part of bacterial clone contigs of
            human chromosome 22, constructed by the Sanger Centre Chromosome 22
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr22
            RP1-101G11 is from the library RPCI-1 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pCYPAC2.
FEATURES             Location/Qualifiers
     source          1. .150236
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="22"
                     /map="q12"
                     /clone="RP1-101G11"
                     /clone_lib="RPCI-1"
ORIGIN

  Query Match             4.9%;  Score 53.2;  DB 9;  Length 150236;
  Best Local Similarity   45.5%;  Pred. No. 1.4;
  Matches  190;  Conservative    0;  Mismatches  228;  Indels    0;  Gaps    0;

Qy        131 ATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGAT 190
              || |   |||  || | | ||||| ||    | | |    |  ||  || | ||  | ||
Db      61703 ATGTACTTTTATTTTATATATATTTATATAAATATAAATATATATTTTTATATATTATAT 61644

Qy        191 CTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACA 250
              | || |  || | ||  |||| |  ||    |    |  | |||| | | |   || |
Db      61643 ATTATATATAATATATATTTTTATATATTATATATAATATATTTTTATATATTATATA 61584

Qy        251 GACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCA 310
              | ||| |  | || || | ||  |    |  | ||| |    ||| | | |    |
Db      61583 TAAATATATATTTTTATATATTATATATAAATATAAATATATAATATATATTTTATATAT 61524

Qy        311 GAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTA 370
              |  |  | ||||| ||   | || | |  |||| | |||   | |  ||||    ||
Db      61523 AAATATAAATATATAATATATATTTTATATATAAATATAAATATAATATATATATTTATA 61464

Qy        371 CTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATC 430
                |   |  | |  |   | | || |  ||   |   |||| ||  |        ||
Db      61463 TAAATATATAAAATATATATTTATAAATATACTATATTTTATAGTATATTAAAAAATACT 61404

Qy        431 GCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATT 490
                |      ||| |  |||    | ||   |       || | | | |        || |
Db      61403 ATAAAATTTATATATTTTATAGTATATAAAAAAACTATAAAATTTTAACTTTTATAAAAT 61344

Qy        491 TTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGTCATTCAGGAGTATGGAA 548
                | |      ||| |||||  ||  | |  | ||||    | || ||  | | |  ||
Db      61343 GTAAAAAGTATATAAAAACTATTTATATTTATATATAAATTTATACATAAATTTATAA 61286


RESULT 90
CR293521/c
LOCUS       CR293521              153572 bp    DNA     linear   HTG 10-OCT-2004
DEFINITION  Danio rerio clone CH211-152F22, WORKING DRAFT SEQUENCE, 7 unordered
            pieces.
ACCESSION   CR293521
VERSION     CR293521.6  GI:54021843
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 153572)
  AUTHORS   McLaren,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 10, 2004 this sequence version replaced gi:45433408.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zC152F22
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 151631 bases at least Q40
            Consensus quality: 152016 bases at least Q30
            Consensus quality: 152290 bases at least Q20
            Insert size: 152972; sum-of-contigs
            Insert size: 158863; 1.5% error; agarose-fp
            Quality coverage: 9.39x in Q20 bases; sum-of-contigs Quality
            coverage: 9.48x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    14764: contig of 14764 bp in length
            *    14765    14864: gap of 100 bp
            *    14865    21642: contig of 6778 bp in length
            *    21643    21742: gap of 100 bp
            *    21743    41316: contig of 19574 bp in length
            *    41317    41416: gap of 100 bp
            *    41417    67009: contig of 25593 bp in length
            *    67010    67109: gap of 100 bp
            *    67110   109819: contig of 42710 bp in length
            *   109820   109919: gap of 100 bp
            *   109920   116506: contig of 6587 bp in length
            *   116507   116606: gap of 100 bp
            *   116607   153572: contig of 36966 bp in length.
FEATURES             Location/Qualifiers
     source          1. .153572
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-152F22"
                     /clone_lib="CHORI-211"
```

```
     misc_feature    1. .14764
                     /note="assembly_fragment:00194
                     fragment_chain:1
                     clone_end:SP6
                     vector_side:left"
     misc_feature    14865. .21642
                     /note="assembly_fragment:00009
                     fragment_chain:1"
     misc_feature    21743. .41316
                     /note="assembly_fragment:00386
                     fragment_chain:1"
     misc_feature    41417. .67009
                     /note="assembly_fragment:00769
                     fragment_chain:2"
     misc_feature    67110. .109819
                     /note="assembly_fragment:01785
                     fragment_chain:2"
     misc_feature    109920. .116506
                     /note="assembly_fragment:00095
                     fragment_chain:2"
     misc_feature    116607. .153572
                     /note="assembly_fragment:01202
                     clone_end:T7
                     vector_side:right"
ORIGIN

  Query Match             4.9%;  Score 53.2;  DB 2;  Length 153572;
  Best Local Similarity   48.6%;  Pred. No. 1.4;
  Matches  174;  Conservative    0;  Mismatches  183;  Indels    1;  Gaps    1;

Qy        107 CATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATATTGATCATTACAAA 166
              | | ||| | | |||   |  ||  || | || || ||| | || | || ||  ||| |
Db      23411 CTTGTCTTGATTTTGTGGAATCAGATGTAAGCTTTGTTTACAACTTTTTATTTTTAGATT 23352

Qy        167 TCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCC 226
              ||  |||    |    | | |||| | || |||||   |   |  |||   | ||| |
Db      23351 TCTTTTAGATTTAGATTTGCAGATTTTATTTTAAATGTAGCTTAATGTGTGTTTATTTAT 23292

Qy        227 TTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAA 286
              ||  | ||   |||  |||   |||| || ||  ||  | | | |   |  | |
Db      23291 TTTTTGTTGGCTAAATGTCTTTTTTACATGAAGTTAAACAAATATATATATATATATATA 23232

Qy        287 AAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAAC 346
              | |  |||| | | ||  ||  | | | | || ||||||       |  | |
Db      23231 TATATATATATATATATATATATATATATATAT-ATATATATATATATATATATAATTTA 23173

Qy        347 ACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAA 406
              |  ||   | ||  | ||  | || ||| | |      |   |  || ||| |    | |
Db      23172 AATAACCAAATGTTTTTTTAAGTATTTATTTTATAGTTAGCATTTATAAACTATTTCCAC 23113

Qy        407 AGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTT 464
              |  | |  |   |  ||  ||      |  ||||| || ||   | |||||   |||
Db      23112 AATTTTTATGTTTAGTTAATTTCTTGCTTCTAAATATAACTTTTTAAAAATTACTTTT 23055


RESULT 91
CR385026
LOCUS       CR385026              327236 bp    DNA     linear   HTG 27-AUG-2004
DEFINITION  Danio rerio clone DKEY-186K21, *** SEQUENCING IN PROGRESS ***, 28
            unordered pieces.
ACCESSION   CR385026
VERSION     CR385026.4  GI:51591744
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 327236)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-AUG-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 27, 2004 this sequence version replaced gi:51571694.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK186K21
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 314746 bases at least Q40
            Consensus quality: 318288 bases at least Q30
            Consensus quality: 320572 bases at least Q20
            Insert size: 324536; sum-of-contigs
            Insert size: 161249; 4.8% error; agarose-fp
            Quality coverage: 4.46x in Q20 bases; sum-of-contigs Quality
            coverage: 9.40x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 28 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    15282: contig of 15282 bp in length
            *    15283    15382: gap of 100 bp
            *    15383    49743: contig of 34361 bp in length
            *    49744    49843: gap of 100 bp
            *    49844    58186: contig of 8343 bp in length
            *    58187    58286: gap of 100 bp
            *    58287    74970: contig of 16684 bp in length
            *    74971    75070: gap of 100 bp
            *    75071    80760: contig of 5690 bp in length
            *    80761    80860: gap of 100 bp
            *    80861    83592: contig of 2732 bp in length
            *    83593    83692: gap of 100 bp
            *    83693    86474: contig of 2782 bp in length
            *    86475    86574: gap of 100 bp
            *    86575    93315: contig of 6741 bp in length
            *    93316    93415: gap of 100 bp
            *    93416    98102: contig of 4687 bp in length
            *    98103    98202: gap of 100 bp
            *    98203   106288: contig of 8086 bp in length
            *   106289   106388: gap of 100 bp
            *   106389   110047: contig of 3659 bp in length
            *   110048   110147: gap of 100 bp
            *   110148   112630: contig of 2483 bp in length
            *   112631   112730: gap of 100 bp
            *   112731   133115: contig of 20385 bp in length
            *   133116   133215: gap of 100 bp
            *   133216   164309: contig of 31094 bp in length
            *   164310   164409: gap of 100 bp
            *   164410   174005: contig of 9596 bp in length
            *   174006   174105: gap of 100 bp
            *   174106   183714: contig of 9609 bp in length
            *   183715   183814: gap of 100 bp
            *   183815   191118: contig of 7304 bp in length
            *   191119   191218: gap of 100 bp
            *   191219   194963: contig of 3745 bp in length
            *   194964   195063: gap of 100 bp
            *   195064   203931: contig of 8868 bp in length
            *   203932   204031: gap of 100 bp
            *   204032   211389: contig of 7358 bp in length
            *   211390   211489: gap of 100 bp
            *   211490   218180: contig of 6691 bp in length
            *   218181   218280: gap of 100 bp
            *   218281   255571: contig of 37291 bp in length
            *   255572   255671: gap of 100 bp
            *   255672   262945: contig of 7274 bp in length
```

```
            *   262946   263045: gap of 100 bp
            *   263046   280944: contig of 17899 bp in length
            *   280945   281044: gap of 100 bp
            *   281045   283852: contig of 2808 bp in length
            *   283853   283952: gap of 100 bp
            *   283953   286488: contig of 2536 bp in length
            *   286489   286588: gap of 100 bp
            *   286589   297811: contig of 11223 bp in length
            *   297812   297911: gap of 100 bp
            *   297912   327236: contig of 29325 bp in length.
FEATURES             Location/Qualifiers
     source          1. .327236
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-186K21"
                     /clone_lib="DanioKey"
     misc_feature    1. .15282
                     /note="assembly_fragment:00732
                     fragment_chain:1"
     misc_feature    15383. .49743
                     /note="assembly_fragment:01653
                     fragment_chain:1"
     misc_feature    49844. .58186
                     /note="assembly_fragment:00542
                     fragment_chain:1"
     misc_feature    58287. .74970
                     /note="assembly_fragment:00842
                     fragment_chain:1"
     misc_feature    75071. .80760
                     /note="assembly_fragment:00236
                     fragment_chain:1"
     misc_feature    80861. .83592
                     /note="assembly_fragment:00070
                     fragment_chain:1"
     misc_feature    83693. .86474
                     /note="assembly_fragment:00046
                     fragment_chain:1"
     misc_feature    86575. .93315
                     /note="assembly_fragment:00152
                     fragment_chain:1"
     misc_feature    93416. .98102
                     /note="assembly_fragment:00178
                     fragment_chain:1"
     misc_feature    98203. .106288
                     /note="assembly_fragment:00302
                     fragment_chain:1"
     misc_feature    106389. .110047
                     /note="assembly_fragment:00108
                     fragment_chain:1"
     misc_feature    110148. .112630
                     /note="assembly_fragment:00052
                     fragment_chain:1"
     misc_feature    112731. .133115
                     /note="assembly_fragment:01097
                     fragment_chain:1"
     misc_feature    133216. .164309
                     /note="assembly_fragment:01444
                     fragment_chain:1"
     misc_feature    164410. .174005
                     /note="assembly_fragment:00596
                     fragment_chain:2"
     misc_feature    174106. .183714
                     /note="assembly_fragment:00491
                     fragment_chain:2"
     misc_feature    183815. .191118
                     /note="assembly_fragment:00267
                     fragment_chain:2"
     misc_feature    191219. .194963
                     /note="assembly_fragment:00093
                     fragment_chain:2"
     misc_feature    195064. .203931
                     /note="assembly_fragment:00395
                     fragment_chain:2"
     misc_feature    204032. .211389
                     /note="assembly_fragment:00442
                     fragment_chain:2"
     misc_feature    211490. .218180
                     /note="assembly_fragment:00205
                     fragment_chain:2"
     misc_feature    218281. .255571
                     /note="assembly_fragment:01883
                     fragment_chain:3"
     misc_feature    255672. .262945
                     /note="assembly_fragment:00348
                     fragment_chain:3"
     misc_feature    263046. .280944
                     /note="assembly_fragment:00960
                     fragment_chain:3"
     misc_feature    281045. .283852
                     /note="assembly_fragment:00060
                     fragment_chain:3"
     misc_feature    283953. .286488
                     /note="assembly_fragment:00127.0"
     misc_feature    286589. .297811
                     /note="assembly_fragment:00653.0"
     misc_feature    297912. .327236
                     /note="assembly_fragment:01243"
ORIGIN

  Query Match             4.9%;  Score 53.2;  DB 2;  Length 327236;
  Best Local Similarity   45.0%;  Pred. No. 1.2;
  Matches  165;  Conservative    0;  Mismatches  199;  Indels    3;  Gaps    2;

Qy        146 AGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACA 205
              | ||||||  ||  ||| | ||   |  |||  | | || || || |     || | | |
Db     210935 AAAAATATAAATATTTATATATATATATATATATATATATAA-ATATATATATATATATA 210993

Qy        206 TTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTC 265
              ||  ||| |  |   |         |        |||| |       | | |  |  ||||
Db     210994 TTTNTTTTTNTANNAAAATAAAAAAAAAAAAAAAAAAAAANNNNNNANANNTNAAAATATTT 211053

Qy        266 ATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATA 325
              || | |||||  ||| ||              |  |||   ||  | | ||| |   |||
Db     211054 ATATTTTTTTTTTTTTTTTTTNTNNNTANNNANNAAAAAAATAAAAAAAAAAAAATTATT 211113

Qy        326 TTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTC 385
              |||      ||  ||     |    ||||  |     | ||   | || ||  |  |  | |
Db     211114 TTATTTTTTTTNAATAANANNTAAAAAAAAAAAAAAATATAAATATTTTTTNNT--AATT 211171

Qy        386 AAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAA 445
              |    ||||    ||    ||        ||  |  | | |||    |    ||| | |
Db     211172 TATATAAAAANAAAAANNATAATNNTNTTTAATTTNTNTTTTAAAAAANAAAAAAAAANA 211231

Qy        446 ATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTA 505
              | |     |   |      || ||| |   ||||  ||| ||| ||| |  | |||||
Db     211232 ANAAAANNNNNTNNNANNNATATAAAAATATATAAAAAATTAAATTTATATTAAATATTA 211291

Qy        506 AAACATT 512
              ||  |||
Db     211292 AATTATT 211298


RESULT 92
PFMAL1P5
LOCUS       PFMAL1P5               14851 bp    DNA     linear   HTG 16-JUN-1999
DEFINITION  Plasmodium falciparum 3D7 chromosome 1, *** SEQUENCING IN PROGRESS
            ***, 7 unordered pieces.
ACCESSION   AL031748
VERSION     AL031748.2  GI:4493864
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Plasmodium falciparum 3D7
  ORGANISM  Plasmodium falciparum 3D7
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1
  AUTHORS   Bowman,S., Churcher,C., Harris,B., Harris,D., Lawson,D., Quail,M.
```

```
            and Barrell,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUN-1999) P.falciparum Genome Sequencing Consortium,
            The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
            CB10 1SA, UK
COMMENT     On Mar 24, 1999 this sequence version replaced gi:4034845.
            For more information about this sequence or the Malaria Project,
            see http://www.sanger.ac.uk/Projects/P_falciparum. IMPORTANT: This
            sequence is unfinished and does not necessarily represent the
            correct sequence. Work on the sequence is in progress and the
            release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc.
            Order of segments is not known; 800 n's separate segments.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2253: contig of 2253 bp in length
            *     2254     3053: gap of 800 bp
            *     3054     4244: contig of 1191 bp in length
            *     4245     5044: gap of 800 bp
            *     5045     6189: contig of 1145 bp in length
            *     6190     6989: gap of 800 bp
            *     6990     9068: contig of 2079 bp in length
            *     9069     9868: gap of 800 bp
            *     9869    10850: contig of 982 bp in length
            *    10851    11650: gap of 800 bp
            *    11651    12643: contig of 993 bp in length
            *    12644    13443: gap of 800 bp
            *    13444    14851: contig of 1408 bp in length.
FEATURES            Location/Qualifiers
     source         1. .14851
                    /organism="Plasmodium falciparum 3D7"
                    /mol_type="genomic DNA"
                    /strain="3D7"
                    /isolate="3D7"
                    /db_xref="taxon:36329"
                    /chromosome="1"
ORIGIN

  Query Match             4.9%;  Score 53;  DB 2;  Length 14851;
  Best Local Similarity   46.5%;  Pred. No. 2.2;
  Matches  205;  Conservative    0;  Mismatches  235;  Indels    1;  Gaps    1;

Qy        286 AAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAA 345
              ||| |  ||| ||  |||  ||  | | ||||   ||||  |     ||  | ||||| |
Db      11755 AAATATATATTAATTAATATATATATAAAAAGTAAATATTATTAATTGTATATAATTATA 11814

Qy        346 CACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTA 405
              | ||       |  ||   |   ||  ||||  |   | ||  |  |||  | |  |   |
Db      11815 AATAATTAATATTTATTAAAATATATATAATTTATAATTCTCATATAATTAATATAATGA 11874

Qy        406 AAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTT 465
              |   ||||  ||    |  |||   | |  || || | | ||     | | ||||| |
Db      11875 ATAATATAACACAATGTAATTTAATTTAATTTAATTACATAATAAATTTATTATTATTAT 11934

Qy        466 GGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATA 525
                || ||||    |   ||   || |||| | ||| || |   |||   |  ||||
Db      11935 AATTATTATCTATTTATTTATTTATTATTATAAATTATAAATATCATTATAATTAAAATC 11994

Qy        526 TTCTGTCATTCAGGAGTATGGAAGTATCGATGGCTTCTTTAAAATGAAGCAGGAGGGTCT 585
              | | | | |  | |    ||  ||  || | |   |||  |  || ||    | |
Db      11995 AATTATTAATTATTAATGATAAATAATTAAT-GATAAATTATCAATAACCAATTAGATTA 12053

Qy        586 GGCAGAGAGTATCTATGAAATAAGTTCCTCTGACCTTCACGCTTAATTTTCTGAATGGAG 645
              |  ||   |  || | ||  |  | | | |  ||     |  || |     | |  |
Db      12054 TTTATCGATATTTAATTATATTATATTATATCATATTATATATATATATATATTATTAT 12113

Qy        646 TGGAGCAAATTACTTCAAGCTTCACTTAACTTGCATATGAAATGAACCGTACAAAAATAC 705
              | | | | |  | ||  |  || | ||  ||| | ||||    |  | |||
Db      12114 TATAAAATTTATTTATAAATATTTGTTTATTTATTTATTTATTGAATAACAATAGAATTA 12173

Qy        706 AAGAGTGTCAGGAGAAAGTTA 726
              || | |||||  | | | | |
Db      12174 AATATTGTCAATAAATAATAA 12194


RESULT 93
AC138466
LOCUS       AC138466               129492 bp    DNA     linear   PRI 28-MAR-2003
DEFINITION  Homo sapiens 12 BAC RP13-977J11 (Roswell Park Cancer Institute
            Human BAC Library) complete sequence.
ACCESSION   AC138466
VERSION     AC138466.12  GI:29336158
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 129492)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Emerling,S.,
            Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P.,
            Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N.,
            Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Han,J., Harris,C., Harris,K., Hart,M., Havlak,P.,
            Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M.,
            Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S.,
            Hume,J., Ioshikhes,I., Jackson,L.E., Jacobson,B., Jia,Y.,
            Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U.,
            King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N.,
            Leal,B., Lee,E., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O.,
            Lieu,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Lu,X.,
            Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P.,
            Marondel,I., Martin,R., Martindale,A., Martinez,E., Massey,E.,
            Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Merscher,S.,
            Metzker,M., Miller,A., Miner,G., Miner,Z., Mitchell,T.,
            Mohabbat,K., Montgomery,K.T., Morgan,M., Morris,S., Moser,M.,
            Neal,D., Nelson,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shim,C., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Kucherlapati,R.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 129492)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JAN-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
```

```
REFERENCE   3  (bases 1 to 129492)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAR-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   4  (bases 1 to 129492)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAR-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Mar 28, 2003 this sequence version replaced gi:29293974.
            INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
            gc-help@bcm.tmc.edu

            CLONE LENGTH: This sequence does not necessarily represent the
            entire insert of this clone. Overlapping regions of clones are only
            sequenced and submitted once, so the sequence for the remainder of
            the insert may be found in the record for the adjacent clones.
            Overlapping clones are noted at the beginning and end of the
            Features listing.

            ANNOTATION OF FEATURES:
               STSs are identified using ePCR (Genome Res. 7:541-550) searches
            of a local database that includes entries from dbSTS, GDB, and
            local mapping efforts.
               Repeats are identified using RepeatMasker (A. Smit and P. Green,
            unpublished.) for Human and Mouse sequences.
               Genes and Region of sequence similarity are identified by BLAST
            (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
            EST and cDNA sequences. Genes demonstrate at least two exons
            flanked by consensus splice sites that maintained sequence
            continuity across the splice junctions. Sequences that are not
            identical matches are annotated as similar.

            SEQUENCING READ COVERAGE:Sequencing is completed to a minimum
            standard of double strand coverage with a minimum of 2 clones and 2
            reads with no ambiguities or 2 chemistries with a minimum of 2
            clones and 3 reads with no ambiguities. If the sequence quality for
            a region does not meet this standard, it will be indicated in the
            annotation as Low Coverage.

            QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality
            standards - estimated error rate less than 1 per 10,000 bases.
            Reports of lowest quality individual bases and measures of base
            quality are listed below. Description of the metrics can be found
            at URL:
            http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annotation.ht
            ml.
FEATURES             Location/Qualifiers
     source          1. .129492
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="12"
                     /clone="RP13-977J11"
     repeat_region   complement(28. .152)
                     /rpt_family="AluSp"
     misc_feature    complement(33. .2012)
                     /note="overlaps bases 1. .1980 of clone AC137590"
                     /function="clone overlap"
     repeat_region   complement(259. .411)
                     /rpt_family="L1M4"
     repeat_region   complement(412. .715)
                     /rpt_family="AluSx"
     repeat_region   complement(716. .795)
                     /rpt_family="L1M4"
     repeat_region   796. .1084
                     /rpt_family="AluJb"
     repeat_region   complement(1085. .1492)
                     /rpt_family="L1M4"
     repeat_region   1510. .1540
                     /rpt_family="AT_rich"
     repeat_region   complement(1776. .2086)
                     /rpt_family="AluY"
     repeat_region   complement(2135. .2583)
                     /rpt_family="L1M4"
     repeat_region   complement(2606. .2903)
                     /rpt_family="AluSg"
     repeat_region   complement(3184. .3332)
                     /rpt_family="FRAM"
     misc_feature    3841. .3890
                     /note="sized restriction digest
                      624 bp"
                     /function="unresolved tandem repeat"
     repeat_region   3841. .3872
                     /rpt_family="C-rich"
     repeat_region   complement(3875. .4171)
                     /rpt_family="AluSx"
     repeat_region   complement(4249. .4560)
                     /rpt_family="AluSp"
     repeat_region   4778. .5320
                     /rpt_family="LTR10A"
     repeat_region   complement(5448. .7103)
                     /rpt_family="L1ME1"
     STS             6025. .6159
                     /standard_name="RH69942"
     repeat_region   complement(7109. .7397)
                     /rpt_family="AluJo"
     repeat_region   complement(7398. .7545)
                     /rpt_family="L1ME1"
     repeat_region   complement(7546. .7779)
                     /rpt_family="MSTA"
     repeat_region   complement(7780. .8071)
                     /rpt_family="AluSx"
     repeat_region   complement(8072. .8255)
                     /rpt_family="MSTA"
     repeat_region   complement(8256. .8383)
                     /rpt_family="L1ME1"
     repeat_region   complement(8383. .8508)
                     /rpt_family="L1ME1"
     repeat_region   complement(8510. .8794)
                     /rpt_family="AluSx"
     STS             9641. .9790
                     /standard_name="STS-H94865"
     repeat_region   complement(9940. .10152)
                     /rpt_family="MER20"
     repeat_region   complement(10349. .10647)
                     /rpt_family="AluSc"
     repeat_region   11513. .11589
                     /rpt_family="MIR"
     repeat_region   11590. .12548
                     /rpt_family="LTR13A"
     repeat_region   12549. .12664
                     /rpt_family="MIR"
     repeat_region   complement(14856. .15162)
                     /rpt_family="AluSx"
     repeat_region   15199. .15227
                     /rpt_family="(TTTTG)n"
     repeat_region   complement(15232. .15477)
                     /rpt_family="AluSg"
     repeat_region   15491. .15963
                     /rpt_family="L1M4"
     repeat_region   complement(16338. .16408)
                     /rpt_family="AluS"
     repeat_region   16442. .16746
                     /rpt_family="AluY"
     repeat_region   18225. .18284

  Query Match             4.9%;  Score 53;  DB 9;  Length 129492;
  Best Local Similarity   46.6%;  Pred. No. 1.6;
  Matches  170;  Conservative    0;  Mismatches  195;  Indels    0;  Gaps    0;

Qy        192 TCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAG 251
              |||| |  ||     | || | |   | ||          | | ||| |     |||
```

```
Db      74298 TCATATATAAATGTATATTATATATGTTTATATATAAATGTATATTATATATATTTATGT 74357

Qy        252 ACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAG 311

Db      74358 ATAAATATATAATATATATTATATATTTATATATAAATATATTTAAATATATGTTTATAT 74417

Qy        312 AGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTAC 371

Db      74418 ATAAATATATTTAAATATATATTTATATATATTTAAATACATAAATATATTTAAATATAT 74477

Qy        372 TTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCG 431

Db      74478 ATTATATATTCATATACATAAATAAATATATTTAAATATATATTATATATATTCATATAC 74537

Qy        432 CACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTT 491

Db      74538 ATAAATAAATATATTTAAATATATATTATATATATTCATATACATAAATAAATATATTTA 74597

Qy        492 TTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGTCATTCAGGAGTATGGAAGTA 551

Db      74598 AATATATATTATATATATTTATATACATAAATAAATATATTATATATGTTTATATATTTA 74657

Qy        552 TCGAT 556

Db      74658 AATAT 74662
```

```
RESULT 94
AC116551/c
LOCUS       AC116551               153751 bp    DNA     linear   INV 21-FEB-2004
DEFINITION  Dictyostelium discoideum chromosome 2 map complement(581327-427576)
            strain AX4, complete sequence.
ACCESSION   AC116551 AC116101 AC116549 AC116965
VERSION     AC116551.2  GI:28829639
KEYWORDS    HTG.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 153751)
  AUTHORS   Gloeckner,G., Eichinger,L., Szafranski,K., Pachebat,J., Dear,P.,
            Lehmann,R., Baumgart,C., Parra,G., April,J.F., Guigo,R., Kumpf,K.,
            Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and
            Noegel,A.A.
  TITLE     Sequence and analysis of chromosome 2 of Dictyostelium discoideum
  JOURNAL   Nature 418 (6893), 79-85 (2002)
  MEDLINE   22092622
   PUBMED   12097910
  REMARK    The Dictyostelium Genome Sequencing Consortium
REFERENCE   2  (bases 1 to 153751)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAR-2002) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   3  (bases 1 to 153751)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   4  (bases 1 to 153751)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   5  (bases 1 to 153751)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-FEB-2004) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
COMMENT     On or before Feb 21, 2004 this sequence version replaced
            gi:28850309, gi:19807752, gi:19920064, gi:19807784.
            CDS predictions from GeneID do not necessarily reflect true genes.
            Further Information is available from IMB Jena, Department of
            Genome Analysis
            (http://genome.imb-jena.de/dictyostelium/)
            and the Univerity Colonge, Institute for Biochemistry I
            (http://www.uni-koeln.de/dictyostelium/project.shtml
            Funding
            Agency : Deutsche Forschungsgemeinschaft (DFG).
FEATURES             Location/Qualifiers
     source          1. .153751
                     /organism="Dictyostelium discoideum"
                     /mol_type="genomic DNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /chromosome="2"
                     /map="complement(581327-427576)"
     CDS             complement(join(<351. .2554,2687. .2807))
                     /note="GeneID exon scores (in order of location ranges):
                     70.76, 0.70 - GSCJ_ID dd_02254"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAO52218.1"
                     /db_xref="GI:28829702"
                     /translation="MNESISDHNEKLFWKVFRNKYLNIEIFKQLESTEIEYDSFTIAS
                     LIKLSVNNSNNNNDNNNKLIEIFKLLMEKEKYLMEGLDLVTISISSRNFEILKILLNK
                     PYSFLVKETDFNDALQKYGDSEIEKSILQLLGNQININTFLIENNFQIKNSIPFFKNT
                     IQFILNNQNNDDNNDDNKNNNNKLIKLLELKIYSIKKDEILNLELIEMQNKLLDLNYF
                     SKENFQNSNIYWFKELYEFKGVKTIFIHLILLVFKIYYKRKSLNIPNEIEIEINEIMN
                     NSLLIKNEEKEKKLLKLLINETQSNLLFKLYYYRYYEIIEGCKLSKIVYRNLKTFNIF
                     KLIESNNNSNNNKEEIIRIIKNFINVYEFKNNENDKQLINKFRNVILNNLNSKEYYHK
                     LIETIGNSQIKTFGLPIEWNINLLDFLLDDEINNKIIIIKDKQVIDWLSNMPRLNNRI
                     FTFDIKTIYYYKRKKNLIQFQSLELIEYAIKKLETNNNNNSNNDNLPPIKLETINWSM
                     VYNNPLKSIIDEFKFQNESIINEDELNDKVTFNLFKFLKFEKNEFTILVQLIKYFIEK
                     VLNKYGQDKSYEILKNFVKIDNLCKFQKETLELLVLSTINLDLIMVMIILNCGIDFTN
                     KQDEILINFWNVHILTNTSIFYIPTINGNINFFKIITHIKPQYIINNIKLWCEQTSLL
                     ECKHLAINHLFKKLIQTKGTTLNQILYIKDFINTINNNNNNNNNNNNNNNNIIKLDHS
                     KFHSIHYLNLKSPKLILYLIESNQIRDWNYLLIPQ"
     CDS             3515. .5095
                     /note="GeneID exon scores (in order of location ranges):
                     144.31 - GSCJ_ID dd_02253"
                     /codon_start=1
                     /product="similar to Arabidopsis thaliana (Mouse-ear
                     cress). Hypothetical 59.5 kDa protein"
                     /protein_id="AAO52156.1"
                     /db_xref="GI:28829640"
                     /translation="MNNQVKPTLSNKAEIINKWLIDNGVRINNKLIKIVYLGKENNFE
                     QTENTTATTTTSERINDSIVSGLGVISLKELKVDDIVAKIPKSIILSIHTSSISNILE
                     KYKIENNIGTSIALIHEASLGEKSKWYGYISSLPRKVDVPILWDSESRKLLKGTAIED
                     VLNDDDILINQVYADVIESILSKNHPEIFGDKELYSIENFKIANSIISSRAFCVDSYH
                     GDSLVPLADIFNHQTAREHVHIESNGDVCNKCGSIKTCKHRKVVTQHHTVNSTKGKRT
                     HKVAGIPSSKKHIHKGNCCSTTTTKTNEEDKDTIIEEDDEHLYIKVVKGVEANKEVYN
                     TYGDHDNAILLSKYGFLEMDNPCDRLSIDKQLVDKELELIGAENQITRMDLFNRISFY
                     AEMFDIDSRNNHAIENDGRLDDALVCSVGIALAPQSIFQSWKSMSEHKREKYFEQLEA
                     EDIVKQSQLVRNTLLSILNKLFSNYNNSNETTATTTIEQDQEKLSTLKNQREIIAQSL
                     KICEKTLINKSINYYKNL"
     CDS             complement(5235. .7895)
                     /note="GeneID exon scores (in order of location ranges):
                     270.65 - GSCJ_ID dd_03185"
                     /codon_start=1
                     /product="similar to essential for initiation of DNA
                     replication; homolog of S. pombe CDC21; Cdc54p
                     [Saccharomyces cerevisiae]"
                     /protein_id="AAO52157.1"
                     /db_xref="GI:28829641"
                     /translation="MDPKETKTSKISFLNTTKKNNNNNNNNNNNNNNNNNNNNNNNN
                     NNNNNNNNDDNNNGENNIDIEPDDIDLLNKNDIDNIVDGIDNNNKSGSNDMKGLVF
                     WGTTVKVEDVKERFRKFIFNFPTPKERKRFNINDINNNNNNNNNNNSEEMNDDDDYYY
                     DEEQDKLIKKPEKVQYLYRDLLKVLNETKKCHLNINMKYLVQFDMELYLQWVSFPNEM
                     IPLLDEEINLIYRELFPITPRNGDDEDDDDDDEDEEDYRIELHPFNLLRKTPMRDLN
                     PSDIDKIISISGLIIRSSSIIPEIKQAFFMCAVCEATFHANVEKGKIQEPSECSNCKS
                     KQSLSIIHNRCLFGDKQYIKLQETPDAIPEGETPHTVALFAYGDLIDIAKPGDRVELT
                     GVFKASPMRAGSNRSLRSIYKTYIDILHIKRTDKGKYDDDDDDHDDNTGGGGTGTGKE
                     TNENLDFEDLDEFDLSEEKEREIIELSKKPDIYDIVTKSIAPNIWELEDIKKGILCQL
                     FGGSKKSYQDYGGKFRGDINILLCGDPGTSKSQLLSYVHKIAPRGIYTSGKGSSAVGL
                     TAYITKDPDTKETVLESGALVLSDKGVCCIDEFDKMNDQTRSILHEVMEQQTVSIAKA
                     GIICTLNARTSILASANPSGSRYMPKLSVVENIQLPPTLLSRFDLIYLVLDKANERSD
```

```
                     RQLARHLVSMYWDETPVSHFTIPKETLTNYIQYARKHINPKLTDDSAKCLVQGYLEMR
                     SMGSSKKTISATPRQLESLIRIAEAHARIRFSEFVEPLDVEEAIRLIKVALQQAAIDP
                     ENGTIDMDLITTGRSASSREAITRLKSHIKQKIGKKHLTLDQLLKLLTIQNQVQTIQQ
                     IEEIKEALRQLQDEEIIQSSGGIDPTYGVV"
CDS                  join(8672. .9027,9184. .9463)
                     /note="GeneID exon scores (in order of location ranges):
                     37.11, 23.19 - GSCJ_ID dd_02251"
                     /codon_start=1
                     /product="similar to cyctochrome c1 heme lyase
                     [Caenorhabditis elegans]"
                     /protein_id="AAO52158.1"
                     /db_xref="GI:28829642"
                     /translation="MGDDNTNNTSITKSSPTDAMGGGCPIVHDINPANHMYKPNQNPH
                     PDQTKPLSTERITSTIPRTEKDNWQYPSPQMFFNAMKKKQYEPKEEDMSVVISIHNTV
                     NEKCWEDVLQWENDYKDVCPNPKLKKFKGKATDFSPKAKFLNTFLGYKLPFDRHDWIV
                     DRNGKEVRYVIDFYEGRINKDSGKSIGIYIDVRPAIDDFSSLKDRVLHFFK"
CDS                  complement(join(9650. .13297,13390. .13536))
                     /note="GeneID exon scores (in order of location ranges):
                     468.70, 11.48 - GSCJ_ID dd_02250"
                     /codon_start=1
                     /product="similar to C25A11.4b.p [Caenorhabditis elegans]"
                     /protein_id="AAO52159.1"
                     /db_xref="GI:28829643"
                     /translation="MNPFEMIQNQNIVIKNAIEVGCIIVNIGAQDLMNATPYLVLGVI
                     WQIIKQGLLSKVNQNANEILEILFEEDSVKEVDNTQNQSDEHSAEEILLRWVNYHLEK
                     EGCERRISNFSEDIQDSVVYSHLFHQLVPIEFENLVSTIHSESNLFSRAELITKACEK
                     IGVKCFLTPSDIALGHPKLNLALVASLFNSEAAIANLRKEREEKHRTDQERLERERLE
                     RERIERERLERERLEQERLERERLERERLEKQRQEMERIEQERIELERLERERLEQER
                     LEQERIERERLERERLEQERLEQERLERERLEQERLECERLERERLEQERLEQERLER
                     ERIELEELERIERERLEYEERERIERERIERERELEEQERLEYERLEQERIEREEME
                     RIERERIEREEFERLEEERIERERIEREEALRLEAERLERELRERERITRQSELERIEYNK
                     EYKICIRIQKRMIDDTSLTDNMTKLIGDITELKRNIIGELRNTTHHQAELKKLTKSID
                     RVLENKNEGKKLKKQKKVIIKPVKGVPNGLDPNKMKNYQSLFYYLQTRPEVMGKLLFL
                     LDRNQMGKTFDNLLLSMFPYEISPREEDLFLKALTVGIEYQIKYSETLQDFLQSESVP
                     MKLLYQYFKKKGFKYLKEIITPLILKVLDQTDLNLNFDPYWIQKEIQIAKELKDGKKS
                     EKKTLDQTFEQLMANNDVVSIFNSRCLQLVELGNAFLERITKSIEQLPVQLRYLCRII
                     AEKSKYAFPTNNTNNHHNGGGISSDESGSTTANITLNNNRSKQLPENREKYISYFLII
                     YRFIIPIISSPDLLNELVLLQHPYITKQSRFNLSQLSRLIQSVFSEKPWMNYFPTVYL
                     NKWIIQNSSSADKFLCQVAQVPDLDELIHYSEYNEIMLENNIKFCSVINISEIIWLHE
                     TILSNFYDLEDIGTLKEVSEKSIQNLAERAINSKLDKGKYVTAARQSTPAIMTSQKSQ
                     NWPLKKYLDEIDFPIQLPEQGDRSIQLSFINRLDQLNEQSIDLWKECKLWMIDILSKS
                     QPPKKNANTIEIIREAIDYFKTGSNSSENLSIINKSNELISRLRTFDRDQCTDGLDGG
                     GPDNFNLFLLDVEKELATILRKNKVMRSELTRLHSLLKSIRIDAAPLEHLTKFQNEQL
                     EKAIVKKQLKTTQLTPLPKDKSTIKPHKFILLDLVKKGMILRCELPESSFSKTKIYIK
                     SLGPGLFEIEAKFGPITKSLQLELDDLLEKSKNHIDEFPLEGITLDVTLTIHLLMKLF
                     SL"
CDS                  complement(join(13685. .13812,14033. .14207))
                     /note="GeneID exon scores (in order of location ranges):
                     6.21, 14.57 - GSCJ_ID dd_02249"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAO52160.1"
                     /db_xref="GI:28829644"
                     /translation="MTYTDSRYNTLNRGSKKMDPPFITQEVLASAKISKKDNILINAN
                     AEDGSVHTFSEEEKVAYSKFITERLIDQEDVLSNYVPIDPSTDQLFVNCTDGVLLK"
CDS                  16650. .17525
                     /note="GeneID exon scores (in order of location ranges):
                     78.29 - GSCJ_ID dd_00829"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAO52161.1"
                     /db_xref="GI:28829645"
                     /translation="METKDKKKDMELLGDRMKSYEDDMKIQIEKNKPFIIRLDGHSFS
                     KFTKNFNKPHDIRIHNAMIETSTVLLKTFMPTCIYTFSDEITMCFPSIDESTLEEGKE
                     IPNLAYSGKVQKLISLSSGLASTVFFKSITNAQYDKDTELNLIKLLETCTPHFDARIF
                     TLPSNQEIVNNLIWRSLVDCKRNSVANLGFAHFTPKQMLGLNNTEVKKKLLEEKSIIY
                     ENEPAWYRFGTYLKKEYYTLTTVSPNEPEKQITAIRSKVNCFSFDITKLSNSLNFISF
                     KTLPEGFNLELIESK"
CDS                  complement(join(17735. .18959,19168. .19640,19750. .20190,
                     20286. .20531))
                     /note="GeneID exon scores (in order of location ranges):
                     130.29, 48.89, 51.03, 20.75 - GSCJ_ID dd_00831"
                     /codon_start=1
                     /product="similar to Dictyostelium discoideum (Slime
                     mold). ALG-2 interacting protein X"
                     /protein_id="AAO52162.1"
                     /db_xref="GI:28829646"
                     /translation="MLSIERKRTEKVDFSKPLTKYIKEQFSKAESDQHETQIATLNGL
                     REDVRNLQERTETSKEMVWKYYSILSSLELRFPISENNVRISFPWTDSYRQRKSTLYS
                     IYFERASVLFNYGSIVSQIASSTNRSNIEGVKKACNQFQLAAGVFNKLREYASLHPEC
                     STSADFSSESLQALVTIMLAQAQECIYEKASMDNLSDSILSKLAAQVAEYYDTFNQLL
                     NSNSLKSIVDRNWNITATVKSYLYKAISLYCHAKGLEQVSQFGEQVSRLMIAVDNINQ
                     SKVNLAKTAPIELKEIVERYVISITRYCESAKKDNDTIYHDTIPPAHKLTPIEKKPLA
```

```
Query Match             4.9%;  Score 53;  DB 3;  Length 153751;
Best Local Similarity   45.5%;  Pred. No. 1.5;
Matches  228;  Conservative    0;  Mismatches  270;  Indels    3;  Gaps    1;

Qy         69 ATGAGTTTTAAAGCCAGATCTTCCTTATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTC 128
              || | |||  |   |    ||   |    ||  |||     |  | | ||      | |
Db      88259 ATTAATTTCCATCTCTTTACTCTTTATATTCATTGCAATTACAATTGGTCCACCCTTTTT 88200

Qy        129 TCATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAG 188
              |  | | | ||   ||||   | |||| ||||    ||    |  | | | |  || |
Db      88199 TTGTTTAATTTGAATTGATTTAAATTGTTCATCTTTAAAATTTGTAAAATCTGCTAAATC 88140

Qy        189 ATCTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTA 248
                  | ||   |  |||| ||| ||| | |  |||    ||  || || ||     ||
Db      88139 TATCATTGTTTTTGCAATTGATTTTTTATTTTTTTCCTAAAATAATTATATAAATAAATA 88080

Qy        249 CAGACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATT 308
              | | ||| | |  |  ||| || |  |   | || |||   ||||| |||| |||
Db      88079 AATATATATATATCATTTTTTTATTATTATAAAATTAATAAATAAATAAATAAATTAATT 88020

Qy        309 CAGAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAA---ATGAGTGCATATTA 365
              | | |   || |   | ||   |     |||   ||   |||   |  | |  |||| |
Db      88019 AATAAATTAATAAATAAATAAAAAAAGATAAAAATAAATTAAAGATAAAAATAAATATAA 87960

Qy        366 CATTACTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTC 425
              |  |  ||||| ||  || ||   |  |   | | | | ||  | |||    |
Db      87959 AAAAAAATAATCCTGTTGTAAATTTTTTATTTTTATCTTTAATTTTTACATGGTTTAAAA 87900

Qy        426 TTATCGCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACA 485
              | |   |  ||| | |||  |  | |||          ||||||   |  |   |
Db      87899 TAAAAAAAAATAAAAAAAAAAAAAATAAAAAAAGAATAAAAAATAATAATAATAGTAA 87840

Qy        486 TAATTTTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGTCATTCAGGAGTATG 545
              |||| || ||       | | || | | |       ||  |||   || || |     ||
Db      87839 TAATAATTAATAATAACTAATAATAATAATAAAATAATAATTACATCTTTAATTGAAATT 87780

Qy        546 GAAGTATCGATGGCTTCTTTA 566
              | | ||    | | ||| | |
Db      87779 GCAATAGTTTTTGATTCATAA 87759
```

```
RESULT 95
CR751546/c
LOCUS       CR751546              156851 bp    DNA     linear   HTG 10-OCT-2004
DEFINITION  Danio rerio clone CH211-53M6, WORKING DRAFT SEQUENCE, 10 unordered
            pieces.
ACCESSION   CR751546
VERSION     CR751546.3  GI:54021755
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 156851)
  AUTHORS   McLaren,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 10, 2004 this sequence version replaced gi:51510848.
            -------------- Genome Center
```

```
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zC53M6
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 153166 bases at least Q40
            Consensus quality: 153587 bases at least Q30
            Consensus quality: 154258 bases at least Q20
            Insert size: 155951; sum-of-contigs
            Insert size: 137574; 0.8% error; agarose-fp
            Quality coverage: 6.91x in Q20 bases; sum-of-contigs Quality
            coverage: 8.00x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     7977: contig of 7977 bp in length
            *     7978     8077: gap of 100 bp
            *     8078    30710: contig of 22633 bp in length
            *    30711    30810: gap of 100 bp
            *    30811    58186: contig of 27376 bp in length
            *    58187    58286: gap of 100 bp
            *    58287    60445: contig of 2159 bp in length
            *    60446    60545: gap of 100 bp
            *    60546    99346: contig of 38801 bp in length
            *    99347    99446: gap of 100 bp
            *    99447   103804: contig of 4358 bp in length
            *   103805   103904: gap of 100 bp
            *   103905   106861: contig of 2957 bp in length
            *   106862   106961: gap of 100 bp
            *   106962   136493: contig of 29532 bp in length
            *   136494   136593: gap of 100 bp
            *   136594   145799: contig of 9206 bp in length
            *   145800   145899: gap of 100 bp
            *   145900   156851: contig of 10952 bp in length.
FEATURES             Location/Qualifiers
     source          1. .156851
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-53M6"
                     /clone_lib="CHORI-211"
     misc_feature    1. .7977
                     /note="assembly_fragment:01486
                     fragment_chain:1"
     misc_feature    8078. .30710
                     /note="assembly_fragment:01758
                     fragment_chain:1"
     misc_feature    30811. .58186
                     /note="assembly_fragment:00363.0"
     misc_feature    58287. .60445
                     /note="assembly_fragment:01268
                     fragment_chain:2"
     misc_feature    60546. .99346
                     /note="assembly_fragment:00747
                     fragment_chain:2"
     misc_feature    99447. .103804
                     /note="assembly_fragment:01325
                     fragment_chain:2"
     misc_feature    103905. .106861
                     /note="assembly_fragment:01298
                     fragment_chain:2"
     misc_feature    106962. .136493
                     /note="assembly_fragment:00001
                     fragment_chain:2"
     misc_feature    136594. .145799
                     /note="assembly_fragment:01622
                     fragment_chain:2"
     misc_feature    145900. .156851
                     /note="assembly_fragment:01367
                     fragment_chain:2
                     clone_end:T7
                     vector_side:right"
ORIGIN

  Query Match             4.9%;  Score 53;  DB 2;  Length 156851;
  Best Local Similarity   45.8%;  Pred. No. 1.5;
  Matches  212;  Conservative    0;  Mismatches  249;  Indels    2;  Gaps    1;

Qy         75 TTTAAAGCCAGATCTTCCTTATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATAT 134
              ||| |    | |   | ||||   | |      || | || |   | || | | | ||
Db      99343 TTTTATATTATAATATATTTATTAATTTATTATATTTTTATTAAATAAATTTTATTATTA 99284

Qy        135 GAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCA 194
              | | |   |    || ||| || |||  || | | |  | |   ||  |    || | |
Db      99283 TATTATTAATTTTTAAAATTTATNATTTAAATTAAATATAAAAAATTAAAATTTATATTA 99224

Qy        195 TCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACA 254
              | ||| | |   | |||  |||||  | |   ||  | |    |   |  ||    | |
Db      99223 TATTATATATTATATTTATTTAATTAAATAAATTATTATAAAATATTAATTTTTATTATA 99164

Qy        255 TACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGA 314
              |      |   ||   | ||  ||  || | || |  |   | ||| |     | |  | |
Db      99163 TTTTTAATTAATTTAAATATTATTTAATTAATAATAAATTAAAATATTAATAATTTTATA 99104

Qy        315 AAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTA 374
              | |     | ||       |  |  ||||     |||| | |    |||  |||| ||
Db      99103 TATAAATATTTTTTTATTTTATATTATTATTATAAAAATTATTATTAATTTAATTATTTT 99044

Qy        375 ATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCAC 434
              || ||  | |  ||  |||  | | ||  || |   ||| ||  |   || |||
Db      99043 ATATTTAATTTTAA--TAATTAATTAAATTATATAAATAATATATTTATTATTAATTATT 98986

Qy        435 TGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTT 494
                |||  ||||       ||||| |||  |||   |      | |  |  ||| ||
Db      98985 ATATTATATTTATTATATATTTTTTTAATTTTAAAAATTTTATTTTATATTTTTTTTATT 98926

Qy        495 ATGTACTATTAAAACATTTACTTTCCAAATATTCTGTCATTCA 537
              |    ||||   | ||||  |||    || || | | ||| |
Db      98925 TTTATTTATTTTNATATTTTTTTTTANTATTTTATATAATTNA 98883


RESULT 96
CR847809/c
LOCUS       CR847809              172777 bp    DNA     linear   HTG 05-OCT-2004
DEFINITION  Danio rerio clone CH211-28F19, *** SEQUENCING IN PROGRESS ***, 12
            unordered pieces.
ACCESSION   CR847809
VERSION     CR847809.2  GI:53828257
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 172777)
  AUTHORS   McLay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-OCT-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 5, 2004 this sequence version replaced gi:53749627.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
```

```
            ---------- Project Information
            Center project name: zC28F19
            ---------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 167687 bases at least Q40
            Consensus quality: 168529 bases at least Q30
            Consensus quality: 169301 bases at least Q20
            Insert size: 171677; sum-of-contigs
            Insert size: 179655; 0.9% error; agarose-fp
            Quality coverage: 9.92x in Q20 bases; sum-of-contigs Quality
            coverage: 9.57x in Q20 bases; agarose-fp
            ----------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 12 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    20768: contig of 20768 bp in length
            *    20769    20868: gap of 100 bp
            *    20869    44969: contig of 24101 bp in length
            *    44970    45069: gap of 100 bp
            *    45070    49740: contig of 4671 bp in length
            *    49741    49840: gap of 100 bp
            *    49841    55693: contig of 5853 bp in length
            *    55694    55793: gap of 100 bp
            *    55794    68728: contig of 12935 bp in length
            *    68729    68828: gap of 100 bp
            *    68829    86978: contig of 18150 bp in length
            *    86979    87078: gap of 100 bp
            *    87079    90802: contig of 3724 bp in length
            *    90803    90902: gap of 100 bp
            *    90903    93295: contig of 2393 bp in length
            *    93296    93395: gap of 100 bp
            *    93396   114910: contig of 21515 bp in length
            *   114911   115010: gap of 100 bp
            *   115011   153007: contig of 37997 bp in length
            *   153008   153107: gap of 100 bp
            *   153108   169736: contig of 16629 bp in length
            *   169737   169836: gap of 100 bp
            *   169837   172777: contig of 2941 bp in length.
FEATURES             Location/Qualifiers
     source          1. .172777
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-28F19"
                     /clone_lib="CHORI-211"
     misc_feature    1. .20768
                     /note="assembly_fragment:00259
                     fragment_chain:1"
     misc_feature    20869. .44969
                     /note="assembly_fragment:01664
                     fragment_chain:1"
     misc_feature    45070. .49740
                     /note="assembly_fragment:00079
                     fragment_chain:2"
     misc_feature    49841. .55693
                     /note="assembly_fragment:00164
                     fragment_chain:2"
     misc_feature    55794. .68728
                     /note="assembly_fragment:00421
                     fragment_chain:2"
     misc_feature    68829. .86978
                     /note="assembly_fragment:01287
                     fragment_chain:2"
     misc_feature    87079. .90802
                     /note="assembly_fragment:00115
                     fragment_chain:2"
     misc_feature    90903. .93295
                     /note="assembly_fragment:00016
                     fragment_chain:2"
     misc_feature    93396. .114910
                     /note="assembly_fragment:00943
                     fragment_chain:2"
     misc_feature    115011. .153007
                     /note="assembly_fragment:02107
                     fragment_chain:2
                     clone_end:T7
                     vector_side:right"
     misc_feature    153108. .169736
                     /note="assembly_fragment:00611
                     fragment_chain:2"
     misc_feature    169837. .172777
                     /note="assembly_fragment:00034
                     fragment_chain:2"
ORIGIN

  Query Match             4.9%;  Score 53;  DB 2;  Length 172777;
  Best Local Similarity   44.8%;  Pred. No. 1.5;
  Matches  194;  Conservative    0;  Mismatches  239;  Indels    0;  Gaps    0;

Qy         94 TATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAGTTTGGTTGAGAAATAT 153
              ||| | | |      | | |  || || |  | | |  |     |||  ||   ||||||
Db      56389 TATATATATTTTTTTTTTTTTTTTTTTATTTATATATTTATTTTTTATTTTTTAAATAT 56330

Qy        154 TGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTG 213
              |  | || |||  |   || |        | || |    |  |  | | | ||| |
Db      56329 ATAAAAATATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAATTATATAAAAATTTTATT 56270

Qy        214 TTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTT 273
              | ||   |    ||||||||  |  |  ||| | | | | |   |  | |  |
Db      56269 ATTATTANATTTAAAATTTTTTTTATATAATATTAAATAAAAAAAAAAAAAAAAATAATAN 56210

Qy        274 TTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAG 333
              | | |   | |   |         |  ||| | | ||| |  | | |||    |
Db      56209 NAATTATATAATNATATTTTTATTTATTTTTAATTTAAAAAAATAAAAAAAATTTTAAAA 56150

Qy        334 TTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAA 393
              | ||||  ||| | ||||  |    | |||| |    |||| |   |   | | |
Db      56149 ATAAAAAAAAAAAAAAAAAAAAAAATAAATTAAAAATATTAATATATTAAAAATAAAAAT 56090

Qy        394 AAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCAC 453
              |||  |   |||    |  |      | |    |  |  || ||  |
Db      56089 AAAATTTTATTAATAAAAATATTATAATAAATTAAATTTTTTATATTTAAAATAAATATT 56030

Qy        454 AAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTT 513
              |||  || || |  | ||| | | | |||| ||||||  | || ||  | |||  |  | |
Db      56029 AAAAATTTTTATTAATATATTTTAATAAAATATAATTAATAATATAAAAATAATAATTAT 55970

Qy        514 ACTTTCCAAATAT 526
              | | |  |||||
Db      55969 ATTATAATAATAT 55957


RESULT 97
AC022281/c
LOCUS       AC022281              186431 bp    DNA     linear   HTG 24-JAN-2002
DEFINITION  Homo sapiens chromosome 10 clone RP11-10E13, WORKING DRAFT
            SEQUENCE, 9 unordered pieces.
ACCESSION   AC022281
VERSION     AC022281.3  GI:13399348
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN;
            HTGS_CANCELLED.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 186431)
  AUTHORS   Smith,D.R.
  TITLE     Genome Therapeutics Corporation Sequencing Center: Human Genome
            Sequence Data
  JOURNAL   Unpublished
```

```
REFERENCE   2  (bases 1 to 186431)
  AUTHORS   Smith,D.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JAN-2000) Genome Therapeutics Corporation, 100 Beaver
            Street, Waltham, MA 02453, USA
COMMENT     On Mar 21, 2001 this sequence version replaced gi:9929643.
            -------------- Genome Center
                           Center: Genome Therapeutics Corporation
                           Center code: GTC
                           Web site: http://www.genomecorp.com/
                           Contact: gtc-seqcenter@genomecorp.com
                         -------------- Project Information
                           Center project name: hg143
                         -------------- Summary Statistics
                           Sequencing vector: N/A
                           Chemistry: Dye-terminator Big Dye; 100% of reads
                           Assembly program: Phrap; version 990315
                           Consensus quality: 395389 bases at least Q40
                           Consensus quality: 405067 bases at least Q30
                           Consensus quality: 410525 bases at least Q20
                           Insert size: 185631; sum-of-contigs
                           Quality coverage: 6.4x in Q20 bases; sum-of-contigs
                         --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1   124725: contig of 124725 bp in length
            *   124726   124825: gap of unknown length
            *   124826   125924: contig of 1099 bp in length
            *   125925   126024: gap of unknown length
            *   126025   128493: contig of 2469 bp in length
            *   128494   128593: gap of unknown length
            *   128594   156515: contig of 27922 bp in length
            *   156516   156615: gap of unknown length
            *   156616   170011: contig of 13396 bp in length
            *   170012   170111: gap of unknown length
            *   170112   176430: contig of 6319 bp in length
            *   176431   176530: gap of unknown length
            *   176531   179019: contig of 2489 bp in length
            *   179020   179119: gap of unknown length
            *   179120   181965: contig of 2846 bp in length
            *   181966   182065: gap of unknown length
            *   182066   186431: contig of 4366 bp in length.
FEATURES             Location/Qualifiers
     source          1. .186431
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="RP11-10E13"
                     /clone_lib="RPCI-11"
     misc_feature    1. .124725
                     /note="assembly_name:Contig243
                     clone_end:SP6"
     misc_feature    124826. .125924
                     /note="assembly_name:Contig204"
     misc_feature    126025. .128493
                     /note="assembly_name:Contig233"
     misc_feature    128594. .156515
                     /note="assembly_name:Contig240"
     misc_feature    156616. .170011
                     /note="assembly_name:Contig239"
     misc_feature    170112. .176430
                     /note="assembly_name:Contig235"
     misc_feature    176531. .179019
                     /note="assembly_name:Contig231"
     misc_feature    179120. .181965
                     /note="assembly_name:Contig232"
     misc_feature    182066. .186431
                     /note="assembly_name:Contig236
                     clone_end:T7"
ORIGIN

  Query Match             4.9%;  Score 53;  DB 2;  Length 186431;
  Best Local Similarity   43.5%;  Pred. No. 1.5;
  Matches  334;  Conservative    0;  Mismatches  426;  Indels    8;  Gaps    3;

Qy         20 CAGTAACTTTCACTCTCTATGTTCTTATAAGCAAATGTTACAATGAGATATGAGTTTTAA 79
              ||  |  |||   | |  |   | | | ||  ||| |||  ||     | | | |
Db     176409 CAAAATTTTTTTTTATAAAATATATAAAAAAAAAAATTTATTATTTTTATTTATTATATA 176350

Qy         80 AGCCAGATCTTCCTTATCTCTCTGCCCCATCTCTAGTTCTTGAAGTGTCTCATATGAGTT 139
                ||  | ||  || | | |         |  | ||| || ||  |     | || | |
Db     176349 TAACAATTATTAATTTTATATTAAAANAANTTTTATTTATTATAAAAAAAAAAATAAATA 176290

Qy        140 TGGTTGAGAAATATTGATCAT-TACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTT 198
              || | || ||||    || || ||||  || ||   | |  | ||    | |
Db     176289 ATTTTNATAANTATTATAAATATATAAATTTATTTATTAATATTAAAAAATATTAAAAAA 176230

Qy        199 AAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACA 258
              | | |  | |||   |  ||||      |  || || ||||| | |  || |   | |
Db     176229 ATAAANTTAATTTATATTTTATATATAATATTTTATTATAAAAAATTATATAATTANAAT 176170

Qy        259 GCTATTCATTTGTTTTTGGTT-TGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAG 317
                |||   ||| ||||  ||    | |     |  ||| | |||  | |     |
Db     176169 TATATAATTTTATTTTAAATTAAANTTATTTTAATAATATATTAATATAATTTTTAATTA 176110

Qy        318 ATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATC 377
              |||||  |   | || | | ||||| | |  || | |  |||    | ||  ||||
Db     176109 TATATATANTNATAAATTTATATTTAAATAAANTATTAATTTATAAATAAATANATAATT 176050

Qy        378 TTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGC 437
              ||  |    |   || || | ||    | | ||     |  ||      |    |
Db     176049 TTATAAAATATAATATNAATTTAAAAATATAATTTAAANTATTATTAAAATTNTATATAT 175990

Qy        438 AAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATG 497
              ||||| || |||    ||| | ||       ||   || |   || | |  | | || ||
Db     175989 AAATAAAATTTAATTAAAAATATA------AATATATATNTTAATATAATAATATTAATA 175936

Qy        498 TACTATTAAAACATTTACTTTCCAAATATTCTGTCATTCAGGAGTATGGAAGTATCGATG 557
              ||    | |  | | |  || ||| ||  |   |  |     ||| |  || ||
Db     175935 TATATAAATATAAATAAAATTATAAAAATAATAAAAAATANNNTTATAATATAATATATA 175876

Qy        558 GCTTCTTTAAAATGAAGCAGGAGGGTCTGGCAGAGAGTATCTATGAAATAAGTTCCTCTG 617
                | || | ||  |   |  |       | | | | ||  |||
Db     175875 TATAATTAATAAAAATTNAATAATAAATAATTTAAATTTTTTAATAAAATTAAAATAAAA 175816

Qy        618 ACCTTCACGCTTAATTTTCTGAATGGAGTGGAGCAAATTACTTCAAGCTTCACTTAACTT 677
              |  || |   ||  ||     | |  | |       |  | | |  ||  |  | || | |
Db     175815 ATTTTAATAATTTTTTAAAATATTATAATTTTAAATTTAAATATAAAATAAANTTTATAT 175756

Qy        678 GCATATGAAATGAACCGTACAAAAATACAAGAGTGTCAGGAGAAAGTTATGCTCTGGTAA 737
                |  ||||      | |||||   |  | | |      || | |   | |  |||
Db     175755 AATAAAAAAATTTTTTAAAAAAAAAATAATTAATTTTTATTATAAATAAAATTTTTTTAA 175696

Qy        738 ATATTTTGCAAAACAGATAAAAGATAATACTAGAGCTCTGTCCTCAAA 785
              ||||||   || | | ||  |    | || | || |  | | |  | |||
Db     175695 ATATTTAAAAATAAAAATNATTAAAAAAATTAAAATTTTTTAATAAAA 175648


RESULT 98
CR318659
LOCUS       CR318659              208029 bp    DNA     linear   HTG 11-MAR-2004
DEFINITION  Danio rerio clone CH211-215M5, WORKING DRAFT SEQUENCE, 9 unordered
            pieces.
ACCESSION   CR318659
VERSION     CR318659.5  GI:45381879
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 208029)
  AUTHORS   Burton,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-MAR-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Mar 11, 2004 this sequence version replaced gi:45238490.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zC215M5
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 204233 bases at least Q40
            Consensus quality: 205033 bases at least Q30
            Consensus quality: 205725 bases at least Q20
            Insert size: 207229; sum-of-contigs
            Insert size: 198045; 20.1% error; agarose-fp
            Quality coverage: 8.70x in Q20 bases; sum-of-contigs Quality
            coverage: 9.27x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 9 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    43951: contig of 43951 bp in length
            *   43952    44051: gap of 100 bp
            *   44052    99129: contig of 55078 bp in length
            *   99130    99229: gap of 100 bp
            *   99230   111372: contig of 12143 bp in length
            *  111373   111472: gap of 100 bp
            *  111473   118963: contig of 7491 bp in length
            *  118964   119063: gap of 100 bp
            *  119064   134802: contig of 15739 bp in length
            *  134803   134902: gap of 100 bp
            *  134903   141922: contig of 7020 bp in length
            *  141923   142022: gap of 100 bp
            *  142023   152611: contig of 10589 bp in length
            *  152612   152711: gap of 100 bp
            *  152712   205404: contig of 52693 bp in length
            *  205405   205504: gap of 100 bp
            *  205505   208029: contig of 2525 bp in length.
FEATURES             Location/Qualifiers
     source          1. .208029
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="CH211-215M5"
                     /clone_lib="CHORI-211"
     misc_feature    1. .43951
                     /note="assembly_fragment:00844
                     clone_end:T7
                     vector_side:left"
     misc_feature    44052. .99129
                     /note="assembly_fragment:02386
                     fragment_chain:1"
     misc_feature    99230. .111372
                     /note="assembly_fragment:00376
                     fragment_chain:1"
     misc_feature    111473. .118963
                     /note="assembly_fragment:00048
                     fragment_chain:1"
     misc_feature    119064. .134802
                     /note="assembly_fragment:00569
                     fragment_chain:1"
     misc_feature    134903. .141922
                     /note="assembly_fragment:00135
                     fragment_chain:1"
     misc_feature    142023. .152611
                     /note="assembly_fragment:00227
                     fragment_chain:1"
     misc_feature    152712. .205404
                     /note="assembly_fragment:01535
                     fragment_chain:1"
     misc_feature    205505. .208029
                     /note="assembly_fragment:00029"
ORIGIN

  Query Match             4.9%;  Score 53;  DB 2;  Length 208029;
  Best Local Similarity   38.8%;  Pred. No. 1.4;
  Matches  276;  Conservative    0;  Mismatches  425;  Indels   11;  Gaps    2;

Qy        131 ATATGAGTTTGGTTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGAT 190
              ||||   | |   | | | ||||  ||   || | ||     ||||    | || |
Db     151901 ATATATATATATATTATATATATATATATATATATATATAAAAATAAAAATATATATATA 151960

Qy        191 CTCATCTTAAAGACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACA 250
              | |   |||| |  || ||   || ||| | |       ||| ||||| |     ||
Db     151961 TTTAAAATAAAAAATTTATTAATTTTATAAAATAAAAATATTATTTTATATTTTATTTTT 152020

Qy        251 GACATACAGCTATTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCA 310
              ||  | |   ||| || | |||||   |    ||||||  | |   | |  ||  |
Db     152021 GAGTGATAAAAATTTATATATTTTTTTAATAAAAAAAAAAAAAAAAATTATATTTAATAAA 152080

Qy        311 GAGAAAGATCATATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTA 370
              | ||| |   ||| |          | ||||||| | ||||    |   ||||   ||
Db     152081 AAAAAAAAATTTATTT----------ATAATTAAAAATAAAAAATATATTTATTTTTTT 152130

Qy        371 CTTAATCTTGCAGTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATC 430
              | | |     |   |||  |||||| | ||   |||   | | ||  |  ||    ||
Db     152131 TTAATTAAAAAAAAAAAATTTAAAAAATAAAAAAAAAAAAAAAATTAAATTTTTATAAATT 152190

Qy        431 GCACTGCAAATAGAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATT 490
              |   |||   |||  |   || |||| ||||   || ||    | |||| | |
Db     152191 TAATAATAAAATAAAAAAAATTAAATTTTTTTTTTTTTTAANAAAANTAAAAAAAAAANAA 152250

Qy        491 TTTTATGTACTATTAAAACATTTACTTTCCAAATATTCTGTCATTCAGGAGTATGGAAGT 550
                |          |  |  |  |   | |||  |       |      | ||   || |
Db     152251 AAAAAAAANNANNNANNANNTAAAAAATAAAAAAAAAAAANAAAAATNATAATAAANAATT 152310

Qy        551 ATCGATGGCTTCTTTAAAATGAAGCAGGAGGGTCTGGCAGAGAGTATCTATGAAATAAGT 610
              ||   |  || |   |||  |   |  |  |  |  |  | | |
Db     152311 ATTNTTATTTTATAANAAAAAANATAAATTNTTNTTTTTTTTTTNTTTTTNNNNNNNNAN 152370

Qy        611 TCCTCTGACCTTCACGCTTAATTTTCTGAATGGAGTGGAGCAAATTACTTCAAGCTTCAC 670
              |          |       |       |          |          |
Db     152371 NNNNAANAANNNAANNNNANNTNAAAAAANANNNANNNNNNANNNNNNANANNNNNNNAN 152430

Qy        671 TTAACTTGCATATGAAATGAACCGTACAAAAATACAAGAG-TGTCAGGAGAAAGTTATGC 729
                | ||  |||   ||  ||         || | |  |  | | |  | ||| ||||
Db     152431 AANANTTAAATANNNAANAAAAANATTNTNAAAATATNATTTATAAAAATAAAATTATAA 152490

Qy        730 TCTGGTAAATATTTTGCAAAACAGATAAAAGATAATACTAGAGCTCTGTCCTCAAAGAGT 789
              | |    ||   |||   | | | | |     || | |   | |    | | |    |
Db     152491 TATTNNNAAATATTTNATATAAATATTATTATAAANATAATTAATTTNNATTAATAANTT 152550

Qy        790 TAAGCAGCTAATCTAAGGAGGTAAACTCTATGTCAGCAGGATGAACTGCTCT 841
              | |  |   ||   |  |       ||       |  || || |   |
Db     152551 TNAATAAANNATAATANTAANANNNNNATANNATTNTANAATAAAATNNNTT 152602

RESULT 99
CR396591
LOCUS       CR396591              297798 bp    DNA     linear   HTG 23-AUG-2004
DEFINITION  Danio rerio clone DKEY-43I4, *** SEQUENCING IN PROGRESS ***, 35
            unordered pieces.
```

```
ACCESSION   CR396591
VERSION     CR396591.4  GI:51510742
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Danio rerio (zebrafish)
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 297798)
  AUTHORS   Sims,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-AUG-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 23, 2004 this sequence version replaced gi:47058863.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK43I4
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 283329 bases at least Q40
            Consensus quality: 286900 bases at least Q30
            Consensus quality: 289553 bases at least Q20
            Insert size: 294398; sum-of-contigs
            Insert size: 214942; 3.3% error; agarose-fp
            Quality coverage: 5.41x in Q20 bases; sum-of-contigs Quality
            coverage: 8.73x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 35 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    31177: contig of 31177 bp in length
            *    31178    31277: gap of 100 bp
            *    31278    39918: contig of 8641 bp in length
            *    39919    40018: gap of 100 bp
            *    40019    62375: contig of 22357 bp in length
            *    62376    62475: gap of 100 bp
            *    62476    72508: contig of 10033 bp in length
            *    72509    72608: gap of 100 bp
            *    72609    75990: contig of 3382 bp in length
            *    75991    76090: gap of 100 bp
            *    76091    78503: contig of 2413 bp in length
            *    78504    78603: gap of 100 bp
            *    78604    81027: contig of 2424 bp in length
            *    81028    81127: gap of 100 bp
            *    81128    83142: contig of 2015 bp in length
            *    83143    83242: gap of 100 bp
            *    83243    88422: contig of 5180 bp in length
            *    88423    88522: gap of 100 bp
            *    88523    90939: contig of 2417 bp in length
            *    90940    91039: gap of 100 bp
            *    91040   105245: contig of 14206 bp in length
            *   105246   105345: gap of 100 bp
            *   105346   112237: contig of 6892 bp in length
            *   112238   112337: gap of 100 bp
            *   112338   142171: contig of 29834 bp in length
            *   142172   142271: gap of 100 bp
            *   142272   147177: contig of 4906 bp in length
            *   147178   147277: gap of 100 bp
            *   147278   151421: contig of 4144 bp in length
            *   151422   151521: gap of 100 bp
            *   151522   155292: contig of 3771 bp in length
            *   155293   155392: gap of 100 bp
            *   155393   206156: contig of 50764 bp in length
            *   206157   206256: gap of 100 bp
            *   206257   208599: contig of 2343 bp in length
            *   208600   208699: gap of 100 bp
            *   208700   211937: contig of 3238 bp in length
            *   211938   212037: gap of 100 bp
            *   212038   214623: contig of 2586 bp in length
            *   214624   214723: gap of 100 bp
            *   214724   216785: contig of 2062 bp in length
            *   216786   216885: gap of 100 bp
            *   216886   220129: contig of 3244 bp in length
            *   220130   220229: gap of 100 bp
            *   220230   223587: contig of 3358 bp in length
            *   223588   223687: gap of 100 bp
            *   223688   226021: contig of 2334 bp in length
            *   226022   226121: gap of 100 bp
            *   226122   241367: contig of 15246 bp in length
            *   241368   241467: gap of 100 bp
            *   241468   266170: contig of 24703 bp in length
            *   266171   266270: gap of 100 bp
            *   266271   268872: contig of 2602 bp in length
            *   268873   268972: gap of 100 bp
            *   268973   271693: contig of 2721 bp in length
            *   271694   271793: gap of 100 bp
            *   271794   274544: contig of 2751 bp in length
            *   274545   274644: gap of 100 bp
            *   274645   278241: contig of 3597 bp in length
            *   278242   278341: gap of 100 bp
            *   278342   282192: contig of 3851 bp in length
            *   282193   282292: gap of 100 bp
            *   282293   287066: contig of 4774 bp in length
            *   287067   287166: gap of 100 bp
            *   287167   289567: contig of 2401 bp in length
            *   289568   289667: gap of 100 bp
            *   289668   292263: contig of 2596 bp in length
            *   292264   292363: gap of 100 bp
            *   292364   297798: contig of 5435 bp in length.
FEATURES             Location/Qualifiers
     source          1. .297798
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-43I4"
                     /clone_lib="DanioKey"
     misc_feature    1. .31177
                     /note="assembly_fragment:00622
                     fragment_chain:1"
     misc_feature    31278. .39918
                     /note="assembly_fragment:02990
                     fragment_chain:1"
     misc_feature    40019. .62375
                     /note="assembly_fragment:00316
                     fragment_chain:1"
     misc_feature    62476. .72508
                     /note="assembly_fragment:02944
                     fragment_chain:1"
     misc_feature    72609. .75990
                     /note="assembly_fragment:02732
                     fragment_chain:1"
     misc_feature    76091. .78503
                     /note="assembly_fragment:02608
                     fragment_chain:1"
     misc_feature    78604. .81027
                     /note="assembly_fragment:02638
                     fragment_chain:2"
     misc_feature    81128. .83142
                     /note="assembly_fragment:02742
                     fragment_chain:2"
     misc_feature    83243. .88422
                     /note="assembly_fragment:02921
                     fragment_chain:2"
     misc_feature    88523. .90939
                     /note="assembly_fragment:02689
```

```
                fragment_chain:2"
  misc_feature  91040. .105245
                /note="assembly_fragment:00001
                fragment_chain:3"
  misc_feature  105346. .112237
                /note="assembly_fragment:03113
                fragment_chain:3"
  misc_feature  112338. .142171
                /note="assembly_fragment:01319
                fragment_chain:3"
  misc_feature  142272. .147177
                /note="assembly_fragment:02898
                fragment_chain:4"
  misc_feature  147278. .151421
                /note="assembly_fragment:02821
                fragment_chain:4"
  misc_feature  151522. .155292
                /note="assembly_fragment:02762
                fragment_chain:4"
  misc_feature  155393. .206156
                /note="assembly_fragment:01741
                fragment_chain:5"
  misc_feature  206257. .208599
                /note="assembly_fragment:02854
                fragment_chain:5"
  misc_feature  208700. .211937
                /note="assembly_fragment:02705
                fragment_chain:6"
  misc_feature  212038. .214623
                /note="assembly_fragment:02783
                fragment_chain:6"
  misc_feature  214724. .216785
                /note="assembly_fragment:02714
                fragment_chain:7"
  misc_feature  216886. .220129
                /note="assembly_fragment:02645
                fragment_chain:7"
  misc_feature  220230. .223587
                /note="assembly_fragment:02772
                fragment_chain:8"
  misc_feature  223688. .226021
                /note="assembly_fragment:02652
                fragment_chain:8"
  misc_feature  226122. .241367
                /note="assembly_fragment:00146.0"
  misc_feature  241468. .266170
                /note="assembly_fragment:00952.0"
  misc_feature  266271. .268872
                /note="assembly_fragment:02573"
  misc_feature  268973. .271693
                /note="assembly_fragment:02659"

  Query Match             4.9%;  Score 53;  DB 2;  Length 297798;
  Best Local Similarity   45.4%;  Pred. No. 1.4;
  Matches  167;  Conservative    0;  Mismatches  201;  Indels    0;  Gaps    0;

Qy        143 TTGAGAAATATTGATCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAG 202
              || |  || |   || ||   |||  |   |    ||   ||      | |  | |||
Db      30800 TTTATTAAAAAATATTATATAAAAAAAAAAAANATNTTANATATTTTTTTTTTTTTTATTA 30859

Qy        203 ACATTGTTTTGTTAATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTA 262
              |  || | || |   |      ||  |||||||  || |    | | | | | |  ||
Db      30860 AATTTTTATTTTATTAAATNNTAATTATTTTTTTTTTAAAAAAAAAAAAAAAAAAAATTA 30919

Qy        263 TTCATTTGTTTTTGGTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCAT 322
               |  ||| |||||  |||        |||   | ||| |||   |   |   ||| |  ||
Db      30920 ATTTTTTTTTTTTTTTTTAAAATTTAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAAT 30979

Qy        323 ATATTAGCCAGTTGAAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCA 382
              | |||       | ||  ||  || |   ||| | |   |||||  | ||  ||   |   |
Db      30980 AAATTTTAATTATAAATATATAATAATTAATAATTTATTATTAAAATAAATATAATAATA 31039

Qy        383 GTCAAAGGTAAAAAGTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATA 442
```

```
              |      ||||||   ||   |||   | |  |          |       |  ||
Db      31040 TTTTTTTATAAAAATAAAAAAAAAAAAAAAANAAAAAAAANAAAAAATTNAAANATNAAATNT 31099

Qy        443 GAAATTACCACAAATTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTA 502
              ||   |  | | ||||    ||  ||| ||       |||  | | | ||||  |  | ||
Db      31100 TAATAAATAAAATATTTATAATTATAAAAAAATATTAAAATAAAATTATTTTTTTAAATA 31159

Qy        503 TTAAAACA 510
              || ||| |
Db      31160 TTTAAAAA 31167


RESULT 100
AX345640/c
LOCUS       AX345640               6759 bp    DNA     linear   PAT 01-FEB-2002
DEFINITION  Sequence 711 from Patent WO0200928.
ACCESSION   AX345640
VERSION     AX345640.1  GI:18493526
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            other sequences; artificial sequences.
REFERENCE   1
  AUTHORS   Olek,A., Piepenbrock,C. and Berlin,K.
  TITLE     Diagnosis of diseases associated with the immune system
  JOURNAL   Patent: WO 0200928-A 711 03-JAN-2002;
            Epigenomics AG (DE)
FEATURES             Location/Qualifiers
     source          1. .6759
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="chemically treated genomic DNA (Homo sapiens)"
ORIGIN

  Query Match             4.8%;  Score 52.8;  DB 6;  Length 6759;
  Best Local Similarity   46.7%;  Pred. No. 2.8;
  Matches  168;  Conservative    0;  Mismatches  192;  Indels    0;  Gaps    0;

Qy        157 TCATTACAAATCAGTTAATAGTTTTGTAGAAGATCTCATCTTAAAGACATTGTTTTGTTA 216
              | |  | |||  |  || || ||    | |  || | |  || || |   | | || | |
Db       2866 TAAAAAAAAAATATATACTATTTAATCAAATAATTTTACTTTTAAAAATATATCTTATAA 2807

Qy        217 ATATACTCCCTTGATTTTTTTAAAAGACCTTACAGACATACAGCTATTCATTTGTTTTTG 276
              | ||||     | |   |  | || |  ||| | | | | |  ||||    | || ||
Db       2806 AAATACAAATATAAACATACTTTAACAAATTATATAAAAAAATTTATTTCAATATTATTC 2747

Qy        277 GTTTGTTCAAAAAAGGTATAAAGAAATGCATTCAGAGAAAGATCATATATTAGCCAGTTG 336
              |||||||    |||||  ||    ||   | || |  ||| || |   | |
Db       2746 AAAATATCAAAAATAAAATAAAATAAATACTTTTTATTAAAAAAATAAATAAATTATTCT 2687

Qy        337 AAAATTAAACACAAAATGAGTGCATATTACATTACTTAATCTTGCAGTCAAAGGTAAAAA 396
              | || | || || ||   |  | ||  | | |||   ||         |||   |||||
Db       2686 ATAAATTAAAACTAACAAATAACTTAAAAAAATACCCCATTAAAATACTAAAAAAAAAAA 2627

Qy        397 GTCAACCTAAAGGTATACTACCTGCTTTCTTATCGCACTGCAAATAGAAATTACCACAAA 456
                | |  |   | | ||| |   | | |||   |   || |  || ||| |
Db       2626 AAACAACATACAAAAAAATACTTATAATATAATCCTTATTTCAAAAATAAATACTAATTT 2567

Qy        457 TTTTATTTTGGAAATAATCTCAGAAAACATAATTTTTTATGTACTATTAAAACATTTACT 516
              || |  |    | ||| | | | | |   | |  ||||  || ||  | || |
Db       2566 TTAAAAATCTATATAAATATAAACACAAACACACATAAACATACTCATATAATCTATAAT 2507


Search completed: April 21, 2005, 17:05:10
Job time : 3071.48 secs
```

This Page Blank (uspto)

This Page Blank (uspto)